THE ROLE OF INDIVIDUAL DIFFERENCES IN L1 AND L2 PROCESSING
OF BRIDGING AND PREDICTIVE INFERENCES

A Dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Linguistics

By

Julie B. Lake, M.S.

Washington, D.C.
May 20, 2014
THE ROLE OF INDIVIDUAL DIFFERENCES IN L1 AND L2 PROCESSING OF BRIDGING AND PREDICTIVE INFERENCES

Julie Beth Lake, M.S.

Co-Dissertation Advisors: Alison Mackey, Ph.D. and Sonja Kotz, Ph.D.

ABSTRACT

Second language acquisition (SLA) and psycholinguistic researchers are interested in how inferences are processed in first language (L1) and second language (L2) reading to better understand how learners negotiate for meaning within a text. Preliminary evidence from brain-based studies suggests that bridging inferences (i.e., connecting pronouns and referents) may rely more heavily on short-term memory (STM) (Burkhardt, 2006) while predictive inferences (i.e., forecasting upcoming events from textual information) may rely more heavily on cognitive control (CC) (Jin et al., 2009). However, despite these advances, it is still not clear whether or not L2 readers have access to the necessary linguistic structures (Horiba, 1996) or, instead, if proficiency level is a confounding variable (Cohen, 1998; Guerrero, 2005; Leontiev, 1981).

To address these gaps, an L1 English reading study and an L2 English reading study were carried out to explore whether or not inferences are processed differently than non-inferences, whether or not bridging and predictive inferences are processed differently from one another, and which cognitive resources are recruited during the processing of bridging and predictive inferences. The L1 and L2 data were compared to see if inference processing differs based on language background. L1 English speakers (n = 50) and L2 English-L1 Spanish speakers (n = 23) completed a cumulative self-paced reading task in a distraction paradigm; participants read scenarios that required them to
make a bridging inference, a predictive inference or no inference with interference (while completing a secondary task that taxed STM or CC) or without interference. Independent measures of STM and CC were also collected to see whether these scores were associated with reading performance.

The results indicated that L1 and proficient L2 readers process inferences similarly. Participants relied on CC for bridging and predictive inferences, supporting a multi-component framework of attentional control capacity (Barnes & Jones, 2000). These results are consistent with research on individual difference in SLA (e.g., Dornyei & Skehan, 2003), in that they suggest that CC may influence L2 classroom learning and account for the variation in L2 reading acquisition. Additionally, the findings have implications for L2 testing, as they pose a threat to the internal validity of standardized reading tests and ultimately the grade-level placement of L2 readers.
ACKNOWLEDGEMENTS

I want to acknowledge and thank several people that helped me reach this point. It has been my experience that it ‘takes a village’ to finish a dissertation, and I have had an exceptional village. I am lucky to have been surrounded by a community of thoughtful, bright, supportive and insightful people. It is challenging to pay literal tribute, but hopefully through inference, you can fill in the gaps.

First, I want to acknowledge my committee members – Dr. Alison Mackey, Dr. Sonja Kotz, Dr. Patrick Rebuschat, and Dr. Cristina Sanz – for their support, advice, and feedback. Alison has been an academic guide since I began my graduate studies and she continually pushes me to think about my research and how it fits into a larger picture. I have evolved as a scholar under her mentorship and I credit Alison for my rigorous approach to research methodology. I visited Sonja’s research laboratory at the Max Planck Institute for Human Cognitive and Brain Sciences several years ago to learn about brain-based methodologies, and I haven’t looked back since. Sonja has helped me in innumerable ways, but she was instrumental in helping me interpret and expand on the full extent of my dissertation results. Patrick founded the Psycholinguistics Lab at Georgetown, which is still running several years later. We spent hours discussing my dissertation topic as I was developing it. Patrick possesses the perfect blend of patience, support and expectation. Cristina was my first professor of SLA, and I enjoyed every minute of her course. In fact, my interest was peaked to pursue individual differences research in her course 6 years ago.

Georgetown University has also provided excellent resources throughout my graduate studies. I benefitted from an exceptional research assistant, Kimberly Ta,
through the Georgetown Undergraduate Research Opportunities Program (GUROP).
Kimberly brought a careful attention to detail to the research, and she helped with stimuli
development, experimental set-up and the piloting process. I also attended two of the
Graduate Student Organization (GSO) Dissertation Boot Camps, where I was able to
focus on my dissertation without distractions. I would also like to thank Dr. Rusan Chen,
Dr. John Norris, and (the soon-to-be) Dr. Luke Amoroso for their patient assistance with
my dissertation statistics.

I appreciate the staff and faculty in the Linguistics Department. Thank you to the
talented professors throughout my graduate career, who have helped shape me as a
scholar. I wanted to extend a special thanks to Dr. Andrea Tyler and Dr. Jeff Connor-
Linton for their guidance and support over the years. The Linguistics Department staff at
Georgetown University – including Manela Diez, Erin Esch Pereira, Jennifer Brusstar,
and Cristian Desmaras – are true professionals. They have worked tirelessly on my behalf.

The students at Georgetown University are some of the finest scholars (and
human beings) I have had the privilege to work with over the years. They have been
generous with their time and expertise, reading countless drafts, offering methodological
and statistical consultations, and discussing the implications of my results. I would like to
thank (in no order of importance) Kaitlyn Tagarelli, Jessi Grieser, Marissa Fond, Jessie
Cox, Ellen Serafini, Phil Hamrick, Nicole Ziegler, Katie Kim, Sheena Shah, Rebecca
Sachs, Jae Goo, Amelia Tseng, Nick Pandža, Mari Sakai, Haein Lauren Park, Laura West,
Sarah Young, Daniel Ginsberg, Laura Vogrie, Hana Jan, Narges Mahpeykar, and Hillary
Harner. Additionally, thanks to Rachel Sawyer and Sophie Siebach for their excellent
work at the GURT conference (go catering!), which enabled me to participate in the GSO Dissertation Boot Camp. Forgive any (unintentional) omissions.

I received support from an NSF-PIRE grant to learn about brain-based research methodologies, which informed my dissertation research questions, and I was fortunate to spend roughly 10 months at the Max Planck Institute (MPI) for Human Cognitive and Brain Sciences. Thanks to Dr. Josef Rauschecker, Dr. Werner Graf, and Eva Graf-Mackrod for administering the funding, and thank you to Claudi Teickner, Corinna Koch, Josefin Bosch, and the other MPI colleagues for your friendship and mentorship.

I would also like to thank my family and friends. To my parents, Edward and Dena Lake, and my brother, Eli Lake, I have appreciated your love and support. Your humor and steadfast belief in me gave me strength to carry on.

For my soon-to-be husband, Michael Harrison, you are my partner in crime, my muppeteer, and my heart. You have been involved in almost every step of the dissertation process, as my confidante, brainstorming partner, editor, experimental guinea pig, and cheerleader. What words are there? Just spend another day with me, and let us wend our way toward the future of endless learning.

Finally, thank you to Georgetown University for giving me the opportunity to pursue my Ph.D.

This dissertation is dedicated to my parents
# Table of Contents

CHAPTER 1: INTRODUCTION

1.1. Background ........................................................................................................... 1
1.2. Theoretical and Practical Applications of Research .............................................. 2
1.3. Useful Terminology .............................................................................................. 4
  1.3.1. Non-Inference versus Inference ................................................................. 4
  1.3.2. Types of Inferences ...................................................................................... 5
  1.3.3. Two-Sentence Scenarios ............................................................................. 6
  1.3.4. Reaction Time Experiment ......................................................................... 6
  1.3.5. Cumulative Self-Paced Reading Task .......................................................... 7
  1.3.6. Visual Distraction Paradigm ....................................................................... 7
1.4. Overview of L1 Study ......................................................................................... 8
1.5. Overview of L2 Study ......................................................................................... 9
1.6. Preview of Upcoming Chapters ......................................................................... 10

CHAPTER 2: INFERENCE BUILDING ........................................................................ 12

2.1. Theories of Inference Generation ...................................................................... 13
2.2. Inference Types ................................................................................................... 15
  2.2.1. Bridging Inferences ................................................................................... 15
  2.2.2. Predictive Inferences ................................................................................ 16
2.3. Models of inference Processing ........................................................................ 17
  2.3.1. Construction-Integration Models ................................................................ 19
  2.3.2. Constructionist Models ............................................................................. 23
2.4. L1 versus L2 processing ................................................................................... 27
2.5. Summary ............................................................................................................ 32
CHAPTER 3: INDIVIDUAL DIFFERENCES AND INFEERENCE PROCESSING 35

3.1. Individual Differences (IDs) ................................................................. 36
  3.1.1. Working Memory ................................................................. 36
  3.1.2. Attentional Models .............................................................. 38
  3.1.3. Cognitive frameworks used in this dissertation ..................... 40
  3.1.3. Bilingual advantage .............................................................. 40

3.2. Predictions from Cognitive Models .................................................. 41
  3.2.1. CI Model: Predictions ............................................................... 41
  3.2.3. Constructionist Model: Predictions .......................................... 42

3.3. Empirical Findings from Behavioral Studies .................................... 43

3.4. Empirical Findings from Neurolinguistic Studies ............................. 45
  3.4.1. Inference Type and Cognitive Resources ................................. 45

3.5. Summary ......................................................................................... 48

3.6. Research Gap .................................................................................. 49

CHAPTER 4: METHODS .............................................................................. 51

4.1. Research Questions ........................................................................... 51

4.2. Hypotheses ....................................................................................... 52

4.3. Participants ....................................................................................... 53
  4.3.1. L1 English speakers ................................................................. 54
  4.3.2. L2 English speakers ................................................................. 55
    4.3.2.1. Proficiency ..................................................................... 55

4.4. Linguistic Stimuli ............................................................................. 59
  4.4.1. Guidelines used to create verbal stimuli ...................................... 62
  4.4.2 Names in verbal stimuli .............................................................. 65
  4.4.3 Pretest of stimuli ...................................................................... 66
4.4.3.1. Pretest #1: Rating study ................................................................. 66
4.4.3.2. Pretest #2: Reaction Time Sentence Pilot ........................................ 69
4.4.4. Final verbal stimuli ........................................................................ 73
4.5. Experimental Tasks and Materials .......................................................... 74
  4.5.1. Recruitment and Informed Consent Procedure .................................... 74
  4.5.2. Background Questionnaire ............................................................... 76
  4.5.3. Overview of Cognitive Measures ...................................................... 77
    4.5.3.1. Stand-alone individual difference battery ...................................... 77
    4.5.3.2. Distractor tasks ......................................................................... 78
    4.5.3.3. Purpose of dual measures of individual differences ...................... 79
    4.5.3.4. Cognitive Measures Pretest ......................................................... 79
  4.5.4 Individual Difference Test Battery ...................................................... 80
    4.5.4.1. Visual patterns task .................................................................. 80
    4.5.4.2. N-Back Task ........................................................................... 84
  4.5.5. Distractor Tasks .............................................................................. 87
    4.5.5.1. Visual Patterns Task ................................................................. 87
    4.5.5.2 N-Back task ............................................................................. 87
  4.5.6. Correlations of Individual Measures .................................................. 88
  4.5.7. Cumulative Self-Paced Reading Task .................................................. 91
  4.6. Procedure and Data Collection ............................................................. 96
  4.7. Summary of variables ...................................................................... 98

CHAPTER 5: RESULTS OF L1 STUDY ............................................................. 99
  5.1. Introduction ....................................................................................... 99
  5.2. Assumptions of parametric statistical methods .................................... 99
    5.2.1. Assumptions of repeated measures analyses of variance ............... 100
    5.2.2. Assumptions for correlational analyses ........................................ 100
6.6. Research question involving bridging inferences versus predictive inferences ..................................................143

6.6.1. Processing of bridging inferences .............................................................................................................144

6.6.1.1. Repeated-Measures ANOVA ......................................................................................................................145

6.6.2. Processing of predictive inferences ..........................................................................................................146

6.6.2.1. Repeated-Measures ANOVA ......................................................................................................................147

6.7. Summary of results and revisit hypotheses ..................................................................................................148

CHAPTER 7: DISCUSSION ........................................................................................................................................151

7.1. Overview ...........................................................................................................................................151

7.2. Discussion ...........................................................................................................................................152

7.2.1. Processing of inferences versus non-inferences ......................................................................................152

7.2.1.1. Non-inferences ......................................................................................................................................152

7.2.1.2. Automatic versus controlled processing ..............................................................................................158

7.2.1.3. Summary ...........................................................................................................................................160

7.2.2. Reliance on cognitive resources for inference processing ........................................................................161

7.2.2.1. Correlational analyses and interpretation from L1 study ..................................................................161

7.2.2.2. Correlational analyses and interpretation from L2 study ..................................................................167

7.2.2.3. Summary ...........................................................................................................................................169

7.2.3. Processing of bridging inferences versus predictive inferences .............................................................169

7.2.3.1. Review of L1 study results for bridging inferences ..............................................................................171

7.2.3.2. Review of L2 study results for bridging inferences ..............................................................................171

7.2.3.3. Review of L1 study results for predictive inferences ............................................................................172

7.2.3.4. Review of L2 study results for predictive inferences ............................................................................172

7.2.3.5. Faster with distraction? .........................................................................................................................173

7.2.4. A unified model of inference processing ................................................................................................174

7.2.5. L1 versus L2 processing .........................................................................................................................180

7.3. Limitations and Future Directions .............................................................................................................182
Appendix F1: Stand-Alone N-Back Task (by condition and scenario type)..................236

Appendix F2: Stand-Alone VPT (by condition and scenario type) .........................237

Appendix G: L2 Statistical Analyses Without Possible Outliers .........................238

References.............................................................................................................248
# Table of Figures

Figure 2.1 Kintsch’s (1998) Levels of Processing Account .......................................................... 22

Figure 3.1 Baddeley’s Model of Working Memory ......................................................................... 37

Table 4.1. L2 language background information and proficiency ................................................. 57

Table 4.2 Sample stimuli, inferences and verification questions .................................................. 60

Table 4.3 Final list of male and female names used in the verbal stimuli ..................................... 66

Figure 4.1. Trial sequence of the reaction-time pilot task ............................................................. 71

Table 4.4. Breakdown of eliminated two-sentence scenarios based on reaction-time sentence pilot ........................................................................................................................................ 73

Table 4.5 Final Stimuli mean reaction times (in ms) by scenario type and block ......................... 74

Table 4.6. Stand-Alone Individual Difference Tasks ..................................................................... 78

Figure 4.2. Congruent and Incongruent Samples of The Visual Pattern Task ......................... 82

Figure 4.3. Sequence for a correct trial in the Visual Patterns task ............................................. 83

Figure 4.4. Visual depiction of the N-back principle for participants ........................................... 85

Figure 4.5. Trial sequence of the N-back task ................................................................................ 86

Table 4.7. Correlations (using Spearman’s rho) of stand-alone and distractor tasks .................... 89

Table 4.8. Experimental Design by Condition and Inference Type ........................................... 92

Figure 4.6. Visual events of a single trial in the control condition ................................................. 94

Figure 4.7. Visual events of a single trial in the STM condition (with the visual patterns task) .... 94

Figure 4.8. Visual events of a single trial in the CC condition (with the N-Back task) .................. 95

Figure 4.9. Overview of research procedure .................................................................................. 97

Table 5.1. Descriptive statistics for the raw accuracy scores in the experimental tasks for the L1 group ........................................................................................................................................ 102

Table 5.2. Goodness-of-fit tests to check for normality of distributions on accuracy of tasks for L1 group ........................................................................................................................................ 103
Figure 5.1. Boxplots of reaction times (in ms) of cognitive tasks for L1 group ........................................106
Figure 5.2. Boxplots of reaction times (in ms) of self-paced reading tasks for L1 group .................................106
Table 5.3. Descriptive statistics for the reaction times (in milliseconds) in the experimental tasks for the L1 group .................................................................................................................................................. 108
Table 5.4. Goodness-of-fit tests to check for normality of distributions on RT of tasks for L1 group........................................................................................................................................................................ 109
Table 5.5. Descriptive statistics for RQ1 for the L1 group .................................................................................. 111
Table 5.6. Mauchly’s test of sphericity assumptions for Research Question 1 ................................................. 111
Table 5.7. RM ANOVA testing for scenario type in L1 group .......................................................................... 112
Table 5.8. Descriptive statistics and tests of homogeneity for each ID measure and reading task by inference and condition ...................................................................................................................................... 115
Table 5.9. Skewness and kurtosis ratios for each measure (overall, L1 & L2) .................................................. 116
Table 5.10. Correlations for the L1 group ........................................................................................................ 117
Table 5.11. Predictions for RT in different conditions by inference type for RQ3 ........................................... 120
Table 5.12. Descriptive statistics for bridging inferences in different conditions ........................................ 121
Table 5.13. Mauchly’s test of sphericity assumptions for bridging inferences .............................................. 121
Table 5.14. RM ANOVA testing for bridging inferences by condition in the L1 group ................................. 122
Table 5.15. Descriptive statistics for predictive inferences in different conditions ................................... 123
Table 5.16. Mauchly’s test of sphericity assumptions for predictive inferences in L1 group ...................... 123
Table 5.17. RM ANOVA testing for predictive inferences by condition in the L1 group ......................... 124
Table 6.1. Descriptive statistics for the raw accuracy scores in the experimental tasks for the L2 group .................................................................................................................................................. 130
Table 6.2. Goodness-of-fit tests to check for normality of distributions on raw accuracy of tasks for L2 group .............................................................................................................................................. 131
Figure 6.1. Boxplots of reaction times (in ms) of cognitive tasks for L2 group ............................................. 134
Figure 6.2. Boxplots of reaction times (in ms) by condition of self-paced reading task for L2 group. .................................................................134

Table 6.3. Descriptive statistics for the reaction times (in milliseconds) in the experimental tasks for the L2 group. ..................................................................................................................136

Table 6.4. Goodness-of-fit tests to check for normality of distributions on RT of tasks for L2 group. .................................................................................................................................137

Table 6.5. Descriptive statistics for RQ1. .................................................................................................................................139

Table 6.6. Mauchly’s test of sphericity assumptions for RQ1. .............................................................................................139

Table 6.7. RM ANOVA testing for scenario type in L2 group. .............................................................................................140

Table 6.9. Correlations for the L2 group by scenario type (condition) and stand-alone cognitive measures. .............................................................................................142

Table 6.8. Predictions for L2 RT in different conditions by inference type for RQ3. ............................................................144

Table 6.10. Descriptive statistics for bridging inferences in different conditions for L2 group. ........................................145

Table 6.11. Mauchly’s test of sphericity assumptions for bridging inferences. ..............................................................145

Table 6.12. RM ANOVA testing for bridging inferences by condition in the L2 group. ..................................................146

Table 6.13. Descriptive statistics for predictive inferences in different conditions. ..............................................................147

Table 6.14. Mauchly’s test of sphericity assumptions for predictive inferences in L2 group. ..........................................147

Table 6.15. RM ANOVA testing for predictive inferences by condition in the L2 group. ..................................................148

Table 7.2. Individual reaction times (in ms) and cognitive control scores (percentage in parentheses) for L2 participants who were identified as slower to process non-inferences. .............................................................................................156

Table 7.1. Properties of automatic processing in resource account and memory-based accounts .................................................................................................................................159

Table 7.3. Correlations from RQ2 in L1 study represented by their positive or negative relationships. .................................................................................................................................163
Table 7.4. Data patterns of RT by inference type and condition for L1 and L2 groups. The text in red denotes divergences from the hypotheses. .................................................................170

Figure 7.1 Updated Inference Model (using Kintsch’s (1998) Levels of Processing Account and Barnes & Jones (2000) concept of executive attention). .................................................................179
CHAPTER 1: INTRODUCTION

1.1. Background

We rarely think about the process of meaning generation because it happens quite automatically. Inferencing is one of the main processes used to connect information that is not explicitly provided in a text, either in written or spoken discourse (Ariel, 2010). The power of inference can be seen in a conversation from an episode of *the Big Bang Theory*, a popular television show. Two scientists, Sheldon and Raj, are discussing a proof on the white board. Sheldon, who considers himself above making mistakes, had previously insisted that the proof Raj had written was incorrect. In the conversation below, Sheldon reconsiders his assessment.

(1)

Sheldon Cooper: I took another look at the board, and I realized you were right.
Raj Koothrappali: So you were wrong.
Sheldon Cooper: I'm not saying that.
Raj Koothrappali: That's the only logical inference.
Sheldon Cooper: I'm still not saying it (Lorre & Prady, 2009).

It is humorous when Sheldon says “I’m still not saying it,” because Raj, and by extension the television audience, come to the conclusion that Sheldon has essentially admitted to being wrong. However, not only does Sheldon never directly state that he was wrong, he also vehemently denies it. To understand language, we fill in information that is not explicitly given to us. In other words, we form an inference. In the exchange above, Raj, and the television audience, infer that Sheldon has admitted to being wrong. We need inferencing to communicate both efficiently and effectively with each other across all levels and all forms of communication. By studying how the inference process works, we can better understand language comprehension in general. (1) only shows one type of inference, a logical inference, that Raj is right and therefore, Sheldon is wrong. However,
there are several kinds of linguistic and pragmatic inferences that go beyond the scope of a logical inference in terms of complexity and purpose. This dissertation focused on an aspect of reading comprehension, specifically on the process of inference generation in text.

Two parallel studies were carried out using the same tasks and stimuli, a study with English first (L1) language adult readers and a study with English second (L2) language adult readers. First, these studies explored whether inferences are processed differently than non-inferences. Second, these studies tried to determine which cognitive resources (or individual differences) were recruited during the processing of bridging and predictive inferences. Third, they looked at two types of inferences, bridging and predictive, to see whether or not these are processed differently from one another or not. Finally, the L1 and L2 studies were compared to see if inference processing differs dependent on a reader’s language background.

1.2. Theoretical and Practical Applications of Research

In large part, this dissertation attempted to understand the theoretical underpinnings of what happens in the mind when readers process inferences. Toward this end, these two studies connect different areas of research, including Linguistics, Psycholinguistics, and Cognitive Science. By testing and updating existing models of reading comprehension, we can uncover what happens during the comprehension process in adult readers. However, inference processing is not purely a theoretical endeavor and has larger implications for educational interventions and language testing. These two studies look at inference generation at the highest level of processing to see what happens in adult readers. This research ultimately provides a look at normative adult processing
that can later be applied to the developmental stages of child reading acquisition, including attention deficit/hyperactivity disorders (ADHD) and autism spectrum disorder (ASD) research (Flores & Ganz, 2007) and individual differences research (Johnston, Barnes, & Desrochers, 2008). Furthermore, the present dissertation has implications for language testing. Educators are required to assess students in five areas that are considered to contribute to successful reading ability, including comprehension, vocabulary, fluency, phonemic awareness, and phonics (U.S. Department of Education, 2002). Inferencing, as a measure of comprehension, is a regularly tested skill in the informal reading inventories (e.g., Analytical Reading Inventory (ARI): Woods & Moe, 2007; Bader Reading and Language Inventory (BRLI): Bader, 2005; Basic Reading Inventory (BRI): Johns, 2005; Classroom Reading Inventory (CRI-SW): Silvaroli & Wheelock, 2004; Comprehensive Reading Inventory (CRI-CFC): Cooter, Flynt, & Cooter, 2007; Informal Reading Inventory (CRI-BR): Burns & Roe, 2007; Qualitative Reading Inventory-4 (QRI-4): Leslie & Caldwell, 2006; Neale Analysis of Reading Ability (NARA II): Neale, 1989; The Critical Reading Inventory (CRI-2): Applegate, Quinn, & Applegate, 2008; the Wechsler Objective Reading Dimensions Test of Reading Comprehension (WORD): Wechsler, 1990), yet it is not well understood. Despite the fact that inference generation is considered a critical higher order skill in reading comprehension assessments, standardized tests inconsistently measure this linguistic skill. Bowyer-Crane and Snowling (2005) compared two reading comprehension tests (The Neale Analysis of Reading Ability (NARA II): Neale, 1989 and the Wechsler Objective Reading Dimensions Test of Reading Comprehension (WORD): Wechsler, 1990) and found that they tap into different higher order inference skills. The NARA relies on
inferences that depend on world knowledge (otherwise called an elaborative inference) while the WORD relies more on inferences that connect literal information provided in the text (one example of which is a bridging inference). There are significant implications for reading level placement if inference processing varies depending on inference type, especially if these inferences are processed differently from one another.

The research questions asked in this dissertation can help bridge the gap between linguistic theory and practice by examining the process of inference generation at the text level.

1.3. Useful Terminology

Terms relating to the stimuli and study design in the present dissertation will be reviewed in this next section to provide clear definitions. I will first discuss the distinction between non-inferences and inferences and the two types of inferences examined in the present dissertation. I will also make a justification for why I considered the scenarios to be at the text-level. Finally, I will review the methods used, including what is meant by a reaction time experiment, a cumulative self-paced reading task, and a distraction paradigm.

1.3.1. Non-Inference versus Inference. Non-inferences and inferences are thought to rely on different types of processing: non-inferences are considered linguistic while inferences are considered pragmatic. Understanding a non-inference requires that a listener or a reader connect different pieces of information that are explicitly provided in the text. By decoding the grammatical and semantic information, the listener or reader gains access to the meaning of the sentence. In contrast, in order to comprehend an inference, a listener or a reader must connect information that is not explicitly provided to
him or her. By filling in the unstated (and intended) information, meaning is created. Inference building is, thus, considered a pragmatic process (Ariel, 2010).

1.3.2. Types of Inferences. The present dissertation explores how two types of inferences are processed, bridging inferences and predictive inferences. Below are abridged definitions of both types to help clarify the scope and breadth of the study, however more comprehensive definitions can be found in Chapter 2.

Bridging inferences require readers to connect two distinct elements in order to correctly interpret the discourse (Clark, 1975). An example of two sentences that result in a bridging inference with anaphora can be seen in (2a-2b).

(2a) Sally ate the entire cake.
(2b) She felt sick all night.
Inference: She is Sally.

In the comprehension process, readers form a backward link between the pronoun, she, and the referent.

Predictive inferences require readers, on the other hand, to use current information in the text to form a hypothesis about upcoming possible events (McKoon & Ratcliff, 1986). To form these inferences, readers need to use real-world knowledge. See example (3a-3b).

(3a) The ice in the middle of the pond was thin.
(3b) Ella began skating across the pond.
Inference: The ice broke, and Ella fell in the pond.

The reader typically anticipates several possible events, and therefore holds multiple interpretations in mind during comprehension until an outcome is revealed.

In theory, these two inferences require different forms of processing. In the case of bridging inferences, readers must hold onto information across sentence boundaries
and connect it back to previously presented entities. Predictive inferences, alternatively, require that readers generate multiple possible future events, which they hold in their minds until the one prediction is confirmed.

1.3.3. Two-Sentence Scenarios. The stimuli used in the present dissertation were considered to be at the text level. The participants in the two studies had to connect information across sentence boundaries when reading the two-sentence scenarios. It is important to mention that the idea of a sentence boundary is an artificial construct, especially in the context of spoken discourse. However, the use of a period (.) in a written text provides readers with a tangible separation of propositions, making them quite distinct from one another in the reader’s mind. Since participants were presented with scenarios that were comprised of two sentences each, they were forced to mentally ‘juggle’ multiple propositions.

The stimuli were comprised of two sentence scenarios that required that participants form a bridging inference (i.e., a pronoun and a referent are linked in connected discourse), a predictive inference (i.e., the anticipation of a future unknown event) or no inference (i.e., the explicit information comprises the meaning of the sentence without filling in unspecified information).

1.3.4. Reaction Time Experiment. Reaction time experiments (see Keele, 1986 for an overview of reaction-time methodology, a method prevalent in linguistic and psycholinguistic research), administered on the computer, present participants with stimuli and measure their reaction time (in milliseconds) based on how quickly or slowly a button is depressed in reaction to these stimuli. These two studies measured participants’ reading times for the two-sentence scenarios and correct responses for various individual
difference tasks. The use of the term ‘reaction time’ and ‘reading time’ are used interchangeably in this dissertation and refer to the same construct, the length of time to read a given text.

1.3.5. Cumulative Self-Paced Reading Task. In the self-paced reading task, a participant read a two-sentence scenario and completed a verification question about that scenario. The participant was presented with the first sentence of the scenario in the center of the screen. Once he or she depressed a button, the second sentence appeared. Since the task was cumulative, both sentences remained on the screen. Then the participant depressed a button and was directed to the next screen. The reaction time was measured for the first and second sentence of the scenarios.

1.3.6. Visual Distraction Paradigm. The cumulative self-paced reading task was carried out in a visual distraction paradigm, in which participants read the scenarios and answered the verification questions while completing a secondary distractor task. Several studies have used an auditory-visual distraction paradigm to investigate the role of individual differences in processing (Escera, Alho, Winkler, & Naatanen, 1998; Escera & Corral, 2007; Lv et al., 2010; SanMiguel et al., 2008). Few studies have used the visual modality both as the distraction and the target. The distractor tasks were visually presented to the participants, and the presence or the absence of the distractor task comprised the three experimental conditions: there was no distraction (the control condition); there was a distractor task that interfered with phonological short-term memory, i.e., the ability to temporary replay information in the mind (the short-term memory condition); or there was a distractor task that interfered with cognitive control, i.e., the regulation of attention (the cognitive control condition).
1.4. Overview of L1 Study

In the L1 study, participants completed a cumulative self-paced reading task in a distraction paradigm. Participants read scenarios in which they had to form a bridging inference, a predictive inference or no inference in order to test how these different scenarios were processed. The presence or absence of a secondary distractor task served as the basis for the experimental conditions in order to test the theory that bridging inferences and predictive inferences have dissociative reliance on distinct cognitive resources. The conditions were a control condition, a short-term memory condition, or a cognitive control condition. Additionally, the reaction time (in milliseconds) for reading could be measured, which provides useful information about text processing. In this way, it was possible to see whether inferences are processed differently than non-inferences by comparing a participant’s reading time for the scenarios with no inferences to that of the scenarios with inferences in the control condition. It was also possible to examine the role of the cognitive resources in processing inferences and non-inferences. In a separate session, participants’ cognitive control and phonological short-term memory were independently assessed in order to run correlational analyses to see if there was a relationship between the performance on the stand-alone individual difference measures and the reading times for the different scenarios across the various conditions. Finally, these studies tried to determine internal resource was being recruited to process bridging and predictive inferences. If participants were markedly slower to process a bridging inference in the short-term memory condition compared to the cognitive control condition, this would indicate that a bridging inference relies on short-term memory. Furthermore, if participants were markedly slower to process a predictive inference in the
cognitive control condition compared to the short-term memory condition, this would indicate that a bridging inference relies on cognitive control.

The role of cognitive resources in inference processing is a fruitful area of research, as few studies have looked at how individual differences affect inference processing, beyond a strict correlational account. In addition, cognitive resources appear to be a confounding variable in many processing studies; cognitive resources likely affect how text is processed and understood, as they govern access to attentional and memory-based resources. To address this gap, these two studies investigate how individual differences contribute to the processing of two inference types, specifically bridging and predictive inferences. Two influential cognitive models of inference comprehension, the constructionist model (Graesser, Singer, & Trabasso, 1994) and the construction-integration model (Kintsch, 1998), predict that bridging inferences and predictive inferences are made at different levels of processing. Bridging inferences are thought to be automatic, made in the early stages of the bottom-up process. Predictive inferences, on the other hand, are considered to be formed at the highest level of processing. While there is preliminary evidence that these inference types may call on distinctive cognitive resources (Burkhardt, 2006, Jin et al., 2009), there is not yet a clear picture of inference building in reading comprehension. It appears that the resources that are being recruited may vary depending on inference type, but results from brain-based studies are subject to multiple interpretations.

1.5. Overview of L2 Study

The L2 study had the same tasks, stimuli, and procedure to the L1 study; the only exception was the participant population. The participants were highly proficient L2
speakers of English (L1 Spanish). The construct of proficiency is discussed in the 
Methods Chapter (see Section 4.3.2.1).

It is not clear from the current L2 sentence-processing research what drives L2 
text processing, and there is relatively little research in this area to explore this gap. 
Specifically, it is not known whether or not L2 readers have access to the necessary 
linguistic structures and processing routines needed in their L2 (Horiba, 1996) or if 
proficiency level is the confounding variable in readers’ access to the linguistic structures 
in the L2 (Cohen, 1998; Guerrero, 2005; Leontiev, 1981). In addition, individual 
differences seem to mediate prior knowledge, past experience and context effects in the 
L2 (Koda, 2005). For these reasons, a parallel study with L2 readers was carried out.

1.6. Preview of Upcoming Chapters

In Chapter 2, the principle research areas in inference processing will be explored, 
including more theoretical accounts of inference and cognitive models of inference 
processing. In addition, since the linguistic stimuli are comprised of bridging and 
predictive inferences, both inference types will be described. Finally, L2 studies of 
inference processing will be presented to gain a picture of how non-native readers 
process and understand inference in order to see the current gaps in the L2 processing 

studies.

Chapter 3 explores the role of individual differences (i.e., cognitive resources) in 
inference processing. First, generally accepted cognitive models for working memory and 
attentional control will be investigated, as memory-based processing is required for 
inference generation. Within these working memory and attentional control frameworks, 
the predictions from the cognitive models in Chapter 2 will be expanded upon and related
to specific individual difference constructs. At the end of Chapter 3, the interaction between individual differences and inference processing will be reviewed in both L1 and L2 behavioral and neurolinguistic studies.

Chapter 4 reviews the methods used in the studies, including the research questions and hypotheses. Relevant information will be reviewed about the participants, linguistic stimuli, pretests and pilots, materials, and data collection.

Chapter 5 presents the results of the L1 study, including the assumptions for the parametric statistics used, the description of the data (including checking for normality), the descriptive statistics, and the inferential statistics. Using the same structure, Chapter 6 presents the results of the L2 study. While some of the information in Chapter 6 may feel repetitive, it is presented as a stand-alone chapter in the case that it is read independently of the L1 study.

Chapter 7 considers the implications of the results for the L1 and the L2 studies separately. Additionally, the L1 and L2 studies are compared to each other to determine trends in processing based on language background. Finally, the limitations and future directions will be discussed. At the end of the dissertation, a general conclusion will be presented which positions the two studies in the context of the literature and reports the contribution to the fields of Applied Linguistics and Psycholinguistics.
CHAPTER 2: INFERENCE BUILDING

While the process of inference building is critical for successful language comprehension, it is largely under-researched. Much of the research agenda that has occupied the field of psycholinguistics has focused on syntactic processing in individual phrases, without ever venturing across sentence boundaries. In response to the volume of studies that focus on syntactic (i.e., semantic) processing to date, Singer (2007) calls for more studies that investigate inference processing. He considers any progress so far to be preliminary, and as such it “should be viewed as setting the stage for further advances in understanding discourse-inference processing, rather than as constituting definitive solutions to these scientific problems” (Singer, 2007, p. 354). It is, of course, important to uncover the mechanisms that allow us to process an individual sentence, for they serve as a necessary foundation to push the field of language processing forward; however we cannot stop there. Understanding sentences in isolation only shows a small part of the picture. Discourse is exceedingly complex – it is comprised of multiple utterances that have related propositions, ideas and interwoven contexts. To this end, Kintsch and Rawson (2007) emphasize the complexity of text comprehension, setting a research course to investigate the various subprocesses involved.

This chapter will review the major research areas in inference processing, including theories of inference and cognitive models of inference processing. Since the linguistic stimuli used in this dissertation require that participants form either bridging inferences or predictive inferences, a finer-grained taxonomy of inference type will be presented. Finally, L2 studies of inference processing will be reviewed in order to show how non-native readers process inference.
2.1. Theories of Inference Generation

There are several theories that seek to explain the assumptions speakers and hearers follow when they engage in verbal communication in order to understand each other. These theories will not be explored in detail (e.g., Grice’s Cooperative Principle and four maxims: Grice, 1975, 1989; Neo-Gricean Pragmatics: Horn, 1984; Levinson, 2000; and Relevance Theory: Sperber & Wilson, 1986/1995), as they are not the focus of this dissertation. Instead, it is worth pointing out there is a growing consensus that comprehension requires more than a literal decoding of truth-conditional meaning, paving the way for an exploration of how a hearer understands a speaker’s intentions and the role that context plays in this process. The distinction between what is grammatically encoded (i.e., code) and what is intended (i.e., inference) has proven essential for understanding comprehension, and the act of trying to bridge the divide between code and inference has served as the basis of most contemporary pragmatic theories. To be explicit, inferencing is principally seen as a pragmatic process (see Ariel, 2010, as well as Grice, 1975 for exceptions such as conventional implicature) in which a listener draws conclusions, whether from background knowledge, context or prior experience, about information that is not explicitly given in a text.

Two conditions are generally true for the hearer when understanding an inference. First, the hearer generates meaning that is not overtly specified in the utterance. Second, the hearer does not have access to the speaker’s intentions and an inference can be wrongly perceived. These points are demonstrated in example (1a)-(1c).

(1a) The apple was cut in half.
(1b) The knife lay beside it on the cutting board.
Inference: The knife was used to cut the apple in half.
(1c) *Then the chef wiped off the saw that she used to cut the apple in half.*

Using his or her background knowledge about common kitchen instruments used to cut apples, the hearer likely makes the inference that the knife was used to cut the apple in half, even though this information was never explicitly stated in (1a)-(1b). These associative leaps are ubiquitous in successful comprehension, and they generally lead to accurate comprehension. Furthermore, these inferences are made so effortlessly that hearers often do not recognize that they have filled in unstated information, but rather experience that they had direct access to what was meant by the speaker. (1c) reveals that the saw, however uncommonly found in a kitchen setting, was the instrument to cut the apple in half, demonstrating that the assumed link is inferred and not explicitly stated.

This unexpected reversal of inference interpretation can occur with inferences that rely on the interaction of background knowledge and context, as in the case of (1a)-(1c), or with inferences that connect propositions that are provided in the text, in the case of anaphora (i.e., the connection of a pronoun and a referent). To illustrate how an inference that operates at the linguistic level can be misunderstood, see example (2), a monologue from the *Tonight Show with Jay Leno*. Note that (2) purposefully misdirects the listener in regard to the referent-anaphor mapping.

(2) *Earlier today, the president of Haiti was at the White House to meet with President Obama. He said the people of his country need jobs, they need places to live, and they need health care. And then the president of Haiti spoke.*

(Leno, 2010)

As with a lot of humor, this unexpected reversal of expectation is what makes the monologue funny (see Grice, 1975, 1989 for linguistic theories on flouted implicatures; see Attardo, 1997, 2008 for the Incongruity-Resolution Model of humor). The set-up of the discourse leads the audience to incorrectly assign the anaphora *he* to the referent *the*
President of Haiti, when, in fact, the referent-anaphora link is between he and President Obama. The presentation of events makes the President of Haiti’s visit salient, making him the seemingly relevant referent to which all the propositions ascribed to he are mapped. Much of the time, hearers and speakers strive for maximal understanding. While inferences are often canceled in everyday communication, the stimuli used in these two studies investigate a more straightforward process of inference generation, where speakers and hearers achieve comprehension without a reversal of meaning. Both (1a)-(1c) and (2) show that inferences are generated to either infer information that is not explicitly provided or to make connections that are not overtly specified.

In summary, inference is thought to be a pragmatic process that requires hearers to connect information that is not expressly stated in the utterance.

2.2. Inference Types

The two major types explored in these studies are bridging inferences and predictive inferences. These differ in terms of the type of processing required to interpret discourse in a native-like manner.

2.2.1. Bridging Inferences. According to Clark (1975), bridging inferences (also called associative determiner phrases [DPs]) require readers to form a dependency relation between two elements for a coherent interpretation of the discourse. While there are many kinds of bridging inferences, the focus of this dissertation will be on a more constrained type, anaphora, where a pronoun (or anaphor) is linked back to a previously mentioned referent (or antecedent). Any mention to a bridging inference, from this point on, refers to formation of a dependency relation in discourse, unless otherwise specified.
An example (also seen in Chapter 1) of two sentences that result in a bridging inference with anaphora can be seen in (3a-3b).

(3a) Sally ate the entire cake.
(3b) She felt sick all night.
Inference: She is Sally.

Critically, to understand the discourse, a backward link must be formed between the pronoun, she, and the referent, Sally; she is a newly introduced entity whereas Sally is the entity already present in the mind of the reader. The bridging inference works on several levels. First, the reader must recognize that the pronoun she is the same entity as the referent Sally. These entities are dependent upon each other and are intrinsically linked in the discourse, based on the discourse context. Second, the reader is required to map the set of information ascribed to one entity onto the other and vice versa. The entity in the larger mental model of the reader, Sally, is made salient when it is introduced, and it serves as a holding place for the reader to metaphorically fasten their attention and relate back to in order to interpret the relevant information. Yet she is also made salient in the same manner. In this way, the information provided about Sally (i.e., ate the entire cake) is mapped onto she, and the information provided about she (i.e., felt sick all night) is mapped onto Sally.

2.2.2. Predictive Inferences. Predictive inferences, a type of elaborative inference, on the other hand, do not rely on dependency relations, nor do they establish coherence between a just-read entity and a previous clause. Predictive inferences forecast upcoming events with the use of current information in the text (McKoon & Ratcliff, 1986). Essentially readers anticipate upcoming events, and this prediction facilitates
faster comprehension of the following sentence due to context (Kutas and Federmeier, 2011). See example (4), also seen in Chapter 1.

(4a) *The ice in the middle of the pond was thin.*  
(4b) *Ella began skating across the pond.*  
Inference: *The ice broke, and Ella fell in the pond.*

Using real-world knowledge, that thin ice often breaks under the weight of a person, the reader anticipates that the ice would break under Ella’s weight, and she would likely fall in. However, the reader does not typically anticipate only one event, and this requires that the reader hold several interpretations in mind until the action is resolved.

These two inference types are of particular interest because, in theory, they require different forms of processing, as discussed above. In the case of bridging inferences, readers are required to hold information across sentence boundaries and link it back to previously presented entities. Predictive inferences, on the other hand, prompt readers to generate several anticipated future events, which are held in mind until the most likely outcome is revealed.

Cognitive models have attempted to identify the subcomponents of the inference process when an inference is successfully (or unsuccessfully) generated. The next section outlines two widely researched cognitive models of inference that have been influential in language processing research.

2.3. Models of inference Processing

In parallel to the linguistic focus on developing a theory of inference, researchers began to systematically develop models of language comprehension (Singer, 2007). The goal was to understand how the mind processes an inference and to identify the relevant resources recruited during that process. Whether from a domain-general or domain-
specific theoretical approach, there is shared agreement that inference involves processing in context. Furthermore, these models tried to define inference building from a cognitive perspective, breaking the process into its component stages and processes.

Cognitive models are useful tools when investigating complex phenomena, as they predict specific outcomes that are falsifiable. Busemeyer and Diederich (2010) put forth a cyclical and iterative process for cognitive modeling. The first steps include specifying the model, identifying the assumptions, and generating the methods. Once the model, its assumptions and the appropriate tools are identified, the model is then tested. After this, the model’s predictive ability is identified against other models. Ultimately, the model is refined, essentially resuming the cycle. This way it is possible to see whether the predictions from each model are confirmed or repudiated. For this reason, two models were chosen as the theoretical framework for these studies using criteria outlined in Leighton and Gierl (2011). First, the processes identified in each of the models are clearly defined, while still being theoretically distinguishable from each other. Second, the models are falsifiable and supported by substantial empirical support. Finally, these two models have had considerable influence in the development of subsequent reading comprehension models.

The two comprehension models chosen as the theoretical foundation are the construction-integration (CI) model (Kintsch, 1998) and the constructionist model (Graesser, Singer, & Trabasso, 1994). The most important distinction between these two models is the mechanism by which information is filled in, or a new representation is created; it can be attributed to maintaining coherence within propositional theories (see Crothers, 1979; Kintsch, 1998; Kintsch & van Dijk, 1978; Rickheit, Schnitz, & Strohner,
1985) in the CI model or updating mental models through a series of questions posed to the text (see Graesser et al., 1994; Warren, Nicholas, & Trabasso, 1979) in the constructionist model.

2.3.1. Construction-Integration Models. Kintsch’s (1998) model fulfills the three main criteria outlined by Leighton and Gierl (2010) as the basis for a theoretical framework and will, therefore, be used as the representative CI model in this dissertation. Not only does Kintsch’s (1998) model provide a navigable road map for looking at the different levels of processing at the text level, it also generates clear predictions. Finally, the assumptions of the model have been empirically tested.

Kintsch’s (1998) model breaks the inference process into processing steps. This can be understood by first looking at a simplified model from Rickheit et al. (1985) based in connectionist theory. Rickheit et al. (1985) provide a practical CI model of inference in discourse comprehension in which the concrete processing steps are generally automatic. According to their CI-approach, comprehension is fueled by what researchers called a bottom-up process; as such knowledge stored in long-term memory is activated (the construction phase) and this activated knowledge is then integrated into working memory (the integration phase). It is important to note that the CI definition of bottom-up differs from that found in cognitive science literature. The claim from CI researchers that comprehension is a bottom-up process depends on the directionality of the order of events (see Figure 2.1 for the levels and directionality in the CI model) – it begins at the word-level (the “bottom”), progresses to the sentence-level and then finally reaches the context level (the “top”).
It is important to identify how the meaning of a text\(^1\) is internally represented, often termed the mental representation in CI models. Craik and Lockhart (1972) advocate for the Levels-of-Processing approach, in which the different levels are connected with memory traces throughout comprehension. It is generally agreed that the stages of processing are comprised of four common levels, though each model details this slightly differently. The levels of processing approach first describes the lens through which information is understood and then shows the progressions from the explicit content of information to the role of context. The first level contains conceptual units or cognitive schemata through which information is gathered. The second level focuses on semantic information, formally referred to as the propositional units that are generated through argument-predicate structure. The third level has mental models or non-semantic information about the world. The final level is composed of super-structure with a conventionalized form that is broken into objectively universal categories. As these memory traces build up and merge with each other, they provide the foundation for the representation(s) in the model itself.

A concrete example of the levels of processing account can be seen in Rickheit et al.’s (1985) model; they propose that discourse comprehension could be divided into several distinguishable subprocesses (pp. 7-8). The first subprocess is decoding (i.e., moving from grapheme level representation to word concepts). The second subprocess is encoding (i.e., moving from word concepts to phonetic level representation). The final subprocess is inference. This allows for an individual process (within a larger complex of

\(^1\) Text is herein used to refer to both spoken and written discourse throughout the dissertation, unless otherwise specified.
processes) of discourse comprehension to be broken down “according to its functions for the entire system” (Rickheit et al., 1985, p. 8). Not only are the independent components identified, but the role of context is also explored. In this model old information (labeled “A”) in a particular context (labeled “C”) is generated into new information (labeled “B”). This definition allows for identification of the psychological representations of A and B, the process by which B is inferred from A, and the role of C in that process.

Building on this interactional definition, Kintsch and Rawson (2007) also maintain that it is not enough to look at the sum of all the processes involved in text comprehension, but also to consider the individual parts of what “arises from their coordinated operations as a system” (pp. 226) and the interactions therein. Kintsch (1998) also advocates processing at different levels, ranging from word recognition to incorporation of a hearer’s mental models (see Figure 2.1). His model proposes a more detailed intermediate process of building up text meaning than Rickheit et al. (1985). He also brings in the role of working memory (see Baddeley, 2010 for an overview) and long-term memory, depending on the level of processing. Kintsch (1998) offers the following four processing levels (see Figure 2.1 for an illustration of the different numbered levels): (i) the linguistic (or surface) level (i.e., words and phrases), in which the reader decodes the graphic symbol through word recognition and parsing (e.g., role assignments); (ii) the semantic analysis for the meaning of a text (i.e., word meanings form idea units or propositions); (iii) textbase (i.e., the explicit meaning of a text; see Figure 2.1 for more detail about micro- and macrostructure); and (iv) the situation model (see Louwerse, 2002 for CI model at the situation level) or mental model of text integrating prior knowledge, objects, emotions and personal experience.
Importantly, comprehension is modeled with the understanding that readers go beyond information that is explicitly provided in the text.

Several studies have provided empirical evidence for Kintsch’s model, including the construction phase (Lorch, 1998; O’Brien & Albrecht, 1991), the integration process (Albrecht & O’Brien, 1993; O’Brien, Rizzella, Albrecht, & Halleran, 1998) and the interaction between the construction and integration phases (Garrod & Terras, 2000; Kintsch, 1988; Kolodner, 1983; Traxler, Sanford, Aked, & Moxey, 1997; Singer, 2006). Mainly studies have centered on the retrieval process of the construction phase and the selection of information in the integration process. Lorch (1998) suggests in his review of memory-based text processing that readers use the current proposition to retrieve antecedent discourse elements from memory, supporting the lower levels of processing in Kintsch’s model. O’Brien and Albrecht (1991) showed that when an item is successfully retrieved, it is also then verified with its antecedent, e.g., *striped animal* and *skunk*,

---

**Figure 2.1** Kintsch’s (1998) Levels of Processing Account.

- **(iv) Situation Model**
  - inferred meaning

- **(iii) Textbase**
  - explicit meaning

- **(ii) Semantic Level**
  - (a) the *microstructure of text* in which coherence relations are formed between propositions;
  - (b) macrostructure of global topics;
  - (c) micro- & macro-structure, i.e., *textbase* or “meaning of the text, as it is actually expressed by the text” (Kintsch & Rawson, 2007, p. 211) for shallow understanding.

- **(i) Linguistic Level**
  - word recognition or parsing
providing evidence for the semantic level of processing. Several researchers (Garrod & Terras, 2000; Kintsch, 1988; Kolodner, 1983; Traxler, Sanford, Aked, & Moxey, 1997; Singer, 2006) have shown that retrieval promotes an interaction between the textbase and the situation model as ideas enter the discourse context. In the integration stage, evidence has been found that once the information has been verified, the situation model is updated (Albrecht & O’Brien, 1993; O’Brien, Rizzella, Albrecht, & Halleran, 1998).

Despite mounting evidence in support of the CI model, it has been pointed out that it cannot unilaterally serve as a model of text processing (Lorch, 1998), as it fails to address how higher-level processes interact with the so-called bottom-up model, such as the effects of genre or strategy-driven processing. The constructionist model addresses some of these concerns and will be discussed as an alternative to the CI model.

**2.3.2. Constructionist Models.** Graesser et al. (1994), in contrast, proposed that strategies are a driving force for comprehension, such as reader goals, coherence and ultimately explanation (for an overview, see Graesser, 2007). Graesser and colleagues’ constructionist model is also a suitable theoretical framework as per Leighton and Gierl’s (2010) criteria and will serve as the representative constructionist model in this dissertation. The model’s comprehension process is clearly outlined with clear predictions that have been empirically tested. Finally, this strategy-based model has impacted the research agenda for sentence processing research.

Graesser distinguishes between online inferences, which are processed in real-time and offline inferences, which are made only with effortful reflection long after a stimulus has been removed; the constructionist model exclusively investigates online language processing. Graesser was concerned with how situation models are constructed
when comprehending narrative text. Similar to Kintsch’s model, the constructionist model operates within several levels of text processing and uses multiple stores of memory, but the role of strategy is made more central than other knowledge structures. Graesser (2007) defines a reading comprehension strategy as cognitive behavior that is carried out within a specific context (p. 6). For example, when a shopper in a supermarket line picks up a *People* magazine and starts to skim an article entitled *Colin Egglesfield Arrested for Disorderly Conduct, Criminal Damage* (Corriston, April 2, 2014), she directs her attention toward what exactly Colin Eggelsfield did to be arrested, ignoring potentially irrelevant information, such as his age or the show in which he stars. It is important to note that the constructionist model borders on strategy research and as such the majority of processing would occur at the level of controlled processing. However, a distinction can be made between automatic and controlled processes (see Stanovich & West, 1983) and a continuum has even been proposed between the two (see Fischler & Bloom, 1980). Putting this debate aside, if strategy is considered central to processing, then readers use their goals in a top-down process to drive their attentional resources toward specific content. To clarify, top-down refers to the directionality of informational processing – in the case of a constructionist approach, the strategic context dictates the flow of information from a goal to the specific information.

This model also relies heavily on the context of the information, and in this vein “[t]he coherence assumption states that readers attempt to construct meaning representations that are coherent at both local and global levels” (Graesser, 2007, p. 13). Local coherence occurs if a current explicit statement can be conceptually connected to a recently presented relevant proposition that resides in working memory (WM). Global
coherence, on the other hand, groups information into higher order chunks, in that the current explicit statement is linked to several sentences that have been previously presented which may or may not be in working memory (WM).

In contrast to the CI approach, the constructionist model considers strategy paramount, essentially directing any comprehension processing; in a CI model, reading strategy is just one of many knowledge structures that are activated from long-term memory. Readers then generate explanations as to why the text had particular events or actions. In this model, proficient readers ask why- and how-questions and generate plausible explanations with the end result of inference building.

Several studies have found empirical support for constructionism; readers take advantage of situational structures based in their own reader goals to track and connect information across text distance (Long, Golding, & Graesser, 1992; Long, Seely, & Oppy, 1996; Richards and Singer, 2001; Singer and Halldorson, 1996; Suh & Trabasso, 1993; van den Broek and Lorch, 1993). Yet there are some criticisms. Albrecht and Myers (1995) found counter-evidence for the constructionist approach in their reading study. Participants were presented with a written text with five sections (i.e., opening, goal, filler, target and conclusion) while their reading time latencies were measured. Longer reaction time latencies were operationalized as participants’ noticing of relevant information. In the opening section, the main character was introduced, as well as the situation. In the goal section, the main character was provided a goal (i.e., Mary needed to make an airline reservation tonight by midnight). The target sentence either satisfied the goal (i.e., Just after she confirmed her reservation, she got a call from her boss) or did not yet satisfy the goal (i.e., Mary got a call from her boss before she could book the
The position of the filler information varied in that it created more or less distance from the presentation of the goal and the target sentence. The target sentence (i.e., She was tired and decided to go to bed) resolved the unsatisfied condition or, in the case of the satisfied condition, provided more information. The assumption is that if participants were operating on a strategy-based approach, then the reader’s internal question (i.e., Did she succeed in booking the airplane reservation?) would make the target sentence more salient in the unsatisfied condition (i.e., Mary went to bed).

However, Albrecht and Myers (1995) found that participants noticed target information only when they were primed to do so with a contextualized phrase about Mary (i.e., Mary sat on the couch) just prior to the target sentence. Given that the constructionist model claims that since readers use a strategy-driven process to ask questions and generate explanations, longer latencies would be expected for all target sentences in the unsatisfied condition, regardless of whether participants were primed to notice or not. Another concern has to do with defining the stages in the constructionist approach (Singer, 2007).

When strategy mediates the reader’s attention, the top level of processing – situational structures – is accessed in a constructionist model. Unlike a textbase, comprised of explicit sentences that have been stitched together, defining situational structures is quite challenging. By and large, there is a body of evidence that seems to indicate that constructionist theory reasonably accounts for inference processing, especially in narrative text when reader goals are present.

To summarize, two influential models of inference processing are the CI model, in which processing advances from the word- to the context-level, and the Constructionist model, in which strategy largely drives processing through reader-internal questions and
answers. Both have garnered empirical support over the years, yet neither fully explains how language background mediates the processes that are required for comprehension. It is of interest how processing works in the L1 and the L2; despite the fact that cognitive models have attempted to describe the subprocesses in the generation of an inference, it is still not entirely clear whether this process is similar or different in L1 and L2 speakers.

2.4. L1 versus L2 processing

Since these two experiments examine reading, this section will mainly review empirical studies, which investigated inferential processing in L1 and L2 comprehension of written text. While the comprehension process is not entirely dissimilar across modalities, there are distinct differences in processing spoken and written text (see Song, 2008). Most recent studies that have looked at L2 text comprehension have almost exclusively focused on lexical inferencing, either in conjunction with proficiency (Zhang & Koda, 2012; Kaivanpanah & Soltani Moghaddam, 2012), background knowledge and proficiency (Pulido, 2009), knowledge sources (Riazi & Babaei, 2008), topic familiarity (Pulido, 2007), the interaction of strategy and word retention (Hu & Nassaji, 2012; Hamada, 2009; Hamada & Park, 2011; Nassaji, 2004) or incidental acquisition of vocabulary (Ehsanzadeh, 2012). While it is valuable to understand the processes by which L2 readers discern unknown words, discourse comprehension is the larger goal of any theory, model or explanation of L2 processing. There is a strong base of literature on the comprehension of discourse (i.e., across sentence boundaries), but little has been done since the 1990s, calling for more research.

It is generally assumed that L1 and L2 text comprehension differ, in that L2 discourse processing requires more effort and is a more complex process (Koda, 2005).
This difference in processing based on language background is reinforced in Zulu’s (2007) study that looked at reading processing in L1 and L2 university law students in South Africa. She found that L2 readers demonstrate lower analytical, critical and inferential reasoning skills. L2 readers seem to expend more effort, and, therefore, often exhibit slower processing.

The question remains whether L2 readers fundamentally process written text differently than L1 readers. A recent study from Schoenpflug & Klische (2010) investigated this very question; they looked at whether monolingual and bilingual children process text using bottom-up or top-down processing. One of their variables was the type of bilingualism, including balanced bilinguals (where the L1s are at relatively equal proficiency) and non-balanced bilinguals (where there is a dominant L1 and a weaker L1). Using Kintsch’s (1998, 2005) levels-of-processing account, they had children listen to a story and then repeat what they heard. If a child recalled what they termed ‘verbatim performance’, he or she was considered to have used bottom-up mechanisms to access the surface structure of the text. In contrast, if a child recalled the gist of the story, the child was thought to have used top-down mechanisms. They found that, for balanced and non-balanced bilinguals, L1 processing exploits bottom-up mechanisms, while L2 processing relies on top-down processing. This result would support a combined account of the two models reviewed, the CI model and the constructionist model, depending on factors such as language background, exposure and practice.

According to research into the CI model, L2 discourse processing, similar to L1, first requires a construction of the propositional textbase (i.e., the explicit meaning of the
text) through the graphic or phonetic information provided (Jenkins, Prior, Ricardo, Wainwright-Sharp, and Bialystok, 1993). A principal focus of L2 reading research has been on whether the L1 or the L2 is used in the construction of the mental model and hence impacts L2 linguistic sophistication of the propositional textbase. Jenkins et al. (1993) replicated Ehrlich and Johnson-Laird’s (1982) study of spatial reasoning, with an additional variable of language familiarity (or exposure). Using measures of classification (i.e., comparison of the mental model and a sample model) and statement recognition (i.e., ranking of similarity of statements used in task when presented post-experimentally) in a post recall task, Jenkins et al. (1993) asserted that only the propositional (i.e., textbase) level, and not the situation model, was possible in the L2. It is important to note this may have been related to proficiency in the L2, as well. Looking at how the textbase is built up, as well as the connections between the propositional units, Horiba, van den Broek and Fletcher (1993) investigated causal chain status of the propositions and story grammar categories in L1 and L2 English texts. While the L1 and L2 participants both demonstrated relational connections between the propositions, upon further analysis Horiba et al. (1993) found that L2 participants focus and rely on cohesion markers to understand unknown words and build up a mental model. These data suggest that L2 readers’ process of meaning construction does not extend past the textbase to reach the situation model. Yet the claim that L2 readers do not access the situation model has startling implications and severely limits what non-native speakers would possibly be able to understand. Without access to context, full comprehension is close to impossible. This research suggests that L2 proficiency is a possible intervening factor in inferential processing ability.
Another strand of research investigated text-coherence, the next level of comprehension in the CI model. Not only have L1 speakers had more exposure and experience with comprehension, but there may also be differences in how the devices used to bootstrap information (e.g., connectives) operate in the L1 versus the L2 (see Koda, 2005, pp. 139-140). The largely agreed-upon conclusion is that individual differences are a confounding variable in this type of study. Essentially the allocation of attentional and memory-based cognitive resources allows for greater or lesser comprehension in regard to co-referential ties (e.g., Demel 1994) and logical connectors (e.g., Goldman & Murray, 1992) as well as construction of the textbase (Horiba, 1996). Goldman and Murray (1992) investigated how L1 and L2 learners of English understood the following connectors: additive (e.g., and), causal (e.g., therefore), adversative (e.g., however) and sequential (e.g., then). Initial results of the cloze results showed that additive and clausal connectors had the highest accuracy rate, but upon further analysis, L2 participants showed evidence that they did not form the proper connections (i.e., relations) between sentences, with a pattern of over-application of clausal relations. The authors point out, though, that this could be due to proficiency level as opposed to processing mechanisms. Another study by Horiba (1996) compared L1 and L2 Japanese readers’ performance with high- and low-coherence texts using concurrent verbal reports across proficiency levels. L1 readers made more connections in the high-coherence texts, generating bridging and predictive inferences. In contrast, L2 readers did not process the two text types differently, with a closer analysis revealing that many of the lower-level processing was also quite delayed. One explanation is that the L2 participants in Horiba’s study may not have been proficient enough to complete the task. Another explanation
might be that L2 readers do not have full access to the situation model. The sum of studies seems to point to processing differences between the L1 and L2, but it is not clear whether proficiency or individual differences are confounding the results.

Finally, the situation model (i.e., the inferred meaning) relies on background knowledge, including formal (i.e., linguistic) and culturally specific knowledge. This reliance on external knowledge-based factors is supported in much of the L1 research of discourse comprehension. General internal factors that have been shown to play a role in L2 discourse comprehension are proficiency in the L2 (e.g., Ridgeway, 1997), the reader’s perception of text difficulty (e.g., Peretz & Shoham, 1990), the reader’s culture-specific knowledge (Abu-Rabia & Feuerverger, 1996; Alptekin, 2006; Kang, 1992; Steffensen, Goetz & Cheng, 1999), and contextually relevant information (e.g., Roller & Matambo, 1992). In addition, familiarity with the topic, referred to as domain knowledge, impacts L2 comprehension (e.g., Chen & Donin, 1997), as does familiarity with the genre of the text (DuBravac & Dalle, 2002). It is evident from the above studies that L1 and L2 comprehension processes rely on several layers of background knowledge. This dependence on contextual factors, whether internal or external, in language processing, lends support for the role of context in theories of inference processing. While individual differences also play a role in language processing, they will be covered in Chapter 3, since they fall under the larger category of cognitive processing.

In review, a few studies have found evidence for processing differences between L1 and L2 readers, but proficiency seems to be a mediating factor. It is clear that L2 learners need to first form propositional links (explicit meaning) before building a mental model (inference) of a text. While it is not entirely clear how proficient L2 readers need
to be in order to fully comprehend a text in their L2, there have been studies that have shown that higher levels of proficiency are correlated with higher levels of thinking in the L2 (Cohen, 1998; Guerrero, 2005; Leontiev, 1981), which could be evidence for the claim that L2 readers have access to the L2 situation model. There is an indication that L2 learners also use these links (at the explicit level) to decipher meaning. Finally, there is strong confirmation that prior knowledge, and/or cultural-specific knowledge are key factors in assisting in L1 and L2 discourse processing, along with other learner factors. Despite the growing body of research surrounding processing and comprehension above the sentence level, we still know very little about L2 reading processing; Koda (2005) claims “there has been little effort to unveil, on a scientific basis, either the specific processes involved in inferential generation among L2 readers or the conditions affecting it” (p. 152). While many studies point to processing differences between L1 and L2 inferential processing, the underlying cause is unclear, meriting further exploration.

2.5. Summary

This chapter surveyed the major strands of research in regard to inference generation and provided evidence that inference generation is a pragmatic process in which meaning and context interact. The theoretical perspective highlights the need for a hearer or reader to ‘read between the lines’ of what is said. Next, bridging and predictive inferences were described, as they are used as the basis for the verbal stimuli. These inference types purportedly require different processing: bridging inferences require that information that is not overtly linked is connected, and predictive inferences require that information that is not explicitly given is filled in via background information. In addition to theories of inference, two formative cognitive models were examined. First, the CI
model, which states that reading comprehension is a bottom-up process that progresses from the word to the context level, was described. Next, the constructionist model was outlined, in which reader strategy is claimed as the driving force in the comprehension process, with the exception of more automatic processing that is incidental. Finally, much of the relevant L2 sentence processing literature has explored the CI model, yet there is also disagreement as to whether L2 readers are able to access the top-most level of processing, the context or the situation model. It is clear that both L1 and L2 readers need to build up a propositional textbase, yet L2 readers do not seem to have access to the same cohesion markers as their L1 counterparts and use other factors, such as background knowledge, etc. to understand discourse.

The body of literature provides a clear picture of initial stages of sentence comprehension, not text comprehension, and there are still several unanswered questions about L1 and L2 reading comprehension. One missing factor is the role of individual differences, which seem to be a confounding variable in most processing studies. There is variation in how well a reader builds inference, based on prior knowledge, context, and internal resources (Koda, 2005; Singer, 2007); this variation in comprehension performance is possibly a result of individual differences. Cognitive resources should impact how a text is processed and understood, as they determine access to the resources that control attention and memory. It is clear that these models certainly provide a useful framework through which to test the inference generation process, however the predictions made from these models need to take into account the role of individual differences. Moreover, despite the substantial empirical support these models have garnered, they do not resolve whether L1 and L2 processing are similar or different. To
build on these questions, the next section will further explore the role of cognitive resources in inference processing in L1 and L2 reading of bridging and predictive inferences.
CHAPTER 3: INDIVIDUAL DIFFERENCES AND INFERENCE PROCESSING

A few studies have investigated the relationship between individual differences, or cognitive resources, and general reading performance (Ahmadi, Ismail, & Abdullah, 2012; Cain, Oakhill & Bryant 2004, Erçetin & Alptekin, 2013). This research specifically investigates metacognitive ability in task completion (i.e., the capacity to evaluate the success or failure of the process while it is ongoing) (Metcalf, 1996). Reading is a higher-order activity in which ongoing understanding is monitored, giving readers a chance to modify their behavior and to, therefore, strive for more successful comprehension. However, metacognitive ability is comprised of various sub-skills, from the ability to retrieve a semantic representation to the ability to manipulate that semantic representation in context. In conjunction with the broader research in the field of linguistics and education, there has also been a resurgence of studies specifically investigating whether inference type impacts processing routines (Beeman, Bowden & Gernsbacher, 2000; Burkhardt, 2006; Ferstl & von Cramon, 2001; Jin et al., 2009; Virtue, Haberman, Clancy, Parrish & Beeman, 2006).

Since much of this research also takes cognitive resources into account, this chapter will first cover common cognitive models for working memory and attentional resources. Second, the predictions from the CI and constructionist models will be explored within the working memory and attentional models. Third, the role of cognitive factors in bridging and predictive inference will be explored, to examine what we know about learner internal processing and the role that cognitive factors play in this process.
3.1. Individual Differences (IDs)

A framework is needed to understand exactly what is meant by ‘cognitive capacity’ or ‘individual difference’ and how it purportedly functions before surveying the relevant literature. A wealth of information is provided to a reader when she first encounters a text. As more text is processed, more information must be stored, connected, disregarded, retrieved or reanalyzed. Without a system to access meaningful representations of information and manipulate these representations, making meaning would be nearly impossible. Two formative models of memory and attention will be reviewed: Baddeley’s (2010) multi-component model of working memory and Barnes and Jones’ (2000) attentional control model, as they reveal important, and complementary aspects of reader internal resources.

3.1.1. Working Memory. The general conceptualization of memory systems has shifted from one comprised of short-term memory (STM) and long-term memory (LTM) (Atkinson & Shiffrin, 1968; Waugh & Norman, 1965; Sternberg, 1966) to a system comprised of working memory and LTM (Miyake & Shah, 1999). The most salient theoretical difference between the two is as follows: memory systems, in the traditional view, were seen as a storage device, as opposed to a processing and storage device, as in the current view (Miyake and Shah, 1999). Unfortunately, there is little agreement among cognitive researchers about how to this construct is organized, nor about how it functions (for overviews, see e.g., Baddeley, 2007; Conway, Jarrold, Kane, Miyake & Towse, 2007; Shah & Miyake, 1999). Theorists do tend to agree that what is needed is “a system of limited attentional capacity, supplemented by more peripherally-based storage systems” (Baddeley, 2003, p. 829).
This dissertation will rely on a well-tested model of working memory from Baddeley and colleagues (2010). Baddeley (2010) proposes a multicomponent model (see Figure 3.1) that consists of (1) a central executive (referred to as cognitive control in other models), which regulates attention; (2) a phonological loop, which serves as a temporary audio replay of information; (3) a visuospatial sketchpad, which serves as a temporary visual replay of information; (4) an episodic buffer, which combines the above three elements temporarily in episodes; and (5) a role for long-term memory. Working memory has been shown to predict a wide range of language abilities in both L1 language (e.g., Gathercole & Adams, 1994; Gathercole & Baddeley, 1989) and L2 language (for an overview, see e.g., Juffs & Harrington, 2011).

![Baddeley's Model of Working Memory](replicated from Baddeley, 2010, p. R138)

*Figure 3.1* Baddeley’s Model of Working Memory (replicated from Baddeley, 2010, p. R138)
The phonological loop is a store that holds memory traces for a limited period of time, at which the traces disappear. This includes rehearsal of information, as shown by the word length effect (between 1-5 syllables) (Baddeley, Thomson & Buchanan, 1975). The ability to hold phonological information is thought to aid in language comprehension and acquisition, whether in spoken or written language. The visuospatial sketchpad is the visual equivalent to the phonological loop and holds roughly three to four objects in the memory store. The central executive – a vital, yet not well understood component of the working memory model (Baddeley, 2003) – directs the information in the memory system. Due to the difficulty specifying the central executive, the episodic buffer was suggested, which has a finite capacity to bind together information in ‘episodes’ to form meaning. Executive functioning is localized in the frontal lobes (Shallice, 1988; Smith & Jonides, 1997; Stuss & Knight, 2002). The N-Back task is a common measure of executive function, as it has been associated with activation in the frontal lobe, among other related regions (Braver et al., 1997; Cohen et al., 1997).

### 3.1.2. Attentional Models

Since Baddeley (2003) readily concedes that executive function is underspecified in his model, the attentional control framework (see Barnes & Jones, 2000; Rueda, Posner & Rothbart, 2005; Kanske, 2008) offers a fuller picture of this construct.

Similar to the working memory models, the general concept of attention can be seen as a single unitary system or separate systems that support distinct functions (i.e., intensity and selectivity; Kanske, 2008). This dissertation advocates for the multi-component system approach. As such, intensity refers to attentional states and selectivity refers to the underlying mechanisms that allow for the distribution of resources.
Executive attention (or attentional control), the purposeful directing of energy or awareness, stands in direct contrast to attentional capture, the immediate reaction to salient or unexpected stimulus change (Barnes & Jones, 2000). Furthermore, executive attention can be broken into low-level stimulus control of attention, “transient bottom-up involuntary” processes (Barnes & Jones, 2000, p. 254) and high-level cognitive control of attention, “voluntary top-down processes that guide attending toward a goal over the long term” (Barnes & Jones, 2000, p. 255).

Rueda et al. (2005) propose that attention requires the functions of alerting (arousal to stimulus), orienting (selection of information from sensory input), and executive control (involvement of mechanisms for resolving conflict among thoughts/feelings/responses). Attention can be seen as a system with networks (or neural areas) that relate to these major functions. Executive attention is connected to the concept of selectivity; Kanske (2008) relates executive control of attention to “the detection of conflict between opposing activations” (p. 15). Put another way, executive attention allows for behavior to stay goal directed, often in the face of distracting stimuli.

Executive control is voluntary and requires cognitive control; it can also possibly be thought of as conscious and evaluative. The mechanisms of control, through which voluntary control of behavior is achieved, are conscious detection, inhibition and conflict resolution. Conscious detection is thought to contribute to the selection of information in the face of distraction, in that it “plays a special role in selecting target stimulus from among alternatives and engages attention in a way that resists interference by other signals” (Rueda et al., 2005, p. 576). Inhibition is involved in orienting and executive functions of attention (attentional selection and executive control); for example ignoring
irrelevant info and withholding inappropriate responses. Conflict resolution is the monitoring and resolution of conflict between incompatible responses (voluntary and attentive control of action, as seen in the Stroop tasks). Importantly the process of “attending” is a dynamic one, in which alerting, orienting and executive control work in concert, as do the three mechanisms of control (conscious detection, inhibition, and conflict resolution). All the aforementioned processes are most likely dependent on each other and do not operate in serial processing.

3.1.3. Cognitive frameworks used in this dissertation. Since these attentional and working memory frameworks are helpful for specifying the different components of a memory system, they were used in the present dissertation. This research will borrow the construct of visuospatial memory (visual short-term memory) from Baddeley’s model. In addition, the different components of executive function or cognitive control, including conscious detection, inhibition, and conflict resolution, from the attentional framework model will serve as a reference. Throughout the dissertation, the higher order functioning will be referred to as cognitive control.

3.1.3. Bilingual advantage. It is worth mentioning that there is an established bilingual advantage in cognitive control tasks when bilinguals are compared to monolinguals (Bialystok, 1998, 1999, 2004, 2009; Morton & Harper, 2007; Martin-Rhee & Bialystok, 2008; Costa et al., 2008, 2009; Hernández et al., 2010; Prior & Gollan, 2011; Poarch & van Hell, 2012). Stocco, Yamasaki, Natalenko and Prat (2014) recently developed a computational model based on the three-way connection found in the cognitive science literature between bilingualism, executive function, and the prefrontal cortex. Their model simulates how information flows from the gating system in the
striatum to the prefrontal cortex, an area designated for task switching. It is posited that the necessity to switch back and forth from one language to another has strengthened executive control in bilinguals (Garbin et al.; 2010; Prior & MacWhinney, 2010). Since executive control is thought to be a domain-general cognitive capacity, this enhanced capacity is useful both in linguistic tasks (the Stroop task) and non-linguistic tasks (the N-back task).

3.2. Predictions from Cognitive Models

The cognitive models reviewed in Chapter 2 underspecified the role of cognitive resources in inference processing. Within the framework of the models of working memory and attentional control, predictions can be made for each of the inference types in regard to the different cognitive resources required.

3.2.1. CI Model: Predictions. In the CI model, bridging inferences most likely are made at the textbase level, as they create explicit meaning through dependency relations, or in Kintsch’s (1998) terminology, through creation of the macrostructure of global topics. On the other hand, predictive inferences would need to be formed at the situation level, as the reader uses non-linguistic factors (i.e., background knowledge, cultural knowledge) to generate an interpretation that is not explicitly stated in the text. While CI models allow for the role of cognitive resources in inference processing (i.e., the activation of knowledge stored in long-term memory [construction phase] and the integration into working memory [integration phase]) the process is still quite vague. The specific cognitive resources that would be required at the different levels of processing in Kintsch’s model are not fully outlined. That said, we can draw on the neurolinguistic studies discussed below (Burkhardt, 2006; Jin et al., 2009; Ferstl & von Cramon, 2001);
since bridging inferences are processed at the textbase level, they would likely rely on phonological short-term memory to connect the two propositions that are anaphorically linked to each other. The process of forming a bridging inference ultimately requires that representations are stored in memory and then associated, a process that would occur in the phonological loop. In contrast, updating the situation model for processing of predictive inferences likely calls on cognitive control; several interpretations are generated from the set-up of a predictive inference, yet ultimately one is chosen while the others are disregarded. The cognitive complexity of the task points to the storage of phonological representations and the manipulation of that information.

3.2.3. Constructionist Model: Predictions. In contrast to CI models, a strict interpretation of Graesser et al.’s (1994) model would point to heavy reliance on cognitive control rather than elements of short-term memory to process bridging inferences and predictive inferences, based on the purported reliance on strategy. Strategy requires that a reader intentionally seek an answer to a specific goal; as we saw in Graesser et al.’s model, a reader asks continuous questions as a reading strategy. If readers deliberately seek out meaning and explanation while processing text, then regardless of whether the reader is forming a bridging inference or a predictive inference, both processes would rely on cognitive control. However, if the processing is not strategy-driven, then it could be more in line with what Graesser, Weimar-Hastings and Wiemer-Hastings (2001) called “passive-activation.” This process is a more automatic process of connecting information at the local coherence level, such that these inferences are “activated and reactivated by multiple sources of information (e.g., words, propositions, contents of working memory, scripts, global macrostructures)” (Graesser et
al., 2001, p. 254). Processing that is driven by strategy would then be separated from that which is incidental. In this case, a bridging inference, which is passively retrieved, would be largely automatic. However, a predictive inference presents a reader with several alternate hypotheses; if a reader were asking questions while reading text, then predictive inferences would rely more on higher level processing.

At this point it has been established that the CI and constructionist models make similar predictions about the processing of the two inference types investigated in this research, bridging and predictive. Bridging inferences would likely rely on phonological short-term memory and predictive inferences on cognitive control. Next, the interaction between individual differences and inference processing will be reviewed in both behavioral and neurolinguistic studies, as it serves as a basis for the current two studies. A view of both strands of research provides a fuller picture of the role of individual differences in inference processing.

3.3. Empirical Findings from Behavioral Studies

While there are few behavioral studies that have investigated the role of individual differences in inference processing, there is a large body of the behavioral research has looked at working memory, specifically the phonological loop in auditory comprehension (e.g., Ehrlich & Johnson-Laird, 1982; Montgomery, Polunenko & Marinellie, 2009). These will not be reviewed, as this research does not directly address inference comprehension and the studies are, by and large, correlational accounts of auditory processing. While behavioral research has looked into the factors that influence whether an inference will be made or not, few have investigated what underlies this processing. Virtue, Parrish, and Jung-Beeman (2008) remark on the lack of research on
“the component processes or neural substrates involved in generating such inferences” (p. 2274).

Walczyk, Marsiglia, Bryan, and Naquin (2001) carried out a study with L1 readers to test compensatory-encoding model of reading, in which low working memory capacity is correlated with less automatic reading, as evidenced by slower reading rate, pauses and reviewing the text. Verbal working memory capacity, measured by a memory probe procedure, was positively correlated with inference-based comprehension. This result offers a fuller picture of L1 reading in that participants with higher the working memory capacity were faster, more automatic and more likely to access meaning in a text.

In an L2 study, Rai, Loschky, Harris, Peck and Cook (2011) looked at intermediate-level Spanish learners’ reading performance on non-inference, bridging inference and pragmatic inference (i.e., an inference that requires the reader use background knowledge to understand intention) dependent on working memory capacity. Using stress as an independent variable, Rai et al. (2011) found that participants with higher working memory were overall more accurate when answering questions about the inferences. The more complex the inference was, the less likely that all participants understood it in the high stress conditions. Interestingly, in high stress conditions participants with higher working memory processed high complexity inferences less efficiently (i.e., slower), but more effectively (i.e., accurately) as opposed to participants with lower working memory. Rai et al. (2011) concluded that though stress interferes with reading comprehension, processing is mediated by reading strategy. However, the picture is made more complicated when taking the role of working memory into account;
the entire comprehension process is dependent on working memory capacity. Rai et al.’s study provides further evidence for the role of working memory capacity in L2 inference processing, in the face of interference. However, these results need to be interpreted carefully, as Rai et al. (2011) measured the reaction time when answering a comprehension question and not during the reading process.

Reviewing the main findings from the small subset of behavioral studies, L1 and L2 readers with higher working memory capacity process inferences more efficiently and accurately than those with lower working memory capacity. L2 readers with high working memory, when distracted, were slower, but more accurate than those with low working memory. These studies reinforce the finding that working memory capacity is a mediating factor in inference processing. However, working memory capacity is a collection of component skills, and it is still not clear how large a role the sub-skills of phonological short-term memory and/or cognitive control contribute to this processing or if this varies depending on inference type. Now, we will look at the empirical findings from neurolinguistic studies for further evidence.

3.4. Empirical Findings from Neurolinguistic Studies

The majority of neurolinguistic studies have gone beyond correlational accounts of individual differences; these studies have investigated the specific role cognitive resources play in inferential processing.

3.4.1. Inference Type and Cognitive Resources. A small number of functional magnetic resonance imaging (fMRI) and event-related brain potential (ERP) studies² to

² Neuroimaging studies are typically conducted with functional magnetic resonance imaging (fMRI), which uses magnetic fields to create 2- or 3-dimensional images of the brain during processing. The relevant areas “light up”
date have investigated the process inference building (Beeman, Bowden & Gernsbacher, 2000; Burkhardt, 2006; Ferstl & von Cramon, 2001; Jin et al., 2009; Virtue, Haberman, Clancy, Parrish & Beeman, 2006) with an unusual slant. These studies look at cognitive resources as a predictive variable and not a co-variable; essentially these researchers take the perspective that cognitive resources are not an intervening variable, but an integral part of inference processing. These studies report tentative evidence that inference building may call on unique brain areas, and, therefore, rely on different cognitive resources, depending on inference type (Burkhardt, 2006; Jin et al., 2009; Ferstl & von Cramon, 2001).

Burkhardt (2006) investigated the construction-integration framework in an ERP study of German speakers, focusing on the integration phase of comprehension. She found that bridging inferences had storage costs and took this as evidence that anaphoric inferences tax storage capacity. Burkhardt was concerned with the new-given information distinction in inferential Determiner Phrases (DPs) (i.e., bridging inferences), as they link previously presented referents and newly presented pronouns to each other. She wanted to test whether the newly presented referent (that connects to prior information in a sentence) is processed as new information or as given information. New information is considered more taxing and has been shown to elicit a P600 (a component thought to indicate reanalysis of information). Given information is, on the other hand, thought to be less taxing, and, thus, elicit an N400 (a component thought to bind information that is already given in short- or long-term memory). Burkhardt found that inferentially bridged

during processing of different stimuli. Electroencephalography (EEG) studies record electrical activity on the scalp to locate a time course of processing; the averaged signal are then made into event-related potentials (ERP) components that are linked back to different types of processing.
DPs show a pattern that first mirrors given information, an attenuated N400, and then mirror new information, an enhanced P600. Burkhardt (2006) shows that bridging inferences call on storage of memory representations, as it “results in an increase of working memory load and storage cost” (p. 160), due to the integration of given and new information (evidenced by the enhanced P600). It is important to note that Burkhardt’s (2006) interpretation considers a rise in an ERP amplitude to reflect processing costs, when enhanced amplitude could reflect facilitation (i.e., less effortful processing). The reader must form a new discourse referent that relates to the previously incorporated entity.

Also investigated within a construction-integration model, Jin et al. (2009) attempted to localize the neural correlates of predictive inference processing in a functional magnetic resonance imaging (fMRI) study with Chinese texts. The authors manipulated the predictability likelihood of a 3-sentence story. According to Jin et al. (2009), predictive inference is “anticipation of events that are about to happen based on current information in the text” (p. 142). The results showed that stories that had a highly predictable inference “evoked significantly stronger activities in two brain regions, the left inferior frontal gyrus (LIFG) and the right lingual gyrus, relative to reading control stories” (Jin et al., 2002, p. 146). The LIFG is postulated to contribute to the selection of information among competing alternatives, particularly for semantic knowledge (Thompson-Schill et al., 1999; Beeman, 2005; Zhang et al., 2004). The construction process in predictive inference seems to involve a selection process in which relevant text information is chosen among irrelevant information to correctly infer predicted events.
These studies point to the fact that different inference types require different allocations of internal resources for successful processing. Burkhardt (2006) found tentative evidence that the process of bridging information and connecting referents relies, at least in part, on memory capacity (i.e., storage). Jin et al. (2009) found support that predictive inferences, in part, rely on the brain region responsible for cognitive control for processing. While these results only provide a preliminary account of how these inference types are processed, they do provide the basis for the set of hypotheses in the present dissertation.

3.5. Summary

In review, the individual difference models provide a useful framework for understanding the necessary subcomponents of the processing system. As such, this research uses the construct of visuospatial memory (visual short-term memory) from Baddeley and colleagues and the detailed description of executive function from Barnes and Jones (2000). Behavioral studies found that working memory was positively correlated with inference generation and increased comprehension, even in challenging conditions and with complex stimuli. Neurolinguistic studies have focused on Kintsch’s (1998) CI model, finding that bridging and predictive inferences call on different internal resources, and the cognitive models seem to agree to this point, yet the results are not robust enough to make definitive claims.

The following section summarizes the gaps identified in the review of the literature on inference building and individual differences and inference processing.
3.6. Research Gap

As the previous chapters reviewed, an important, yet not well understood, variable related to inference processing is individual differences. In parallel to studies that look at the correlational relationships of cognitive resources and inference processing, there is a growing area of research that has been looking at the direct role of specific cognitive resources dependent on inference type. It has been shown that inferences are processed differently dependent on type, specifically bridging inferences and predictive inferences. Stemming from predictions from influential cognitive models (the CI model and the constructionist model), neurolinguistic researchers found evidence that bridging inferences and predictive inferences rely on different cognitive resources; bridging inferences rely on short-term memory while predictive inferences rely on cognitive control. The research thus far is preliminary. Given that so few studies have investigated the role of cognitive resources used to process different inference types, this research will investigate the role of cognitive resources in inference generation of predictive and bridging inferences. To complicate the picture, there are relatively few studies that explore the role of language background (L1 versus L2) in readers’ processing of inference at the text level. It has been shown that L1 and L2 readers process text differently from one another. The literature review on inference building revealed that there is debate as to whether L2 readers have access to the top-most level of processing (the situation model in the CI framework). However, it is not clear whether the participants in these studies were at a high enough proficiency to process the texts. Many studies that compare L1 and L2 inference generation point to diverse factors that may impact processing, such as proficiency, cultural-specific knowledge, internal resources
and prior knowledge. These factors all point to linguistic and cultural familiarity. In other words, the more familiar an L2 speaker is with their second language and culture, the more easily they are able to process inference. It will look at two types of inferences (bridging and predictive) in first language (L1) and second language (L2) reading comprehension. The current research ultimately seeks to examine what role cognitive factors play in inference processing dependent on inference type and language background.
CHAPTER 4: METHODS

4.1. Research Questions

At the end of the last chapter, language background and cognitive resources were shown to be important variables to understanding inference processing. Additionally, it became clear that different inference types may require different types of processing. To explore the role of cognitive resources in the processing of bridging and predictive inferences, the two studies investigated the role of individual differences during the processing of bridging and predictive inference building in L1 and L2 reading. Recall that the experimental design, materials and procedure were identical for the L1 and the L2 studies; therefore this chapter will discuss the general methodology.

The research questions for the L1 study are:

1a. Are inferences processed differently from non-inferences in the L1?

2a. Do L1 readers rely on different cognitive resources to successfully process inferences?

3a. Are bridging inferences processed differently from predictive inferences in the L1?

The research questions for the L2 study are:

1b. Are inferences processed differently from non-inferences in the L2?

2b. Do L2 readers rely on different cognitive resources to successfully process inferences?

3b. Are bridging inferences processed differently from predictive inferences in the L2?

Finally, the L1 and L2 studies will be compared in the last research question:
4. Do the strength of the relationships change based on language background?

4.2. Hypotheses

To review, these two experiments aim to examine the role of cognitive resources (short-term memory and cognitive control), inference type (bridging inference and predictive inference) and language background (L1 versus L2) in inference building. These experiments ultimately explored if inference processing relied on different cognitive resources depending on inference type and if the relationship between cognitive resource and inference type changed based on whether participants were reading in their L1 or their L2. In a within-subject design (in which all participants completed all the tasks) using a visual distraction paradigm, participants’ read two-sentence scenarios and answered questions about them while completing a secondary non-linguistic task. As reaction-time experiments, participants’ reading times were measured. The two-sentence scenarios in the studies required that participants form a bridging inference, a predictive inference, or no inference. The distraction paradigm included three conditions: control (i.e., no distractor task), short-term memory distractor (i.e., a task that taxes either the concepts held in the visual short-term memory), and cognitive control distractor (i.e., a task that taxes cognitive control). Measures of short-term memory and cognitive control were also collected to see the relationships between these different capacities and the reading times.

In the case of the first set of directional hypotheses, the literature suggests that bridging inferences rely on memory capacity for processing whereas predictive inferences rely on cognitive control.
**H1:** L1 readers will rely on memory capacity over and above cognitive control for bridging inferences. Based on Burkhardt’s (2006) study, it is plausible that visual/phonological short-term memory is taxed more than cognitive control in bridging inferences.

**H2:** L1 readers will rely on cognitive control over and above visual/phonological short-term memory in for predictive inferences. If Jin et al. (2009)’s findings bear out, then participants will rely on cognitive over visual/phonological short-term memory.

The third directional hypothesis addresses the fact that L2 readers have enhanced cognitive control.

**H3:** Based on research on a bilingual advantage for executive or cognitive control (see Bialystok, 1998, 1999, 2004, 2009), L2 speakers will modulate the effects of cognitive control distraction in predictive inference due to enhanced cognitive control.

### 4.3. Participants

The participants who participated in the L1 study were L1 English speakers, and the participants who participated in the L2 study were highly proficient L2 English speakers with L1 Spanish (see Section 4.3.2.1. for how proficiency was operationalized). Of the 78 participants who volunteered for these studies, 52 were assigned to the L1 study and 26 were assigned to the L2 study. Three participants were removed from the study because they did not complete the second session (L1 group: \( n = 1 \) and L2 group: \( n = 2 \)), leaving 75 participants who completed the study. Language background information was collected via a background questionnaire (see Appendix E) on the first day of data
collection. Based on participants’ responses, 1 participant was discarded for reported hearing issues (L2 group: n = 1), and 1 participant was discarded for language background ineligibility, i.e., one L1 English speaker reported not having completed schooling in an English-speaking setting (L1 group: n = 1). Of the remaining participants, 73 participants (53 female and 20 male) between the ages of 18 and 37 completed the study and most, if not all, of their data were included in the analyses. The mean age was 21.2 (SD = 4.16).

Data were collected from undergraduate or graduate students attending Georgetown University (see section 4.5.1. on recruitment practices). All participants were right-handed (both in writing and cutting) with normal hearing and normal to corrected vision, and they did not receive any compensation for participation.

4.3.1. L1 English speakers. The L1 English group had 50 participants (n = 50), and their ages ranged from 18 to 27 years with a mean age of 19.6 (SD = 1.7); 38 were female, and 12 were male. All participants listed English as their native language; 47 reported English as their only native language and 3 reported an additional native language (Hindi: n = 1, French: n = 1, German: n = 1). All participants verified that the majority of their schooling had been in an English-speaking setting. The majority of the participants had knowledge of one or more second languages (i.e., Arabic, ASL, Chinese, French, German, Greek, Italian, Japanese, Latin, Portuguese, Russian, Spanish, Swedish and Urdu) to a greater or lesser extent; since expertise in other languages was not directly related to the research questions for the L1 group, this is reported more broadly: 49 participants spoke a second language; 24 participants spoke a third second language; 4 participants spoke a fourth second language; and 2 spoke a fifth second language.
4.3.2. L2 English speakers. The L2 group had 23 participants \((n = 23)\), and their ages ranged from 18 to 37 years with a mean age of 24.8 \((SD = 5.6)\). There were 15 female and 8 male participants. All participants spoke the same L1 (Spanish) and one had a second L1 (Zapotec: \(n = 1\)). Participants came from various countries (i.e., Argentina, Chile, Costa Rica, Cuba, Colombia, Ecuador, El Salvador, Guatemala, Mexico, Panama, Spain, and Venezuela). The majority of the participants had knowledge of one or more second languages (i.e., Catalan, French, German, Hebrew, Italian, Korean, Portuguese, and Russian) to a greater or lesser extent; since expertise in other languages was also not directly related to the research questions for the L2 group, this is reported more generally: 21 participants spoke a second language, 9 participants spoke a third second language, 4 participants spoke a fourth second language, and 1 spoke a fifth second language.

4.3.2.1. Proficiency. All the participants in the L2 group had highly proficient L2 English, especially in reading comprehension, as that is the focus of the study. It is challenging to operationalize proficiency for L2 English learners, as there is little agreement as to what qualifies as “low”, “mid” or “high” proficiency based on the various standardized tests (such as the TOEFL, IELTS, Pearson Test of English, etc.). Proficiency is based on the entry-level admittance required for Georgetown University’s business school (Admissions, Frequently Asked Questions, n.d.). Since these standards are meant to ensure that entering students are able to read at a high academic level, it is assumed that the students who either earn these scores or complete a degree at an English-speaking university are skilled in reading above the surface level of a text and making inferences from information not explicitly provided. In addition, the proficiency exams are typically administered in the written modality, lending further support to the
claim that participants’ proficiency in reading would be adequate. L2 participants were considered proficient based on one of the following two criteria: a test score (including a computer-based TOEFL [cBT] score of at least 250/300, a paper-based TOEFL [pBT] score of 600/677, an internet-based TOEFL [iBT] score 100/120, an IELTS academic module of 7.5/9, or Pearson Test of English of 68/90) or a degree from a U.S. academic institution (high school or university). To reiterate, all the L2 participants were enrolled in an undergraduate or graduate program at Georgetown University and demonstrated adequate levels of proficiency to attend college or graduate school in an English-language academic environment.

Proficiency was also measured using additional information collected from the Background Questionnaire (see section 4.5.2). The information collected assessed self-rated proficiency (i.e., 1 = not at all confident; 2 = somewhat confident; 3 = mostly confident; 4 = very confident), Age of Acquisition (AoA), years of instruction, setting of learning (i.e., formal, informal or a mix of informal and formal) and time abroad in English-speaking countries. The L2 group had lived or been living in an English speaking country for 1 to 16 years, with a mean of 5.5 (SD = 4.6). See Table 4.1 for details of language background information and proficiency measures in the L2 group. The L2 participants’ background questionnaires placed them as advanced speakers of English and therefore at a high enough level to complete the reading tasks in the experiment.
Table 4.1. L2 language background information and proficiency.

<table>
<thead>
<tr>
<th>ID #</th>
<th>Age</th>
<th>Speaking</th>
<th>Reading</th>
<th>Writing</th>
<th>Listening</th>
<th>AoA</th>
<th># of Years Learned</th>
<th>Learning Setting</th>
<th>Years in English-Speaking Setting</th>
<th>Proficiency Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>S01</td>
<td>20</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>13</td>
<td>Mixed</td>
<td>16</td>
<td>U.S. school</td>
</tr>
<tr>
<td>S02</td>
<td>18</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>14</td>
<td>Formal</td>
<td>16</td>
<td>U.S. school</td>
</tr>
<tr>
<td>S03</td>
<td>18</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>12</td>
<td>Mixed</td>
<td>1</td>
<td>U.S. school</td>
</tr>
<tr>
<td>S04</td>
<td>28</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>22</td>
<td>Mixed</td>
<td>8</td>
<td>TOEFL iBT 115</td>
</tr>
<tr>
<td>S05</td>
<td>32</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>9</td>
<td>23</td>
<td>Mixed</td>
<td>6</td>
<td>TOEFL cBT 283</td>
</tr>
<tr>
<td>S06</td>
<td>32</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>14</td>
<td>Mixed</td>
<td>1</td>
<td>U.S. school</td>
</tr>
<tr>
<td>S07</td>
<td>18</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>14</td>
<td>Mixed</td>
<td>1</td>
<td>TOEFL iBT 115</td>
</tr>
<tr>
<td>S08</td>
<td>18</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>15</td>
<td>Mixed</td>
<td>1</td>
<td>U.S. school</td>
</tr>
<tr>
<td>S09</td>
<td>19</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>13</td>
<td>Formal</td>
<td>6</td>
<td>TOEFL iBT 115</td>
</tr>
<tr>
<td>S10</td>
<td>27</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>23</td>
<td>Formal</td>
<td>3</td>
<td>TOEFL iBT 118</td>
</tr>
<tr>
<td>S11</td>
<td>28</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>23</td>
<td>Formal</td>
<td>5.5</td>
<td>TOEFL iBT 112</td>
</tr>
<tr>
<td>S14</td>
<td>28</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>16</td>
<td>12</td>
<td>Mixed</td>
<td>2</td>
<td>IELTS 7.5</td>
</tr>
<tr>
<td>S15</td>
<td>19</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>16</td>
<td>Mixed</td>
<td>1</td>
<td>TOEFL iBT 109</td>
</tr>
<tr>
<td>S16</td>
<td>29</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>18</td>
<td>Mixed</td>
<td>6</td>
<td>TOEFL iBT 119</td>
</tr>
<tr>
<td>S17</td>
<td>23</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>19</td>
<td>Formal</td>
<td>11</td>
<td>U.S. school</td>
</tr>
<tr>
<td>S18</td>
<td>30</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>11</td>
<td>15</td>
<td>Mixed</td>
<td>3</td>
<td>College in U.S.</td>
</tr>
<tr>
<td>S19</td>
<td>19</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>14</td>
<td>Mixed</td>
<td>10</td>
<td>U.S. school</td>
</tr>
<tr>
<td>S20</td>
<td>23</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>15</td>
<td>Mixed</td>
<td>5</td>
<td>U.S. school</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>----</td>
</tr>
<tr>
<td>S21</td>
<td>37</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>23</td>
<td>Mixed</td>
<td>7</td>
<td>TOEFL iBT 103</td>
</tr>
<tr>
<td>S23</td>
<td>23</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>16</td>
<td>Formal</td>
<td>5</td>
<td>U.S. school</td>
</tr>
<tr>
<td>S24</td>
<td>25</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>18</td>
<td>Mixed</td>
<td>2</td>
<td>TOEFL iBT 99</td>
</tr>
<tr>
<td>S25</td>
<td>26</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>23</td>
<td>Mixed</td>
<td>4</td>
<td>TOEFL iBT 114</td>
</tr>
<tr>
<td>S26</td>
<td>30</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>22</td>
<td>Mixed</td>
<td>16</td>
<td>TOEFL iBT 110</td>
</tr>
</tbody>
</table>

* 1 = not at all confident; 2 = somewhat confident; 3 = mostly confident; 4 = very confident
4.4. Linguistic Stimuli

The goal of the experiment is to test whether bridging inferences and predictive inferences are processed similarly or differently dependent on inference type and language background. Discourse requires a reader to understand and connect information across sentence boundaries (Singer, 2007). As such, the two-sentence scenarios used as the basis of the stimuli in the present dissertation are considered to be at the discourse level. There were a total of 150 two-sentence scenarios across three conditions: 60 that required a bridging inference, 60 that required a predictive inference, and 30 that required no inference.

The two sentence scenarios were presented as follows: first participants read sentence #1; next the participants read sentence #2; then the participants saw a verification question. Each verification question had two possible responses, a ‘yes’ response or a ‘no’ response. Table 4.2 shows a sample two-sentence scenario and verification question for each of the three types of stimuli in the finalized set.
Table 4.2 Sample stimuli, inferences and verification questions

<table>
<thead>
<tr>
<th>Bridging Inference</th>
<th>(1a) Ashley loved to drink coffee in the morning.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1b) She liked a special brew of coffee from Guatemala.</td>
</tr>
<tr>
<td><strong>Inference:</strong></td>
<td>She refers to Ashley.</td>
</tr>
<tr>
<td><strong>Verification Questions:</strong></td>
<td></td>
</tr>
<tr>
<td>- Did Ashley enjoy Guatemalan/Ethiopian* coffee? (1a-1b)</td>
<td></td>
</tr>
<tr>
<td>- Did she like to have coffee in the morning/evening*? (1b-1a)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predictive Inference</th>
<th>(2a) Kaitlyn had an early flight in the morning.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(2b) Kaitlyn forgot to set an alarm.</td>
</tr>
<tr>
<td><strong>Inference:</strong></td>
<td>Kaitlyn missed the flight.</td>
</tr>
<tr>
<td><strong>Verification Questions:</strong></td>
<td></td>
</tr>
<tr>
<td>- Did Kaitlyn miss/make* the flight?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No Inference</th>
<th>(3a) The dean only saw students in the morning.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(3b) The dean had a cold office.</td>
</tr>
<tr>
<td><strong>Inference:</strong></td>
<td>None.</td>
</tr>
<tr>
<td><strong>Verification Questions:</strong></td>
<td></td>
</tr>
<tr>
<td>- Did the dean counsel students in the morning/evening*? (3a)</td>
<td></td>
</tr>
<tr>
<td>- Did the dean keep the office cold/warm*? (3b)</td>
<td></td>
</tr>
</tbody>
</table>

* The words in *italics* represent the *yes/no* responses; fifty percent of the verification questions had a *yes* response and fifty percent of the verification questions had a *no* response.
The verification questions were designed so that participants were required to pay attention to both sentences and the relationships between the propositions to prevent them from developing response strategies during the experiment. These were counterbalanced throughout the pretests and the main experiment.

For the two-sentence scenarios with bridging inferences, the verification questions required that participants map the information in two directions. This yielded the following four questions for sentences (1a) and (1b): (a) the information ascribed to the anaphora *she* was correctly remapped onto the antecedent *Ashley*, for example, Did *Ashley enjoy Guatemalan coffee*? (yes); (b) the information ascribed to the anaphor *she* was incorrect, yet still mapped onto the antecedent *Ashley*, for example, Did *Ashley enjoy *Ethiopian coffee*? (no); (c) the information ascribed to the antecedent *Ashley* was correctly remapped onto the anaphor *she*, for example, Did *she like to have coffee in the morning*? (yes); or (d) the information ascribed to the antecedent *Ashley* is incorrect, yet still mapped onto the anaphor *she*, for example, Did *she like to have coffee in the *evening*? (no).

For the two-sentence scenarios with predictive inferences, the verification questions refer to a future consequence, so this principle was not applied; therefore there were only two possible questions for sentences 2a and 2b: those with a ‘yes’ response, for example, Did *Kaitlyn miss the flight*?, and those with a ‘no’ response, for example, Did *Kaitlyn *make the flight*?.

For the two-sentence scenarios with no inferences, the participants were asked verification questions about both sentences, yielding four total possible questions: (a) the information in sentence (3a) was correctly represented, for example, Did *the dean
counsel students in the morning? (yes); (b) the information in sentence (3a) was incorrectly represented, for example, *Did the dean counsel students in the evening?* (no); (c) the information in sentence (3b) was correctly represented, for example, *Did the dean keep the office cold?* (yes); or (d) the information in sentence (3b) was incorrectly represented, for example, *Did the dean keep the office warm?* (no).

The two-sentence scenarios that required no inference were semi-related in their content: the sentences had the same referent (e.g., the dean) but unrelated propositions (e.g., counsel students in the morning or keep the office cold). This minimal connection was provided in an effort to prevent participants from looking for unwanted connections in adjacent pieces of text (see Sperber & Wilson’s 1986 work on Relevance Theory) and displaying longer latencies in the reaction time.

4.4.1. **Guidelines used to create verbal stimuli.** Initially, 300 two-sentence scenarios were created for piloting purposes: 120 that required bridging inferences, 120 that required predictive inferences, and 60 that did not involve any type of inferences. Several guidelines were followed in order to create unbiased and valid stimuli. All the two-sentence scenarios were reviewed in order to make sure that the scenarios adhered to the guidelines by the researcher and her research assistant.

First, efforts were made to base inferences on general world knowledge, and not on assumptions rooted in North American culture. Second, the two-sentence scenarios were consistently sequenced from past to present, so that the participants could rely on a future order of events throughout the experiment. This maintained uniformity reduced a possible confounding variable of stimuli difficulty. For example, (4a-b) would be acceptable, as it sequences the past and then the present:
(4a) Sarah had just received a promotion.
(4b) She was excited.

Third, the sentences contained only single matrix clauses and not subordinate clauses, to ensure comparable complexity across sentences. Fourth, the two-sentence scenarios incorporated one and only one type of inference, to isolate inference processing. This ensured that the scenarios with predictive inferences only tested hypothesis generation and the scenarios with bridging inference only tested anaphor-antecedent mapping or the antecedent-anaphor mapping. For example, (5a-b) would not be an acceptable story because it contains both a predictive and a bridging inference:

(5a) The hamster would not eat or drink.
(5b) The veterinarian could not help it.

The two-sentence scenario contains the predictive inference (i.e., the hamster died) and the bridging inference (i.e., it refers to the hamster). By replacing it with the hamster, the two-sentence scenario then contains only the predictive inference. In the same vein, (6a-b) further illustrates this point:

(6a) Jenny studied hard for the test.
(6b) She was about to take the test.

Since the second sentence sets up the reader to start hypothesis generation, the two-sentence scenario contains a predictive inference (i.e., Jenny/she would do well on the test) and a bridging inference (i.e., she refers to Jenny). This two-sentence scenario would easily conform to the guidelines by changing the second sentence to She received a passing grade. By completing the arc of action, hypothesis generation is less likely. Fifth, all two-sentence scenarios with bridging inferences only contained one instance of antecedent to anaphora mapping, and not multiple instances, with a consistent
presentation order of antecedent (e.g., Sarah, the director) and then anaphora (e.g., she, he, it).

In addition, the questions were carefully constructed along several dimensions. The verification questions paraphrased the content of the two-sentence scenarios and when possible, altered the sentence structure from the two-sentence scenario in the question. Reformulating the language ensured that the readers processed the meaning of the sentences and prevented a priming effect for memory in which readers simply matched propositional phrases. The following two-sentence scenario with a bridging inference demonstrates this point:

(7a) Robert rescued a puppy from the shelter.
(7b) He got the puppy to stop peeing in the house.

Instead of asking the content with the same wording, Did Robert get the puppy to stop peeing in the house? the question would be paraphrased Did Robert successfully potty train the puppy? Also, the questions with a ‘no’ response were improbable, but not impossible, to force readers to pay attention to the two-sentence scenarios. The logic is that impossible questions are easily discerned without reading the previously presented content. All the stimuli were checked at each stage of development along the above standards, and a theoretical linguist was consulted at the final stage of stimuli development to ensure the validity.

Individual sentences had 5 to 10 words, and the final stimuli had an average of 7.2 words per sentence. Several t-tests were run in order to show that the sentence length was balanced. The overall length of sentence #1 ($M_{\text{length}} = 7.3$ words, $SD = 1.0$) and sentence #2 ($M_{\text{length}} = 7.2$ words, $SD = 1.0$) did not differ significantly for the complete set of stimuli, $p = .26$. The length also did not differ significantly between sentence #1 ($M_{\text{length}} =
7.0 words, $SD = 1.0$) and sentence #2 ($M_{length} = 7.2$ words, $SD = 1.0$) for sentence pairs with bridging inferences ($p = .88$), between sentence #1 ($M_{length} = 7.4$, $SD = 2.0$) and sentence #2 ($M_{length} = 7.6$ words, $SD = 1.0$) for sentence pairs with predictive inferences ($p = .69$) or between sentence #1 ($M_{length} = 7.1$ words, $SD = 1.0$) and sentence #2 ($M_{length} = 7.1$ words, $SD = 2.0$) for sentence pairs with no inferences ($p = .70$).

4.4.2 Names in verbal stimuli. A total of 30 two-syllable names were used in the verbal stimuli (15 male and 15 female). All of the male and female names were names obtained from the U.S. Social Security database (Popular Baby Names, n.d.), in which names from social security applications are recorded and then ranked in terms of popularity. The researcher accessed the fifty most popular names from the past 15 years; the following lists were downloaded: 2000, 2006, and 2012. When necessary, the names were modified to fit the two-syllable structure (e.g., Alexander was changed to Alex). The names chosen occurred in at least one of the aforementioned lists and were also roughly equally distributed across the alphabet for males and females (A-E: Males = 5, Females = 4; F-J: Males = 3, Females = 3; K-O: Males = 5; Females = 5; P-S: Males = 1; Females = 3; T-Z: Males = 1; Females = 0).
Table 4.3 Final list of male and female names used in the verbal stimuli.

<table>
<thead>
<tr>
<th>MALE</th>
<th>FEMALE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Andrew</td>
</tr>
<tr>
<td>2</td>
<td>Alex</td>
</tr>
<tr>
<td>3</td>
<td>David</td>
</tr>
<tr>
<td>4</td>
<td>Ethan</td>
</tr>
<tr>
<td>5</td>
<td>Evan</td>
</tr>
<tr>
<td>6</td>
<td>Henry</td>
</tr>
<tr>
<td>7</td>
<td>Joseph</td>
</tr>
<tr>
<td>8</td>
<td>Justin</td>
</tr>
<tr>
<td>9</td>
<td>Kevin</td>
</tr>
<tr>
<td>10</td>
<td>Michael</td>
</tr>
<tr>
<td>11</td>
<td>Matthew</td>
</tr>
<tr>
<td>12</td>
<td>Noah</td>
</tr>
<tr>
<td>13</td>
<td>Nathan</td>
</tr>
<tr>
<td>14</td>
<td>Robert</td>
</tr>
<tr>
<td>15</td>
<td>Tommy</td>
</tr>
</tbody>
</table>

4.4.3 Pretest of stimuli. Prior to the main experiments, two pretests were carried out on two distinct groups of participants who had no prior knowledge of the verbal stimuli and who did not take part in the main study. The pretests ensured that the verbal stimuli were valid and reliable, both in the content of the inferences (pretest #1) and the reaction time to read the sentences and to answer questions about the sentences (pretest #2). All pretest experiments recruited L1-English participants.

4.4.3.1 Pretest #1: Rating study. In order to test the validity (or lack thereof) of the different types of inferences, native speakers of English completed an untimed rating study of the two-sentence scenarios delivered through SurveyMonkey software, http://www.surveymonkey.com (last visited [March 1, 2014]). Participants confirmed or disconfirmed the researcher’s intuition about the inference or lack thereof in the two-sentence scenarios. Sentence pairs that had a verification rate of below 90% were
eliminated from the study. In an effort to reduce fatigue and disinterest, the stimuli were split into two versions so that each version took approximately 30 minutes to complete.

Thirty native speakers of English (19 women and 11 men, $M_{age} = 43.3$ years, $SD_{age} = 15.05$) were recruited to take part in the rating study. All participants listed English as their only native language and verified that the majority of their schooling had been in an English-speaking setting. Participants were presented with different versions of the rating study based on random criteria (i.e., when their birthday falls in the year): 15 participants were presented with one half of the stimuli in Version A and 15 participants were presented with the other half of the stimuli in Version B. The two groups did not differ significantly across age, $p > .05$ ($M_{age} = 38$ years for Version A versus $M_{age} = 49$ years for Version B), or gender, $p > .05$ (6 males for Version A versus 5 males for Version B). Participants did not receive any compensation for participation.

Each version of the study contained 60 predictive inferences, 60 bridging inferences, and 30 unrelated, for a total of 150 two-sentence scenarios in each version.

After a verbal informed consent and brief background survey, participants completed the online rating study (see Appendix A for directions in the rating study). There were three different tasks, based on the inference type or lack thereof.

For the Bridging Inference section of the rating study, participants read a two-sentence scenario, for example, (8a-b), and were presented with four indicative statements. The instructions were to choose two options from a set of four choices based on what best represents the content of what participants just read, for example, (8i-8iv). The correct answers were (8i) and (8iii).

(8a) Greta had read all the great classics.
(8b) She loved to discuss literature with friends.
(8i) She had read many classic books.
(8ii) She had read many modern books.
(8iii) Greta enjoyed talking about the books she read.
(8iv) Greta enjoyed talking about politics.

For the Predictive Inference section of the rating study, participants read a two-sentence scenario, for example, (9a-b), and they were then presented with four possible outcomes. Participants were asked to choose the most logical outcome from the set of four choices. The instructions were to choose only one option from a set of four choices based on their intuition about what outcome would most likely occur. The correct answer was (9i).

(9a) The dog ran into the middle of the street
(9b) The driver did not see the dog
(9i) The driver hit the dog.
(9ii) The driver hit a tree
(9iii) The driver missed the dog
(9iv) The driver fell asleep

For the No Inference section, participants read three sentences and chose the sentence that did not fit with the others, for example, (10a-c). The correct answer was (10a).

(10a) The students learned how to make prints for art.
(10b) The students learned that the Egyptians built pyramids.
(10c) The students were doing a play about ancient Egypt.

Note that all three sentences have the same subject referent (i.e., the students), but two of them have propositions with related content (i.e., learned that the Egyptians built pyramids, were doing a play about ancient Egypt) versus a proposition with unrelated content (i.e., learned how to make prints for art). Using the same referent (i.e., the students) eliminates the chance that a bridging inference will be formed, as there is no
anaphor that can be connected to the referent. Providing sentences with unrelated, additional information prevents the generation of a predictive inference, as it does not set up the reader to start forming hypotheses about the students’ Egypt unit. The sentence was chosen because it was deemed not to “fit” could, therefore, be paired with either of the remaining two, as it was unrelated in content.

Of the 120 two-sentence scenarios with bridging inferences tested, 6 were eliminated (5%), as their endorsement rate was below 90%. This left 114 two-sentence scenarios and questions. Of the 120 two-sentence scenarios with predictive inferences tested, 13 were eliminated (11%), as their endorsement rate was below 90%. This left 107 two-sentence scenarios and questions. Of the 60 two-sentence scenarios with no inference tested, 5 were revised (8%) and then verified by an outside source.

Once participants verified the content of the inferences, or the lack of inference, in the case of the stimuli with no inferences, a second pretest was carried out with different participants to see how long it took to read the two-sentence scenarios and answer verification questions.

4.4.3.2. Pretest #2: Reaction Time Sentence Pilot. To ensure that stimuli were reliable and valid, a self-paced reading reaction time experiment was carried out in which participants read two-sentence scenarios and answered a verification question about each story. The self-paced reading task let the participants read at their own pace; participants were presented with the first sentence of the scenario, and they had to push a button to reveal the second sentence of the scenario. The participants’ reaction time was their reading time for the two-sentence scenarios and verification questions; reaction time was measured in milliseconds (ms) from the onset of the stimulus (i.e., sentence #1,
sentence #2, or a verification question) and when participants pushed a button to move on to the next stimulus. Of the 281 two-sentence scenarios that were approved or revised from the rating study, the stimuli were further tested: 106 predictive inferences; 1 was used for a practice item, 113 bridging inferences; 1 was used for a practice item, and 59 unrelated; 1 was used for a practice item. Reaction time and accuracy were measured. For example, participants read a two-sentence scenario: (i) *Alex was the top employee in the section.* (ii) *Alex was scheduled for an annual review.* They were then asked a verification question to ensure that they were paying attention: *Did Alex get a good review?* The self-paced reading study took approximately 45 minutes to complete.

Fifteen native speakers of English (14 women and 1 man, $M_{age} = 19.1$ years, $SD_{age} = 1.16$) participated in the reaction-time experiment. All participants listed English as their only native language and verified that the majority of their schooling had been in an English-speaking setting; three participants had a second native language (Ewe, Japanese and Spanish) and all fifteen spoke a second language and sometimes a third language to a greater or lesser extent (including French, Latin, German, Spanish, Hebrew and Ancient Greek). None of the participants had any hearing problems and also had normal to corrected vision. All participants were right-handed (both in writing and cutting), and they did not receive any compensation for participation.

Half of the two-sentence scenarios presented a question that required a “yes” answer and half presented a question that required a “no” answer, so that participants did not develop a response bias. Participants indicated a yes-answer with their right hand. Recall that it was described that different types of verification questions were developed to prevent participants from developing a strategy while completing the task (see Table
4.2 for examples of each type of inference and the accompanying verification questions. The questions for the predictive inferences were split 50% yes and 50% no. The questions for the bridging inferences were split 25% yes (1a-1b), 25% yes (1b-1a), 25% no (1a-1b), and 25% no (1b-1a). The questions for the no inference were split 25% yes (3a), 25% yes (3b), 25% no (3a), and 25% no (3b). Each name (see section 4.4.2) occurred five times in the main experiment, including the practice items. After a verbal informed consent and brief background survey, participants completed the self-paced reading reaction time experiment.

The self-paced reading and verification procedure were as follows, and all the trials in the sentence pilot were approximately 9750 milliseconds (ms) per trial (4500 ms of fixed events + 5250 ms of events that measured reaction time). See Figure 4.1 for a visual display of events for a trial in the reaction-time experiment.

![Figure 4.1. Trial sequence of the reaction-time pilot task.](image-url)
Participants were presented with a blank screen before and after the presentation of the sentences to mimic the conditions of the main experiment where the distractor task would occur.

The goal of the pilot was to assess the reaction time for the two-sentence scenarios and to eliminate two-sentence scenarios in which the incorrect inferences were chosen. For this reason, only two-sentence scenarios stimuli in which the verification questions were correctly answered were analyzed; incorrect responses were considered to be improperly processed or to not have been processed at all. The average reading reaction time was significantly different for the first sentence ($M = 1698$ ms, $SD = 862$ ms) and the second sentence ($M = 2138$ ms, $SD = 1055$ ms), $p < .05$; this would be expected due to wrap-up effects (Just & Carpenter, 1980; Rayner, Kambe, & Duffy, 2000). The average reaction time to answer a verification question correctly was $1451$ ms ($SD = 437$ ms). Stimuli were removed based on a three-tier system. All reaction time outliers were removed based on guidelines from Poarch and van Hell (2012): data points were eliminated if the reaction time was less than 200 ms or more than 2.5 standard deviations (SDs) from the participant’s mean in the task for correct responses only. In the first tier, stimuli were removed if the reaction time for sentence #1 or sentence #2 was above or below 2.5 standard deviations for the individual and the group. In the second tier, stimuli were removed if the reaction time for the questions was above or below 2.5 standard deviations for the individual and the group. In the third tier, stimuli were removed if the error rate on the questions was above 14% (i.e., more than 3/15 wrong) per item. Given the multiple factors, one or more violations resulted in the removal of stimuli.
A total of 78 sentence pairs were eliminated in the sentence reaction time experiment (28%), leaving 200 sentence pairs. The breakdown of eliminated two-sentence scenarios by type can be seen in Table 4.4.

Table 4.4. Breakdown of eliminated two-sentence scenarios based on reaction-time sentence pilot. This includes the total number before the pilot, the number removed (percentage in parentheses) and the number remaining after the pilot.

<table>
<thead>
<tr>
<th></th>
<th># Scenarios Before Pilot</th>
<th># Removed (% Removed)</th>
<th># Scenarios After Pilot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bridging Inferences</td>
<td>112</td>
<td>35 (31%)</td>
<td>78</td>
</tr>
<tr>
<td>Predictive Inferences</td>
<td>105</td>
<td>26 (25%)</td>
<td>79</td>
</tr>
<tr>
<td>No Inference</td>
<td>60</td>
<td>17 (29%)</td>
<td>43</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>277</strong></td>
<td><strong>78 (28%)</strong></td>
<td><strong>200</strong></td>
</tr>
</tbody>
</table>

To review, after the reaction-time sentence pilot, there were 78 two-sentence scenarios with bridging inferences, 79 two-sentence scenarios with predictive inferences, and 43 two-sentence scenarios with no inference.

4.4.4. Final verbal stimuli. Of the verbal stimuli that were approved from the two pretests (200 in total), 186 two-sentence scenarios were used for in the main experiment. There were 36 practice trials (12 with bridging inferences, 12 with predictive inferences and 12 with no inferences). See Appendix B for the finalized verbal stimuli. There were 150 experimental trials (60 with bridging inferences, 60 with predictive inferences and 30 with no inference). Since each experimental block (i.e., condition: control [CONT], STM [short-term memory], CC [cognitive control]) had a mix of different two-sentence scenarios, the mean reaction time (and standard deviation) for
each scenario type was checked by block. See Table 4.5 for a list of mean reaction times and standard deviation by scenario type and block.

*Table 4.5* Final Stimuli mean reaction times (in ms) by scenario type and block. Standard deviations are in parentheses.

<table>
<thead>
<tr>
<th>Scenario Type</th>
<th>CONT Block (no distractor)</th>
<th>STM Block (VPT distractor)</th>
<th>CC Block (N-Back distractor)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bridging inference</td>
<td>3583 (528)</td>
<td>3531 (631)</td>
<td>3566 (572)</td>
</tr>
<tr>
<td>Predictive Inference</td>
<td>3789 (498)</td>
<td>3790 (414)</td>
<td>3741 (390)</td>
</tr>
</tbody>
</table>

It is important to mention that bridging inferences are not directly compared with predictive inferences in the second and third research questions; therefore the 200 ms difference in reaction times between scenario types is not of consequence. Instead, it was important to ensure that the reaction times had the same relative variance and mean reaction time for scenario type across condition.

### 4.5. Experimental Tasks and Materials

#### 4.5.1. Recruitment and Informed Consent Procedure. These two studies were approved on the basis of exempt status by Georgetown University’s Institutional Review Board (IRB) prior to recruitment or testing of participants.

The recruitment process was three-tier for the L1 English speaking participants: I contacted professors of undergraduate linguistics classes via email to ask if I could visit their classes, recruited students in the classes and then scheduled appointments with the individual students. Initially, I contacted professors by email; I described the goals of the study, detailed the criteria for participation (i.e., a right-handed L1 English speaker or an L2 English-L1 Spanish speaker) and the expected requirements of participation (i.e., two 45-minute sessions where participants would complete computerized tasks) and requested
permission to either visit the class to recruit participants for the study. Second, if a professor agreed to let me visit their class, I followed a recruitment script as follows: I explained that I was a Ph.D. student in the Applied Linguistics concentration researching what happens when we read and understand a text. I talked about the importance of understanding what happens when we read, as it has an impact on how we teach children and adult second language speakers reading strategies and how we test reading comprehension. The recruitment script did not detail the exact goals of the study to prevent the participants from adopting strategies for the different research tasks. I went on to describe criteria for participation (detailed above). Finally, I explained that if they chose to participate, they would complete two 45-minute sessions in a lab at Georgetown University in which they would do “fun” activities on the computer, with frequent breaks. I fielded any questions and then passed out a sign-up sheet with dates and times for sessions. Finally, I sent a follow-up message to the participants to confirm their appointments by email or text.

The recruitment process was two-tier for the L2 English participants. First, I contacted undergraduate and graduate departments with Spanish speaking majors via email to ask if they could send out an email to their listserv about my study; I included the information discussed above from the professor initial contact and the recruitment script and requested that the department send out an email about the study to students who fit the criteria. Second, if a student emailed me to volunteer, I verified that they fit the criteria and scheduled an appointment. I sent a follow-up message to the participants to confirm their appointments by email or text.
At the first meeting, I carried out a verbal informed consent procedure (IC) (see Appendix C) before collecting any information, as per the IRB-approved informed consent procedure. In this procedure, participants were instructed to read the IRB IC form; once they finished I reiterated that participation is voluntary and that their information would be kept confidential. I also explained that the participant was essentially giving his or her consent to participate by starting the first experimental task. I offered to give the participant a copy of the IC form.

4.5.2. Background Questionnaire. After giving consent, participants filled out a brief background questionnaire on the computer via SurveyMonkey software, http://www.surveymonkey.com (last visited [March 1, 2014]). Each participant was identified by a unique participant code, so that no information could be personally linked back to him or her. The native (L1) English version and the non-native (L2) English version took roughly the same amount of time to complete. A full version of both questionnaires can be found in Appendices D1 & D2.

The questionnaire for the L1 speakers of English had ten questions and took roughly 3 minutes. It collected standard biodata, such as gender, age, first (L1s) language(s) and second (L2s) languages with self-rated proficiency and amount of use. Finally, participants verified that most of their schooling was in their native language (English).

The questionnaire for the L2 speakers of English had seventeen questions and took roughly 5 minutes; in addition to the information above, it also elicited information about their English usage and proficiency. L2 participants reported their score on an English proficiency exam and the year taken. They also rated their proficiency in the four
areas (reading, writing, listening and speaking), and reported when and how long they
had been studying English and the environment in which they learned English. Lastly, L2
participants gauged their motivation level to master English and how much they enjoyed
speaking in English.

Since L1 participants were native speakers of English who had attended the
majority of their schooling in English, I did not require that they report on their English
usage or proficiency.

4.5.3. Overview of Cognitive Measures. In the two experiments, different
components of participants’ cognitive capacity were both measured independently and as
embedded distractor tasks within a self-paced reading task. All the cognitive measures
used in this dissertation have reportedly high face validity and are commonly used in
cognitive science. The justification for using the tasks, as well as the descriptions of each
of the tasks, will be discussed next.

4.5.3.1. Stand-alone individual difference battery. In the first session of
the main experiment, participants were given a battery of individual difference measures
to triangulate their cognitive capacity along two dimensions. The first dimension was
type of cognitive capacity, tapping into participants’ short-term memory capacity, i.e., the
finite storage capacity used in processing incoming information, and cognitive control
capacity, i.e., the ability to manipulate information by selecting relevant information,
ignoring irrelevant information and comparing contexts for the best fit. The second
dimension was what type of processing was being used, whether language-dependent
cognitive capacity, i.e., processing that is reliant on verbal information or language-
independent cognitive capacity, i.e., processing that is reliant on spatial processing; in the
end, these measures were partially correlated with the non-verbal measures, so they were not used in analyses for this dissertation. Information can be found in Appendix D about the verbal measures, their construct validity, and their normality. In the second session, the main experiment used a distraction paradigm; the language-independent tasks were used for this purpose. A summary of the independent individual difference tasks is listed in Table 4.6.

Table 4.6. Stand-Alone Individual Difference Tasks

<table>
<thead>
<tr>
<th></th>
<th>Short-Term Memory</th>
<th>Cognitive Control</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Language-Independent</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task</td>
<td>Visual Patterns</td>
<td>N-back Task</td>
</tr>
</tbody>
</table>

The task used to measure short-term memory capacity was the Visual Patterns task. The task used to measure cognitive control was an N-back task.

4.5.3.2. Distractor tasks. In the second session, participants completed a self-paced reading experiment using a distraction paradigm. This section focuses on the distractor tasks used in the main experiment. These tasks only varied along one dimension, the type of cognitive capacity. Similar to the stand-alone measures of individual differences, the tasks tapped into participants’ short-term memory capacity and cognitive control capacity. All tasks were language independent, so as not to interfere with the main goal of the experiment, reading a story.

The task used to measure short-term memory capacity in a language-independent manner was the Visual Patterns Task and the task used to measure cognitive control along language-independent dimensions was an N-back Task. The same visual stimuli
were used in the stand-alone measures, however the order of presentation was altered to prevent participants from mastering the task.

4.5.3.3. **Purpose of dual measures of individual differences.** Participants’ performance on these stand-alone measures of individual differences provided an independent measure of participants’ capacity to remember and retrieve stored information, as well as to manipulate that information. These stand-alone measures ultimately allowed for the researcher to triangulate the short-term memory and cognitive control capacities. In addition, assessing participants’ cognitive skills using non-verbal stimuli ensured that reading was still possible with interference. The performance on the individual difference measures and the distractor tasks were correlated to ensure that the distractor tasks were tapping into the same underlying construct as the individual difference measures.

4.5.3.4. **Cognitive Measures Pretest.** A pretest of all the cognitive measures was carried out with fifteen native speakers of English (9 women and 6 men, \( M_{\text{age}} = 19.7 \) years, \( SD_{\text{age}} = 1.79 \)) to test face and construct validity. These tasks were used as the stand-alone individual difference measures and as distractor tasks in the main experiment. Participants completed two tasks; each lasted roughly 5-8 minutes, and the entire session lasted approximately 20 minutes. Participants were first led through the informed consent procedure and directed to fill out the background questionnaire. The participant background information is briefly outlined in this section.

All participants listed English as their only native language and verified that the majority of their schooling had been in an English-speaking setting; four participants had a second native language (either Vietnamese or Spanish) and all fifteen spoke a second or
third language (including French, Spanish, Russian, Japanese, Chinese and Portuguese). None of the participants had any hearing problems, and all had normal to corrected vision. All participants were right-handed (both in writing and cutting), and they did not receive any compensation for participation. For reaction time data, outliers were removed according to guidelines from Poarch and van Hell (2012): data points were eliminated if the reaction time was less than 200 ms or more than 2.5 standard deviations (SDs) from the participant’s mean in the task for correct responses only.

The procedure and results of the pretest will be described for each individual task in their respective sections.

4.5.4 Individual Difference Test Battery.

4.5.4.1. Visual patterns task. Language-independent short-term memory capacity, or spatial memory, was measured using the Visual Patterns Task (Della Sala, Gray, Baddeley, & Wilson, 1997). It is also referred to as visual short-term memory or visuospatial short-term memory. In the visual patterns task, participants saw a figure after which they were presented with a second figure; their task was to say whether the figures matched or not by pressing a button. The figures used in this task were 3 x 3 grids with three shaded-in squares. There were two differences from the procedure in Della Sala et al. (1997). First, the number of squares did not increase throughout the task; it was fixed at nine squares total (a 3 x 3 grid). This modification was employed to accommodate the processing demands in the dual-task paradigm. The difficulty level of the distractor task (i.e., the visual patterns task) could not be so difficult that the participants were unable to complete the primary task (i.e., reading the sentences), but not so easy that it did not serve the purpose of interrupting the participants’ spatial short-term memory. Second, in
Della Sala et al. (1997), the participants reproduced what they saw on paper after seeing the figure. In the version used in this dissertation, participants indicated whether it matched or did not match by a button press. Similar to the reason stated above, reproducing the figure on paper would have been time-consuming and overly taxing to participants’ attention, making the primary task extremely difficult to complete.

The visual patterns task assesses a participant’s ability to perceive, encode and retrieve spatial information. Spatial memory is a cognitive process that allows for recall of different locations, as well as spatial relations between objects. This fits into a number of different memory-based models, and can be situated in Baddeley’s (2000) multi-component working memory model with the master component (i.e., the central executive) and three slave systems (i.e., visuospatial sketchpad, phonological loop and the episodic buffer) or unitary models of memory (see Jones, Stuart & Morris, 1995). It is important to note that few tasks are purported to measure visuospatial short-term memory alone (Pickering, 2001); most have a component of working memory. This can be understood at the level of intuition, as processing requires the perception and encoding of information, but it also requires the retrieval and use of said information. Despite the difficulty isolating this sub-skill, the visual patterns task is largely accepted and used to measure visual short-term memory.

The visual patterns task was developed on the experiment-generating package, SuperLab (Version 4.0; Cedrus Corporation; San Pedro, CA) and presented via this software. Participants sat in front of a computer screen and were presented with a figure (presentation screen). After a blank screen, participants saw a second figure (verification screen). They indicated whether the figure matched or did not match with a button press.
Reaction time and accuracy were measured. The task took roughly 10 minutes to complete. See Figure 4.2 for an example of the figures in a correct (i.e., congruent) and an incorrect (i.e., incongruent) trial. All the visual stimuli for the visual patterns task were created using Adobe Illustrator; each figure was 222 x 222 pixels and placed in the center of the screen. All the figures were a 3 x 3 matrix of squares with three shaded-in squares. The 3 x 3 grids used in the visual patterns task were modeled from a standard template to ensure that the grids had the same size and placement. 62 different grids were created; in the trials that did not match, only pairings that were easily discernable were chosen.

There were 60 randomized trials in the task, 30 congruent and 30 incongruent, and two practice trials.

![Figure 4.2. Congruent and Incongruent Samples of The Visual Pattern Task](image)

The procedure was as follows for all the trials in the visual patterns task, approximately 6500 milliseconds (ms) per trial (4500 ms of fixed events +/- reaction time for button press). See Figure 4.3 for the trial sequence of the visual patterns task.
Figure 4.3. Sequence for a correct trial in the Visual Patterns task.

The pretest results showed that the overall accuracy for the task was quite high; on average, 56.9 out of 60 items were correctly responded to \((M = .95, SD = .03)\). Only correct answers were analyzed; the reaction time for correctly responded to items was 788 ms \((SD = 114)\). All reaction time data were treated in that any that fell above or below 2.5 standard deviations was removed \((M_{\text{items removed}} = 2.47, SD_{\text{items removed}} = .95)\).

The task was perceived as fairly straight-forward, but based on the pretest, the instructions were modified to specify that the orientation of the 3 x 3 grid was important; if the image was a mirror image, then the participants were instructed that the figures did not match.
In contrast to the tasks described above, in which either phonological or visual representations need to be held in memory, the cognitive control task also require that participants keep track of current stimuli and compare it to previous stimuli, with conflict resolution of divergent interpretations. In the next sections, the tasks that measured cognitive control will be described.

4.5.4.2. N-Back Task. The task used to measure language-independent cognitive control was the N-Back task (Kirchner, 1958). In the task, participants were shown a series of non-verbal stimuli and were asked to press a button when the stimulus was the same as one that was n steps previous. In this task, participants were asked to remember two (2) back from the stimulus. Kane, Conway, Miura and Colflesh (2007) tested the construct validity in an experimental-correlational study and found weak correlations to the working memory span test. Yet, when looking at a wider collection of literature, correlations between the N-back task and other recognized measures of working memory capacity have ranged from .13 to .55 (see Schmeichel, Volokov & Demaree for an overview; individual studies were carried out by Kane, Conway, Miura, & Colflesh, 2007; Oberauer, 2005; Shelton, Metzger, & Elliott, 2007). The N-Back task was considered a valid measure of the cognitive control construct, as it required that participants select relevant information and simultaneously inhibit irrelevant information.

The N-back task was developed on the experiment-generating package, SuperLab (Version 4.0; Cedrus Corporation; San Pedro, CA) and presented via this software. The entire task took approximately 5 minutes to complete. Participants sat in front of a computer screen and were presented with a series of stimuli. The shapes used were a circle, a square and a triangle. Participants verified whether the shape on the screen was
the same as the one they saw two (2) back. See Figure 4.4 for a visual depiction of the n-back principle; please note this graphic was provided to the participants in the task instructions. All the shapes used in the n-back task were of comparable shape and at a central placement on the screen (322 x 233 pixels).

All the visual stimuli for the visual patterns task and the n-back distractor tasks were created using Adobe Illustrator. They were placed in the center of the screen.

Your job is to press the “/” key each time the shape is the same as the shape you saw two screens back.

Press the SPACE BAR to continue.

Figure 4.4. Visual depiction of the N-back principle for participants. The graphic was used in the task instructions to demonstrate when it was necessary to push the button. There were a total number of 60 trials (and 3 practice trials). The participants were required to press the button a total of 19 times (32% of the trials) to avoid a response bias. The shapes all appeared at the same proportion (the circle appeared 20 times, the triangle appeared 19 times, and the square appeared 21 times).
The procedure was as follows for all the trials in the N-back task, approximately 4500 milliseconds (ms) per trial (2500 ms of fixed events +/- reaction time for button press). See Figure 4.5 for the trial sequence of the N-Back task.

![Diagram of trial sequence](image)

*Figure 4.5. Trial sequence of the N-back task.*

The pretest revealed that the mean accuracy for the task was quite low, perhaps because the task is inherently more cognitively complex: 50.5 out of 60 items were responded to correctly ($M = .84$, $SD = .08$). Only correct answers were analyzed; the mean reaction time for correctly recognizing that the figure is the same 2-back was 593 ms ($SD = 113$). All reaction time data were treated in that any that fell below 200 ms or above 2.5 standard deviations was removed ($M_{items\_removed} = 1.08$, $SD_{items\_removed} = .28$).

The N-back task was changed to clarify instructions for the participants; the purpose of the fixation point (+) was explained, so participants did not think it was one of the shapes they had to monitor.
In sum, the tasks used in the individual difference battery were as follows: the visual pattern task to measure visual short-term memory capacity, and the N-back task to measure non-verbal cognitive control.

In the next section, the tasks used in the distraction paradigm will be described.

**4.5.5. Distractor Tasks.** All the tasks used as distractor tasks were language independent, so as not to interfere with the reading process. The distractor tasks used were the Visual Patterns Task and the N-back task. Each task used the same visual stimuli as its corresponding individual difference measure, but new versions were created to avoid task effects.

**4.5.5.1. Visual Patterns Task.** The task used to interfere with visual short-term memory (i.e., the storage of representations) during the reading task was the visual patterns task. There were a total number of 50 randomized trials to remain consistent with the number of trials per block in the main study. The visual stimuli used in the distractor task were identical to those used in the individual difference measure described above (i.e., there were 62 3 x 3 grids with three squares shaded in); the only modification was that a new version was created with different pairings of 3 x 3 figures to create the congruent and incongruent trials. In this way, even though the 3 x 3 grids were similar across tasks, the pairings were different. The complete procedure will be described in the section on the cumulative self-paced reading task.

**4.5.5.2 N-Back task.** The task used to interfere with language-independent cognitive control during the reading task was the N-back task. This task also had 50 trials to match the number of trials per block in the main study. The visual stimuli used in the distractor task were also a square, a triangle and a circle (identical to the individual
difference task); yet a new version was created with a new presentation order. The participants were required to press the button a total of 16 times (32% of the trials) to avoid response bias. The shapes all appeared at roughly the same proportion (the circle appeared 16 times, the triangle appeared 17 times and the square appeared 17 times). The procedure will be detailed in the section on the cumulative self-paced reading task.

In summary, the distractor tasks used were the visual pattern task to interrupt language-independent short-term memory capacity and the N-back task to interrupt language-independent cognitive control.

**4.5.6. Correlations of Individual Measures.** Independent measures were collected to assess each participant’s capacity to remember, retrieve, and manipulate stored information along cognitive control and short-term memory dimensions. The distraction paradigm did not permit the use of verbal tasks as distractors because participants had to complete two tasks within the same trial; participants needed to process and understand the two-sentence scenarios while completing a secondary distractor task. For this reason, a non-verbal cognitive control task and a non-verbal short-term memory task were used as distractors for the self-paced reading task. As section 4.5.3.3. described, participants’ performance on the stand-alone measures of individual differences were collected to gauge the construct validity of the cognitive control distractor (the N-Back task) and the construct validity of the short-term memory distractor (the Visual Patterns task). Participants also completed verbal (i.e., language dependent) measures of short-term memory and cognitive control. Of the verbal measures collected, only the stand-alone verbal and non-verbal cognitive control measures were correlated (see Appendix D3) and the verbal and non-verbal short-term memory tasks
were not at all correlated; therefore they were not used in any subsequent analyses for the
two studies in this dissertation. The entire participant population \((N = 73)\) completed all
the tasks. Therefore, all the participants were used in the correlational analyses to
increase the power of the test.

Bivariate correlations shed light on whether two variables are associated by
investigating the degree that two continuous variables covary. The five assumptions of
parametric correlations are that the data are independently collected (i.e., one person’s
data does not influence the data of another), the data are assumed to be interval, the data
are assumed to be normally distributed, the relationship between the variables is linear,
and finally, the data should have homoscedasticity (i.e., the variances of their residuals
are comparable). Since the accuracy and reaction time data were checked for the
assumptions of correlations in chapter 5, it will not be reviewed here. The data do not
adhere to the assumptions of parametric data; therefore Spearman’s Rho will be used for
the correlations.

The construct validity of the stand-alone tasks and the embedded secondary tasks
was checked via correlations of the stand-alone measure and its respective distractor task.
Table 4.7 shows the correlations between the stand-alone and distractor tasks.

*Table 4.7. Correlations (using Spearman’s rho) of stand-alone and distractor tasks.*

<table>
<thead>
<tr>
<th>Stand-Alone Tasks</th>
<th>Distractor Tasks</th>
<th>N-Back Accuracy</th>
<th>N-Back RT</th>
<th>VPT Accuracy</th>
<th>VPT RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-Back Accuracy</td>
<td>0.2 (0.095)</td>
<td>-0.261 (0.027)</td>
<td>0.258 (0.029)</td>
<td>-0.137 (0.252)</td>
<td></td>
</tr>
<tr>
<td>N-Back RT</td>
<td>0.076 (0.524)</td>
<td>0.268 (0.022)</td>
<td>-0.029 (0.809)</td>
<td>0.318 (0.006)</td>
<td></td>
</tr>
<tr>
<td>VPT Accuracy</td>
<td>0.215 (0.074)</td>
<td>-0.243 (0.041)</td>
<td>0.209 (0.08)</td>
<td>-0.252 (0.034)</td>
<td></td>
</tr>
<tr>
<td>VPT RT</td>
<td>-0.201 (0.093)</td>
<td>0.384 (0.001)</td>
<td>-0.142 (0.234)</td>
<td>0.175 (0.141)</td>
<td></td>
</tr>
</tbody>
</table>
The stand-alone tasks were correlated with the distractor tasks, in that each stand-alone task had some relationship with its distractor equivalent. The stand-alone N-Back Accuracy was negatively correlated with the distractor N-Back RT, $r = -.261, p < .05$; this indicates that the faster participants were on the distractor N-Back task, the less accurate they were on the stand-alone N-Back task. In addition, the stand-alone N-Back RT was positively correlated with the distractor N-Back RT, $r = .268, p < .05$, showing similar patterns between the two sets of reaction times. Similar to what we find with the N-Back task, the stand-alone VPT accuracy was negatively correlated with the distractor VPT RT, $r = -.252, p < .05$; this shows that the faster participants were to complete the distractor task, the less accurate they were on the stand-alone task.

However, the same types of relationship were also found for the stand-alone VPT accuracy and the distractor N-Back RT, $r = -.243, p < .05$ as well as the stand-alone VPT RT and the distractor N-Back RT, $r = .384, p < .05$. In addition, the stand-alone N-Back accuracy was positively correlated with the distractor VPT Accuracy, $r = .258, p < .05$, indicating that participants who were accurate on the stand-alone N-Back task were also accurate on the distractor VPT accuracy. Finally, there was a positive correlation between the stand-alone N-Back RT and the distractor VPT RT, $r = .318, p < .05$.

Taken together, this indicates that the stand-alone and distractor tasks are related to each other. The correlations must be interpreted carefully, as the magnitude of the shared variance is small (between .243-.384). It should be noted that the distraction paradigm doubly taxed participants’ cognitive resources; they needed to complete the distractor task and read two-sentence scenarios with the aim of answering a verification question. Nevertheless, since the stand-alone and distractor tasks for each cognitive
construct were all but identical in stimuli and directions, barring the order of the stimuli, they are thought to tap into the same underlying construct. These correlations were run to see the strength of the relationships; it appears that the dual task paradigm affected the reaction time and accuracy of the participants. Despite these small cross-task correlations, the stand-alone N-Back task and the distractor N-Back were correlated and appear to be tapping into the same underlying construct, as were the stand-alone VPT and the distractor VPT. Implications regarding the construct validity of the two stand-alone cognitive measures are discussed in the limitations section in Chapter 7.

4.5.7. Cumulative Self-Paced Reading Task. Using a distraction paradigm, participants completed a cumulative self-paced reading task (see Appendix B for a complete set of stimuli) with verification questions. The distraction paradigm functions as follows: participants read two-sentence scenarios and answered verification questions while completing a secondary task within the same trial. The secondary task acts as a distractor, ultimately splitting the participants’ cognitive resources between two tasks. The cumulative self-paced reading task allowed participants to read at their own speed: for each of the two-sentence scenarios, participants were first presented with one sentence; upon a button press, a second sentence was presented. Participants’ reaction times for reading time on each sentence were measured. As such, the L1 and L2 experiments were designed to investigate how readers perform when distracted along different dimensions, in order to test the hypotheses that bridging inferences rely primarily on visual or phonological short-term memory, and predictive inferences rely primarily on cognitive capacity. The test stimuli consisted of two-sentence scenarios that required bridging inferences, predictive inferences or sentence pairs without inferences to
be understood under three conditions: control (i.e., no distractor task),
visual/phonological short-term memory distractor, a task that taxes the concepts stored in
the visuospatial sketchpad (i.e., the visual patterns task) and cognitive control distractor, a
task that taxes cognitive control (i.e., the N-back task). A summary of the distractor tasks
is listed in Table 4.8.

*Table 4.8.* Experimental Design by Condition and Inference Type

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Short-Term Memory Distractor</th>
<th>Cognitive Control Distractor</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bridging Inference</strong></td>
<td>No Secondary Task</td>
<td>Visual Patterns Task</td>
<td>N-back Task</td>
</tr>
<tr>
<td><strong>Predictive Inference</strong></td>
<td>No Secondary Task</td>
<td>Visual Patterns Task</td>
<td>N-back Task</td>
</tr>
<tr>
<td><strong>No Inference</strong></td>
<td>No Secondary Task</td>
<td>Visual Patterns Task</td>
<td>N-back Task</td>
</tr>
</tbody>
</table>

The conditions were presented in blocks, with randomized trials (of inference
type). The conditions were not mixed within blocks as the cognitive control distractor
task occurs across trials and cannot be presented with other conditions without making
the entire block a strain on cognitive control. The conditions were presented in blocks of
50 trials (with a practice session of 12 trials) in which participants read two sentences in
succession and answered a verification question at the end of each trial. The pilot of the
main task procedure (N = 5) revealed that participants found it more manageable to start
with the control block (without distractor tasks) and then to add secondary tasks in the
subsequent blocks. Therefore, the presentation of the blocks was pseudo randomized for
each participant; each participant started with the control block and then was randomly
presented with either the cognitive control distractor block or the short-term memory
distractor block.
The placement of the distractor tasks, before the presentation of the two-sentence scenarios, was meant to interrupt the processing of the sentences to maximally see the effect of the different distractor tasks. It is important to note that the distractor tasks functioned quite differently. One difference between the tasks was that the cognitive control task only presented visual stimuli before sentence presentation and a blank afterward, as participants are required to remember information across several trials. An additional way that the distractor tasks were dissimilar was whether the task was resolved within one trial or across several trials. The short-term memory distractor was presented within a trial. The cognitive control distractor was presented across several trials.

All sentences were displayed in black on a white background in size 28 Times New Roman Font. The experiment was developed and presented using the experiment-generating package, SuperLab (Version 4.0; Cedrus Corporation; San Pedro, CA) and run on a Macintosh laptop with pseudo-randomized presentation. Responses to the verification question were entered via a keyboard; fifty percent had a yes response and fifty percent had a no response. Participants used their writing hand for the yes response. The trials were randomized for each participant to prevent associations between consecutive items, except in the cognitive control distractor that had a fixed order due to the nature of the task. Reaction times were measured for the reading screens and the verification question. See Figure 4.6 for the visual events of a single trial in the control condition (without distraction), Figure 4.7 for the visual events of a single trial in the STM condition (with the distraction of the visual patterns task), and Figure 4.8 for the visual events of a single trial in the CC condition (with the distraction of the N-Back task).
Figure 4.6. Visual events of a single trial in the control condition

Figure 4.7. Visual events of a single trial in the STM condition (with the visual patterns task)
Figure 4.8. Visual events of a single trial in the CC condition (with the N-Back task)

Based on the average reaction times from the pilots, the self-paced reading procedure is further broken down for each condition; a trial in each condition was 8500-9000 milliseconds. The self-paced reading and verification procedure was as follows for a trial in the control condition, approximately 8500 milliseconds (ms) per trial:

- Fixation cross (+) in middle of screen (1000 ms)
- Presentation of first sentence (until button press, $M = 1698$ ms)
- Presentation of second sentence (until button press, $M = 2138$ ms)
- Blank screen (2000 ms)
- Yes/no verification question (until response/5000 ms, $M = 1451$ ms)

The self-paced reading and verification procedure was as follows for a trial in the short-term memory distractor condition, approximately 9000 ms per trial:

- Fixation cross (+) in middle of screen (1000 ms)
• Introduction of distractor task (2000 ms)
• Presentation of first sentence (until button press, $M = 1698$ ms)
• Presentation of second sentence (until button press, $M = 2138$ ms)
• Completion of distractor task (until button press/5000 ms, $M = 788$ ms)
• Yes/no verification question (until response/5000 ms, $M = 1451$ ms)

The self-paced reading and verification procedure was as follows for a trial in the cognitive control distractor condition, approximately 9000 ms per trial:

• Fixation cross (+) in middle of screen (1000 ms)
• Introduction of distractor task (2000 ms / button press, $M = 593$ ms)
• Presentation of first sentence (until button press, $M = 1698$ ms)
• Presentation of second sentence (until button press, $M = 2138$ ms)
• Blank screen (2000 ms)
• Yes/no verification question (until response/5000 ms, $M = 1451$ ms)

4.6. Procedure and Data Collection

The entire study took approximately 55 minutes for both the L1 and the L2 study; participants took part in two sessions, a 20-minute session and 35-minute session, on a Macintosh computer. In the first session, participants gave verbal informed consent, filled out the Language Background Form and completed four stand-alone individual difference tasks. In the second session, participants completed the main experiment with integrated practice before each block. Between all the tasks, participants were offered breaks. Since the sessions were fairly short, most participants took a 1-2 minute break between each task. Figure 4.9 shows the schedule of tasks the participants completed.
Figure 4.9. Overview of research procedure
4.7. Summary of variables

The L1 study and the L2 study have the same set of research questions, focusing on three issues: (1) whether inferences are processed differently than non-inferences; (2) whether cognitive resources are related to the processing of inferences and non-inferences; and (3) whether bridging inferences rely on different cognitive capacity than predictive inferences.

The first research question has the independent variable of scenario type (non-inference, bridging inference and predictive inference), and the dependent variable is the participants’ reaction times in the control block, when participants were not distracted. The second research question has two independent variables, scenario type (no inference, bridging inference and predictive inference) and condition (control/none [CONT], cognitive control distractor [CC] and short-term memory distractor [STM]). The dependent variable was reaction time. The third research question had two independent variables, type of distractor task used in each condition (control/none [CONT], cognitive control distractor [CC] and short-term memory distractor [STM]) and type of inference scenario (bridging or predictive), while the dependent variable was reaction time for reading. The co-variables were the independent non-verbal measures of cognitive control and short-term memory (reaction time and accuracy on the N-Back task and Visual Patterns task).

Chapters 5 and 6 presents the results of the analyses that were conducted to answer the research questions for the L1 group and the L2 group respectively. Chapter 7 presents a discussion of the results and some possibilities for future analysis, along with a conclusion.
5.1. Introduction

This chapter presents the results of the statistical analyses used to address the research questions in the L1 study. The software used to carry out these analyses was PAW (Predictive Analytics SoftWare) Statistics 21, from SPSS: An IBM company; the alpha level was set at .05 for all tests. This chapter first examines the data for normality. Next, each research question is reviewed, including the assumptions and parametric statistics used to answer the research questions, the alternate nonparametric statistical method used (only when the assumptions were not met for parametric tests), the descriptive statistics (e.g., mean, range, standard deviations) for all the predictor and dependent variables, and the results of the inferential statistics. Nonparametric statistics serve as acceptable alternatives to parametric statistics if their limitations are understood; they are valuable because they rely on fewer assumptions while approaching the exactness of parametric procedures, yet they require a larger sample size than similar parametric tests to reach the same power and often discard useful information that a parametric procedure would otherwise use (Siegel & Castellan, 1988).

5.2. Assumptions of parametric statistical methods

Prior to analysis, the reaction time and accuracy data for the stand-alone measures (i.e., N-back, Visual Patterns), the self-paced reading task in all three conditions (i.e., control, short-term memory distractor and cognitive control distractor), and the embedded distractor tasks (i.e., N-back and Visual Patterns tasks) were preprocessed, pretreated, and checked for normality. The following statistical tests were used in the study: repeated-measures analyses of variance (ANOVA) and correlational analyses.
5.2.1. **Assumptions of repeated measures analyses of variance.** It is appropriate to use repeated measures ANOVA when participants contribute more than one measure of the dependent variable and two or more means are being compared. Repeated-measures ANOVA tests have several assumptions, including normal distribution of data, and interval data and sphericity (Tabachnick & Fiddell, 1996). Normal distribution implies that the data falls on a normal curve. Interval data can be ordered on a scale while having equal intervals on that scale (i.e., on a scale of 1-100, the distance between 3 to 4 and 70-71 are the same). Sphericity can be claimed when the variances of the differences between different pairs of groups are equal. Since all the research questions use repeated-measures ANOVA, the predictor and dependent variables will be checked for all assumptions for repeated-measures ANOVA in section 5.3.

5.2.2. **Assumptions for correlational analyses.** The correlation coefficient can be used when data meets four assumptions: it should be normally distributed, there should be a linear relationship between the two (or more) variables, there are no significant outliers and the data are interval. Since several research questions use correlation, the predictor and dependent variables will be checked for all assumptions for correlational analyses in the sections with the relevant research questions.

5.3. **Data Screening**

All the accuracy and reaction time data were checked for normality; the additional assumptions for the parametric statistics were tested within the research question itself (i.e., sphericity, homogeneity of variance).

5.3.1. **Accuracy Data.** Only trials where the task stimulus was answered correctly were analyzed because incorrect button presses indicate that the stimulus was either not
correctly processed in context or that it was not processed at all. It is possible that participants accidentally pressed the wrong button, however since it is not possible to determine whether the stimulus was understood or not in such cases, all incorrect responses were discarded. For the cognitive tasks, correct trials were when the participant indicated the correct match (in the case of the visual patterns task) or when the participant pressed a button for the shape that was seen two back (in the case of the N-back task). For the self-paced reading task, trials where the verification question were answered correctly were analyzed, as it shows that the sentences were read and comprehended.

First, the assumption of normality was investigated using visual inspection, descriptive statistics and z-scores for skew and kurtosis. Table 5.1 shows the means, ranges, standard deviations, standard errors, and 95% confidence intervals for each measure on accuracy for the L1 group. To further explore the data’s normality, Table 5.2 shows skewness and kurtosis values/ratios and the results of Kolmogorov-Smirnov and Shapiro-Wilkes goodness-of-fit tests for the L1 group.
Table 5.1. Descriptive statistics for the raw accuracy scores in the experimental tasks for the L1 group. Percentages in parentheses.

<table>
<thead>
<tr>
<th>Task</th>
<th>Condition</th>
<th>Mean (%)</th>
<th>Range</th>
<th>SD</th>
<th>SE</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-Back Task*</td>
<td></td>
<td>51.82 (86%)</td>
<td>14</td>
<td>3.27</td>
<td>0.46</td>
<td>51 - 53</td>
</tr>
<tr>
<td>VPT*</td>
<td></td>
<td>56.55 (94%)</td>
<td>7</td>
<td>2.07</td>
<td>0.30</td>
<td>56 - 57</td>
</tr>
<tr>
<td>Self-Paced Reading Task Control</td>
<td></td>
<td>47.08 (94%)</td>
<td>10</td>
<td>2.43</td>
<td>0.34</td>
<td>47 - 48</td>
</tr>
<tr>
<td>Self-Paced Reading Task STM</td>
<td></td>
<td>46.62 (93%)</td>
<td>9</td>
<td>2.22</td>
<td>0.31</td>
<td>46 - 47</td>
</tr>
<tr>
<td>Self-Paced Reading Task CC</td>
<td></td>
<td>47.32 (95%)</td>
<td>7</td>
<td>1.86</td>
<td>0.26</td>
<td>47 - 48</td>
</tr>
<tr>
<td>N-Back Task**</td>
<td></td>
<td>39.06 (78%)</td>
<td>24</td>
<td>5.88</td>
<td>0.83</td>
<td>37 - 41</td>
</tr>
<tr>
<td>VPT**</td>
<td></td>
<td>47.98 (96%)</td>
<td>9</td>
<td>1.84</td>
<td>0.26</td>
<td>47 - 49</td>
</tr>
</tbody>
</table>

* Stand-alone task ** Distractor-task
Table 5.2: Goodness-of-fit tests to check for normality of distributions on accuracy of tasks for L1 group

<table>
<thead>
<tr>
<th>Task</th>
<th>Condition</th>
<th>Skew</th>
<th>SE</th>
<th>Ratio</th>
<th>Kurtosis</th>
<th>SE</th>
<th>Ratio</th>
<th>Statistic</th>
<th>p-value</th>
<th>Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Skewness</td>
<td></td>
<td>Kurtosis</td>
<td></td>
<td></td>
<td></td>
<td>Kolmogorov-Smirnov (K-S)⁺</td>
<td></td>
<td>Shapiro-Wilk (S-W)</td>
<td></td>
</tr>
<tr>
<td>N-Back Task*</td>
<td></td>
<td>-0.98</td>
<td>0.34</td>
<td>-2.90</td>
<td>0.64</td>
<td>0.66</td>
<td>0.97</td>
<td>0.20</td>
<td>0.00</td>
<td>0.90</td>
<td>0.00</td>
</tr>
<tr>
<td>VPT*</td>
<td></td>
<td>-0.79</td>
<td>0.34</td>
<td>-2.31</td>
<td>-0.19</td>
<td>0.67</td>
<td>-0.28</td>
<td>0.17</td>
<td>0.00</td>
<td>0.89</td>
<td>0.00</td>
</tr>
<tr>
<td>Self-Paced Reading Task</td>
<td>Control</td>
<td>-0.90</td>
<td>0.34</td>
<td>-2.68</td>
<td>0.31</td>
<td>0.66</td>
<td>0.47</td>
<td>0.19</td>
<td>0.00</td>
<td>0.90</td>
<td>0.00</td>
</tr>
<tr>
<td>Self-Paced Reading Task</td>
<td>STM</td>
<td>-1.12</td>
<td>0.34</td>
<td>-3.31</td>
<td>0.71</td>
<td>0.66</td>
<td>1.08</td>
<td>0.23</td>
<td>0.00</td>
<td>0.86</td>
<td>0.00</td>
</tr>
<tr>
<td>Self-Paced Reading Task</td>
<td>CC</td>
<td>-0.33</td>
<td>0.34</td>
<td>-0.98</td>
<td>-0.76</td>
<td>0.66</td>
<td>-1.15</td>
<td>0.17</td>
<td>0.00</td>
<td>0.94</td>
<td>0.01</td>
</tr>
<tr>
<td>N-Back Task**</td>
<td></td>
<td>-0.39</td>
<td>0.34</td>
<td>-1.14</td>
<td>-0.52</td>
<td>0.66</td>
<td>-0.79</td>
<td>0.08</td>
<td>0.20</td>
<td>0.97</td>
<td>0.17</td>
</tr>
<tr>
<td>VPT**</td>
<td></td>
<td>-1.74</td>
<td>0.34</td>
<td>-5.11</td>
<td>4.01</td>
<td>0.67</td>
<td>6.00</td>
<td>0.25</td>
<td>0.00</td>
<td>0.82</td>
<td>0.00</td>
</tr>
</tbody>
</table>

* Stand-alone task       ** Distractor-task       + Performed with the Lilliefors correction
The accuracy data shows the symmetry of the scores around the mean were clustered at the high end of the scale; this is evidence that participants were, by and large, more accurate than not on the experimental tasks. Since the data are negatively skewed, it is not considered to be normally distributed. This can be seen in all accuracy scores for the stand-alone tasks (See Tables 5.1 and 5.2); the means are quite high with highly constrained standard deviations for most tasks. It is important to note that the measures were constructed to assess a skill that should have been readily accessible to the participants. The K-S test showed that the L1 group’s accuracy data were not normally distributed for the majority of the tasks: the N-back task (N-Back**), $D(48) = 0.20, p < .05$; the stand-alone visual patterns task (VPT**), $D(48) = 0.17, p < .05$; the control condition of the self-paced reading task, $D(48) = 0.19, p < .05$; the short-term memory distractor (STM) condition of the self-paced reading task, $D(48) = 0.23, p < .05$; the cognitive control distractor (CC) condition of the self-paced reading task, $D(48) = 0.17, p < .05$; and the distractor visual patterns task (VPT*), $D(48) = 0.25, p < .05$. One task appears to be normally distributed: the distractor N-Back task, $D(48) = 0.08, p > .05$.

Another way to assess whether the data are normally distributed is to convert the skew and the kurtosis values into a z-score; this is calculated by dividing the skewness/kurtosis value by its standard error. This not only shows how several measures that potentially use different scales compare, but also the likelihood of skewness or kurtosis occurring. Field (2009) suggests that larger sample sizes are prone to Type I error and, as such, even small deviations from normality can yield significant values. For this reason, any values below 2.58 will be considered normally distributed; in the case that a value is above the threshold, it will be visually inspected. Note that many of the converted skewness or
kurtosis z-scores were above or approaching 3.29; upon visual inspection, and based on their performance on the above tests, they violate this measure of normality. For this reason, all inferential statistics using accuracy data will use nonparametric statistics.

5.3.2. Reaction Time (RT) Data. Reaction times (RT) were measured for all tasks. The RTs of the reading time of the first sentence and the second sentence of the two-sentence scenarios in the self-paced reading task were measured; they represent the amount of time (in milliseconds) from the onset of the first sentence and the depressing of the space bar to display the onset of the second sentence and the onset of the second sentence and the depressing of the space bar to display the next screen. In this way the reaction time for the first and second sentence were separately measured. For the cognitive tasks, whether stand-alone or distractor tasks, the participant pressed a button to indicate that the stimulus matched a stimulus (or not), the reaction time represents the amount of time (in milliseconds) from the onset of the presentation of the stimulus and the depressing of the correct button. Reaction time that were +/- 2.5 standard deviations from the mean were removed for each participant, resulting in a loss of 2.6% reaction time data from the L1 group on all tasks.

To address the assumption of normality, data were first visually inspected using histograms and probability-probability (P-P) plots. Figures 5.1 and 5.2 provide a view of how the data are distributed.
Upon visual inspection, the cognitive measures tend to have fairly symmetrical distribution. In contrast to the accuracy data, the reaction times tended toward the lower (i.e., faster) end of the distribution, showing that the majority of the participants
responded fairly quickly to the stimulus. This can be seen in all the stand-alone tasks (see Table 5.3). Participants were instructed to respond as quickly as possible to the stimuli, and faster reaction times would be expected.

The self-paced reading task also shows that the conditions are relatively in the same range, with a positive skew toward higher RTs. Table 5.3 shows the means, ranges, standard deviations, standard errors, and 95% confidence intervals for each measure, as well as the data removed for the L1 group by task. To further explore whether the data meets normality, Table 5.4 shows skewness and kurtosis values/ratios and the results of Kolmogorov-Smirnov goodness-of-fit test.
Table 5.3. Descriptive statistics for the reaction times (in milliseconds) in the experimental tasks for the L1 group

<table>
<thead>
<tr>
<th>Task</th>
<th>Condition</th>
<th>Mean</th>
<th>Range</th>
<th>SD</th>
<th>SE</th>
<th>95% Confidence Interval</th>
<th>% RT data removed</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-back*</td>
<td></td>
<td>632</td>
<td>838</td>
<td>189</td>
<td>27</td>
<td>571 - 682</td>
<td>1.5%</td>
</tr>
<tr>
<td>VPT*</td>
<td></td>
<td>709</td>
<td>516</td>
<td>118</td>
<td>17</td>
<td>674 - 741</td>
<td>2.7%</td>
</tr>
<tr>
<td>Self-Paced Reading</td>
<td>Control</td>
<td>3004</td>
<td>2575</td>
<td>108</td>
<td>749</td>
<td>3580 - 3998</td>
<td>5.0%</td>
</tr>
<tr>
<td>Self-Paced Reading</td>
<td>STM</td>
<td>2590</td>
<td>2214</td>
<td>93</td>
<td>647</td>
<td>3164 - 3541</td>
<td>2.1%</td>
</tr>
<tr>
<td>Self-Paced Reading</td>
<td>CC</td>
<td>4280</td>
<td>2775</td>
<td>132</td>
<td>923</td>
<td>4047 - 4584</td>
<td>4.5%</td>
</tr>
<tr>
<td>N-back**</td>
<td></td>
<td>885</td>
<td>955</td>
<td>185</td>
<td>26</td>
<td>853 - 958</td>
<td>1.1%</td>
</tr>
<tr>
<td>VPT**</td>
<td></td>
<td>826</td>
<td>814</td>
<td>154</td>
<td>22</td>
<td>785 - 875</td>
<td>2.3%</td>
</tr>
</tbody>
</table>

* Stand-alone task  ** Distractor-task
Table 5.4. Goodness-of-fit tests to check for normality of distributions on RT of tasks for L1 group

<table>
<thead>
<tr>
<th>Task</th>
<th>Condition</th>
<th>Skew</th>
<th>SE</th>
<th>Ratio</th>
<th>Kurtosis</th>
<th>SE</th>
<th>Ratio</th>
<th>Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-back*</td>
<td></td>
<td>0.94</td>
<td>0.34</td>
<td>2.80</td>
<td>0.66</td>
<td>0.66</td>
<td>1.00</td>
<td>0.13</td>
<td>0.07</td>
</tr>
<tr>
<td>VPT*</td>
<td></td>
<td>0.31</td>
<td>0.34</td>
<td>0.91</td>
<td>0.15</td>
<td>0.66</td>
<td>0.23</td>
<td>0.06</td>
<td>0.20</td>
</tr>
<tr>
<td>Self-Paced Reading</td>
<td>Control</td>
<td>0.11</td>
<td>0.34</td>
<td>0.31</td>
<td>-0.45</td>
<td>0.67</td>
<td>-0.67</td>
<td>0.07</td>
<td>0.20</td>
</tr>
<tr>
<td>Self-Paced Reading</td>
<td>STM</td>
<td>0.48</td>
<td>0.34</td>
<td>1.40</td>
<td>-0.49</td>
<td>0.67</td>
<td>-0.73</td>
<td>0.11</td>
<td>0.20</td>
</tr>
<tr>
<td>Self-Paced Reading</td>
<td>CC</td>
<td>0.62</td>
<td>0.34</td>
<td>1.81</td>
<td>-0.06</td>
<td>0.67</td>
<td>-0.09</td>
<td>0.12</td>
<td>0.09</td>
</tr>
<tr>
<td>N-back**</td>
<td></td>
<td>0.21</td>
<td>0.34</td>
<td>0.61</td>
<td>0.47</td>
<td>0.66</td>
<td>0.71</td>
<td>0.09</td>
<td>0.20</td>
</tr>
<tr>
<td>VPT**</td>
<td></td>
<td>0.41</td>
<td>0.34</td>
<td>1.21</td>
<td>0.74</td>
<td>0.66</td>
<td>1.12</td>
<td>0.11</td>
<td>0.20</td>
</tr>
</tbody>
</table>

* Stand-alone task  ** Distractor-task  + Performed with the Lilliefors correction
The K-S test showed that the L1 group’s RT data were normally distributed for all the tasks: the N-back task (N-Back**), $D(45) = 0.13, p > .05$; the stand-alone visual patterns task (VPT**), $D(45) = 0.06, p > .05$; the control condition of the self-paced reading task, $D(45) = 0.07, p > .05$; the short-term memory distractor (STM) condition of the self-paced reading task, $D(45) = 0.11, p > .05$; the cognitive control distractor (CC) condition of the self-paced reading task, $D(45) = 0.12, p > .05$; the distractor N-back task (N-Back*), $D(45) = 0.09, p > .05$; and the distractor visual patterns task (VPT*), $D(45) = 0.11, p > .05$.

$Z$-scores of the skewness and kurtosis were calculated by dividing the skewness/kurtosis value by its standard error. Note that all but one of the converted skewness or kurtosis $z$-scores were below 2.58; the N-Back task had a $z$-score for skewness at 2.8. Upon visual inspection, the RTs for this task were clustered at the lower end of the spectrum, but the distribution appears normal. The cut-off point of 2.58 is considered a moderate measure of normality and significant at $p < .01$, especially in a larger sample size ($n = 50$).

Now that the data has been reviewed for normality, the research questions will be reviewed, along with any additional assumption of the parametric statistics.

5.4. Research question involving inferences versus non-inferences

The first research question in the L1 study looked at whether first language readers process inferences differently than non-inferences. For non-inferences, the items were two-sentence scenarios that did not require readers to form an inference. For inferences, the test items were two-sentence scenarios that required readers to form a connection that was not explicitly provided, either a bridging inference or predictive inference. In order to address this research question, a repeated-measures analyses of variance (ANOVA) was carried out. The predictor variable was the three scenario types.
(filler/no inference, bridging inference and predictive inference) in the control condition, where there was no interruption to the reading process, and the dependent variable was reaction time. If the reaction time is significantly different between the inferences and the non-inferences, then this shows that these stimuli are processed differently from one another. Note that the two inference types were not combined into a general ‘inference’ category (in which the reaction times for bridging and predictive inferences were averaged together). Forming an overall comparison of inference versus non-inference would have likely washed out the effects of the comparison based on the predictions from the literature that bridging and predictive inferences are differently processed.

5.4.1. Repeated-Measures ANOVA. Revisiting the assumptions of the repeated-measures, the reaction time data were normally distributed and interval data, satisfying the two assumptions of the repeated-measures ANOVA. See Table 5.5 for descriptive statistics of the three two-sentence scenario types: no inference (N), bridging (B), and predictive (P).

Table 5.5. Descriptive statistics for RQ1 for the L1 group.

<table>
<thead>
<tr>
<th>Two-Sentence Scenario Type</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Inference (N)</td>
<td>4012</td>
<td>865</td>
</tr>
<tr>
<td>Predictive (P)</td>
<td>3929</td>
<td>788</td>
</tr>
<tr>
<td>Bridging (B)</td>
<td>3538</td>
<td>698</td>
</tr>
</tbody>
</table>

Next, the assumption of sphericity was checked. Mauchly’s test indicated that the assumption of sphericity was met, ($\chi^2(2) = 0.957$, $p > .01$) (see Table 5.6).

Table 5.6. Mauchly’s test of sphericity assumptions for Research Question 1.

<table>
<thead>
<tr>
<th>Within-Subjects Effect</th>
<th>Mauchly's Test of Sphericity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Machly's W</td>
</tr>
<tr>
<td>Story Type</td>
<td>0.957</td>
</tr>
</tbody>
</table>
The results of the repeated-measures ANOVA show that reaction time was significantly affected by the type of two-sentence scenario read, $F(2, 48) = 35.7, p < .05$. Partial eta$^2$ was .432 and observed power was 1.0. Repeated-measures t-tests (using a Bonferroni adjustment, $\alpha = .05/3 = .017$) showed that subjects were significantly faster to read on the two-sentence scenarios with bridging inferences than they were for the two-sentence scenarios with predictive inferences or the two-sentence scenarios with no inference (B versus P: $t(48) = -7.414, p < 0.01$; B versus N: $t(47) = 7.303, p < 0.01$), but that there was no significant difference in reading time between the two-sentence scenarios with no inferences and the two-sentence scenarios with predictive inferences (N versus P: $t(47) = 1.368, p = .178$, not significant). See Table 5.7 for the repeated-measures ANOVA for the scenario type in the L1 group.

Table 5.7. RM ANOVA testing for scenario type in L1 group.

<table>
<thead>
<tr>
<th></th>
<th>Type III SS</th>
<th>df</th>
<th>$F$</th>
<th>$p$-value</th>
<th>Partial Eta$^2$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario Type</td>
<td>6161976.798</td>
<td>2</td>
<td>35.709</td>
<td>0</td>
<td>0.432</td>
<td>1</td>
</tr>
<tr>
<td>Error (Scenario Type)</td>
<td>8110377.122</td>
<td>94</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Participants took longer to read the two-sentence scenarios with no-inference ($M = 4012$) and with predictive inferences ($M = 3929$) than those with the bridging ($M = 3538$) inferences. These data show patterns that are counter to the hypothesis that non-inferences are processed differently than inferences. In fact, it seems to indicate that there may be similar processing underlying the reading of predictive inferences and non-inferences while processing is different for bridging inferences. Since bridging inferences are thought to be a more automatic process, it would logically follow that they would be processed faster than predictive inferences. What is unexpected is the finding that
predictive inferences appear to be processed similarly to non-inferences. This will be further explored in the discussion chapter through a closer look at participant data and the context of the experimental task.

5.5. Research question involving role of cognitive resources

The second research question in the L1 study looked at whether first language readers rely on cognitive resources to process inferences. As stated above, the non-inferences were two-sentence scenarios that did not require readers to form an inference and the inferences were two-sentence scenarios that required readers to form a connection that was not explicitly provided, either a bridging inference or predictive inference. In order to address this research question, several correlations were carried out. The predictor variables were the type of distractor task used in each condition (i.e., control/none [CONT], cognitive control distractor [CC] and short-term memory distractor [STM]) and type of inference scenario (bridging or predictive), while the dependent variable was reaction time for reading. The co-variables were the independent measures of cognitive control and short-term memory. If there are significant correlations between the stand-alone cognitive measures and the reaction times for the inferences in the conditions with distractor tasks, then this shows that the cognitive resources are being recruited to process those inferences.

5.5.1. Correlations. Bivariate correlations shed light on whether two variables are associated by investigating the degree that two continuous variables covary. The five assumptions of parametric correlations are that the data are independently collected (i.e., one person’s data does not influence the data of another), the data are assumed to be interval, the data are assumed to be normally distributed, the relationship between the
variables is linear, and finally, the data should have homoscedasticity (i.e., the variances of their residuals are comparable). The L1 and L2 data were checked for all assumptions in this chapter, as it was more efficient, however results of the L2 correlations will be discussed in chapter 6.

The distributions of the participants’ accuracy and reaction time data were previously checked for normality; it was shown that the overall reaction time data determined to be normally distributed, however, the accuracy data across all the measures were non-normally distributed. This section will look at the reaction time data in a finer-grained manner, by scenario type and condition. Table 5.8 presents the descriptive statistics (means and standard deviations) for the stand-alone measures of individual differences again and for the scenario type by condition for the first time. The last column in this table displays the results of Levene’s test of homogeneity of variances. Homogeneity of variance was checked, despite it not being an underlying assumption of correlation, to ensure that the eventual between group analyses for the L1 and L2 groups is valid. The amount of variance for each measure provides more information about these correlations. Table 5.9 shows the ratios for skewness and kurtosis for L1, L2 and for the overall group for the finer-grained reaction time data of the reading times. Highlighted in red are any ratios that are +/- 2.58, as they are considered skewed or lepto-/platykurtic.
Table 5.8. Descriptive statistics and tests of homogeneity for each ID measure and reading task by inference and condition.

<table>
<thead>
<tr>
<th>Measure (Condition)</th>
<th>L1 Group (n = 50 except VPT accuracy [49] and all two-sentence scenarios [49])</th>
<th>L2 Group (n = 22 except N-Back RT and predictive and control scenarios [23])</th>
<th>All (N = 72 except N-Back RT [73] and all VPT accuracy, bridging scenarios for control and CC and predictive scenarios for CC [71])</th>
<th>Levene statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-Back Accuracy</td>
<td>51.82 (3.27)</td>
<td>52 (3.30)</td>
<td>51.87 (3.25)</td>
<td>0.335</td>
</tr>
<tr>
<td>N-Back RT</td>
<td>632 (189)</td>
<td>683 (187)</td>
<td>648 (189)</td>
<td>0.069</td>
</tr>
<tr>
<td>VPT Accuracy</td>
<td>56.55 (2.07)</td>
<td>56.86 (1.75)</td>
<td>56.65 (1.97)</td>
<td>1.147</td>
</tr>
<tr>
<td>VPT RT</td>
<td>709 (118)</td>
<td>751 (135)</td>
<td>722 (124)</td>
<td>0.773</td>
</tr>
<tr>
<td>Bridging (CONT)</td>
<td>3571 (730)</td>
<td>4178 (904)</td>
<td>3759 (831)</td>
<td>1.25</td>
</tr>
<tr>
<td>Bridging (STM)</td>
<td>3341 (698)a</td>
<td>4287 (1012)a</td>
<td>3643 (919)</td>
<td>7.426*</td>
</tr>
<tr>
<td>Bridging (CC)</td>
<td>4008 (870)b</td>
<td>5203 (1337)b</td>
<td>4378 (1168)</td>
<td>8.91*</td>
</tr>
<tr>
<td>Predictive (CONT)</td>
<td>3964 (817)</td>
<td>5008 (1093)</td>
<td>4297 (1030)</td>
<td>2.266</td>
</tr>
<tr>
<td>Predictive (STM)</td>
<td>3486 (780)</td>
<td>4660 (1041)</td>
<td>3861 (1025)</td>
<td>3.517</td>
</tr>
<tr>
<td>Predictive (CC)</td>
<td>4464 (1002)c</td>
<td>5755 (1521)c</td>
<td>4864 (1321)</td>
<td>6.157*</td>
</tr>
</tbody>
</table>

Notes: Means are presented with the standard deviations in parentheses.

Statistically significant results for Levene’s tests are indicated with asterisks; they suggest violations of the assumption of homogeneity of variances.

Superscripts (a, b, c) indicate where post-hoc tests identified differences in the means.
Table 5.9. Skewness and kurtosis ratios for each measure (overall, L1 & L2).

<table>
<thead>
<tr>
<th>Scenario Type (Condition)</th>
<th>L1 Group</th>
<th></th>
<th>L2 Group</th>
<th></th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Skew</td>
<td>Kurtosis</td>
<td>Skew</td>
<td>Kurtosis</td>
<td>Skew</td>
</tr>
<tr>
<td>Bridging (Control)</td>
<td>0.115</td>
<td>-0.650</td>
<td>1.147</td>
<td>-0.466</td>
<td>1.537</td>
</tr>
<tr>
<td>Bridging (STM)</td>
<td>3.071</td>
<td>1.112</td>
<td>0.751</td>
<td>-1.015</td>
<td>3.367</td>
</tr>
<tr>
<td>Bridging (CC)</td>
<td>1.597</td>
<td>0.057</td>
<td>0.521</td>
<td>-0.993</td>
<td>2.926</td>
</tr>
<tr>
<td>Predictive (Control)</td>
<td>1.274</td>
<td>-0.199</td>
<td>1.472</td>
<td>-0.150</td>
<td>2.799</td>
</tr>
<tr>
<td>Predictive (STM)</td>
<td>2.885</td>
<td>1.969</td>
<td>1.297</td>
<td>0.082</td>
<td>3.279</td>
</tr>
<tr>
<td>Predictive (CC)</td>
<td>2.671</td>
<td>1.451</td>
<td>0.224</td>
<td>-0.816</td>
<td>2.996</td>
</tr>
</tbody>
</table>

Next, the assumption of linearity was checked through visual inspection of scatterplots plotting the relationship of the cognitive measures and the reaction time data for the self-paced reading tasks by inference and condition using regression and Loess lines. According to Jacoby (2000), scatterplots allow for visual inspection of the relationships between variables and ultimately determine whether their relationship is linear or not, and, as such, “functional dependence exists when points that have different coordinates on one scale axis of the scatterplot also tend to exhibit systematically different coordinates on the other scale axis” (pp. 578-579).

Given that several of the measures had visible deviations in the Loess lines compared to the straight regression lines in the scatterplots, the relationships between the cognitive measures and the reaction time scores were determined to be non-linear. See Appendices F1 and F2 for scatterplots of the cognitive measures and the reaction time data by condition and scenario type. Also, upon inspection, the accuracy data for the cognitive measures were not considered to be normally distributed (see above).
assumptions of normality, linearity, and homoscedasticity were not met. When data does not meet the assumptions of parametric correlations (i.e., Pearson’s $r$), Field (2009) recommends the use of Spearman’s rho as a nonparametric alternative; as such Spearman’s rank order correlations were used for all analyses investigating the relationships of the variables. See Table 5.10 for the correlations. Since $p$-values have been included, the levels of statistical significance will not be denoted. With correlations, the difference in the magnitudes of the correlations is most valuable, rather than which correlation is more or less significant.

Table 5.10. Correlations for the L1 group by scenario type (and condition) and stand-alone cognitive measures.

<table>
<thead>
<tr>
<th>ID measure</th>
<th>Scenarios with bridging inferences</th>
<th></th>
<th>Scenarios with predictive inferences</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(CONT)</td>
<td>(STM)</td>
<td>(CC)</td>
<td>(CONT)</td>
</tr>
<tr>
<td>N-Back Accuracy</td>
<td>0.21</td>
<td>.287</td>
<td>.358</td>
<td>0.195</td>
</tr>
<tr>
<td></td>
<td>(0.074)</td>
<td>(0.023)</td>
<td>(0.006)</td>
<td>(0.089)</td>
</tr>
<tr>
<td>N-Back RT</td>
<td>-0.087</td>
<td>-0.152</td>
<td>0.095</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>(0.276)</td>
<td>(0.149)</td>
<td>(0.259)</td>
<td>(0.463)</td>
</tr>
<tr>
<td>VPT Accuracy</td>
<td>0.084</td>
<td>0.066</td>
<td>.354</td>
<td>0.118</td>
</tr>
<tr>
<td></td>
<td>(0.284)</td>
<td>(0.328)</td>
<td>(0.007)</td>
<td>(0.212)</td>
</tr>
<tr>
<td>VPT RT</td>
<td>0.169</td>
<td>.304</td>
<td>0.077</td>
<td>0.108</td>
</tr>
<tr>
<td></td>
<td>(0.123)</td>
<td>(0.017)</td>
<td>(0.298)</td>
<td>(0.229)</td>
</tr>
</tbody>
</table>

Spearman’s Rho, a non-parametric correlation statistic, was used to compute these statistics.

Before looking at the correlations in detail, it is important to note that any claims need to be interpreted with caution; despite the significance of the correlations reported, the $r$-squared values provide more valuable information. The magnitude of the correlations suggests that there is not a lot of shared variance between any two values
reported. Therefore, any significant correlations describe a small portion of the role of the
cognitive resources in the processing of the two inference types.

It is interesting that the cognitive measures did not have any relationship with the
reading times in the control condition, regardless of scenario type. There was, however, a
significant relationship between the accuracy on the N-back task, a measure of cognitive
control, and the reaction time for all the scenarios in all the conditions with distraction,
including scenarios with bridging inferences in the short-term memory condition ($r
= .287), p\text{ (one-tailed)} < .02, scenarios with bridging inferences in the cognitive control
condition ($r = .358), p\text{ (one-tailed)} < .00, scenarios with predictive inferences in the
short-term memory condition ($r = .317), p\text{ (one-tailed)} < .01, and scenarios with
predictive inferences in the cognitive control condition ($r = .294), p\text{ (one-tailed)} < .02.
This indicates that the more accurate participants were on the stand-alone cognitive
control task, the slower they were to process scenarios while distracted, regardless of
inference type or distractor type. There was also a significant relationship between
accuracy on the Visual Patterns task, a measure of short-term memory, and the reaction
time for scenarios in the cognitive control conditions, including scenarios with bridging
inferences in the cognitive control condition ($r = .354), p\text{ (one-tailed)} < .00 and scenarios
with predictive inferences in the cognitive control condition ($r = .331), p\text{ (one-tailed)}
< .01, showing that the more accurate participants’ short-term memory, the slower they
were to process both inference types in the cognitive control condition. Finally, there was
a significant relationship between the reaction time on the Visual Patterns task, a measure
of short-term memory, and the reaction time for all the scenarios in all the short-term
memory conditions, including scenarios with bridging inferences in the short-term
memory condition \((r = .304), p \text{ (one-tailed)} < .02\), and scenarios with predictive inferences in the short-term memory condition \((r = .248), p \text{ (one-tailed)} < .04\). Since the relationship was positive, the slower participants were to complete the visual patterns task, the slower they were to process scenarios with both inference types in the short-term memory condition. Taken together, these correlations suggest that L1 readers relied on both cognitive control and short-term memory to process both inference types.

5.6. Research question involving bridging inferences versus predictive inferences

The third research question looked at how the participants processed the scenarios with bridging inferences and predictive inferences in the different conditions; the directional hypotheses from the literature suggest that L1 English readers would rely more on cognitive control than short-term memory when processing predictive inferences and L1 English readers would rely more on short-term memory than cognitive control when processing bridging inferences (Burkhardt, 2006; Jin et al., 2009; Ferstl & von Cramon, 2001). As we saw in Chapter 4, cognitive control was operationalized by an N-back distractor task and short-term memory was operationalized by a Visual Patterns distractor task in the self-paced reading task. By distracting the participants along each cognitive dimension while having them process both inference types, it is possible to see which cognitive resources are used to process the different stimuli. The predictor variables are the type of distractor task used in each condition (i.e., control/none [CONT], cognitive control distractor [CC] and short-term memory distractor [STM]) and the type of scenario (bridging or predictive) and the dependent variables are reaction time. For correlational analyses, the non-verbal independent measures of cognitive control and short-term memory were used as predictor variables. If participants processed predictive
inferences with approximately the same speed in the control and short-term memory conditions, but were slower to read predictive inferences when distracted along the cognitive control dimension, this would indicate that cognitive control was more heavily relied upon to process this stimulus type. Similarly, if participants processed bridging inferences with approximately the same speed in the control and cognitive control conditions, but were slower to read bridging inferences when distracted along the short-term memory dimension, this would indicate that short-term memory was more heavily relied upon to process this stimulus type. See Table 5.11 for a visual description of the predictions.

*Table 5.11. Predictions for RT in different conditions by inference type for RQ3.*

<table>
<thead>
<tr>
<th>Inference Type</th>
<th>Control Condition</th>
<th>CC Condition</th>
<th>STM Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bridging</td>
<td>Faster RT</td>
<td>Faster RT</td>
<td>Slower RT</td>
</tr>
<tr>
<td>Predictive</td>
<td>Faster RT</td>
<td>Slower RT</td>
<td>Faster RT</td>
</tr>
</tbody>
</table>

5.6.1. **Processing of bridging inferences.** Next, we will look at the processing of bridging inferences in the different conditions more directly. Essentially, this is a question of whether participants were slower to form bridging inferences when memory capacity is interfered with as compared to when it is not interfered with. The sub-research question is: Do L1 English readers rely more on memory capacity than cognitive control to process bridging inferences? If participants are relying on their short-term memory to form bridging inferences, then a secondary task that taxes their short-term memory would result in slower reaction times when processing these inference types.

5.6.1.1. **Repeated-Measures ANOVA.** The statistical test used to investigate this question was a repeated-measures ANOVA, which carries the assumptions of normal distribution of interval data and sphericity. As we saw above, the
reaction time data were normally distributed continuous data. See Table 5.12 for descriptive statistics of the bridging inferences in the three conditions: control (CONT), short-term-memory (STM) and cognitive control CC.

*Table 5.12.* Descriptive statistics for bridging inferences in different conditions.

<table>
<thead>
<tr>
<th>Two-Sentence Scenario Type</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control (CONT)</td>
<td>3546</td>
<td>715</td>
</tr>
<tr>
<td>Short-term memory (STM)</td>
<td>3298</td>
<td>634</td>
</tr>
<tr>
<td>Cognitive control (CC)</td>
<td>4000</td>
<td>877</td>
</tr>
</tbody>
</table>

Next, the assumption of sphericity was checked. Mauchly’s test indicated that the assumption of sphericity was met, ($\chi^2(2) = 0.875, p > .01$) (see Table 5.13).

*Table 5.13.* Mauchly’s test of sphericity assumptions for bridging inferences.

<table>
<thead>
<tr>
<th>Mauchly’s Test of Sphericity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within-Subjects Effect</td>
</tr>
<tr>
<td>Condition</td>
</tr>
</tbody>
</table>

The results of the repeated-measures ANOVA show that reaction time was significantly affected by the condition, $F(2, 48) = 26.0, p < .05$. Partial eta$^2$ was .356 and observed power was 1.0. Repeated-measures $t$-tests (using a Bonferroni adjustment, $\alpha = .05/3 = .017$) showed that participants were significantly faster to read the two-sentence scenarios with bridging inferences in the short-term memory (STM) condition than they were for the cognitive control condition (CC) or the control condition (CONT) (STM versus CC: $t(47) = -6.778, p < 0.01$; STM versus CONT: $t(48) = 2.863, p < 0.01$), and that there was also a significant difference in reading time between the two-sentence scenarios with bridging inferences in the cognitive control (CC) condition and the control condition (CONT) (CC versus CONT: $t(47) = -4.125, p < 0.01$). See Table 5.14 for the repeated measures ANOVA of bridging inference reaction times (in ms) by condition.
Table 5.14. RM ANOVA testing for bridging inferences by condition in the L1 group.

<table>
<thead>
<tr>
<th></th>
<th>Type III SS</th>
<th>df</th>
<th>F</th>
<th>p-value</th>
<th>Partial Eta²</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td>4948100.193636</td>
<td>1</td>
<td>17.02</td>
<td>0.266</td>
<td>.981</td>
<td></td>
</tr>
<tr>
<td>Error (Condition)</td>
<td>13667673.964009</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This seems to indicate that bridging inferences do not rely on short-term memory capacity for their processing, as interference did not affect processing time in the short-term memory condition. It appears that cognitive control plays a larger role in the processing of bridging inferences than hypothesized.

5.6.2. Processing of predictive inferences. Next, we will look at the processing of predictive inferences in the different conditions. As we investigated above, this is a question of whether participants were slower to form predictive inferences when cognitive control capacity is interfered with as compared to when it is not interfered with. The sub-research question is: Do L1 English readers rely more on cognitive control than short-term memory capacity to process predictive inferences? If participants are relying on their cognitive control to form predictive inferences, then a secondary task that taxes their cognitive control would result in slower reaction times when processing these inference types.

5.6.2.1. Repeated-Measures ANOVA. The statistical analysis used to investigate this question is a repeated-measures ANOVA, which carries the assumptions of normal distribution of interval data and sphericity. As we saw above, the reaction time data were normally distributed continuous data. See Table 5.15 for descriptive statistics of the predictive inferences in the three conditions: control (CONT), short-term-memory (STM) and cognitive control CC.)
Table 5.15. Descriptive statistics for predictive inferences in different conditions.

<table>
<thead>
<tr>
<th>Two-Sentence Scenario Type</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control (CONT)</td>
<td>3923</td>
<td>773</td>
</tr>
<tr>
<td>Short-term memory (STM)</td>
<td>3443</td>
<td>727</td>
</tr>
<tr>
<td>Cognitive control (CC)</td>
<td>4421</td>
<td>967</td>
</tr>
</tbody>
</table>

Next, the assumption of sphericity was checked. Mauchly’s test indicated that the assumption of sphericity was violated, \( \chi^2(2) = 0.738, p < .01 \) (see Table 5.16).

Table 5.16. Mauchly’s test of sphericity assumptions for predictive inferences in L1 group.

<table>
<thead>
<tr>
<th>Mauchly's Test of Sphericity</th>
<th>Within-Subjects Effect</th>
<th>Machly's W</th>
<th>Approx. Chi-Square</th>
<th>df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td>0.738</td>
<td>13.976</td>
<td>2</td>
<td></td>
<td>0.001</td>
</tr>
</tbody>
</table>

Since the assumption of sphericity was violated, multivariate tests are reported \( \varepsilon = .792 \), using a Greenhouse-Geisser correction. The results of the repeated-measures ANOVA show that reaction time was significantly affected by the condition, \( F(1.58, 38) = 39.54, p < .05 \). Partial eta\(^2\) was .457 and observed power was 1.0. Repeated-measures t-tests (using a Bonferroni adjustment, \( \alpha = .05/3 = .017 \)) showed that participants were significantly slower to read the two-sentence scenarios with predictive inferences in cognitive control (CC) condition than they were in the short-term memory (STM) condition or the control (CONT) condition (CC versus STM: \( t(47) = -7.637, p < 0.01 \); CC versus CONT: \( t(47) = -4.239, p < 0.01 \)); however, the reading time for the two-sentence scenarios with predictive inferences in the control condition (CONT) was significantly slower than in the short-term memory (STM) condition (CONT versus STM: \( t(48) = 6.252, p < 0.01 \)). See Table 5.17 for the repeated-measures ANOVA.
Table 5.17. RM ANOVA testing for predictive inferences by condition in the L1 group.

<table>
<thead>
<tr>
<th></th>
<th>Type III SS</th>
<th>df</th>
<th>F</th>
<th>p-value</th>
<th>Partial Eta²</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td>22977905.144</td>
<td>4</td>
<td>1.58</td>
<td>39.54</td>
<td>0.457</td>
<td>1</td>
</tr>
<tr>
<td>Error (Condition)</td>
<td>27311788.994</td>
<td>38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Performed with Greenhouse-Geisser corrections.

This could indicate that predictive inferences rely on cognitive control capacity for their processing, as interference significantly affected processing time in the cognitive control condition. Since this effect was seen across scenario type, it seems to indicate that cognitive control is relied upon for bridging and predictive inferences. However, it is counter intuitive that it would take less time to read sentences with no distraction than a short-term memory distractor task; this was most likely due to practice effect and the pseudo-randomized blocks.

5.7. Summary of results and revisiting hypotheses

To summarize, the results of the L1 study, the first research question investigated whether participants processed inferences differently than non-inferences.

- It appears that bridging inferences are processed more quickly, and possibly more automatically, than predictive inferences.
- Unexpectedly, the non-inferences took the longest to process in uninterrupted processing (i.e., the control condition). Possible reasons for this will be discussed in the Discussion Chapter.

The second research question explored whether participants rely on cognitive resources to process inferences. The correlational analyses tell a complex story in which both inference types rely the use of short-term memory and cognitive control; it is
important to interpret the correlations carefully, as the amount of shared variance in the significant correlations was on the lower end of the spectrum (between .287 and .358).

- Overall, L1 readers seem to recruit cognitive resources to process inferences.
- It appears that whether reading a scenario with bridging or predictive inferences, participants rely on cognitive control ability and short-term memory, but with slightly different patterns.
- There seems to be a relationship between cognitive control ability (as operationalized by the N-Back task accuracy) and effortful processing of both inference types in the face of distraction. Only scenarios with correctly answered verification questions were analyzed; this seems to show that the cost effect when processing taps into cognitive control ability.
- There is also tentative evidence that short-term memory is accessed for both inference types in both conditions. The more accurate participants were on the stand-alone short-term memory task, the slower they were to process both inference types, but only in the cognitive control condition.
- Also, the reaction time in stand-alone measure of short-term memory was positively correlated with the reaction time in the short-term memory condition for both inference types.
- All together, this points to an interrelated system of cognitive resources for the processing of inferences.

The third research question explored how participants processed two-sentence scenarios for bridging inferences and predictive inferences in the distraction paradigm; this shed light on whether L1 readers would rely on more on short-term memory or
cognitive control on each inference type. Each inference type was separately explored. Looking at bridging inferences, we can ask whether L1 English readers rely more on short-term memory capacity than cognitive control to process bridging inferences. The sub-question for predictive inferences was the inverse, whether L1 English readers rely more on cognitive control than on short-term memory capacity to process predictive inferences. Recall that the directional hypotheses suggested that L1 readers will rely on memory capacity over and above cognitive control for bridging inferences (Burkhardt, 2006) and L1 readers will rely on cognitive control over and above visual/phonological short-term memory for predictive inferences (Jin et al., 2009). These can be collapsed according to the results.

- There was a global effect for cognitive control interference regardless of condition or inference type. This seems to point to the fact that cognitive control is integral to inference processing.

In Chapter 6 the results of the L2 study will be reviewed.
CHAPTER 6: RESULTS OF L2 STUDY

6.1. Introduction

This chapter presents the results of the statistical analyses used to address the research questions in the L2 study. As described in chapter 5, PAW (Predictive Analytics SoftWare) Statistics 21, from SPSS: An IBM company was used to carry out these analyses. The alpha level was set at .05 for all tests. This chapter repeats much of the information in Chapter 5, so it can be read on its own; as such it first examines the data for normality. Next, each research question is reviewed, including the assumptions and parametric statistics used to answer the research questions, the alternate nonparametric statistical method used (only when the assumptions were not met for parametric tests), the descriptive statistics (e.g., mean, range, standard deviations) for all the predictor and dependent variables, and the results of the inferential statistics. Nonparametric tests are an acceptable alternative to parametric statistics, however they do have less power (Siegel & Castellan, 1988).

6.2. Assumptions of parametric statistical methods

The reaction time and accuracy data for the stand-alone measures (i.e., N-back and Visual Pattern tasks), the self-paced reading task in all three conditions (i.e., control, short-term memory distractor and cognitive control distractor), and the embedded distractor tasks (i.e., N-back and Visual Patterns tasks) were preprocessed and pretreated and checked to see if they meet the assumptions of the normality. The statistical tests used in the L2 study were repeated-measures analyses of variance (ANOVA) and correlational analyses.
6.2.1. Assumptions of repeated measures analyses of variance. Repeated measures ANOVA are suitable when participants provide more than one measure of the dependent variable and two or more means are being compared. Repeated-measures ANOVA tests have several assumptions: normal distribution of data, homogeneity of variances, interval data, and sphericity (Tabachnick & Fiddell, 1996). Normal distribution implies that the data represents a normal curve. Homogeneity of variance assumes that the variances should be the same throughout the data. Interval data can be ordered on a scale while having equal intervals on that scale. Sphericity is when the variances of the differences between different pairs of groups are equal. The predictor and dependent variables will be checked for all assumptions for repeated-measures ANOVA in section 6.3 as all the research questions use repeated-measures ANOVA.

6.2.2. Assumptions for correlational analyses. The correlation coefficient is fitting when data are normally distributed, there is a linear relationship between the two (or more) variables, there are no significant outliers and data are interval. The predictor and dependent variables will be checked for all assumptions for correlational analyses in section 6.3 since several questions use correlation.

6.3. Data Screening

As we saw in Chapter 5, the accuracy and reaction time data were checked for normality; the additional assumptions for the parametric statistics were tested within the research question itself (i.e., sphericity, homogeneity of variance).

6.3.1. Accuracy Data. Only trials where the task stimulus was correctly answered were analyzed; incorrect button presses are thought to indicate that the stimulus was not processed or not correctly processed. It is always a possibility that participants
unintentionally pressed the wrong button; that said, all incorrect responses were discarded because determining whether the stimulus was understood or not in such cases is near impossible. Concerning the cognitive tasks, trials were considered correct when the participant indicated the correct match (in the case of the visual patterns task) or pressed a button when she saw the shape that was seen two back (in the case of the N-back task). For the self-paced reading task, only trials where the verification question were correctly answered were analyzed; this indicated that the sentences were read and comprehended.

Once the data were visually inspected, descriptive statistics and z-scores for skew and kurtosis were checked for normal distribution. Table 6.1 shows the means, ranges, standard deviations, standard errors, and 95% confidence intervals for each measure on accuracy for the L2 group. Table 6.2 shows skewness and kurtosis values/ratios and the results of Kolmogorov-Smirnov and Shaipro-Wilkes goodness-of-fit tests for the L2 group to further investigate whether or not the data are normally distributed.
Table 6.1. Descriptive statistics for the raw accuracy scores in the experimental tasks for the L2 group. Percentages in parentheses.

<table>
<thead>
<tr>
<th>Task</th>
<th>Condition</th>
<th>Mean (%)</th>
<th>Range</th>
<th>SD</th>
<th>SE</th>
<th>95% Confidence Interval</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-Back Task*</td>
<td></td>
<td>52 (87%)</td>
<td>11</td>
<td>3.3</td>
<td>0.706</td>
<td>50.63-53.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VPT*</td>
<td></td>
<td>56.86 (95%)</td>
<td>6</td>
<td>1.8</td>
<td>0.374</td>
<td>56.04-57.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Paced Reading Task</td>
<td>Control</td>
<td>46 (92%)</td>
<td>13</td>
<td>3.4</td>
<td>0.7</td>
<td>44.45-47.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Paced Reading Task</td>
<td>STM</td>
<td>45.7 (91%)</td>
<td>8</td>
<td>1.9</td>
<td>0.405</td>
<td>44.75-46.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Paced Reading Task</td>
<td>CC</td>
<td>45.65 (91%)</td>
<td>11</td>
<td>2.3</td>
<td>0.473</td>
<td>44.5-46.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-Back Task**</td>
<td></td>
<td>36.7 (73%)</td>
<td>20</td>
<td>5.9</td>
<td>1.233</td>
<td>35.09-39.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VPT**</td>
<td></td>
<td>47.52 (95%)</td>
<td>10</td>
<td>2.6</td>
<td>0.544</td>
<td>46.36-48.78</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Stand-alone task  ** Distractor-task
Table 6.2. Goodness-of-fit tests to check for normality of distributions on raw accuracy of tasks for L2 group.

<table>
<thead>
<tr>
<th>Task</th>
<th>Condition</th>
<th>Skew</th>
<th>SE</th>
<th>Ratio</th>
<th>Kurtosis</th>
<th>SE</th>
<th>Ratio</th>
<th>Statistic (K-S)$^+$</th>
<th>p-value</th>
<th>Statistic (S-W)$^+$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-Back Task*</td>
<td></td>
<td>-0.54</td>
<td>0.49</td>
<td>-1.10</td>
<td>-1.01</td>
<td>0.95</td>
<td>-1.06</td>
<td>0.24</td>
<td>0.00</td>
<td>0.88</td>
<td>0.02</td>
</tr>
<tr>
<td>VPT*</td>
<td></td>
<td>-0.41</td>
<td>0.49</td>
<td>-0.84</td>
<td>-0.44</td>
<td>0.95</td>
<td>-0.46</td>
<td>0.17</td>
<td>0.12</td>
<td>0.91</td>
<td>0.05</td>
</tr>
<tr>
<td>Self-Paced Reading Task Control</td>
<td></td>
<td>-1.47</td>
<td>0.48</td>
<td>-3.05</td>
<td>2.01</td>
<td>0.94</td>
<td>2.15</td>
<td>0.21</td>
<td>0.02</td>
<td>0.83</td>
<td>0.00</td>
</tr>
<tr>
<td>Self-Paced Reading Task STM</td>
<td></td>
<td>-1.08</td>
<td>0.48</td>
<td>-2.25</td>
<td>1.86</td>
<td>0.94</td>
<td>1.98</td>
<td>0.18</td>
<td>0.09</td>
<td>0.90</td>
<td>0.03</td>
</tr>
<tr>
<td>Self-Paced Reading Task CC</td>
<td></td>
<td>-0.70</td>
<td>0.48</td>
<td>-1.46</td>
<td>2.33</td>
<td>0.94</td>
<td>2.49</td>
<td>0.20</td>
<td>0.03</td>
<td>0.90</td>
<td>0.04</td>
</tr>
<tr>
<td>N-Back Task**</td>
<td></td>
<td>-0.57</td>
<td>0.48</td>
<td>-1.18</td>
<td>-0.19</td>
<td>0.94</td>
<td>-0.20</td>
<td>0.22</td>
<td>0.01</td>
<td>0.87</td>
<td>0.01</td>
</tr>
<tr>
<td>VPT**</td>
<td></td>
<td>-1.52</td>
<td>0.48</td>
<td>-3.17</td>
<td>2.24</td>
<td>0.94</td>
<td>2.39</td>
<td>0.28</td>
<td>0.00</td>
<td>0.80</td>
<td>0.00</td>
</tr>
</tbody>
</table>

* Stand-alone task  ** Distractor-task  $^+$ Performed with the Lilliefors correction
The accuracy data indicates that participants were more accurate than not on the experimental tasks; the data are negatively skewed and not normally distributed. This is evidenced in the experimental tasks (See Tables 6.1 and 6.2); the means and corresponding percentage scores are at the higher end of the distribution. As mentioned in Chapter 5, the measures were created to assess a skill that the participants should be able to easily access. The K-S test showed that the L2 group’s accuracy data were not normally distributed for all the tasks: the N-back task (N-Back**), $D(22) = 0.24, p < .05$; the control condition of the self-paced reading task, $D(22) = 0.21, p < .05$; the cognitive control distractor condition (CC) of the self-paced reading task, $D(22) = 0.20, p < .05$; the distractor N-Back task, $D(22) = 0.22; p < .05$; and the distractor visual patterns task (VPT*), $D(22) = 0.28, p < .05$. Two tasks appear to be normally distributed: the stand-alone visual patterns task (VPT**), $D(22) = 0.12, p = .12$; and the short-term memory distractor (STM) condition of the self-paced reading task, $D(22) = 0.18, p = .09$; two out of the three measures were close to significant, though. In addition, many of the converted skewness or kurtosis z-scores (calculated by dividing the skewness/kurtosis value by its standard error) were above or approaching 3.29 (see Field, 2009 for a description of measures of normality using z-scores). This is evidence that the accuracy data violates normality. For this reason, all inferential statistics using accuracy data will use nonparametric statistics.

6.3.2. Reaction Time (RT) Data. Reaction times (RT) were measured for all tasks. The RTs of the reading time of the first sentence and the second sentence of the two-sentence scenarios in the self-paced reading task were measured; they represent the amount of time (in milliseconds) from the onset of the first sentence and the depressing
of the space bar to display the onset of the second sentence and the onset of the second sentence and the depressing of the space bar to display the next screen. The reaction times for the first and second sentence were separately measured. For the cognitive tasks, whether stand-alone or distractor tasks, the participant pressed a button to indicate that the stimulus matched a stimulus (or not); the reaction time represents the amount of time (in milliseconds) from the onset of the presentation of the stimulus and the depressing of the correct button. The reaction times that were +/- 2.5 standard deviations from the mean were removed for each participant, resulting in a loss of 2.8% reaction time data from the L2 group on all tasks.

The initial step was to visually inspect the data using histograms and probability-probability (P-P) plots. Figures 6.1 and 6.2 provide a view of how the data are distributed.
Figure 6.1. Boxplots of reaction times (in ms) of cognitive tasks for L2 group

Figure 6.2. Boxplots of reaction times (in ms) by condition of self-paced reading task for L2 group.
The boxplots reveal that the cognitive measures have a somewhat symmetrical distribution. Unlike the accuracy data, which were clustered at the higher end of the scale, the reaction times tended toward the lower (i.e., faster) end of the distribution; this shows that the majority of the participants responded fairly quickly to the stimulus. This is reflected in all of the stand-alone tasks (see Table 6.3). Participants were instructed to respond as quickly as possible to the stimuli, and therefore faster reaction times would be expected.

The self-paced reading task also shows that the conditions are relatively in the same range, with slightly higher RTs; the cognitive control condition had a larger range and larger variance than the other two conditions. Table 6.3 shows the means, ranges, standard deviations, standard errors, and 95% confidence intervals for each measure, as well as the data removed for the L2 group by task. To further explore whether the data meets normality, Table 6.4 shows skewness and kurtosis values/ratios and the results of Kolmogorov-Smirnov goodness-of-fit test.
Table 6.3. Descriptive statistics for the reaction times (in milliseconds) in the experimental tasks for the L2 group.

<table>
<thead>
<tr>
<th>Task</th>
<th>Condition</th>
<th>Mean</th>
<th>Range</th>
<th>SD</th>
<th>SE</th>
<th>95% Confidence Interval</th>
<th>% RT Data removed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
<td>Upper Bound</td>
</tr>
<tr>
<td>N-back*</td>
<td></td>
<td>683</td>
<td>707</td>
<td>187</td>
<td>39</td>
<td>593</td>
<td>775</td>
</tr>
<tr>
<td>VPT*</td>
<td></td>
<td>751</td>
<td>472</td>
<td>135</td>
<td>29</td>
<td>686</td>
<td>817</td>
</tr>
<tr>
<td>Self-Paced Reading</td>
<td>Control</td>
<td>4741</td>
<td>3724</td>
<td>1023</td>
<td>218</td>
<td>4193</td>
<td>5087</td>
</tr>
<tr>
<td>Self-Paced Reading</td>
<td>STM</td>
<td>4402</td>
<td>3303</td>
<td>988</td>
<td>206</td>
<td>3835</td>
<td>4605</td>
</tr>
<tr>
<td>Self-Paced Reading</td>
<td>CC</td>
<td>5509</td>
<td>5014</td>
<td>1371</td>
<td>292</td>
<td>4860</td>
<td>6117</td>
</tr>
<tr>
<td>N-back**</td>
<td></td>
<td>939</td>
<td>865</td>
<td>250</td>
<td>52</td>
<td>849</td>
<td>1079</td>
</tr>
<tr>
<td>VPT**</td>
<td></td>
<td>881</td>
<td>671</td>
<td>188</td>
<td>39</td>
<td>821</td>
<td>991</td>
</tr>
</tbody>
</table>

* Stand-alone task  ** Distractor-task
Table 6.4. Goodness-of-fit tests to check for normality of distributions on RT of tasks for L2 group.

<table>
<thead>
<tr>
<th>Task</th>
<th>Condition</th>
<th>Skew</th>
<th>SE</th>
<th>Ratio</th>
<th>Kurtosis</th>
<th>SE</th>
<th>Ratio</th>
<th>Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-back*</td>
<td></td>
<td>1.03</td>
<td>0.48</td>
<td>2.15</td>
<td>0.81</td>
<td>0.94</td>
<td>0.87</td>
<td>0.17</td>
<td>0.13</td>
</tr>
<tr>
<td>VPT*</td>
<td></td>
<td>0.26</td>
<td>0.49</td>
<td>0.53</td>
<td>-0.49</td>
<td>0.95</td>
<td>-0.51</td>
<td>0.09</td>
<td>0.20</td>
</tr>
<tr>
<td>Self-Paced Reading</td>
<td>Control</td>
<td>0.65</td>
<td>0.49</td>
<td>1.32</td>
<td>-0.30</td>
<td>0.95</td>
<td>-0.32</td>
<td>0.11</td>
<td>0.20</td>
</tr>
<tr>
<td>Self-Paced Reading</td>
<td>STM</td>
<td>0.60</td>
<td>0.48</td>
<td>1.25</td>
<td>-0.59</td>
<td>0.94</td>
<td>-0.63</td>
<td>0.17</td>
<td>0.13</td>
</tr>
<tr>
<td>Self-Paced Reading</td>
<td>CC</td>
<td>0.15</td>
<td>0.49</td>
<td>0.30</td>
<td>-0.51</td>
<td>0.95</td>
<td>-0.53</td>
<td>0.10</td>
<td>0.20</td>
</tr>
<tr>
<td>N-back**</td>
<td></td>
<td>-0.24</td>
<td>0.48</td>
<td>-0.50</td>
<td>-0.69</td>
<td>0.94</td>
<td>-0.73</td>
<td>0.12</td>
<td>0.20</td>
</tr>
<tr>
<td>VPT**</td>
<td></td>
<td>0.47</td>
<td>0.48</td>
<td>0.99</td>
<td>-0.43</td>
<td>0.94</td>
<td>-0.46</td>
<td>0.20</td>
<td>0.03</td>
</tr>
</tbody>
</table>

* Stand-alone task  ** Distractor-task  + Performed with the Lilliefors correction
The K-S test showed that the L2 group’s data were normally distributed for the following tasks: the N-back task (N-Back**), $D(22) = 0.17, p > .05$, the stand-alone visual patterns task (VPT**), $D(22) = 0.09, p > .05$, the control condition of the self-paced reading task, $D(22) = 0.11, p > .05$; the short-term memory distractor (STM) condition of the self-paced reading task, $D(22) = 0.17, p > .05$; the cognitive control distractor (CC) condition of the self-paced reading task, $D(22) = 0.10, p > .05$; and the distractor N-back task (N-Back*), $D(22) = 0.12, p > .05$. The K-S test showed that the L2 group’s data in the distractor visual patterns task (VPT*) were not normally distributed, $D(22) = 0.20, p < .05$; upon visual inspection, the RTs fell into a bimodal distribution at the 650 ms and 1050 ms mark, accounting for the significant result. Also, skewness $z$-scores (ratio in Table 6.4) fell below 2.58. These data are considered to be normally distributed.

Now that the data has been reviewed for normality, the research questions will be reviewed, along with any additional assumption of the parametric statistics.

6.4. Research question involving inferences versus non-inferences

The first research question in the L2 study looked at whether L2 readers process inferences differently than non-inferences. For non-inferences, the items were two-sentence scenarios that did not require readers to form an inference. For inferences, the test items were two-sentence scenarios that required readers to form a connection that was not explicitly provided, either a bridging inference or predictive inference. In order to address this research question, a repeated-measures analyses of variance (ANOVA) was carried out with the reaction times for the three item types (filler/no inference, bridging inference and predictive inference) in the control condition, where there was no interruption to the reading process. As before, if the reaction time is significantly
different between the inferences and the non-inferences, then this shows that these stimuli are processed differently from one another. As noted in Chapter 5, a general ‘inference’ category (in which the reaction times for bridging and predictive inferences were averaged together) was not formed. This overall comparison of inference versus non-inference would have likely washed out the effects of the comparison based on the predictions from the literature that bridging and predictive inferences are differently processed.

6.4.1. Repeated-Measures ANOVA. Revisiting the assumptions of the repeated-measures, the reaction time data were shown to be interval data that was normally distributed. See Table 6.5 for descriptive statistics of the three two-sentence scenario types: no inference (N), bridging (B) and predictive (P). Note that the reaction times were slower than those in the L1 study.

Table 6.5. Descriptive statistics for RQ1.

<table>
<thead>
<tr>
<th>Two-Sentence Scenario Type</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Inference (N)</td>
<td>5154</td>
<td>1328</td>
</tr>
<tr>
<td>Predictive (P)</td>
<td>4892</td>
<td>963</td>
</tr>
<tr>
<td>Bridging (B)</td>
<td>4178</td>
<td>904</td>
</tr>
</tbody>
</table>

Next, the assumption of sphericity was checked. Mauchly’s test indicated that the assumption of sphericity was violated, \( \chi^2(1.4, ) = 0.572, p < .01 \) (see Table 6.6).

Table 6.6. Mauchly’s test of sphericity assumptions for RQ1.

<table>
<thead>
<tr>
<th>Mauchly's Test of Sphericity</th>
<th>Within-Subjects Effect</th>
<th>Machly's W</th>
<th>Approx. Chi-Square</th>
<th>df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Story Type</td>
<td>0.572</td>
<td>11.186</td>
<td>2</td>
<td></td>
<td>0.004</td>
</tr>
</tbody>
</table>

Since the assumption of sphericity was violated, multivariate tests are reported \( (\varepsilon = .7) \), using a Greenhouse-Geisser correction. The results of the repeated-measures ANOVA
show that reaction time was significantly affected by the type of two-sentence scenario read, $F(1.4, 15.4) = 30.4, p < .05$. Partial $\eta^2$ was .4592 and observed power was 1.0. Repeated-measures t-tests (using a Bonferroni adjustment, $\alpha = .05/3 = .017$) showed that subjects were significantly faster to read on the two-sentence scenarios with bridging inferences than they were for the two-sentence scenarios with predictive inferences or the two-sentence scenarios with no inference (B versus P: $t(15.4) = -9.34, p < 0.01$; B versus N: $t(15.4) = 6.674, p < 0.01$), but that there was no significant difference in reading time between the two-sentence scenarios with no inference and the two-sentence scenarios with predictive inferences (N versus P: $t(15.4) = 1.262, p = .220$, not significant). See Table 6.7 for the repeated-measures ANOVA of scenario type for the L2 group.

**Table 6.7.** RM ANOVA testing for scenario type in L2 group.

<table>
<thead>
<tr>
<th>Tests of Within-Subjects Effects*</th>
<th>Type III SS</th>
<th>df</th>
<th>F</th>
<th>p-value</th>
<th>Partial Eta²</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario Type</td>
<td>11230250.788</td>
<td>1.4</td>
<td>30.43</td>
<td>0</td>
<td>0.592</td>
<td>1</td>
</tr>
<tr>
<td>Error (Scenario Type)</td>
<td>7750957.920</td>
<td>15.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Performed with Greenhouse-Geisser corrections

As discussed in chapter 5 in regard to the L1 results, these results are contrary to the hypothesis. Participants took longer to read the two-sentence scenarios with no-inference ($M = 5154$) and predictive inferences ($M = 4892$) than those with bridging inferences ($M = 4178$). Similar to the L1 study, these data do demonstrate that bridging inferences are processed more quickly, and therefore more automatically, than predictive inferences.

**6.5. Research question involving role cognitive resources**

The second research question in the L2 study looked at whether second language readers rely on cognitive resources to process inferences. As we learned in Chapter 5, the
non-inferences were two-sentence scenarios that did not require readers to form an
inference and the inferences were two-sentence scenarios that required readers to form a
connection that was not explicitly provided, either a bridging inference or predictive
inference. In order to address this research question, several correlations were carried out.
The predictor variables were the type of distractor task used in each condition (i.e.,
control/none [CONT], cognitive control distractor [CC] and short-term memory distractor
[STM]) and type of inference scenario (bridging or predictive), while the dependent
variable was reaction time for reading. The co-variables were the independent measures
of cognitive control and short-term memory. If there are significant correlations between
the stand-alone cognitive measures and the reaction times for the inferences in the
conditions with distractor tasks, then this shows that the cognitive resources are being
recruited to process those inferences.

6.5.1. Correlational Analyses. Bivariate correlations shed light on whether two
variables are associated by investigating the degree that two continuous variables covary.
The five assumptions of parametric correlations are that the data are independently
collected (i.e., one person’s data does not influence the data of another), the data are
assumed to be interval, the data are assumed to be normally distributed, the relationship
between the variables is linear, and finally, the data should have homoscedasticity (i.e.,
the variances of their residuals are comparable). We saw in Chapter 5 that the L1 and L2
data were checked for all assumptions.

The assumptions of normality, linearity, and homoscedasticity were not met for
the L2 data, and as such Spearman’s rank order correlations were used for all analyses
investigating the relationships of the variables. See Table 6.9 for the correlations. Since
p-values have been included, the levels of statistical significance will not be denoted.

With correlations, the difference in the magnitudes of the correlations is most valuable, rather than which correlation is more or less significant.

*Table 6.9. Correlations for the L2 group by scenario type (condition) and stand-alone cognitive measures.*

<table>
<thead>
<tr>
<th>ID measure</th>
<th>Scenarios with bridging inferences</th>
<th>Scenarios with predictive inferences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(CONT) (STM) (CC)</td>
<td>(CONT) (STM) (CC)</td>
</tr>
<tr>
<td>N-Back Accuracy</td>
<td>-0.297 (0.096) -0.165 (0.232) 0.182 (0.215)</td>
<td>-0.353 (0.054) -0.324 (0.07) 0.182 (0.215)</td>
</tr>
<tr>
<td>N-Back RT</td>
<td>0.163 (0.234) 0.273 (0.104) 0.243 (0.138)</td>
<td>-0.024 (0.457) 0.25 (0.125) 0.303 (0.085)</td>
</tr>
<tr>
<td>VPT Accuracy</td>
<td>0.062 (0.394) 0.114 (0.307) 0.025 (0.457)</td>
<td>0.047 (0.418) -0.114 (0.307) 0.21 (0.181)</td>
</tr>
<tr>
<td>VPT RT</td>
<td>0.275 (0.114) 0.357 (0.051) 0.178 (0.22)</td>
<td>0.316 (0.076) .411 (0.029) 0.113 (0.313)</td>
</tr>
</tbody>
</table>

Spearman’s Rho, a non-parametric correlation statistic, was used to compute these statistics.

As mentioned in Chapter 5, these correlations need to be interpreted with caution, as the r-squared values indicate the magnitude of the correlations. These results suggest that there is not a lot of shared variance between any two values reported. Additionally, the L2 data were less revealing, perhaps due to a smaller sample size. That said, there was a significant relationship between the reaction time on the Visual Patterns task, a measure of short-term memory, and the reaction time for the scenarios with predictive inferences in the short-term memory condition ($r = .411$), $p$ (one-tailed) < .05. Since the relationship was positive, the slower participants were to complete the visual patterns.
task, the slower they were to process scenarios with predictive inferences in the short-term memory condition. This lends support to the link between short-term memory capacity and processing of predictive inferences, which will be discussed more in Chapter 7. These results do not definitively address whether L2 readers rely on cognitive resources for the processing of inferences.

6.6. Research question involving bridging inferences versus predictive inferences

The third research question looked at how the participants processed the bridging inferences and predictive inferences in the different conditions; the literature suggests that L2 English readers would rely more on cognitive control than short-term memory when processing predictive inferences and L2 English readers would rely more on short-term memory than cognitive control when processing bridging inferences (Burkhardt, 2006; Jin et al., 2009; Ferstl & von Cramon, 2001). However, the L2 bilingual advantage (Bialystok, 1998, 1999, 2004, 2009; Morton & Harper, 2007; Martin-Rhee & Bialystok, 2008; Costa et al., 2008, 2009; Hernández et al., 2010; Prior & Gollan, 2011; Poarch & van Hell, 2012) might complicate this picture. First, the literature on the bilingual advantage suggests that L2 readers’ enhanced cognitive control is a mediating factor in the face of related distractor tasks (the N-Back task). Additionally, L2 readers might rely more heavily on their cognitive control capacity for inference processing, regardless of inference type. The cognitive control was operationalized by an N-back distractor task in the self-paced reading task and short-term memory was operationalized by a Visual Patterns distractor task in the self-paced reading task. By distracting the participants along each cognitive dimension while having them process both inference types, we can see which cognitive resources are used to process the different stimuli. The predictor
variable is the type of distractor task used in each condition (i.e., control/none [CONT], cognitive control distractor [CC] and short-term memory distractor [STM]) and the dependent variables are reaction time and accuracy. For correlational analyses, the non-verbal independent measures of cognitive control and short-term memory were used as predictor variables. If participants processed predictive inferences with approximately the same speed in the control and short-term memory conditions, but were slower to read predictive inferences when distracted along the cognitive control dimension, this would indicate that cognitive control was more heavily relied upon to process this stimulus type. Similarly, if participants processed bridging inferences with approximately the same speed in the control and cognitive control conditions, but were slower to read bridging inferences when distracted along the short-term memory dimension, this would indicate that short-term memory was more heavily relied upon to process this stimulus type. See Table 6.8 for a visual description of the predictions.

Table 6.8. Predictions for L2 RT in different conditions by inference type for RQ3.

<table>
<thead>
<tr>
<th>Inference Type</th>
<th>Control Condition</th>
<th>CC Condition</th>
<th>STM Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bridging</td>
<td>Faster RT</td>
<td>Faster RT</td>
<td>Slower RT</td>
</tr>
<tr>
<td>Predictive</td>
<td>Faster RT</td>
<td>Slower RT</td>
<td>Faster RT</td>
</tr>
</tbody>
</table>

6.6.1. Processing of bridging inferences. The third research question in the L2 study will investigate the processing of bridging inferences in the different conditions and whether participants were slower to form bridging inferences when memory capacity is interfered with as compared to when it is not interfered with. The sub-research question is: Do L2 English readers rely more on memory capacity than cognitive control to process bridging inferences? If participants rely on their short-term memory to form
bridging inferences, then a secondary task that taxes their short-term memory would result in slower reaction times when processing these inference types.

**6.6.1. Repeated-Measures ANOVA.** The statistical test used to investigate this question was a repeated-measures ANOVA, which carries the assumptions of normal distribution of interval data and sphericity. As we saw above, the reaction time data were normally distributed continuous data. See Table 6.10 for descriptive statistics of the bridging inferences in the three conditions: control (CONT), short-term-memory (STM) and cognitive control (CC).

*Table 6.10.* Descriptive statistics for bridging inferences in different conditions for L2 group.

<table>
<thead>
<tr>
<th>Two-Sentence Scenario Type</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control (CONT)</td>
<td>4101</td>
<td>850</td>
</tr>
<tr>
<td>Short-term memory (STM)</td>
<td>4118</td>
<td>884</td>
</tr>
<tr>
<td>Cognitive control (CC)</td>
<td>5117</td>
<td>1306</td>
</tr>
</tbody>
</table>

Next, the assumption of sphericity was checked. Mauchly’s test indicated that the assumption of sphericity was violated, ($\chi^2(2) = 0.422$, $p < .01$) (see Table 6.11).

*Table 6.11.* Mauchly’s test of sphericity assumptions for bridging inferences.

<table>
<thead>
<tr>
<th>Mauchly's Test of Sphericity</th>
<th>Within-Subjects Effect</th>
<th>Machly's W</th>
<th>Approx. Chi-Square</th>
<th>df</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td></td>
<td>0.422</td>
<td>16.399</td>
<td>2</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Since the assumption of sphericity was violated, multivariate tests are reported ($\varepsilon = .634$), using a Greenhouse-Geisser correction. The results of the repeated-measures ANOVA show that reaction time was significantly affected by the condition, $F(1.3, 13.9) = 16.98$, $p < .05$. Partial $\eta^2$ was .459 and observed power was 0.991. Repeated-measures t-tests (using a Bonferroni adjustment, $\alpha = .05/3 = .017$) showed that participants were
significantly slower to read the two-sentence scenarios with bridging inferences in the
cognitive control condition (CC) condition than they were in the short-term memory
condition (STM) or in the control condition (CONT) (CC versus STM: t(21) = -4.892, p
< 0.01; CC versus CONT: t(20) = -4.074, p < 0.01), but there was also not a significant
difference in reading time between the two-sentence scenarios with bridging inferences in
the short-term memory (STM) condition and the control condition (CONT) (STM versus
CONT: t(21) = -0.354, p = 0.72, not significant). See Table 6.12 for the repeated
measures ANOVA of bridging inference reaction times (in ms) by condition.

Table 6.12. RM ANOVA testing for bridging inferences by condition in the L2 group.

<table>
<thead>
<tr>
<th>Tests of Within-Subjects Effects*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type III SS</td>
</tr>
<tr>
<td>Condition</td>
</tr>
<tr>
<td>Error (Condition)</td>
</tr>
</tbody>
</table>

*Performed with Greenhouse-Geisser corrections.

This could indicate that bridging inferences do not rely on short-term memory capacity
for their processing, as interference did not affect processing time in the short-term
memory condition. This points to the fact that cognitive control is more influential in role
the processing of bridging inferences than originally thought.

6.6.2. Processing of predictive inferences. Next, we will look at the processing
of predictive inferences in the different conditions. As we investigated above, this is a
question of whether participants were slower to form predictive inferences when
cognitive control capacity is interfered with as compared to when it is not interfered with.
The sub-research question is: Do L2 English readers rely more on cognitive control than
short-term memory capacity to process predictive inferences? If participants are relying
on their cognitive control to form predictive inferences, then a secondary task that taxes
their cognitive control would result in slower reaction times when processing these inference types.

6.6.2.1. Repeated-Measures ANOVA. The statistical analysis used to investigate this question is a repeated-measures ANOVA, which carries the assumptions of normal distribution of interval data and sphericity. As we saw above, the reaction time data were normally distributed continuous data. See Table 6.13 for descriptive statistics of the predictive inferences in the three conditions: control (CONT), short-term-memory (STM) and cognitive control CC).

Table 6.13. Descriptive statistics for predictive inferences in different conditions.

<table>
<thead>
<tr>
<th>Two-Sentence Scenario Type</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control (CONT)</td>
<td>4956</td>
<td>1089</td>
</tr>
<tr>
<td>Short-term memory (STM)</td>
<td>4584</td>
<td>998</td>
</tr>
<tr>
<td>Cognitive control (CC)</td>
<td>5755</td>
<td>1521</td>
</tr>
</tbody>
</table>

Next, the assumption of sphericity was checked. Mauchly’s test indicated that the assumption of sphericity was violated, \( \chi^2(2) = 0.544, p < .01 \) (see Table 6.14).

Table 6.14. Mauchly’s test of sphericity assumptions for predictive inferences in L2 group.

<table>
<thead>
<tr>
<th>Mauchly's Test of Sphericity</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Within-Subjects Effect</td>
<td>0.544</td>
<td></td>
</tr>
<tr>
<td>Machly's W</td>
<td></td>
<td>12.177</td>
</tr>
<tr>
<td>Approx. Chi-Square</td>
<td></td>
<td></td>
</tr>
<tr>
<td>df</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>p-value</td>
<td></td>
<td>0.002</td>
</tr>
</tbody>
</table>

Since the assumption of sphericity was violated, multivariate tests are reported (\( \varepsilon = .687 \)), using a Greenhouse-Geisser correction. The results of the repeated-measures ANOVA show that reaction time was significantly affected by the condition, \( F(1.4, 15) = 13.45, p < .05 \). Partial eta\(^2\) was .39 and observed power was 0.97. Repeated-measures t-tests (using a Bonferroni adjustment, \( \alpha = .05/3 = .017 \)) showed that participants were significantly
slower to read the two-sentence scenarios with predictive inferences in cognitive control (CC) condition than they were in the short-term memory (STM) condition or the control (CONT) condition (CC versus STM: \( t(21) = -4.754, p < 0.01 \); CC versus CONT: \( t(21) = -2.825, p < 0.01 \)); the reading time for the two-sentence scenarios with predictive inferences in the control condition (CONT) was not significantly different than in the short-term memory (STM) condition (CONT versus STM: \( t(22) = 2.593, p = 0.017 \)), albeit just barely. See Table 6.15 for the repeated-measures ANOVA of predictive inferences by condition for the L2 group.

*Table 6.15. RM ANOVA testing for predictive inferences by condition in the L2 group.*

<table>
<thead>
<tr>
<th></th>
<th>Type III SS</th>
<th>df</th>
<th>( F )</th>
<th>( p )-value</th>
<th>Partial Eta(^2)</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td>15750636.590</td>
<td>9</td>
<td>13.449</td>
<td>0</td>
<td>0.390</td>
<td>.978</td>
</tr>
<tr>
<td>Error (Condition)</td>
<td>24592956.557</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Performed with Greenhouse-Geisser corrections.

This could indicate that predictive inferences rely on cognitive control capacity for their processing, as interference significantly affected processing time in the cognitive control condition. However, since this effect was seen across scenario type, it is a less robust finding. Although there was not a significant difference in reading time for scenarios with no distraction and a short-term memory distractor task, this was a marginal finding, at best, and seems to mirror the L1 group; perhaps this was due to practice effect.

6.7. **Summary of results and revisit hypotheses**

To summarize the results of the L2 study, the first research question investigated whether L2 readers processed inferences differently than non-inferences. The results mirrored those in the L1 study.
• Bridging inferences are processed more quickly, and possibly more automatically, than predictive inferences.

• Unpredictably, the non-inferences took the longest to process in uninterrupted processing (i.e., the control condition). Possible reasons for this will be discussed in the Discussion chapter.

The second research question explored whether inferences rely on cognitive resources for the processing using correlational analyses; again these should be interpreted with caution.

• There was one positive correlation between the stand-alone short-term memory tasks’ reaction time and the reaction time for predictive inferences in the short-term memory condition.

• The sample size was not large enough ($n = 23$) to yield significant correlations.

The third research question explored how participants processed two-sentence scenarios for bridging inferences and predictive inferences in the distraction paradigm; this investigated whether L2 readers would rely more on short-term memory or cognitive control on each inference type. Looking at bridging inferences, we can look at whether L2 English readers rely more on short-term memory capacity than cognitive control to process bridging inferences. For predictive inferences, the question is whether L2 English readers rely more on cognitive control than on short-term memory capacity to process predictive inferences. The directional hypotheses suggested that L2 readers would follow a similar pattern to L1, i.e., they would rely on memory capacity over and above cognitive control for bridging inferences (Burkhardt, 2006) and on cognitive control over
and above visual/phonological short-term memory for predictive inferences (Jin et al., 2009). However, it was also hypothesized that enhanced cognitive control might modulate the effects of the cognitive control distractor. These can be collapsed according to the results.

- Similar to the L1, there was a global effect for cognitive control interference regardless of condition or inference type. This seems to indicate that cognitive control is relied on most heavily in inference processing.
CHAPTER 7: DISCUSSION

7.1. Overview

At the beginning of this dissertation, I discussed how it is of interest to SLA and psycholinguistic researchers how inferences are processed in first language (L1) and second language (L2) reading. It is clear that attention and memory-based resources are integral to the comprehension process. However, few studies have addressed this question directly. In recent years, researchers have begun to look at whether pragmatic processing varies depending on the inference type. Additionally, it is not clear whether L2 readers have access to the necessary linguistic structures needed for processing in their L2 (Horiba, 1996) or, instead, if proficiency level is a confounding variable in readers’ access to the linguistic structures in the L2 (Cohen, 1998; Guerrero, 2005; Leontiev, 1981). By studying both L1 and L2 English readers, I sought to uncover the process of inference generation in reading comprehension, specifically in regard to: (a) whether inferences are processed differently than non-inferences, (b) which cognitive resources, or individual differences, were recruited for inferences and non-inferences, and (c) whether or not two inference types – bridging and predictive – are processed differently from one another. In the following sections, I will summarize and discuss the results of each research question. In addition, the limitations will be discussed in the context of future methodological innovations. Also, future implications, possible expansions on this research will be explored. Finally, I will summarize the general conclusions from this research.
7.2. Discussion

To address the fourth research question, whether L1 and L2 inference processing differ or not, the results of the L1 and L2 studies will be compared in each section.

7.2.1. Processing of inferences versus non-inferences. The first research question looked at whether inferences are processed differently than non-inferences in L1 and L2 reading. The results of both experiments in the present dissertation showed that L1 readers processed all the scenario types faster than the L2 readers, but despite the difference in speed, these results reveal a similar pattern for both groups of readers. Essentially, L1 and L2 participants were faster to read scenarios with bridging inferences than those with predictive inferences or no inference. However, L1 and L2 participants took the same amount of time to read scenarios with predictive inferences and no inferences. These results suggest that readers, regardless of language background, process bridging inferences differently than predictive inferences and non-inferences; however, predictive inferences and non-inferences appear to be processed similarly.

7.2.1.1. Non-inferences. The results related to the scenarios with no-inferences (i.e., the filler items) were somewhat unexpected. The two-sentence scenarios with no inference presented straightforward, declarative statements (e.g., *The dean kept a cold office, The dean only saw students in the morning*) about the same referent (e.g., *the dean*) without requiring participants to form an inference. L1 and the L2 participants processed these scenarios slower than the other scenario types.

The body of sentence-processing research offers contradicting predictions about the mechanisms and time course of processing for the scenario types (non-inferences versus inferences). Additionally, few, if any, studies have investigated the reading
process for inferences versus non-inferences. In an effort to gain a clearer picture of how semantic information (i.e., a non-inference) is processed when compared to pragmatic information (i.e., an inference), empirical studies that examine written and spoken modalities will be reviewed.

There is a debate about the nature of sentence comprehension and whether or not the time course is serial or interactive (Pickering, 2007). This discussion has implications for how non-inferences and inferences would be processed. One set of studies says that meaning is processed serially (e.g., Clifton et al., 2003; Frazier & Rayner, 1982). In the serial model, there is a linear progression of processing stages: the semantic information is first accessed and retrieved, and only when the semantic information is processed, does context come into play. However, the other set of studies disputes this claim. Instead of a serial process, semantic information is thought to interact with meaning at the early stages of processing (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Stanovich, & West, 1983; Trueswell, Tanenhaus, & Garnsey, 1994). (1) below illustrates the difference between the two accounts.

(1a) You broke the dog!
(1b) Clean it up!

*Background information*: The dog is a small statuette.

In a serial account, a reader would access all the likely meanings of the words in the sentences (e.g., a *dog* is a furry canine animal, *broke* means that something does not work). Then the reader would take the non-linguistic information (e.g., context and background information) into account. In this way, the concept of *dog* would be reinterpreted based on the background knowledge that *dog* in this context refers to a small statuette. In addition, the reader would access the meaning of *broke* in which
something shatters, and discard the meaning of something does not work. Importantly, in the serial account, only at the later stage of the process is any contextual meaning accessed. In the interactive account, the reader would access the semantic representations at the same time as the context, and as soon as the reader encountered the word *broke*, the reader would reevaluate the meaning of the *dog* and access the background information that the *dog* is actually a statuette. This reassessment occurs early in the comprehension process. These accounts offer different predictions for inferences and non-inferences. Recall that inferences are made when semantic meaning is accessed in context. Non-inferences do not need context to be understood. Thus, in the serial account, a reader would predict that non-inferences (in which meaning does not rely on context) would be processed fairly quickly, while inferences would take longer to process. On the other hand, the interactive account would predict that inferences and non-inferences would possibly be processed at the same speed; it is not clear if one would be understood more quickly than the other. This is relevant to the present dissertation, as the dependent measure of processing was reading time. These results can be positioned within the interactive account, as there is a growing consensus that the constraint-based model provides a more compelling explanation of sentence processing (see Trueswell & Gleitman, 2007; Van Gompel, 2006; Van Gompel & Pickering, 2007 for reviews).

In combination to the predictions made by the interactive account, there is another strain of auditory processing research using an eye-tracking paradigm that found evidence that inferences take longer to process than non-inferences for adults and children. Using a picture-identification task with scalar implicatures (inference) and quantifiers (non-inference) as the target, Huang & Snedeker (2009a, 2009b, 2011)
showed that semantic-based processing has shorter latencies (i.e., faster processing) than inference-based processing. Three recent eye-tracking studies investigated how scalar implicatures (i.e., The girl ate some of the ice cream sandwiches) and quantifiers (i.e., The girl ate all of the ice cream sandwiches) are understood (Huang & Snedeker, 2009a, 2009b, 2011). In essence, it took participants longer to infer information than to access lexical meaning. While inferences and non-inferences displayed different reaction-time latencies in Huang and Snedeker’s auditory processing studies, any comparisons to the reading process should be made with caution. First, Huang and Snedeker’s dependent measure was length of the fixation in an eye-tracking paradigm; as such, they measured participants’ fixation on the target as a proportion of their total fixation time. Eye movements are a possibly related, however, distinct, measure of attentional resources. Finally, scalar implicatures are a different type of inference than bridging or predictive, and they may be processed differently. Along these lines, Rai et al. (2011) looked at reaction times for factual (non-inference), bridging, and pragmatic (predictive) questions about a text, and found that no-inference questions were the most quickly responded to. However, Rai et al. (2011) analyzed the reaction time for verification questions rather than the reading time of the text. The processes of reading a text and verifying information about a text are inherently different from one another.

While the reading times for non-inferences and predictive inferences in the present dissertation were not statistically different, it is worth mentioning that the non-inferences had the largest (i.e., slowest) reaction times. The larger variance ($SD_{L1} = 865$; $SD_{L2} = 1328$), especially in the L2 group, may have accounted for the lack of significance in the repeated measures t-tests. Boxplots of the reading times for non-inferences for the
L1 study and the L2 study did not uncover any participant outliers, however a boxplot of the reaction time data for the non-inferences in the control condition across all participants (N = 73) revealed three participants (S04, S21, & S26) who had disproportionately higher reading times for the non-inferences as opposed to the inferences. It is possible that these three participants may have become overloaded and adopted a different processing strategy within the control block. Table 7.2 displays the means for three scenario types in the control condition and the scores on cognitive control. 

*Table 7.2. Individual reaction times (in ms) and cognitive control scores (percentage in parentheses) for L2 participants who were identified as slower to process non-inferences.*

<table>
<thead>
<tr>
<th>Participant #</th>
<th>Non-Inferences</th>
<th>Bridging Inferences</th>
<th>Predictive Inferences</th>
<th>Cognitive Control Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>S04</td>
<td>7237</td>
<td>6666</td>
<td>5486</td>
<td>49 (82%)</td>
</tr>
<tr>
<td>S21</td>
<td>6845</td>
<td>4760</td>
<td>4393</td>
<td>45 (75%)</td>
</tr>
<tr>
<td>S26</td>
<td>8600</td>
<td>6146</td>
<td>5785</td>
<td>54 (90%)</td>
</tr>
</tbody>
</table>

If a participant were over-challenged, this could impact their processing regardless of scenario type. The repeated-measures ANOVA was rerun without the three participants identified above and the results did not change for the larger group (see Appendix G for results of all the L2 group’s statistical analyses without the three outliers; note that the results remained the same for the L2 group in the analyses with and without the three outliers, with the exception of the correlational analyses). It is also relevant that the scores on the independent measure of cognitive control, N-Back task, did not show discernable patterns, as the scores ranged from 75%-90%.
However, participants’ behavior is likely best explained by the probability of distribution in the experimental task set-up. The scenario types were purposefully mixed within the control block (across the conditions): in a total of 50 two-sentence scenarios, there were 10 non-inferences (20% of the total), 20 bridging inferences (40% of the total) and 20 predictive inferences (40% of the total). Lundstedt et al. (1998) discuss the characteristics of fillers in the context of chemical experimentation; while the experimentation differs, the purpose still holds. Fillers can have several characteristics in an experimental design: they are not of primary interest for the study, they are present across all conditions, and they are proportional to the other tested elements (Lundstedt et al., 1998, p. 35). As such, the filler stimuli served an important purpose in this mixed-factor design; they prevented participants from correctly discerning the purpose of the task by providing non-target stimuli. In contrast to the conventional wisdom from Lundstedt et al. (1998) above, they were also designed to serve as a secondary control, and therefore tested element. Non-inferences and inferences were compared in this study. It is not unexpected that participants would experience cognitive overload in the interleaved design (as opposed to a block design where one stimuli type is presented in isolation). Since each interleaved block required that participants form an inference, whether bridging or predictive, eighty percent of the time, participants were likely conditioned to look for and form unstated connections and this impacted the reading time on the non-inference scenarios. When participants become over-challenged, they often adopt strategies for processing the stimuli, typically using top-down regulation of resources to minimize distractors (see Scalf, Torralbo, Tapia, & Beck, 2013 for a discussion of this in the framework of perception load and dilution theories). In turn, the
top-down regulation can have global effects irrespective of stimuli type within an experimental block. It is possible that while L1 and L2 participants were in an ‘inference’ mindset, they looked to form connections for the scenarios with no-inferences that were not there to be made. In the case of these studies, a participant’s reaction times would be directly impacted. This may be part of the reason that the non-inferences were not processed faster.

The current results cannot answer the question whether or not readers process inferences differently than non-inferences. It is likely that participants’ reaction times for the filler stimuli (i.e., non-inferences) were impacted by the interleaved design. It is possible that non-inferences might very well be processed more slowly, or just as slowly, as inferences. If non-inferences are processed similarly to inferences, this would be telling evidence for the serial/interactive debate. This could be explored further in a future study.

7.2.1.2. Automatic versus controlled processing. The finding that bridging inferences are processed quickly and predictive inferences are processed slowly is consistent with the literature on inference building and individual differences in inference building; the construction-integration model (Kintsch, 1998) and the constructionist model (Graesser et al., 2001) suggest that bridging inferences are more automatically processed while the predictive inferences require more effort and control of internal resources. It is advantageous to position these predictions within the literature on automaticity, as the question of whether processing would be faster or slower can be directly addressed. The underlying structures and allocated resources used in automatic and controlled processing are disputed in automaticity theories, but nevertheless,
automatic processing is considered to be faster, effortless, and most likely obligatory when compared to controlled processing (see Logan, Taylor & Etherton, 1999 for a review of automatic and controlled processing). To briefly review the two dominant opposing perspectives on automaticity: the resource account considers automatic processing as independent from attention (e.g., LaBerge & Samuels, 1974; Shiffrin & Schneider, 1977), and, as such, claims that it requires no effort, while the memory-based account considers both automatic and controlled processing to be highly dependent on attention (e.g. Logan, 1988). Table 7.1 shows the specific predictions and underlying claims for the three properties of automatic processing according to the resource and memory-based accounts.


<table>
<thead>
<tr>
<th>Properties of Automatic Processing</th>
<th>Faster</th>
<th>Effortless</th>
<th>Obligatory</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Resource Account</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does not use processing resources.</td>
<td></td>
<td>Does not require effort.</td>
<td>Driven solely by stimulus presentation; no choice whether to respond or not.</td>
</tr>
<tr>
<td><strong>Memory-Based Account</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strong memories that have been often used can be quickly retrieved.</td>
<td>Requires less effort than algorithmic (i.e., controlled) performance.</td>
<td>Retrieval process is obligatory and necessary for memory-based explanation of automatic processing.</td>
<td></td>
</tr>
</tbody>
</table>

Irrespective of the underlying processes, automatic processing would present as faster at the surface level, while controlled processing would be discernibly slower; for the purposes of these two studies, these differences would be visible in the reaction times between scenario types.
Revisiting the specific predictions from the two cognitive models, Kintsch (1998) considers bridging inferences at the textbase level, where explicit meaning for the entire text is formed. Graesser et al. (2001) believe the connecting of information to form local coherence, such as the dependency relations found in bridging inferences, to be an automatic process, termed ‘passive-activation’ (p. 254). Both comprehension models consider bridging inferences to be automatic and to be formed early in the processing of text; based on the claims from cognitive theories on automaticity, bridging inferences would be processed faster, with less effort (compared to controlled processing) and obligatorily. In contrast, predictive inferences are thought to rely on cognitive control. Kintsch (1998) claims that predictive inferences are processed at the level of the situation model, in which background knowledge, memory resources, and context interact; cognitive control would likely be required for successful processing. Graesser et al. (2001) consider strategy to be the driving force of comprehension, in which a reader asks and answers internal questions about the text; cognitive control is necessary to disregard irrelevant information and select relevant information. In line with controlled processing, the processing of predictive inferences would be slower, requiring more effort (compared to automatic processing), and be optional. To review the results, bridging inferences were processed significantly faster then predictive inferences irrespective of language background, supporting the idea that bridging inferences are automatically processed while predictive inferences are processed in a controlled manner.

7.2.1.3. Summary. The scenarios with bridging and predictive inferences were processed differently from one another; the bridging inferences appear to be processed automatically and the predictive inferences appear to be processed using
controlled processing. However, these results need to be interpreted with caution, as direct comparisons of the inference types would need to be better controlled in a future study (see the Limitations and Future Directions section). The results from the L1 and the L2 studies do not show a global difference between processing of inferences and non-inferences. In addition, the reading times for the scenarios with non-inferences were not significantly different than those with predictive inferences in the L1 and the L2. The analyses with and without the outliers did not reveal different patterns. It is reasonable to think that there was a task effect due to the interleaved design.

The results of the first research question indicate that L1 readers and L2 readers process bridging inferences more quickly and with less effort than predictive inferences, which are slower and more effortful. It is not clear at this point how non-inferences are processed and if this differs from inferences.

7.2.2. Reliance on cognitive resources for inference processing. The second research question explored whether or not L1 and L2 readers rely on cognitive resources to process inferences. These results inevitably address the third research question, as well, and, as such, the interpretations will be discussed in relation to general reliance on cognitive resources for inferences in general and possible dissociations dependent on inference type.

7.2.2.1. Correlational analyses and interpretation from L1 study. In order to get a sense of the bigger picture, several correlations were run to see whether there was a relationship between the stand-alone cognitive measures (the N-Back task, a measure of cognitive control and the Visual Patterns task, a measure of short-term
memory)\textsuperscript{3} and the reading times for the bridging and predictive inferences in the three conditions. Since there were no patterns in the correlations by scenario type, the results of bridging inferences and predictive inferences will be reviewed together. Table 7.3 shows the direction of the significant correlations (positive or negative) in the L1 study to give a visual overview of the correlational results. The specific magnitudes and \textit{p}-values can be found in Table 5.11 (see Chapter 5). While these correlations provide a view into the role of cognitive resources in inference processing, it is important to note that the magnitudes of all the correlations reported in this dissertation are quite small, and therefore, offer just a fraction of the total picture. That said, since the L1 study offers a unique approach to inference generation, the correlations will be discussed below. Recall that the conditions were based on the presence or absence of a distractor task; reaction time was measured for how long it took to read the two-sentence scenarios in each of the conditions. There was no distractor task in the control (CONT) condition. There was a distractor task that taxed a participant’s cognitive control (the N-Back task) in the cognitive control condition (CC). There was a task that taxed a participant’s short-term memory (the Visual Patterns task) in the short-term memory condition (STM). The reaction time was correlated with the stand-alone individual difference measures for the N-Back task and Visual Patterns task (VPT). While Table 7.3 indicates the directionality of the correlations (i.e., positive [+] or negative [-]), it is relevant to note that there were no significant negative correlations.

\textsuperscript{3} As discussed in the methods section, the stand-alone measures used in this set of analyses, collected in a separate session, were also correlated with the distractor tasks in the self-paced reading task.
Table 7.3. Correlations from RQ2 in L1 study represented by their positive or negative relationships.

<table>
<thead>
<tr>
<th>ID measure</th>
<th>Scenarios with bridging inferences</th>
<th>Scenarios with predictive inferences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Condition</td>
<td>(CONT)</td>
</tr>
<tr>
<td>N-Back Accuracy</td>
<td>N/A</td>
<td>+</td>
</tr>
<tr>
<td>N-Back RT</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>VPT Accuracy</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>VPT RT</td>
<td>N/A</td>
<td>+</td>
</tr>
</tbody>
</table>

+ indicates a positive correlation and - indicates a negative correlation

The first interesting finding was that there were no significant correlations for the cognitive measures for bridging or predictive inferences in the control condition. That means that there is no relationship to how the participants scored on the stand-alone individual difference tasks and their reading time without distraction. All together, this points to the finding that L1 participants’ reading performance in the control block was not mediated by individual differences. This is entirely expected, as reading without distraction provides an optimal processing environment in which a participant’s cognitive resources are not compromised.

The rest of the picture is more complex. Predictive inferences are thought to rely on cognitive control. As such, it was predicted that the accuracy on the N-Back task would have been negatively correlated with the reading time for the scenarios with predictive inferences in the cognitive control condition. This would have meant that the less accurate L1 participants were on the stand-alone cognitive control task, the slower they would be to read predictive inferences in a distraction paradigm. This would indicate that participants’ cognitive control is being ‘doubly’ taxed (i.e., by the distractor task in
the self-paced reading task and by the act of forming predictive inferences). However, the
correlational analyses with the N-Back task accuracy tell quite a different story; accuracy
on the N-Back task, a measure of cognitive control, was positively correlated with
reading time for all scenarios across all conditions in which participants were distracted:
in the short-term memory condition for bridging inferences and for predictive inferences,
and in the cognitive control condition for scenarios with bridging inferences and
predictive inferences. Essentially, it does not matter which inference the participant
formed or what kind of distraction participants experienced; these correlations indicate
that the more accurate participants were on stand-alone cognitive control task, the slower
they were to process all inference-based scenarios while distracted. There are three
possible interpretations for this finding. First, this might indicate that participants were
more careful and more in control. Enhanced control would contribute to accuracy.
Second, it also stands to reason that the introduction of a secondary distractor task may
have incidentally activated participants’ cognitive control when it would have been
normally inactive for some stimuli. Finally, it is possible that both inference types rely
more on cognitive control for their processing than hypothesized. Botvinick, Braver,
Barch, Carter, & Cohen’s (2001) account of the conflict monitoring hypothesis provides
support for incidental activation of cognitive control in the face of distracting stimuli. In
the conflict monitoring hypothesis, the ‘control’ attempts to prevent conflict from
occurring by minimizing ‘crosstalk’ between parallel processing streams (see Allport,
1987; Cohen, Dunbar, & McClelland, 1990; Mozer, 1990; Mozer & Sitton, 1998; Navon,
1985; Navon & Miller, 1987; Schneider & Detweiler, 1987) while use of that conflict
serves as the ‘demand’ (i.e., level) of control (Schneider & Detweiler, 1987). An area
commonly activated in cognitive control, the anterior cingulate cortex (e.g., D’Esposito, et al., 1995; LaBerge, 1990; Mesulam, 1981; Posner & DiGirolamo, 1998), has also been recruited in conflict monitoring (Botvinick et al., 2001). In an ERP study in which participants matched pictures and sentences, van de Meerendonk, Kolk, Vissers, & Chwilla (2010) suggest that the P600 effect (normally considered syntactic repair or restructuring) may index a more domain-general ability to check for possible errors in processing, indexing conflict monitoring.

Finally, there might be a combination of all three of these interpretations, in that bridging and predictive inferences rely on cognitive control, and that the experimental set-up further activated participants’ cognitive control, resulting in enhanced accuracy. The N-Back task purportedly measures not only by the ability to access, but also to manipulate information. However, the ability to successfully manipulate information does not imply that participants would be faster (for accounts of controlled processing, see LaBerge & Samuels, 1974; Logan, 1988; 1974; Shiffrin & Schneider, 1977), only that they can successfully manage several concepts or interpretations simultaneously. It is noteworthy that the reaction time for the N-Back task was not correlated with reading time across all conditions and all scenario types; this is counter to the hypothesis that reaction time on the stand-alone measure of cognitive control would be positively correlated with reading time for predictive inferences in the cognitive control condition and negatively correlated with all other scenario types and conditions.

The next set of correlations relates to the Visual Patterns task, a measure of short-term memory, which did not pattern as expected. It was hypothesized that there would be a negative relationship between accuracy on the Visual Patterns task, as a measure of
short-term memory, and the reading time for scenarios with bridging inferences. In this way, participants with lower scores on a measure of short-term memory would be slower to process bridging inferences when distracted along short-term memory capacity. Bridging inferences were predicted to rely on short-term memory for processing. If this were the case, then participants’ short-term memory would be overloaded by the competing demands from the short-term memory task and from processing the bridging inference. The lower that capacity, the more cost in processing the inference. However, accuracy on the Visual Patterns task was instead positively correlated to the reaction time for scenarios with bridging inferences and predictive inferences, but only in the cognitive control condition. This suggests that the more accurate the participants’ short-term memory, the slower they were to process both inference types when distracted along the cognitive control dimension. This follows logically, considering that a sub-skill of the N-Back task, a cognitive control distractor, requires participants to retrieve conceptual information. In fact, having this skill may be a prerequisite for the subsequent manipulation of the information. This might explain why participants were accurate, albeit slower, to process these inference types when distracted along cognitive control dimensions. Finally, there was a significant relationship between the reaction time on the Visual Patterns task and the reading time for scenarios with bridging inferences and predictive inferences in the short-term memory condition. Since the relationship was positive, the slower participants were to complete the visual patterns task, the slower they were to process scenarios with both inference types in the short-term memory condition (and vice versa in regard to being faster). If the reaction time for the Visual Patterns task, a short-term memory task, was positively correlated only for bridging inferences in the
short-term memory condition, it would conform to the prediction that bridging inferences rely on short-term memory for their processing. However, predictive inferences follow the same pattern, which lends support to the link between short-term memory capacity as a necessary sub-skill for processing bridging and predictive inferences; after all, both inference types require that participants are able to access and retrieve conceptual information. Put simply, readers may need to access the concepts in their short-term memory (i.e., retrieve information from memory) before engaging their cognitive control.

7.2.2.2. Correlational analyses and interpretation from L2 study. The research questions were identical to those from the L1 study, however the hypotheses were slightly different. L2 readers were hypothesized to recruit short-term memory capacity for the processing of bridging inferences and cognitive control for the processing of predictive inferences. However, as reviewed in Chapter 6, an L2 reader’s enhanced cognitive control was considered a possible mediating factor when the secondary distractor task (the N-Back task) taxed cognitive control. Cognitive control was hypothesized to attenuate the effect of the distraction. Additionally, L2 readers were thought to possibly rely more heavily on their cognitive control capacity for inference processing, regardless of inference type. Interestingly, the correlational results were the only analyses in which the results differed between the intact L2 group (including the three outliers) and the partial L2 group (excluding the three outliers).

In the L2 study with intact group there was one significant relationship: it was between the reaction time on the Visual Patterns task, a measure of short-term memory, and the reaction time for the scenarios with predictive inferences in the short-term memory condition. However, there were three significant correlations in the partial L2
group (see Appendix G for the results). There was a significant relationship between the reaction time on the Visual Patterns task, a measure of short-term memory, and the reaction time for the scenarios with bridging and predictive inferences in the short-term memory condition. Looking at the analyses for both the intact and partial L2 groups, these results indicates that the slower participants were to complete the visual patterns task, the slower they were to process inference scenarios in the short-term memory condition. Similar to what we saw in the L1 study, short-term memory capacity seems to be a necessary sub-skill for processing bridging and predictive inferences, for readers appear to possibly serially access the concepts in their short-term memory (i.e., retrieve information from memory) before engaging their cognitive control. The partial L2 group also had a significant correlation between the reaction time on the Visual Patterns task and the reaction time for the scenarios with predictive inferences in the control condition; this finding is difficult to interpret. There were no other significant relationships between the stand-alone cognitive measures and the reading times, irrespective of condition or scenario type. It is noteworthy that there were relatively few correlations compared to the L1 study; with a larger L2 sample size, this would be remedied. Recall that three participants were removed at the outset of the study, and then three were identified as outliers and left out of the intact data set, leaving twenty participants for these correlational analyses. Future research would need to use larger sample sizes based on a priori power analyses to determine a sufficient sample size. Given the current results, it is not clear whether the non-significant results are due to large Type II error or not; in this way these set of results are not considered generalizable. A larger sample size might lead
to more telling results in terms of individual differences. This will be discussed in the limitations section below.

7.2.2.3. Summary. A synthesis of these correlational analyses points to interrelated recruitment of cognitive resources in the processing of bridging and predictive inferences. Since the L2 results were not robust, interpretations will be discussed in light of the L1 results. First, when L1 readers are allowed to read free from interruption, there is not a relationship between individual differences and inference processing, regardless of inference type. Second, short-term memory, as a sub-skill of cognitive control, appears to be employed to access stored representations and connect necessary information. Finally, cognitive control seems to play a large role in the processing of bridging and predictive inferences in the face of distraction. There was an overall effect by condition for all scenario types. While this was an unexpected result based on the predictions found in the literature, it suggests that cognitive control may play a larger role in processing than initially thought, even for automatic processing. This will be discussed in more detail below, in an elaborated combined model of inference processing and the purported resources used for that processing (see Figure 7.1). In addition, I would argue that the ability to access and manipulate information is integral when trying to complete two tasks simultaneously (in the distraction paradigm).

7.2.3. Processing of bridging inferences versus predictive inferences. The third research question investigated how L1 and L2 readers processed the scenarios with bridging inferences and predictive inferences. In the distraction paradigm, participants read two-sentence scenarios and answered verification questions about the scenarios with or without interference from a distractor task; the secondary distractor task determined
the condition: the control condition did not have a distractor task, so a reader’s uninterrupted processing could be measured; the short-term memory condition had the visual patterns distractor task (to tax a reader’s visual short-term memory capacity); and cognitive control condition had the N-Back distractor task (to tax a reader’s cognitive control capacity). As such, the predictor variables were (1) the type of condition in the distraction paradigm: control (CONT), short-term memory (STM), and cognitive control (CC); and (2) the type of two-sentence scenario: bridging inference or predictive inference. The dependent variable was the reaction time (in milliseconds), or reading time, of the two-sentence scenarios. Since the purpose of these two studies was to see how each inference type was processed in the face of distraction, a separate analysis was run for each inference type; in other words bridging and predictive inferences were not directly compared to each other. Instead the reading times for a scenario type – bridging or predictive – were compared across the different conditions in order to see how the readers coped with the different distractions. Table 7.4 revisits the predictions from the studies, with the results from the repeated-measures ANOVAs.

*Table 7.4.* Data patterns of RT by inference type and condition for L1 and L2 groups. The text in red denotes divergences from the hypotheses.

<table>
<thead>
<tr>
<th>Group</th>
<th>Inference Type</th>
<th>Condition</th>
<th>CONT</th>
<th>CC</th>
<th>STM</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>Bridging</td>
<td>Fast RT*</td>
<td>Slow RT*</td>
<td>Fastest RT*</td>
<td></td>
</tr>
<tr>
<td>L2</td>
<td></td>
<td>Fast RT</td>
<td>Slow RT*</td>
<td>Fast RT*</td>
<td></td>
</tr>
<tr>
<td>L1</td>
<td>Predictive</td>
<td>Slow RT*</td>
<td>Slow RT</td>
<td>Fastest RT*</td>
<td></td>
</tr>
<tr>
<td>L2</td>
<td></td>
<td>Slow RT*</td>
<td>Slow RT</td>
<td>Fast RT*</td>
<td></td>
</tr>
</tbody>
</table>

Next, the L1 and L2 results for each inference type will be reviewed.
7.2.3.1. Review of L1 study results for bridging inferences. The directional hypothesis forecasted that bridging inferences rely more on short-term memory than cognitive control in the L1. This would surface in the reaction times – participants would be faster to process bridging inferences in the control and cognitive control conditions than in the short-term memory condition. The results tell the exact opposite story; L1 readers process bridging inferences slowest in the cognitive control condition and fastest in the short-term memory condition. This seems to indicate that bridging inferences do not rely on short-term memory capacity for their processing, as interference in the short-term memory condition did not affect processing time. Cognitive control seems to play a larger role in the processing of bridging inferences than originally hypothesized.

An unanticipated finding is that L1 readers were significantly faster to process bridging inferences with distraction (in the short-term memory condition) as opposed to without distraction (in the control condition).

7.2.3.2. Review of L2 study results for bridging inferences. The directional hypothesis forecasted that bridging inferences rely more on short-term memory than cognitive control in the L2. It was also predicted that L2 readers might demonstrate overreliance on enhanced cognitive control. If this hypothesis bears out, then participants should be faster to process bridging inferences in the control and cognitive control conditions than in the short-term memory condition. L2 readers process bridging inferences slowest in the cognitive control condition and equally as fast in the short-term memory and control conditions. These results suggest that bridging inferences do not rely on short-term memory capacity for their processing, as the participants were faster to
process bridging inferences when distracted by the Visual Patterns task, a short-term memory measure, than by the N-Back task, a measure of cognitive control. While this could show over-reliance on cognitive control in the L2, it is more likely a function of an overall pattern: cognitive control seems to be instrumental in the processing of bridging inferences, regardless of language background.

7.2.3.3. Review of L1 study results for predictive inferences. The directional hypothesis anticipated that predictive inferences rely more on cognitive control than short-term memory in the L1. This would be seen in the reaction times – participants would be faster to process predictive inferences in the control and short-term memory conditions than in the cognitive control condition. The results support this hypothesis; L1 readers process predictive inferences in the cognitive control condition significantly slower than the other conditions. Additionally, L1 readers process bridging inferences in the control condition slower than in the short-term memory condition. Again, we see the unanticipated finding is that L1 readers were significantly faster to process predictive inferences with distraction (in the short-term memory condition) as opposed to without distraction (in the control condition). These results seem to indicate that predictive inferences rely on cognitive capacity for their processing, as interference in the cognitive control condition negatively impacted processing time.

7.2.3.4. Review of L2 study results for predictive inferences. The directional hypothesis forecasted that predictive inferences rely more on cognitive control than short-term memory in the L2, with possible attenuated reaction times due to enhanced cognitive control. All participant reaction times would be faster when processing predictive inferences in the control and short-term memory conditions than in
the cognitive control condition. The results support this hypothesis; L2 readers process predictive inferences in the cognitive control condition significantly slower than in the other conditions. Additionally, L2 readers do not process bridging inferences in the control condition significantly differently than in the short-term memory condition. These reading times demonstrate an effect for cognitive capacity interference in the processing of predictive inferences.

7.2.3.5. Faster with distraction? An unexpected finding was that L1 and L2 readers were significantly slower to process inferences in the control condition (without distraction) as opposed to in the short-term memory condition (with distraction). It is reasonable to expect that participants would process inferences in the control block the fastest, as their attention is undivided and can be applied to the singular task of reading comprehension. However, the conflicting result is most likely due to the pseudo-randomized presentation of blocks. Recall that when the procedure and tasks were piloted, participants found it quite challenging to complete the tasks accurately in the distraction paradigm when they were first presented with a block that had distractor tasks (either the Visual Patterns task or the N-Back task) and the target task (the cumulative self-paced reading task and verification questions). It is worth mentioning that all the blocks in the cumulative self-paced reading task were preceded by twelve practice trials. This, however, does not seem to have been sufficient practice. Based on the pilot participant feedback, all L1 and L2 participants in the main study were first presented with the control block, so they could acclimate to the main study task (reading scenarios and answering related questions). The tasks demanded that participants learn which buttons needed to be pressed for each event, including indicating that sentence #1 and sentence
were read (the spacebar), indicating that the verification question about the scenario was correct or not ("/" for yes and "z" for no), and indicating that the distractor task was correct or not ("/" for correct and "z" for incorrect). While this not inherently complex, participants needed some practice to familiarize themselves to the artificial constraints of the experiment. The SuperLab software codes correct and incorrect answers based on whether or not specific keys are depressed. It was vital that each participant was precise in pressing the correct buttons at the correct time or their results could not be used in the analyses. This necessary methodological adaption seemed to, unfortunately, impact the results; there were task effects because the control block was consistently presented first. Participants were likely slower to read and understand inferences in the control condition (as opposed to the short-term memory condition), as they were getting used to the task demands in first block.

To summarize, it appears that the cognitive control condition, as expected, was the most demanding for participants. The inferences in the short-term memory condition would probably be have been processed slower than the control condition had the blocks been fully randomized. Rather than interpret that the short-term memory condition was less cognitively demanding then the control condition, this result is undoubtedly due to a practical compromise in the study procedure. Strategies for handling this methodological caveat in future studies will be discussed further in the limitations section.

7.2.4. A unified model of inference processing. The hypotheses that bridging inferences would rely on short-term memory and predictive inferences would rely on cognitive control did not bear out in the two studies. Recall that participants had longer processing times when reading bridging inferences and predictive inferences in the
condition where their cognitive control was doubly taxed; yet, when short-term memory was interfered with, participants read both inference types fairly quickly. This indicates that inferences may require cognitive control for their successful comprehension, regardless of type.

Given that cognitive control was shown to be integral in the processing of bridging inferences, these results need to be viewed in a larger model of inference processing and to be further contextualized. These findings were also supported by the correlational analyses of the independent measures of individual differences and the reading times across the conditions. It is important to keep in mind that the correlational analyses from the L1 study also suggest that short-term memory was a significant resource when processing both inference types. It is not, in my mind, a contradiction to claim that inference processing uses on an interrelated system of cognitive resources with an overall reliance on cognitive control. A possible – and less theoretically interesting – explanation for the combined results is that the experimental set-up may have triggered a participant’s cognitive control in all the distractor conditions, washing out the effects of the short-term memory distractor. Correlational analyses are a more sensitive measure than ANOVAs, however, they also require larger sample sizes to make definitive claims. Given the small magnitudes of the correlations, in that the shared variance was between 25%-35% for all the significant correlations, these interpretations are, at best, tentative. A more theoretically interesting implication examines the concept of cognitive control as an

---

4 The correlational analyses reveal a relationship for short-term memory and cognitive control when participants process both inference types. However, when this is tested specifically across the three conditions, inferences are processed slowest in the cognitive control condition. This is preliminary evidence that cognitive control plays a larger role in inference processing, but it does not mean that inferences do not also rely their short-term memory.
amalgamation of processing sub-skills. It is clear that accessing short-term memory, in other words, the act of retrieving conceptual information, is a prerequisite for cognitive control, the subsequent manipulation of that information. Therefore, the cognitive control measure may be more predictive of inference processing, since it includes short-term memory activation. Together, the results of the second and the third research questions from the L1 and L2 studies suggest that cognitive control may play a significant role in the processing of inferences, regardless of type. This claim offers a unique contribution to sentence processing literature, however the present dissertation only looks at two types of inferences. This phenomenon would need to be tested with several inference types before saying definitively whether or not cognitive control is a mediating factor for inference processing as a whole.

Reviewing the two cognitive models surveyed in this dissertation, the construction-integration model (Kintsch, 1998) and the constructionist model (Graesser et al., 2001), recall that both forecasted that bridging and predictive inferences would be processed differently from one another. Graesser et al. (2001) and Kintsch (1998) consider bridging inferences to be processed early, and involuntarily, while predictive inferences are processed later, and with greater levels of attention. The analyses from the first research question, comparing the reading times for the different scenarios in the control condition, support the claim that bridging inferences may be processed more automatically while predictive inferences seem be processed in a controlled manner (Logan et al., 1999), especially in unencumbered processing (i.e., in the control condition). These comparisons need to be interpreted carefully, though, given that the reading times in the reaction-time sentence pilot yielded larger latencies for the set of
predictive inferences as opposed to the set of bridging inferences. Kintsch’s model supports a serial model of sentence processing (see the Levels-of-Processing account by Craik and Lockhart, 1972). As discussed above, the serial model has been largely discarded in favor of an interactive one (Trueswell & Gleitman, 2007; Van Gompel, 2006; Van Gompel & Pickering, 2007). However, it is still unclear what underlying mechanisms are at play in inference processing.

This does bring up the question as to whether the process of making a predictive inference is inherently more cognitively costly than a bridging inference. After all, the reader generates several interpretations for the forecasted event before selecting one likely interpretation. Compared to pronoun-referent mapping, this is a more complex process. Barnes and Jones’ (2000) multi-component approach to executive attention provides a useful framework for understanding the results of this dissertation. Throughout the present dissertation, I have referred to the construct that the N-Back task taps into as ‘cognitive control’. However the concept is labeled, whether cognitive control, executive attention, or executive control, the idea that executive attention has at least two modes might explain why it was activated irrespective of condition or inference type. Barnes and Jones (2000) believe the first mode is an involuntary, bottom-up process, while the second mode is a voluntary, top-down process (pp. 254-255). In the same vein, Logan et al. (1999) discuss the concept of voluntary versus involuntary processing; in light of this theory, bridging inferences were considered to tap into automatically processing and predictive inferences controlled processing.

As a linguistic skill, there is converging evidence that bridging inferences occur automatically. It may not be possible to suppress the pronoun-referent mapping that
occurs when she is linked to Sally. In contrast, predictive inferences may very well occur at a more conscious level. The multiple forecasted events that are generated at the outset of a predictive inference require a reader to monitor and track the goal of overall comprehension. Additionally, this would rely on the mechanisms of control specified in Rueda et al. (2005), such as conscious detection, inhibition, and conflict resolution. In the face of distracting information (i.e., the multiple readings), a reader must stay goal oriented (see Kanske, 2008). This interpretation supports both inference-based models covered in the literature review, the construction-integration model (Kintsch, 1998) and the constructionist model (Graesser et al., 2001), however, only in the context of an interactive processing account (see Trueswell & Gleitman, 2007; Van Gompel, 2006; Van Gompel & Pickering, 2007). Based on the results from this dissertation, I have updated Kintch’s (1998) model using the concept of executive attention to shed light on how bridging and predictive inferences are processed. At this point, I have created a revised model of inference generation that represents my best description of L1 inferences processing; the sample size for the L2 population was not large enough to make definitive claims and would need to be further tested. See Figure 7.1 for a visual depiction of the process of inference generation for bridging and predictive inferences; bridging inferences still are considered to be processed at the textbase level and predictive inferences are thought to be processed at the situation model.
Figure 7.1 Updated Inference Model (using Kintsch’s (1998) Levels of Processing Account and Barnes & Jones (2000) concept of executive attention).

Note that in contrast to Kintch’s serial model, inference processing is interactive, where the situation model can come into play at any level of processing. When the reader needs to access the situation model for background information or context, then the second mode of attention would be activated, a voluntary, top-down processing. Up until that point, more automatic processing would occur at the textbase level, activating the first mode of executive attention, the involuntary, bottom-up processing. Given the results of the present dissertation, this revised model only addresses inference generation for bridging and predictive types. Further research would need to be done in order to investigate other inference types.
To summarize, the results fit into an interpretation in which cognitive control, as a multi-component construct (see Barnes & Jones, 2000), accounts for the processing of bridging and predictive inferences in the L1 and the L2. Bridging inferences rely on automatic, bottom-up executive attentional processing, while predictive inferences rely on top-down, more voluntary executive attentional processing. These results support the inference-based comprehension models proposed by Kintsch (1998) and Graesser et al. (2001). However, rather than a serial account, processing is interactive (Trueswell & Gleitman, 2007; Van Gompel, 2006; Van Gompel & Pickering, 2007).

7.2.5. L1 versus L2 processing. Contrary to the conclusions found in the presented literature, the present dissertation did not find significant differences between L1 and L2 processing, with the exception of overall processing speed. L2 readers were discernably slower than L1 readers across all stimuli and conditions. However, both groups show similar processing patterns.

It is important to first address the finding that L2 readers were slower than L1 readers. This is a critical question when considering complex language processing as a phenomenon that relies on non-linguistic factors. It is likely that L2 participants read more slowly because they have had less exposure and practice reading in their L2 as opposed to L1 participants reading in their L1. In my opinion, this finding is not related to participants’ overall cognitive control ability, and might be better understood in a follow-up study when exposure to genre is considered (see DuBravac & Dalle, 2002).

Several studies have found that a high degree of proficiency is related to higher-order thinking (Cohen, 1998; Guerrero, 2005; Leontiev, 1981). This hypothesis was largely supported, as the participants in this study were able to comprehend fairly
complex texts. Since the L2 population was comprised of advanced speakers, it appears that proficiency is an important factor in L2 processing. While this dissertation did not compare different proficiency levels, the participants showed similar processing patterns regardless of language background. The higher level of proficiency possibly mediated the effects seen in previous studies.

Another area where speakers are thought to diverge is the mechanisms underlying processing. Schoenpflug and Klische (2010) found, in their study of monolingual, unbalanced bilingual, and balanced bilingual children, that L1 auditory processing differs from L2 auditory processing. Within the framework of the levels-of-processing account, Schoenpflug and Klische (2010) claim that L1 speakers use ‘bottom-up’ mechanisms and L2 speakers use ‘top-down’ mechanisms (not to be confused with the terms ‘bottom-up’ and ‘top-down’ in attentional research; see section 2.3.1 for clarification). It is important to mention here that the two experiments in this dissertation did not specifically investigate top-down versus bottom-up processing, per se. However, both inferences relied on different levels of cognitive control, which is principally both bottom-up and top-down. What is noteworthy is that the L1 and L2 readers followed the same pattern of processing.

Additionally, there is a question as to whether or not L2 speakers access the situation model when processing their second language (Horiba, 1996). In the present dissertation, the L1 and L2 participants correctly answered verification questions about the text that required them to progress beyond the surface level. Recall that the verification questions, when possible, were paraphrased so that participants could not
only access the surface structure of the text to answer the question. The questions forced the readers to process the meaning of the text, showing their access to the situation model.

Finally, much of the research on bilingual processing shows an advantage for cognitive control tasks when compared to monolinguals (Bialystok, 1998, 1999, 2004, 2009; Morton & Harper, 2007; Martin-Rhee & Bialystok, 2008; Costa et al., 2008, 2009; Hernández et al., 2010; Prior & Gollan, 2011; Poarch & van Hell, 2012). An established enhanced capacity did not bear out in the L2 study. This is likely due to the participants’ individual experience with different text genres, as well as how much time they read in English (i.e., their usage). This result is counter-intuitive, and is likely an artifact of the between group comparison and lack of fine-grained assessment of individual differences, such as genre and usage.

To summarize, there were no concrete processing differences found in the present dissertation between L1 and L2 readers. However, as discussed in the Limitation section, this is counter to majority of research on a bilingual advantage for cognitive control. These results need to be further verified with a within-subjects study of the L1 and L2 groups with finer-grained measures of exposure to genre and reading in general.

7.3. Limitations and Future Directions

The present dissertation has limitations that should be considered in future studies that investigate the role of individual differences in inference processing, including issues of methodology and construct validity. It is noteworthy that this research takes a novel approach to inference processing; as an early attempt, many methodological ‘lessons’ were learned.
7.3.1. Methodological caveats. The first methodological caveat is the sample size of the two studies, especially for the L2 group, threatening the generalizability of the results. In a between-group comparison, this would need to be remedied. Recall that a total of six participants were removed from the L2 analyses leaving a sample size of twenty participants. The L2 sample size was sufficient for the analysis of the reading time data in the repeated measures ANOVAs, but it was not sufficient for the correlational analyses. Also, while the L1 correlations of individual differences and reading times across the conditions and scenario types were significant, they only tell a fraction of the total picture based on the shared variance of the significant correlations. Furthermore, the standard deviations of the reading times in the cumulative self-paced reading task were quite large, which is entirely expected, given the parameters of the distraction paradigm. Participants’ attention was split between reading scenarios and answering questions about them and completing a secondary task. Individual differences likely influenced participants’ ability to manage this distraction; however, this did not present itself clearly in the results of the present dissertation. It is essential for future research to carry out these studies with larger sample sizes to see how individual differences mediate the participants’ ability to manage distraction when processing different scenario types. Additionally, the sample sizes of the two groups were not equal; the L1 group had fifty participants and the L2 group had twenty (once the outliers were identified). The groups would need to be of comparable size for a future study to provide generalizable results comparing the two groups.

An additional methodological caveat relates to the overall study design and the subsequent generalizability of the L1-L2 group comparison. The participant populations
in the two studies were different, L1 English speakers in the L1 study and L2 English-L1 Spanish speakers in the L2 study), required between-group comparisons to address any questions about the role of language background. Any generalizations about the role of language in inference generation are, at best, tentative. This is a particularly true for the finding that L2 readers did not exercise enhance cognitive control. A within-subject design – in which bilinguals read in their L1 and then their L2 – may yield more reliable results, and it would better control the variability of the individual differences. This, in turn, would offer more solid evidence about how L1 and L2 readers process inference and the role of cognitive resources in that processing. A future study could implement a design with built-in control populations. The study could use comparable Spanish and English verbal stimuli and test L1 Spanish-L2 English bilinguals, L1 English-L2 Spanish bilinguals, L1 Spanish monolinguals, and L1 English monolinguals. This would provide an appropriate comparison of the different groups using a within-subject design and give more definitive evidence about the role of cognitive control across the two populations.

Additionally, this study does not adequately separate possible confounds in the L2 participant population, including exposure to English, exposure to text genre, and types of bilingualism. The biodata form used to assess participants’ language background information did not provide a full enough picture of the participants’ language background to assess L2 inference processing. One issue had to do with the schooling, as there was a diversity of experience. Some of the L2 participants attended the majority of their classes in English, at American schools in their country, others were exposed to English as a foreign language for one class a day, and some attended bilingual schools. The factor of exposure to English was not completely controlled, and would need to be
better isolated in future studies. An additional concern is participants’ experience with text genre. Finer-grained measures of exposure to genre are needed to tease apart what role L2 participants’ experience with different types of text plays in their subsequent reading times. Finally, it would be useful to get a sense of how much time participants spend reading versus speaking. In a series of future studies with different types of proficiency and bilinguals and with a more comprehensive biodata form, stronger claims could be made about L2 sentence processing.

It is also important to mention task artificiality as a methodological caveat. Compromises are necessary in research, and the task design made it possible to investigate the research questions posed in the present dissertation. That said, the act of reading has many interrelated aspects, and the task design in the present dissertation reduces this complexity considerably. There were several ways that the task differed from a more natural reading process. One concession that I made was in regard to scenario length. Recall that each scenario was comprised of two-sentences, and while this allowed me to test processing across sentence boundaries, it did so in a limited context. This was a first step in inference processing research, and further studies could investigate this processing using longer, more involved texts. It would be important to control for matrix and sub clauses, so that the stimuli were comparable in terms of complexity and reading time. The second concern is that the cumulative self-paced reading task, in which participants read a sentence and had to push a button in order to reveal the second sentence, does not replicate normative reading. Readers typically have access to a complete text, and they can jump forward or backward with ease. The experimental design required participants to read the first sentence in isolation, without the added
benefit of being able to preview the upcoming text. Since it was not necessary to separately analyze the first and the second sentences of the scenarios in the present dissertation, the complete text could be presented in a future study without using a cumulative self-paced reading paradigm. A third concession was that participants not only read a text, but they also had to complete additional tasks – answering a verification question and completing a non-verbal memory-based task. While readers may internally ask questions of a text, they are generally not required to answer questions as part of the reading process. The verification questions were not analyzed in the two studies, as they were meant to ensure that the scenarios were processed and understood, but they still could have influenced the natural reading process. In addition, the secondary task in the distraction paradigm certainly made it possible to see the role of short-term memory and cognitive control in processing inferences, but it also provided a consistent, and contrived distraction for the reader. The use of the additional tasks was necessary, but they must also be considered when interpreting the results.

The next methodological caveat concerns the procedure employed in the two studies. Recall that the presentation of the experimental blocks for the self-paced reading task (which represented the conditions in the task: control, short-term memory, and cognitive control) was not fully randomized. Participants from the mini pilot reported cognitive overload when they were first presented with the short-term memory or cognitive control blocks (i.e., two tasks at once). This was remedied by a pseudo-randomized design in which participants initially completed the control block followed by a randomized presentation of the remaining blocks with the distractor tasks. Given the choice to either fully randomize the presentation of the blocks (with more chance of
cognitive overload) or to pseudo randomize the presentation of the blocks (with less chance of cognitive overload), I decided for the latter. Experimental design can be compared to a tightrope walk; when you control for one possible confounding variable, another comes in its place. In an effort to maximize participants’ ability to successfully complete the tasks in the distraction paradigm, I made a principled decision to semi-randomize the presentation of the blocks. This, in turn, impacted the participants’ reaction times on the control block. It is clear, though, that the initial presentation of the control block impacted the reading times (i.e., RTs for the short-term memory condition were faster than for the control condition), rendering them an unreliable control for the repeated-measures ANOVA analyses. It is relevant that participants were given 12 practice trials before each block to become accustomed to the task; however, in future research, participants could complete a separate longer block of the self-paced reading task that is not included in the analysis to habituate them to the task, followed by a fully randomized presentation of the control, short-term memory and cognitive control condition blocks.

Another methodological caveat concerns the fact that the blocks had mixed (i.e., interleaved) scenario types. It is important to mention that the study design intentionally presented the scenario types with and without inferences together in an interleaved block for each condition. Nevertheless, as mentioned in the discussion, participants were significantly slower to process non-inferences as compared to inferences. Recall that the use of non-inferences served as filler stimuli (to prevent a participant from discerning the aim of the study) and as control stimuli (to compare processing of inferences and non-inferences). It is likely that since a participant encountered scenarios with and without
inferences in the same block, they were operating in an inference mindset. In response to the finding that scenarios with non-inferences were processed relatively slowly, a future study could require a secondary control, in which the scenario types are not interleaved, but presented in separate blocks, to prevent participant strategies due to cognitive overload from confounding the results.

Another methodological caveat to the study design relates to the choice of language-independent distractor tasks. There is indirect evidence that participants might have adopted different processing strategies in the cognitive control and the short-term memory conditions. First, participants related that the short-term memory condition seemed to help facilitate faster comprehension of the sentences; this is likely an artifact of the pseudo-randomization discussed above, in that participants were accustomed to the task by the time they encountered the short-term memory block. Additionally, L1 and L2 participants reported that the block with the short-term memory condition (with the Visual Patterns distractor task) was considerably easier for them in comparison to the cognitive control condition (with the N-Back distractor task). It is worth mentioning that the N-Back task was designed to be more cognitively demanding than the Visual Patterns task in that participants had to track information across trials. This also bears out in the accuracy data across the two studies; L1 and L2 participants were by and large more accurate on the Visual Patterns task than the N-Back task for the stand-alone and the distractor tasks. However, it points to concerns with executive control measures that purport to assess a domain-general skill using linguistic stimuli (see Hernández et al., 2010; Costa et al., 2008). The stimuli used in the N-Back task are verbalizable (e.g., the concept of a square has a semantic equivalent) whereas the stimuli used in the Visual
Patterns task are not (e.g., the figure with three shaded squares has no semantic equivalent). As such, the participants in the cognitive control condition could have potentially adopted a strategy in which they repeated the names for the shapes in their minds to try to remember their order. In contrast, the Visual Patterns task does not allow for the same strategy and therefore does not introduce verbal interference. Future research could develop stimuli for the N-Back task that cannot be captured with verbal language. One suggestion is to use simplified, identifiable 3x3 grids, similar to those in the Visual Patterns task.

An additional methodological caveat concerns the necessary adaptations made to the cognitive tasks, stemming from practical concerns of the dual-task paradigm in the main experiment. It was a challenge to locate tasks that purportedly measured the requisite sub-skills and still functioned in the reading experiment. All of the tasks were modeled after reputable measures with high construct- and face-validity in cognitive science. However, since the language-independent tasks (i.e., Visual Patterns and N-Back Tasks) were embedded in a self-paced reading experiment, it was necessary to adapt them for the dual-task paradigm. In their original conception, all the tasks used in this research were meant to serve as stand-alone tasks, and therefore needed to be quite cognitively demanding. However, as a secondary distractor task in this research, the tasks needed to be just challenging enough to distract, but not interfere, with the reading process. As noted in the results section, the accuracy on the Visual Patterns task was quite high for the stand-alone and distractor versions across L1 and L2 participants; admittedly the design used in this experiment was not as complex as the one conceived and carried out
by Della Sala et al. (1997). Future research might create more and less cognitively demanding distractor tasks to see their relative impact on inference processing.

The final methodological caveat to the present dissertation is direct comparison of the different scenarios to each other used in the first research question. Recall that the reading times in the sentence reaction-time pilot yielded larger reaction time latencies for the predictive inferences than the bridging inferences across the blocks; in other words, it took participants longer to read the predictive inferences than the bridging inferences. While the verbal stimuli were carefully controlled in several different ways (e.g., the sentences across the scenario types had the same average number of words per sentence, the same types of topic areas, etc.), there could still be frequency effects at play. It is also possible that bridging inferences take longer to process than predictive inferences. More definitive claims about the processing of the different inference types could be made with stimuli that have similar reaction times across scenario types.

7.3.2. Construct validity. The next set of limitations concerns the construct validity of the tasks and stimuli, in other words, whether or not these measured what they were designed to measure. It appears that the tasks and stimuli in the present dissertation had fairly high construct validity, as evidenced by the pretests. However, these two studies point to more theoretical concerns.

The first theoretical concern is related to the act of measuring cognitive capacity. It is quite challenging to design a task that assesses a singular, or pure, form of a cognitive ability, as processing undeniably occurs at the macro level (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Pickering, 2001). To support this claim further, Conway et al. (2002) maintain that all span tasks, whether simple or complex, “will tap
each capacity to some extent” (p. 165). By that same logic, other measures of cognitive capacity follow this model, in that measuring phonological short-term memory incidentally accesses the entire working memory system. The body of literature on cognitive measures finds that many of the tasks that purport to measure an isolated cognitive skill often correlate with general measures of working memory, but not with other ‘like’ measures. This bears out in the literature with the Stroop task (see Appendix D). Kane and Engle (2003) found correlations between the Stroop task and general working memory capacity, even though it is thought to measure cognitive control, a sub-skill of working memory. The present dissertation found correlations between the stand-alone and distractor measures for cognitive control and short-term memory, but also across the different cognitive capacities (see Section 4.5.6). In other words, the Visual Patterns task (a measure of short-term memory) was significantly correlated with the N-Back task (a measure of cognitive control). There is a need for more nuanced measures of cognitive control that assess the bottom-up and top-down features of a multi-component system. Future studies might also try different types of distractor tasks for each cognitive capacity to triangulate the measures.

In addition, the cumulative self-paced reading experiment assessed inference building at the time of the reading of the two sentences. It is, unfortunately, not possible to know whether the participants processed the inference at the time of reading the two-sentence scenarios or at the time of answering the verification question. The verification question was employed only to make sure that participants read and processed the two-sentence scenarios. Brain-based studies might be able to shed light on this kind of question, but the task demands would also need to be more constrained.
In light of these methodological and experimental caveats, some suggestions for future research are discussed in the Future Directions section. Before addressing these, I will discuss the implications for this research within the fields of SLA and psycholinguistics.

7.4. Implications

As previously discussed, the results need to situated within a larger research agenda. As such, the implications for the results of this research will be discussed in reference to the field mentioned in the introduction in this section, SLA research and psycholinguistic research.

7.4.1. Implications for SLA research. Inference processing has immediate applications for educational interventions and language testing. The present dissertation suggests that bridging and predictive inferences rely on cognitive control for their processing in normative adult processing. It has been found that emergent (i.e., struggling) readers are often misdiagnosed as having diminished cognitive control capacity, when, in reality, there are several possible comorbid conditions that could be contributing to the developmental delay, such as ADHD, decreased phonemic awareness, and lower levels of verbal working memory capacity (e.g., Willcutt et al., 2001). The accompanying conditions, however, have been linked to cognitive control deficits. Furthermore, the catchall phrase of ‘working memory’ includes a person’s ability not only to store information, but also to manipulate it; this is the definition of cognitive control. Cognitive control is especially apropos in light of the results of these two studies and it can, furthermore, apparently be developed through brain training exercises (Hussey & Novick, 2012; Novick, Hussey, Teubner-Rhodes, Harbison, & Bunting, 2013). If a
developing reader is struggling with inference generation in reading comprehension, then an intervention utilizing activities that strengthen cognitive control could assist with the reading process.

These results also have implications for language testing, especially in light of construct-irrelevant variance (CIV), in other words the mistaken “inflation or deflation of test scores due to certain types of uncontrolled or systematic measurement error” (Downing, 2002, p. 235). A possible measurement error could result from testing methods. Inference generation is considered a benchmark for comprehension in the majority of formal and informal reading inventories, as repeating propositional content verbatim does not demonstrate comprehension, but rather access to memory resources. Nonetheless, many of the testing procedures require a student to read a text and answer questions without being allowed to visually refer back the text. Essentially, this procedure tests the working memory of the students and not their comprehension ability. If cognitive control plays a larger role than initially though in inferencing, then a student could have high cognitive control, but low proficiency, yet falsely present as high proficiency. Conversely, if a student has low cognitive control, but high proficiency, this might falsely present as low proficiency By replicating these conditions, the student is at a disadvantage in the testing process.

7.4.2. Implications for psycholinguistics. The theoretical implication of the results of the present dissertation, as discussed above, is that both inference types rely on cognitive control (using the construct of the two modes of executive attention) for their successful comprehension. If cognitive control possibly accounts for inference processing in general, then it would be important to see the relationship of cognitive
control and general reading ability in future studies. Furthermore, the role of cognitive control may need to be operationalized differently in future studies – specifically the specific aspect of cognitive control, automatic or controlled – in research and processing.

The present L1 and L2 studies have not successfully tested non-inferences as a basis of comparison for inferences. The construction-integration model (Kintsch, 1998) and the constructionist model (Graesser et al., 2001) are supported by these results for bridging and predictive inferences, however, only when processing is interactive (Trueswell & Gleitman, 2007; Van Gompel, 2006; Van Gompel & Pickering, 2007).

7.5. Future Directions

The present dissertation was designed to uncover the role of individual differences when bridging and predictive inferences are processed. As a preliminary attempt, this research was able to identify cognitive control as a factor in inference generation, at least for the two types explored in these two studies. Nevertheless, the scope of these studies was intentionally narrow, which leaves room for future studies. Below are several ideas on how this research could be expanded.

While it is clear that cognitive control, as understood within the attentional control model, plays a role in the processing of bridging and predictive inference, there are two interesting questions that were not completely teased apart in this dissertation: Which individual differences are used to process bridging and predictive inferences? and How does a higher or lower cognitive degree of control capacity mediate this processing? This could be researched by replicating the study – using the methodological innovations in the Limitations and Future Directions section – with larger sample sizes for the L1 and the L2 studies. It is likely that a larger participant pool would have more
inherent variation in processing and, hence, show more discernable patterns. Moreover, the distractor tasks could be adapted in two different ways. First, two blocks could be added to the study design with a secondary cognitive control distractor and secondary short-term distractor. This would allow for triangulation of the measures. A second innovation might introduce a third distractor task that does not interfere with short-term memory or cognitive control to serve as a ‘control’ distractor block. All of these adaptations would strengthen the validity of the current study design.

Another expansion of this research could incorporate the methodological improvements while examining how different types of stimuli would be processed. First, a future study could investigate how other inference types, aside from bridging and predictive inferences, are processed using the same experimental design. This would either provide evidence for or against the finding that cognitive control was a part of general inference processing. Along the same lines, it would be worthwhile to research how more and less reliably generated inferences are processed. The verbal stimuli that were developed and tested in this research were constrained to inferences that are reliably made; this is a small fraction of a much larger picture of inference processing. As stated in the introductory chapter, many utterances are purposefully made confusing and end up misconstrued or with reversed meaning for comic effect. It would be interesting for future research to investigate the role individual differences plays in the processing of more and less likely made bridging and predictive inferences.

To address the question of proficiency, a future study could look at higher and lower levels of linguistic ability using the same methodology. This would offer definitive support for or against the claim that L2 readers have access to the situation level of text
processing. It would be important to recruit participants who have enough proficiency to understand the stimuli in the experiment. With more participants, factors that contribute to proficiency, such as usage, exposure, motivation, etc. could be explored using regression analysis. Also, an additional task could be added to the study, in which participants are required to verbally summarize a subset of the scenarios to see if they have surface-level or gist-level understanding.

Finally, this research could be expanded to look at auditory processing, using an auditory-visual distraction paradigm. Scenarios could be auditorally presented for participants, along with the verification questions, and the distractor tasks could be visually presented. It is even possible to allow participants control on the presentation of the auditory stimuli, using a button press. It is important to mention that this information might be less reliable, as the recording is set at a certain speed for presentation. That said, it still offers valuable information about processing.

The innovations presented above offer several possible directions for future research in the area of inference processing.

7.6. Conclusion

The present dissertation, comprised on an L1 reading study and an L2 reading study, contributes to several areas of sentence processing literature. It first investigated the question of whether inferences are processed differently than non-inferences, situated within an interactive processing account (see Trueswell & Gleitman, 2007; Van Gompel, 2006; Van Gompel & Pickering, 2007 for reviews). While the established literature in auditory processing suggest that non-inferences are less cognitively demanding than inferences, and therefore require less processing time (Huang & Snedeker 2009a, 2009b,
2011), the present studies were not able to resolve this question. Participants went into an ‘inference’ mode of processing in the interleaved stimuli presentation, and it is hypothesized that participants were cognitively overloaded (Scalf, Torralbo, Tapia, & Beck, 2013 discuss this in the context of perception load and dilution theories). Therefore, L1 and L2 participants adopted processing strategies when processing non-inferences in the control block, which presented in longer reaction time latencies.

This dissertation also explored whether or not L1 and L2 readers rely on cognitive resources to process inferences. The correlational analyses revealed that there is interrelated recruitment of cognitive resources in the processing of bridging and predictive inferences, with significant relationships for cognitive control and, to a lesser extent, short-term memory. It is worth mentioning that the experimental design may have incidentally activated participants’ cognitive control, resulting in enhanced control. This result can be understood within the context of the conflict monitoring effect (see Botvinick et al., 2001), in that cognitive control was incidentally activated in the face of distracting stimuli. However, in light of the results from the third research question, it is more likely that cognitive control plays a larger role in the processing of inferences than initially assumed. In a related question, the present studies also researched how L1 and L2 readers specifically processed the scenarios with bridging inferences and predictive inferences. The current studies’ results support the claim that bridging and predictive inferences recruit a multi-component cognitive control capacity for their processing (see Barnes and Jones, 2000, for a view of executive attention in which there is involuntary, bottom-up and voluntary top-down control of attention). Based on the sum of the results in the present dissertation, I updated Kintsch’s (1998) inference processing model in
reference to L1 processing, with a few modifications. A significant change in the model is that rather than a serial sentence processing account (e.g., Clifton et al., 2003; Frazier & Rayner, 1982), I suggest an interactive processing account (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Stanovich, & West, 1983; Trueswell, Tanenhaus, & Garnsey, 1994) for bridging and predictive inferences. As such, context and background knowledge can be accessed at any stage of processing. Importantly, this updated inference comprehension model still considers bridging inferences to be processed at the textbase level and predictive inferences to be processed at the situation model. However, the cognitive resources used to process these inference types would vary. Bridging inferences are hypothesized to rely on automatic processing and, therefore, occur at the textbase level. Hence, they activate the first mode of executive attention: involuntary, bottom-up processing. In contrast, predictive inferences are hypothesized to rely on controlled processing, and therefore, occur at the situation model. They trigger the second mode of executive attention: voluntary, top-down processing.

The final research question explored how L1 and L2 readers process non-inferences, bridging inferences and predictive inferences. Contrary to the findings in the literature, the present dissertation did not find significant differences between L1 and L2 processing comparing the trends, with the exception of overall processing speed. L2 readers were discernably slower than L1 readers across all stimuli and conditions; however, both groups show similar processing patterns. The study would need to be replicated with an within-subject design to draw more definitive conclusions in regard to language background.
Several methodological innovations were also proposed, as well as ideas for future research, to expand this research further and address overall inference processing and the roles of cognitive resources and proficiency in L1 and L2 auditory and reading comprehension.
APPENDICES

Appendix A: Rating Study Directions

Appendix A1: Bridging Inferences

DIRECTIONS

In this section, you will see a two-sentence scenario and be presented with four statements about the stories. You will then choose two statements whose content best represents what you just read.

Read the two-sentence scenario below:

- Greta had read all the great classics.
- She loved to discuss literature with friends.

Now look at the statements below. Which two sentences best state what you just read from the four choices below?

1. Greta enjoyed talking about the books she read.
2. Greta enjoyed talking about politics.
3. She had read many modern books.
4. She had read many classic books.

The best options are *Greta enjoyed talking about the books she read* and *She had read many classic books*. It does not make sense to say that *Greta enjoyed talking about politics*, as she likes to discuss literature, not politics. Nor does it make sense to say *She had read many modern books*, as she reads classical not modern literature. This means you would choose options 1 and 4.

For the next set of 60 two-sentence scenarios, please choose two options from a set of four choices based on your intuition about what statements best represent the content of the two-sentence scenarios.

As before, there will be three screens for this task. This is the second to last task.

Let's begin.

*** *** *** *** *** *** *** ***
Appendix A2: Predictive Inferences

DIRECTIONS

In this section, you will see a two-sentence scenario and be presented with four possible outcomes. You will then choose the most logical outcome from a set of four choices.

Read the two-sentence scenario below:
- The dog ran into the middle of the street.
- The driver did not see the dog

Now look at the possible outcomes. Which is the most likely of the four choices below?

1. The driver fell asleep.
2. The driver missed the dog.
3. The driver hit the dog.
4. The driver hit a tree.

The most likely option is The driver hit the dog because the driver did not see the dog that ran into the middle of the street. This means you would choose option 3.

For the next set of 60 two-sentence scenarios, please choose only one option from a set of four choices based on your intuition about what outcome would most likely occur.

Let's begin.
Appendix A3: No Inference

DIRECTIONS

In this section, you will see a set of three sentences. You will then choose the sentence that does not fit from the three options.

Read the set of three sentences below:

1. The students learned how to make prints for art.
2. The students learned that the Egyptians built pyramids.
3. The students were doing a play about ancient Egypt.

Which sentence does not fit with the others?

The sentence that does not fit is The students learned how to make prints for art because sentences 2 and 3 have to do with Egypt. This means you would choose option 1.

For the next set of 30 sets of sentences, please choose only one option from a set of three choices based on your intuition about what sentence does not fit.

This is the last task.

Let's begin.

*** *** *** *** *** *** *** ***
### Appendix B: Final Verbal Stimuli

#### Appendix B1: Two-Sentence Scenarios with Bridging Inferences

**KEY:** yes questions (i-ii) for items 1-15, yes questions (ii-i) for items 16-30, and no questions (i-ii) for items 31-45, no questions (ii-1) for items 46-60.

<table>
<thead>
<tr>
<th>#</th>
<th>Sentence 1</th>
<th>Sentence 2</th>
<th>Inference</th>
<th>Verification Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ashley loved to drink coffee in the morning.</td>
<td>She liked a special brew of coffee from Guatemala.</td>
<td>She refers to Ashley.</td>
<td>Did Ashley enjoy Guatemalan coffee?</td>
</tr>
<tr>
<td>2</td>
<td>Becca invited the co-workers to a happy hour.</td>
<td>She was happy the turnout was so good.</td>
<td>She refers to Becca.</td>
<td>Did Becca feel pleased so many people came?</td>
</tr>
<tr>
<td>3</td>
<td>Kaitlyn was excited about turning ten.</td>
<td>She planned a big birthday party.</td>
<td>She refers to Kaitlyn.</td>
<td>Did Kaitlyn organize a large birthday party?</td>
</tr>
<tr>
<td>4</td>
<td>The car was going over the speed limit.</td>
<td>It was pulled over by the police.</td>
<td>It refers to the car.</td>
<td>Did the police stop the car?</td>
</tr>
<tr>
<td>5</td>
<td>Andrew stayed up all night to finish the project.</td>
<td>He fell asleep in class the next day.</td>
<td>He refers to Andrew.</td>
<td>Did Andrew nap in class the following day?</td>
</tr>
<tr>
<td>6</td>
<td>Michael said a bad word to the teacher.</td>
<td>He was sent to time out during recess.</td>
<td>He refers to Michael.</td>
<td>Did Michael have to sit in time out during free time?</td>
</tr>
<tr>
<td>7</td>
<td>Matthew was talented at programming computers.</td>
<td>He was always called to solve problems.</td>
<td>He refers to Matthew.</td>
<td>Did Matthew always get consulted to deal with problems?</td>
</tr>
<tr>
<td>8</td>
<td>Noah refused to go on an airplane.</td>
<td>He was deathly afraid of flying.</td>
<td>He refers to Noah.</td>
<td>Did Noah have a fear of flying?</td>
</tr>
<tr>
<td>9</td>
<td>The cook cut a finger badly.</td>
<td>She was rushed to the emergency room.</td>
<td>She refers to Leah.</td>
<td>Did the cook get taken to the hospital?</td>
</tr>
<tr>
<td>10</td>
<td>Megan packed a lot of luggage for the trip.</td>
<td>She screamed at the flight attendant about the lost bags.</td>
<td>She refers to Megan.</td>
<td>Did Megan react badly about the lost bags?</td>
</tr>
<tr>
<td>11</td>
<td>Ella attended every class every semester.</td>
<td>She graduated at the top of the class.</td>
<td>She refers to Ella.</td>
<td>Did Ella graduate with good grades?</td>
</tr>
<tr>
<td>12</td>
<td>Hannah applied to study in Turkey in the fall.</td>
<td>She anxiously awaited the decision.</td>
<td>She refers to Hannah.</td>
<td>Did Hannah nervously wait for an answer?</td>
</tr>
<tr>
<td>13</td>
<td>Julie liked to borrow clothes from friends.</td>
<td>She often did not give the clothes back.</td>
<td>She refers to Julie.</td>
<td>Did Julie seldom return the clothes?</td>
</tr>
<tr>
<td>14</td>
<td>The chocolate was being sold for half price.</td>
<td>It was sold out by noon.</td>
<td>It refers to the chocolate.</td>
<td>Did the chocolate get bought out by lunch?</td>
</tr>
<tr>
<td>15</td>
<td>The baker wrote the wrong name on the cake.</td>
<td>She was fired the following day.</td>
<td>She refers to the baker.</td>
<td>Did the baker get let go the next day?</td>
</tr>
<tr>
<td>16</td>
<td>Alex sat in front of the television.</td>
<td>He loved to watch movies on Sundays.</td>
<td>He refers to the Alex.</td>
<td>Did he sit in front of the television?</td>
</tr>
<tr>
<td>17</td>
<td>Ella ate the entire cake.</td>
<td>She felt sick all night.</td>
<td>She refers to Ella.</td>
<td>Did she eat the whole cake?</td>
</tr>
<tr>
<td>18</td>
<td>Hannah was afraid of flying.</td>
<td>She spent many vacations traveling domestically.</td>
<td>She refers to Hannah.</td>
<td>Did she worry about flying?</td>
</tr>
<tr>
<td></td>
<td>Julie waited tables at many weddings.</td>
<td>She especially enjoyed the summer weddings.</td>
<td>She refers to Julie.</td>
<td>Did she work as a waitress at several weddings?</td>
</tr>
<tr>
<td>---</td>
<td>-------------------------------------</td>
<td>---------------------------------------------</td>
<td>---------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>20</td>
<td>The museum had many visitors.</td>
<td>It had a diverse collection of art.</td>
<td>It refers to museum.</td>
<td>Did it have several visitors?</td>
</tr>
<tr>
<td>21</td>
<td>Nicole loved jogging in the park on Sundays.</td>
<td>She even went in the rain.</td>
<td>She refers to Nicole.</td>
<td>Did she enjoy running in the park on the weekends?</td>
</tr>
<tr>
<td>22</td>
<td>Nathan called a friend to say happy birthday.</td>
<td>He got the answering machine.</td>
<td>He refers to Nathan.</td>
<td>Did he call to wish a friend happy birthday?</td>
</tr>
<tr>
<td>23</td>
<td>Rachel was a business major at the university.</td>
<td>She also majored in East Asian government.</td>
<td>She refers to Rachel.</td>
<td>Did she major in business at school?</td>
</tr>
<tr>
<td>24</td>
<td>Robert rescued a puppy from the shelter.</td>
<td>He got the puppy to stop peeing in the house.</td>
<td>He refers to Robert.</td>
<td>Did he save a puppy from the shelter?</td>
</tr>
<tr>
<td>25</td>
<td>Tommy had eaten three cookies in bed.</td>
<td>He left cookie crumbs in the sheets.</td>
<td>He refers to Tommy.</td>
<td>Did he eat cookies in the bed?</td>
</tr>
<tr>
<td>26</td>
<td>Ethan frequently flew to visit family in Manila.</td>
<td>He could not go last year due to work.</td>
<td>He refers to Ethan.</td>
<td>Did he often travel to Manila to spend time with family?</td>
</tr>
<tr>
<td>27</td>
<td>Evan often organized football games outside during recess.</td>
<td>He got upset on rainy days.</td>
<td>He refers to Evan.</td>
<td>Did he usually put together football games during recess?</td>
</tr>
<tr>
<td>28</td>
<td>Henry worked as a bartender on weekends.</td>
<td>He was an expert at mixing drinks.</td>
<td>He refers to Henry.</td>
<td>Did he work weekends at the bar?</td>
</tr>
<tr>
<td>29</td>
<td>The travel agent worked until late most days.</td>
<td>She dreamed of a vacation in Hawaii.</td>
<td>She refers to the travel agent.</td>
<td>Did she work many long hours?</td>
</tr>
<tr>
<td>30</td>
<td>Lauren was in Paris for the first time.</td>
<td>She did not want to see the Eiffel Tower.</td>
<td>She refers to Lauren.</td>
<td>Did she visit Paris for the first time?</td>
</tr>
<tr>
<td>31</td>
<td>David studied often at the library.</td>
<td>He focused well on homework at the library.</td>
<td>He refers to David.</td>
<td>Did David concentrate poorly at the library?</td>
</tr>
<tr>
<td>32</td>
<td>Jenny studied hard for the math test.</td>
<td>She received a good grade.</td>
<td>She refers to Jenny.</td>
<td>Did Jenny get a bad grade?</td>
</tr>
<tr>
<td>33</td>
<td>Ethan forgot to bring his lunch to school.</td>
<td>He borrowed lunch money from the teacher.</td>
<td>He refers to Ethan.</td>
<td>Did the bus driver lend Ethan money to buy lunch?</td>
</tr>
<tr>
<td>34</td>
<td>Evan was off from work on Saturdays.</td>
<td>He spent the day doing house chores.</td>
<td>He refers to Evan.</td>
<td>Did Evan go to the park all day?</td>
</tr>
<tr>
<td>35</td>
<td>The bride missed the flight.</td>
<td>She was forced to rent a car.</td>
<td>She refers to the bride.</td>
<td>Did the bride have to rent a jet?</td>
</tr>
<tr>
<td>36</td>
<td>Sarah started a food fight during lunch.</td>
<td>She was punished by the principal.</td>
<td>She refers to Sarah.</td>
<td>Did Sarah get into trouble with the janitor?</td>
</tr>
<tr>
<td>37</td>
<td>Sophie forgot to lock the front door.</td>
<td>She came back to the door wide open.</td>
<td>She refers to Sophie.</td>
<td>Did Sophie return to a closed door?</td>
</tr>
<tr>
<td>38</td>
<td>The bus arrived late in New York.</td>
<td>It was filled with angry passengers.</td>
<td>It refers to bus.</td>
<td>Did the bus have many happy customers?</td>
</tr>
<tr>
<td>39</td>
<td>Andrew met several college friends for dinner.</td>
<td>He was happy to reconnect with old friends.</td>
<td>He refers to Andrew.</td>
<td>Did Andrew feel annoyed to see old friends again?</td>
</tr>
<tr>
<td>40</td>
<td>Alex had recently divorced parents.</td>
<td>He stayed half time with each parent.</td>
<td>He refers to Alex.</td>
<td>Did Alex live full time with the mother?</td>
</tr>
<tr>
<td>41</td>
<td>Joseph memorized the alphabet in class.</td>
<td>He recited the alphabet perfectly.</td>
<td>He refers to Joseph.</td>
<td>Did Joseph say the multiplication</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>42</td>
<td>Leah took a taxi to the hotel.</td>
<td>She refused to tip the driver.</td>
<td>She refers to Leah.</td>
<td>Did Leah pay the driver a large tip?</td>
</tr>
<tr>
<td>43</td>
<td>Justin was accepted to Hunter College.</td>
<td>He was excited about living in New York City.</td>
<td>He refers to Justin.</td>
<td>Did Justin dislike the idea of moving to New York City?</td>
</tr>
<tr>
<td>44</td>
<td>Kevin went on a field trip to the zoo.</td>
<td>He most wanted to catch a glimpse of the tigers.</td>
<td>He refers to Kevin.</td>
<td>Did Kevin look forward to seeing the elephants?</td>
</tr>
<tr>
<td>45</td>
<td>Michael loved taking care of animals.</td>
<td>He found a job at a pet store.</td>
<td>He refers to Michael.</td>
<td>Did Michael find employment at a clothing store?</td>
</tr>
<tr>
<td>46</td>
<td>Lauren always went on vacation in March.</td>
<td>She decided to travel to Rome this year.</td>
<td>She refers to Lauren.</td>
<td>Did she always take trips in September?</td>
</tr>
<tr>
<td>47</td>
<td>Henry often put in overtime at the office.</td>
<td>He was praised for excellent performance on company projects.</td>
<td>He refers to Henry.</td>
<td>Did he often leave early?</td>
</tr>
<tr>
<td>48</td>
<td>Joseph frequently visited the coffee house.</td>
<td>He eventually asked the barista on a date.</td>
<td>He refers to Joseph.</td>
<td>Did he go to the bookstore a lot?</td>
</tr>
<tr>
<td>49</td>
<td>Justin went camping regularly with friends.</td>
<td>He liked the atmosphere away from the city.</td>
<td>He refers to Justin.</td>
<td>Did he often go out dancing with friends?</td>
</tr>
<tr>
<td>50</td>
<td>Kevin was the leader in a rock band.</td>
<td>He was an excellent singer.</td>
<td>He refers to Kevin.</td>
<td>Did he lead an orchestra?</td>
</tr>
<tr>
<td>51</td>
<td>David brought a pie to the dinner.</td>
<td>He was responsible for bringing dessert.</td>
<td>He refers to David.</td>
<td>Did he bring drinks?</td>
</tr>
<tr>
<td>52</td>
<td>Ashley waited tables at a French cafe.</td>
<td>She liked to practice speaking French with the customers.</td>
<td>She refers to Ashley.</td>
<td>Did she wait tables at an Italian cafe?</td>
</tr>
<tr>
<td>53</td>
<td>The clinic had more patients than ever.</td>
<td>It did not have enough chairs for the patients.</td>
<td>It refers to the clinic.</td>
<td>Did it have few sick people to treat?</td>
</tr>
<tr>
<td>54</td>
<td>Becca regularly showed up late to the office.</td>
<td>She was reprimanded by the manager.</td>
<td>She refers to Becca.</td>
<td>Did she usually come on time to work?</td>
</tr>
<tr>
<td>55</td>
<td>Emma enjoyed visiting museums on her vacations.</td>
<td>She enjoys art museums most of all.</td>
<td>She refers to Emma.</td>
<td>Did she enjoy going to the beach on holiday?</td>
</tr>
<tr>
<td>56</td>
<td>Megan went on a vacation to Hong Kong.</td>
<td>She was excited to see the sights.</td>
<td>She refers to Megan.</td>
<td>Did she take a work trip to Hong Kong?</td>
</tr>
<tr>
<td>57</td>
<td>Matthew spent the entire day in the sun.</td>
<td>He enjoyed the cruise immensely.</td>
<td>He refers to Matthew.</td>
<td>Did he spend the day inside?</td>
</tr>
<tr>
<td>58</td>
<td>Nicole always mowed the lawn on Sunday.</td>
<td>She liked the smell of cut grass.</td>
<td>She refers to Nicole.</td>
<td>Did she do laundry on Sundays?</td>
</tr>
<tr>
<td>59</td>
<td>Noah got on the school bus.</td>
<td>He found a seat next to the teacher.</td>
<td>He refers to Noah.</td>
<td>Did he miss the bus?</td>
</tr>
<tr>
<td>60</td>
<td>Nathan punched another boy in the schoolyard.</td>
<td>He got in trouble with the teacher.</td>
<td>He refers to Nathan.</td>
<td>Did he play football with a boy?</td>
</tr>
</tbody>
</table>
### Appendix B2: Two-Sentence Scenarios with Predictive Inferences

**KEY:** yes questions for items 1-30 and no questions for items 31-60.

<table>
<thead>
<tr>
<th>#</th>
<th>Sentence 1</th>
<th>Sentence 2</th>
<th>Inference</th>
<th>Verification Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Tommy was the top employee in the section.</td>
<td>Tommy was scheduled for an annual review.</td>
<td>Tommy got a good review.</td>
<td>Did Tommy get a good review?</td>
</tr>
<tr>
<td>2</td>
<td>Megan loved being in the wilderness.</td>
<td>Several friends were planning to go camping this weekend.</td>
<td>Megan went on the camping trip.</td>
<td>Did Megan go on the camping trip?</td>
</tr>
<tr>
<td>3</td>
<td>Sarah screamed a curse at another student.</td>
<td>The teacher overheard Sarah cursing.</td>
<td>Sarah got into trouble for cursing.</td>
<td>Did Sarah get into trouble for cursing?</td>
</tr>
<tr>
<td>4</td>
<td>The conference was scheduled for Saturday morning.</td>
<td>It snowed three feet on Friday night.</td>
<td>The conference was cancelled.</td>
<td>Did the conference get canceled?</td>
</tr>
<tr>
<td>5</td>
<td>Robert had almost finalized the guest list for the wedding.</td>
<td>Robert suddenly remembered an old friend from school.</td>
<td>Robert added the old friend to the guest list.</td>
<td>Did Robert add the old friend to the guest list?</td>
</tr>
<tr>
<td>6</td>
<td>The thief planned a bank robbery.</td>
<td>The thief unknowingly tripped the silent alarm.</td>
<td>The thief was caught by the police.</td>
<td>Did the thief get caught?</td>
</tr>
<tr>
<td>7</td>
<td>The vase was sitting on the edge of the counter.</td>
<td>Nathan flung a book bag across the counter.</td>
<td>The book bag pushed the vase off the counter.</td>
<td>Did the book bag push the vase off the counter?</td>
</tr>
<tr>
<td>8</td>
<td>Noah was reading text messages while walking.</td>
<td>Noah did not notice the open sewer grate.</td>
<td>Noah fell into the sewer.</td>
<td>Did Noah fall into the sewer?</td>
</tr>
<tr>
<td>9</td>
<td>Rachel was suddenly very hungry.</td>
<td>Rachel walked toward the sandwich shop.</td>
<td>Rachel ordered a sandwich.</td>
<td>Did Rachel order a sandwich?</td>
</tr>
<tr>
<td>10</td>
<td>Matthew needed to pass every course to graduate on time.</td>
<td>Matthew failed the Biology 101 final.</td>
<td>Matthew graduated late.</td>
<td>Did Matthew graduate late?</td>
</tr>
<tr>
<td>11</td>
<td>The phone on the desk rang twice.</td>
<td>The secretary reached for the phone.</td>
<td>The secretary answered the phone.</td>
<td>Did the secretary answer the phone?</td>
</tr>
<tr>
<td>12</td>
<td>Joseph bought a diamond engagement ring.</td>
<td>Joseph made a reservation for two at a fancy restaurant.</td>
<td>Joseph proposed to the girlfriend.</td>
<td>Did Joseph propose marriage to the girlfriend?</td>
</tr>
<tr>
<td>13</td>
<td>The laptop battery only had 5% charge left.</td>
<td>Andrew kept working on the computer.</td>
<td>The computer died.</td>
<td>Did the computer die?</td>
</tr>
<tr>
<td>14</td>
<td>Henry wanted to get more involved in the office community.</td>
<td>The boss needed an employee to organize the picnic.</td>
<td>Henry volunteered to organize the picnic.</td>
<td>Did Henry volunteer to organize the picnic?</td>
</tr>
<tr>
<td>15</td>
<td>Jenny took the leaky boat out on the water.</td>
<td>Jenny forgot to bring a bucket.</td>
<td>The boat filled up with water.</td>
<td>Did the boat fill up with water?</td>
</tr>
<tr>
<td>16</td>
<td>Julie was in terrible traffic heading to work.</td>
<td>The company meeting was beginning in a few minutes.</td>
<td>Julie was late for the meeting.</td>
<td>Did Julie come late for the meeting?</td>
</tr>
<tr>
<td>17</td>
<td>Evan routinely listened to loud rock music at work.</td>
<td>The boss called Evan in for a talk.</td>
<td>The boss asked Evan to lower the music.</td>
<td>Did the boss ask Evan to lower the music?</td>
</tr>
<tr>
<td>18</td>
<td>The art teacher loved photography.</td>
<td>Hannah asked for a new project.</td>
<td>The teacher suggested a photography project.</td>
<td>Did the teacher suggest a photography project?</td>
</tr>
<tr>
<td>19</td>
<td>The ice in the middle of the pond was thin.</td>
<td>Ella began skating across the pond.</td>
<td>The ice broke.</td>
<td>Did the ice break?</td>
</tr>
<tr>
<td>No.</td>
<td>Statement 1</td>
<td>Statement 2</td>
<td>Statement 3</td>
<td>Question</td>
</tr>
<tr>
<td>-----</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>----------</td>
</tr>
<tr>
<td>20</td>
<td>Ethan did not read well.</td>
<td>Ethan worked with a tutor after school every day.</td>
<td>Ethan improved in reading.</td>
<td>Did Ethan improve in reading?</td>
</tr>
<tr>
<td>21</td>
<td>Rachel saw a beautiful necklace at the souvenir shop.</td>
<td>Rachel had just enough money to buy the necklace.</td>
<td>Rachel bought the necklace.</td>
<td>Did Rachel buy the necklace?</td>
</tr>
<tr>
<td>22</td>
<td>Nicole had not eaten the entire day.</td>
<td>Nicole passed a favorite Thai restaurant on the way home.</td>
<td>Nicole stopped for Thai food.</td>
<td>Did Nicole stop for Thai food?</td>
</tr>
<tr>
<td>23</td>
<td>The wedding was a week away.</td>
<td>The woman saw the fiancé kissing another woman.</td>
<td>The woman cancelled the wedding.</td>
<td>Did the woman cancel the wedding?</td>
</tr>
<tr>
<td>24</td>
<td>Tommy worked until midnight the past week.</td>
<td>Tommy was exhausted upon returning home.</td>
<td>Tommy went straight to sleep.</td>
<td>Did Tommy go straight to sleep?</td>
</tr>
<tr>
<td>25</td>
<td>Leah was the best director in New York.</td>
<td>The theater needed someone to direct the play.</td>
<td>The theater asked Leah to direct the play.</td>
<td>Did the theater ask Leah to direct the play?</td>
</tr>
<tr>
<td>26</td>
<td>Robert recently bought an ant farm.</td>
<td>The teacher asked for a volunteer for show and tell.</td>
<td>Robert presented the ant farm to the class.</td>
<td>Did Robert present the ant farm to the class?</td>
</tr>
<tr>
<td>27</td>
<td>The truck was running extremely low on fuel.</td>
<td>The next exit was 40 miles away.</td>
<td>The truck ran out of gas on the road.</td>
<td>Did the truck run out of gas on the road?</td>
</tr>
<tr>
<td>28</td>
<td>The professor did not tolerate lateness.</td>
<td>The student came ten minutes before class ended.</td>
<td>The professor stopped the student from entering class.</td>
<td>Did the professor stop the student from entering class?</td>
</tr>
<tr>
<td>29</td>
<td>The man filled the watering can.</td>
<td>The man took the watering can to the flower garden.</td>
<td>The man watered the flower garden.</td>
<td>Did the flowers get watered?</td>
</tr>
<tr>
<td>30</td>
<td>The campers did not put out the fire.</td>
<td>A strong wind blew through the campsite.</td>
<td>The fire spread to the forest.</td>
<td>Did the fire spread to the forest?</td>
</tr>
<tr>
<td>31</td>
<td>Nicole was hungry for dinner.</td>
<td>Nicole looked through the drawer with take-out menus.</td>
<td>Nicole ordered take-out for dinner.</td>
<td>Did Nicole skip dinner?</td>
</tr>
<tr>
<td>32</td>
<td>The waitress saw the mug break on the floor.</td>
<td>The waitress walked toward the broom closet.</td>
<td>The waitress cleaned up the broken mug.</td>
<td>Did the waitress leave the broken mug?</td>
</tr>
<tr>
<td>33</td>
<td>Michael ran out of money the first day of vacation.</td>
<td>Michael came from a wealthy family.</td>
<td>Michael called home to get more money.</td>
<td>Did Michael come home early from vacation?</td>
</tr>
<tr>
<td>34</td>
<td>Leah was feeling sleepy all morning at work.</td>
<td>The coffee cart started to come around.</td>
<td>Leah ordered a cup of coffee.</td>
<td>Did Leah decide not to get coffee?</td>
</tr>
<tr>
<td>35</td>
<td>Kevin had to catch an early train to Boston.</td>
<td>Kevin overslept the 6:30am alarm.</td>
<td>Kevin missed the train.</td>
<td>Did Kevin catch the train?</td>
</tr>
<tr>
<td>36</td>
<td>Lauren started heating a pot of stew.</td>
<td>Lauren forgot about the stew.</td>
<td>The stew burned.</td>
<td>Did the stew cook perfectly?</td>
</tr>
<tr>
<td>37</td>
<td>Justin was the fastest runner in the class.</td>
<td>Justin was challenged to race during recess.</td>
<td>Justin won the race.</td>
<td>Did Justin lose the race?</td>
</tr>
<tr>
<td>38</td>
<td>The printer was running low on ink.</td>
<td>Henry printed a 100-page document.</td>
<td>The printer ran out of ink.</td>
<td>Did the printer ink last?</td>
</tr>
<tr>
<td>39</td>
<td>The science teacher loved class pets.</td>
<td>The class asked for a pet rat.</td>
<td>The science teacher got a pet rat.</td>
<td>Did the science teacher refuse to get a class pet?</td>
</tr>
<tr>
<td>40</td>
<td>Kaitlyn had an early flight in the morning.</td>
<td>Kaitlyn forgot to set an alarm.</td>
<td>Kaitlyn missed the flight.</td>
<td>Did Kaitlyn make the flight?</td>
</tr>
<tr>
<td>41</td>
<td>Emma had almost recovered from the flu.</td>
<td>Emma got caught in a cold rain.</td>
<td>Emma got sick again.</td>
<td>Did Emma get better?</td>
</tr>
<tr>
<td>42</td>
<td>Becca desperately wanted to go to Berlin.</td>
<td>Becca saw a half price ticket to Berlin online.</td>
<td>Becca booked the flight to Berlin.</td>
<td>Did Becca continue working?</td>
</tr>
</tbody>
</table>
43 David left the house wearing a warm coat. It was warm and humid outside. David took off the coat. Did David leave on the coat?

44 Soccer was a popular sport at the university. Megan posted flyers for soccer try-outs around campus. Many students signed up for the try-outs. Did few students sign up for try-outs?

45 Alex was trying to lose weight before the cruise. Alex ran 3 miles every morning for a month. Alex lost weight for the cruise. Did Alex gain weight?

46 It was almost one in the morning. Ashley had only 10 pages left in the book. Ashley finished the book. Did Ashley go to sleep?

47 Andrew had a chemistry test on Friday. Andrew knew the material well. Andrew did well on the chemistry test. Did Andrew do poorly on the chemistry test?

48 Soccer was a popular sport at the university. Megan posted flyers for soccer try-outs around campus. Many students signed up for the try-outs. Did few students sign up for try-outs?

49 Alex was trying to lose weight before the cruise. Alex ran 3 miles every morning for a month. Alex lost weight for the cruise. Did Alex gain weight?

50 It was almost one in the morning. Ashley had only 10 pages left in the book. Ashley finished the book. Did Ashley go to sleep?

51 Andrew had a chemistry test on Friday. Andrew knew the material well. Andrew did well on the chemistry test. Did Andrew do poorly on the chemistry test?

52 Soccer was a popular sport at the university. Megan posted flyers for soccer try-outs around campus. Many students signed up for the try-outs. Did few students sign up for try-outs?

53 Alex was trying to lose weight before the cruise. Alex ran 3 miles every morning for a month. Alex lost weight for the cruise. Did Alex gain weight?

54 It was almost one in the morning. Ashley had only 10 pages left in the book. Ashley finished the book. Did Ashley go to sleep?

55 Andrew had a chemistry test on Friday. Andrew knew the material well. Andrew did well on the chemistry test. Did Andrew do poorly on the chemistry test?

56 Soccer was a popular sport at the university. Megan posted flyers for soccer try-outs around campus. Many students signed up for the try-outs. Did few students sign up for try-outs?
## Appendix B3: Two-Sentence Scenarios with No Inference

**KEY**: yes questions (i) for items 1-8, yes questions (ii) for items 9-15, and no questions (i) for items 16-22, no questions (ii) for items 23-30.

<table>
<thead>
<tr>
<th>#</th>
<th>Sentence 1</th>
<th>Sentence 2</th>
<th>Inference</th>
<th>Verification Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Andrew met a wonderful girl in Germany.</td>
<td>Andrew got a new job.</td>
<td>None</td>
<td>Did Andrew get to know a girl in Germany?</td>
</tr>
<tr>
<td>2</td>
<td>David bought a pair of blue jeans.</td>
<td>David baked a delicious cake.</td>
<td>None</td>
<td>Did David buy new pants?</td>
</tr>
<tr>
<td>3</td>
<td>Ella went to a birthday party on Saturday.</td>
<td>Ella was a talented ballet dancer.</td>
<td>None</td>
<td>Did Ella attend a birthday party over the weekend?</td>
</tr>
<tr>
<td>4</td>
<td>Justin traveled to many countries in the summer.</td>
<td>Justin loved American culture and cuisine.</td>
<td>None</td>
<td>Did Justin visit many countries in the summer?</td>
</tr>
<tr>
<td>5</td>
<td>Kaitlyn played charades with the guests.</td>
<td>Kaitlyn lost money last week in the poker game.</td>
<td>None</td>
<td>Did Kaitlyn participate in a game of charades with the guests?</td>
</tr>
<tr>
<td>6</td>
<td>Ethan liked walking the dog in the mornings.</td>
<td>Ethan went to yoga class at the gym.</td>
<td>None</td>
<td>Did Ethan enjoy taking the dog for a walk in the morning?</td>
</tr>
<tr>
<td>7</td>
<td>Hannah lived off-campus with friends.</td>
<td>Hannah enjoyed philosophy classes the most.</td>
<td>None</td>
<td>Did Hannah reside off-campus with friends?</td>
</tr>
<tr>
<td>8</td>
<td>The student traveled to Beijing for the summer.</td>
<td>The student majored in art history.</td>
<td>None</td>
<td>Did the student go to Beijing for the summer?</td>
</tr>
<tr>
<td>9</td>
<td>The band had a lead singer from Germany.</td>
<td>The band enjoyed traveling in the tour bus.</td>
<td>None</td>
<td>Did the band like traveling by bus?</td>
</tr>
<tr>
<td>10</td>
<td>The guests arrived late to the party.</td>
<td>The guests danced the entire time.</td>
<td>None</td>
<td>Did the guests spend the party dancing?</td>
</tr>
<tr>
<td>11</td>
<td>Ashley majored in business and marketing.</td>
<td>Ashley was talented at sign language.</td>
<td>None</td>
<td>Did Ashley have a talent for sign language?</td>
</tr>
<tr>
<td>12</td>
<td>Kevin went skiing many winters.</td>
<td>Kevin liked to surf with friends.</td>
<td>None</td>
<td>Did Kevin like going surfing with friends?</td>
</tr>
<tr>
<td>13</td>
<td>The cruise ship was sailing to the Bahamas.</td>
<td>The cruise ship was a new model.</td>
<td>None</td>
<td>Did the ship recently get upgraded?</td>
</tr>
<tr>
<td>14</td>
<td>Evan could not get a passport to go abroad.</td>
<td>Evan enjoyed spending weekends with friends.</td>
<td>None</td>
<td>Did Evan enjoy weekends with friends?</td>
</tr>
<tr>
<td>15</td>
<td>The firm catered lunch every Wednesday.</td>
<td>The firm was started less than a year ago.</td>
<td>None</td>
<td>Did the firm start in the last year?</td>
</tr>
<tr>
<td>16</td>
<td>The dean only saw students in the morning.</td>
<td>The dean had a cold office.</td>
<td>None</td>
<td>Did the dean counsel students in the evening?</td>
</tr>
<tr>
<td>17</td>
<td>The house had the best garden in town.</td>
<td>The house was the most expensive on the block.</td>
<td>None</td>
<td>Did the house have a beautiful display?</td>
</tr>
<tr>
<td>18</td>
<td>Becca rode the bus to school.</td>
<td>Becca had many friends in the class.</td>
<td>None</td>
<td>Did Becca walk to school?</td>
</tr>
<tr>
<td>19</td>
<td>The director was pleased about the large audience.</td>
<td>The director enjoyed the sunny weather.</td>
<td>None</td>
<td>Did the director feel worried that many people came to the show?</td>
</tr>
<tr>
<td>20</td>
<td>The house had a faulty air conditioning system.</td>
<td>The house was protected as a historical landmark.</td>
<td>None</td>
<td>Did the house have working air-conditioning?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>21</td>
<td>Henry went to Rome in July.</td>
<td>Henry enjoyed taking walks in the rain.</td>
<td>None</td>
<td>Did Henry travel to Paris in July?</td>
</tr>
<tr>
<td>22</td>
<td>Noah was the best at music class.</td>
<td>Noah always stared at a girl in the art class.</td>
<td>None</td>
<td>Did Noah have an aptitude for history class?</td>
</tr>
<tr>
<td>23</td>
<td>Emma drove most days to work.</td>
<td>Emma loved the company employees.</td>
<td>None</td>
<td>Did Emma dislike spending time with the other employees?</td>
</tr>
<tr>
<td>24</td>
<td>Alex worked close to home.</td>
<td>Alex went out for lunch every Friday.</td>
<td>None</td>
<td>Did Alex eat lunch at work every Friday?</td>
</tr>
<tr>
<td>25</td>
<td>Robert loved meeting new people.</td>
<td>Robert disliked traveling to different countries.</td>
<td>None</td>
<td>Did Robert like traveling to other countries?</td>
</tr>
<tr>
<td>26</td>
<td>Lauren met a handsome guy at the park.</td>
<td>Lauren did not drink any type of caffeine.</td>
<td>None</td>
<td>Did Lauren drink caffeine?</td>
</tr>
<tr>
<td>27</td>
<td>Megan applied to universities in New York City.</td>
<td>Megan enjoyed seeing historical sites.</td>
<td>None</td>
<td>Did Megan prefer watching movies?</td>
</tr>
<tr>
<td>28</td>
<td>The house was guarded attentively by a dog.</td>
<td>The house had a beautiful sculpture out front.</td>
<td>None</td>
<td>Did the house have a swimming pool?</td>
</tr>
<tr>
<td>29</td>
<td>Sarah had lunch at a local cafe.</td>
<td>Sarah had a great view from the ninth floor.</td>
<td>None</td>
<td>Did Sarah have a terrible view?</td>
</tr>
<tr>
<td>30</td>
<td>Sophie lived in an all girls’ dormitory.</td>
<td>Sophie rode a bike to class every day.</td>
<td>None</td>
<td>Did Sophie walk to class?</td>
</tr>
</tbody>
</table>
Appendix C: Informed Consent

INFORMED CONSENT SCRIPT

You are invited to participate in a research study titled “The role of individual differences in L1 and L2 reading”. This study is being conducted by Julie Beth Lake, a Ph.D. Candidate at Georgetown University in the Linguistics Department. This is for her Ph.D. dissertation and is being conducted to order to better understand how we process language when we read. Results from this study may be presented at academic conferences and published in academic journals.

Participation in this study is entirely voluntary at all times. You can choose not to participate at all or to leave the study at any time. Regardless of your decision, there will be no effect on your relationship with the researcher or have any other consequences. You are being asked to take part in this study because you are a native or an advanced speaker of the English language.

If you agree to participate, you will be asked to complete several computerized tasks, where you will respond to either text or figures on the screen and press a button in response. This session should last (i) for 80 minutes (two 40 minutes sessions) for the main experiment or (ii) 45 minutes in one session for the pilot experiments. The session will take place on Georgetown University campus in a data acquisition collection lab (Intercultural Center 233 or 201) at a time that is convenient for the participant.

Any data collected will remain anonymous and cannot be linked to you in any way. No identifying information about you will be collected at any point during the study, and your data will be identified only with a random number.

Study data will be kept in digital format in a password-protected computer only accessed by the researcher.

There are no risks associated with this study. While you will not experience any direct benefits from participation, information collected in this study may benefit others in the future by helping to better understand the process of reading.

If you have any questions about your rights as a research participant, please contact the Georgetown University IRB at (202) 687-6553 or irboard@georgetown.edu.

By taking part in the computerized tasks, you are indicating your consent to participate in this study. Thank you!

Julie Beth Lake
Ph.D. Candidate
Georgetown University
jbl3@georgetown.edu
Appendix D: Verbal Measures: Stimuli, Task Design, and Construct Validity

Appendix D1: English Non-Word Repetition Task Stimuli


<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Glistow</td>
</tr>
<tr>
<td>2</td>
<td>Altupatory</td>
</tr>
<tr>
<td>3</td>
<td>Sepretennial</td>
</tr>
<tr>
<td>4</td>
<td>Confrantually</td>
</tr>
<tr>
<td>5</td>
<td>Ballop</td>
</tr>
<tr>
<td>6</td>
<td>Fenneriser</td>
</tr>
<tr>
<td>7</td>
<td>Detratapillic</td>
</tr>
<tr>
<td>8</td>
<td>Perplisteronk</td>
</tr>
<tr>
<td>9</td>
<td>Pristoractional</td>
</tr>
<tr>
<td>10</td>
<td>Pennerriful</td>
</tr>
<tr>
<td>11</td>
<td>Glistering</td>
</tr>
<tr>
<td>12</td>
<td>Sladding</td>
</tr>
<tr>
<td>13</td>
<td>Versatrationist</td>
</tr>
<tr>
<td>14</td>
<td>Voltnularity</td>
</tr>
<tr>
<td>15</td>
<td>Rubid</td>
</tr>
<tr>
<td>16</td>
<td>Defermication</td>
</tr>
<tr>
<td>17</td>
<td>Barrazon</td>
</tr>
<tr>
<td>18</td>
<td>Empliforvent</td>
</tr>
<tr>
<td>19</td>
<td>Doppelate</td>
</tr>
<tr>
<td>20</td>
<td>Woogalamic</td>
</tr>
<tr>
<td>21</td>
<td>Commeecitate</td>
</tr>
<tr>
<td>22</td>
<td>Bannow</td>
</tr>
<tr>
<td>23</td>
<td>Hampent</td>
</tr>
<tr>
<td>24</td>
<td>Reutterpation</td>
</tr>
<tr>
<td>25</td>
<td>Bannifer</td>
</tr>
<tr>
<td>26</td>
<td>Blonterstaping</td>
</tr>
<tr>
<td>27</td>
<td>Frescovent</td>
</tr>
<tr>
<td>28</td>
<td>Commerine</td>
</tr>
<tr>
<td>29</td>
<td>Tafflest</td>
</tr>
<tr>
<td>30</td>
<td>Underbrantuand</td>
</tr>
<tr>
<td>31</td>
<td>Contramponist</td>
</tr>
<tr>
<td>32</td>
<td>Trumpetine</td>
</tr>
<tr>
<td>33</td>
<td>Stopographic</td>
</tr>
<tr>
<td>34</td>
<td>Thickery</td>
</tr>
<tr>
<td>35</td>
<td>Prindle</td>
</tr>
<tr>
<td>36</td>
<td>Diller</td>
</tr>
<tr>
<td>37</td>
<td>Loddenapish</td>
</tr>
<tr>
<td>38</td>
<td>Skiticult</td>
</tr>
<tr>
<td>39</td>
<td>Pennel</td>
</tr>
<tr>
<td>40</td>
<td>Brasterer</td>
</tr>
</tbody>
</table>
Appendix D2: Spanish Non-Word Repetition Task Stimuli

Nonwords from Mackey & Sachs (2012).

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>termar – baldor</td>
</tr>
<tr>
<td>2</td>
<td>irdol – urda</td>
</tr>
<tr>
<td>3</td>
<td>taurete – patanco</td>
</tr>
<tr>
<td>4</td>
<td>permalla – laderal</td>
</tr>
<tr>
<td>5</td>
<td>movilido – sucursina</td>
</tr>
<tr>
<td>6</td>
<td>antiembre – desboroto</td>
</tr>
<tr>
<td>7</td>
<td>arstancio – doporsidad</td>
</tr>
<tr>
<td>8</td>
<td>apardapal – rabirosa</td>
</tr>
<tr>
<td>9</td>
<td>emitancia – sifimedero</td>
</tr>
<tr>
<td>10</td>
<td>santificorado – embulicio</td>
</tr>
<tr>
<td>11</td>
<td>eliminicio – conespidente</td>
</tr>
<tr>
<td>12</td>
<td>proseguida – abastalogia</td>
</tr>
<tr>
<td>13</td>
<td>decacuadard-refeexo – trasora-naderio</td>
</tr>
<tr>
<td>14</td>
<td>curtillo-barajento – autido-desampato</td>
</tr>
<tr>
<td>15</td>
<td>crosar-partiferencia – tironano-civinista</td>
</tr>
<tr>
<td>16</td>
<td>ampato-debicario – mungual-solteramente</td>
</tr>
</tbody>
</table>
Appendix D3: Description of Verbal Measures

In the first session of the main experiment, participants were given a battery of individual difference measures to triangulate their cognitive capacity along two dimensions: first, cognitive capacity: whether short-term memory capacity (i.e., the finite storage capacity used in processing incoming information) or cognitive control capacity, (i.e., the ability to manipulate information by selecting relevant information, ignoring irrelevant information and comparing contexts for the best fit) and second, type of processing type: language-dependent cognitive capacity (i.e., processing that is reliant on verbal information) or language-independent cognitive capacity (i.e., processing that is reliant on spatial processing). This appendix will focus on the second dimension, the verbal and non-verbal stand-alone tasks. The verbal cognitive tasks provided an independent measure of a participant’s capacity to remember and retrieve stored information and to manipulate that information in a language-dependent manner. It is particularly important to assess participants’ verbal capacity to store and manipulate information in the context of a reading comprehension study. Each participant’s performance on the verbal and non-verbal individual difference measures were correlated to ensure that the verbal tasks were tapping into the same underlying construct as the non-verbal tasks, which were ultimately used as distractor tasks in the main experiment. However, the language dependent, or verbal, measures were not fully correlated to the non-verbal measures (see Table A7). For this reason, they were not used in the main studies. A summary of the verbal individual difference tasks is listed in Table A1.

Table A1. Verbal Stand-Alone Individual Difference Tasks
To review, the task used to measure short-term memory capacity in a language-dependent manner was the Non-Word Repetition task. The task used to measure cognitive control along language-dependent dimensions was the Stroop task.

Note that the verbal stand-alone tasks were piloted in the pretest of all the cognitive measures described in the dissertation; the participant information can be found in Section 4.5.3.4. The task procedure and results of the pretest will be described for the two tasks in their respective sections.

**Non-Word Repetition Task**

Language-dependent short-term memory capacity, also known as phonological short-term memory, was measured using the non-word repetition (NWR) task (French & O’Brien, 2008). In this task, participants heard non-words with syllable length that varied across items and were asked to repeat them back. Typically, performance declines as syllable length increases, yet this effect is mediated when participants have high phonological short-term memory. The NWR task is, by nature, a verbal task, as it relies on the phonotactics of the participant’s language with word-like stimuli; if the task used real words, it would conflate prior knowledge and semantic memory. As such, this was considered to tap into language-dependent processing, as it measures how much phonological information a person can remember. The NWR task has been shown to be a reliable measure of phonological processing (see Baddeley, Gathercole & Papagno, 1998; Gathercole 2006). While the NWR task taps into general phonological
skills, such as awareness of phonemes, it has been correlated with established measures that do not rely on these processes, such as the digit span task (Baddeley et al., 1998).

Participants sat in front of a computer and the NWR task was presented via PowerPoint software. First, participants read and heard a recording of the task directions; they were instructed that they would hear ‘nonsense’ words and have to repeat them after hearing a beep. At this point, if the participant was clear about the task, the researcher left the room. Next, participants stared at a fixation cross while hearing the non-words and beep sounds from the computer speakers, and repeated them back into an MP3 recorder. There was a 200 millisecond delay between the presentation of the non-word and the beep. The task took roughly 6 minutes. For the L1 participants, all the non-words used in the experiment were taken directly from French and O’Brien (2008), and were based in English phonotactics. A native English speaker recorded the words and instructions. For the L2 participants, the non-words used were from Mackey and Sachs (2012), and were based on Spanish phonotactics. Mackey and Sachs (2012) based their nonwords on those found in Speciale, Ellis & Bywater (2004), and the directions and nonwords were adapted, translated, and recorded by Dr. Ana Maria Nuevo, native Spanish speaker, to better follow the phonotactics of Spanish. The only difference between the two tasks was that rather than individual words, the Spanish NWR task had participants repeat pairs of words. The Spanish NWR task was not pretested, as the exact recorded version of the task in the L2 study was used by Mackey and Sachs (2012). See Appendices E1 and E2 for a complete list of non-words based in English and Spanish phonotactics.
The results of the pretest for the NWR task demonstrate that the task was valid and reliable. An item was only considered correct if all the phonemes in the non-word were repeated correctly; if one phoneme was erroneously repeated then the item was marked as incorrect. With adjusted scores (as per the change listed below), the accuracy rate was 36.6 out of 40 ($M = .92, SD = .005$).

The task was modified slightly based on the pretest results. The initial accuracy rate was 35 out of 40 items ($M = .88, SD = .052$). An item analysis revealed that two of the words had an accuracy rate far below the mean: tafflest ($M_{item} = .13$) and fenneriser ($M_{item} = .40$). The echo in the room was likely responsible for participants mishearing the place and manner of articulation. An example of the place of articulation can be seen in the word fenneriser; instead of hearing the labiodental $[f]$ (i.e., $[\text{fənərəizər}]$), participants heard an alveolar $[s]$ (i.e., $[\text{sənərəizər}]$). An example of the manner of articulation can be seen in the word tafflest; instead of hearing the fricative $[f]$ (i.e., $[\text{tæflɛst}]$), participants heard a stop $[p]$ (i.e., $[\text{tæplɛst}]$). This was resolved by providing earphones to the participants. The directions were slightly revised to maximize understanding, as well.

The overall accuracy rate had less variation than expected, so the length of the pause between the aural presentation of the non-word and the prompt to repeat the non-word (indicated by a beep) was increased from 200 ms to 2000 ms. This more accurately assessed the phonological short-term memory, as participants were required to hold the nonwords in phonological short-term memory for a longer period before repeating it.
Stroop Task

The task used to measure cognitive (or executive) control along language-dependent dimensions was the Stroop task (Stroop, 1935). The task design was based on Schroeter, Zyseet, Kupka, Kruggel and von Cramon (2002); participants saw a string of letters that were congruent, incongruent or neutral to the font color in which they were presented. Their task was to indicate what color the letters appeared in by pressing a color on the keyboard (i.e., red, green, blue or yellow).

The Stroop task presents its participants with a conflict they have to resolve (the color word versus the text color), and therefore is thought to tap into cognitive control. Kane and Engle (2002) position Stroop effects within an executive-attention framework, and through a review of brain-based research claim that the dorsolateral PFC circuitry is critical to executive-attention functions. It is important to note that while this task has been long associated as a reliable measure of cognitive control, there is a lack of consensus about the scope of the claims that can be made from Stroop-like tasks. This debate is between the classic lexical-competition account (see work of La Heij and colleagues, 2006) and the alternate response-selection (see work of Finkbeiner & Caramazza, 2006). Finkbeiner and Caramazza (2006) suggest that the semantic interference effect typically associated with Stroop-like tasks is task-specific and does not carry over to more general lexical processing in meaning-based situations. This would impact the interpretation of the task results for lexical access and hence measurement of phonological short-term memory. La Heij, Kuipers and Starreveld (2006) provide a convincing explanation of how the data from the literature fits into the classic lexical-competition model, but also show how Finkbeiner and Caramazza’s
alternative account is underspecified. Even with the debate on classic and alternative views of lexical processing, Stroop tasks are thought to tap into cognitive control. For example, certain studies that found the ability to control visual attention (i.e., the ability to focus the eyes away from a noticeable visual stimulus) has also been linked to working memory capacity (Kane, Bleckley, Conway, & Engle, 2001; Kane & Engle, 2003; de Fockert et al., 2001). This seems to indicate that the Stroop task is a valid measurement of cognitive control, and therefore it was chosen for the dissertation.

The Stroop task was developed on the experiment-generating package, SuperLab (Version 4.0; Cedrus Corporation; San Pedro, CA) and presented via this software. The entire task took roughly 5 minutes to complete. Participants sat in front of a computer screen and were presented with a string of letters in three conditions: in the neutral trials, “XXXX”, appeared on the screen in one of four font colors (i.e., red, green, blue or yellow); in the congruent trials, the words “RED”, “GREEN”, “BLUE” or “YELLOW” appeared on the screen in the font color that corresponded with the word, i.e., “GREEN” appeared in green font; and in the incongruent trials, the words “RED”, “GREEN”, “BLUE” or “YELLOW” appeared on the screen in the font color that did not correspond with the word, i.e., “YELLOW” appeared in blue font; the aim was to create interference between the color word and the color name. The task had 60 randomized trials (20 neutral, 20 congruent, and 20 incongruent). The participant pushed a key covered with a colored tab (red on “1”, green on “2”, blue on “9” and yellow on “0”), so participants could rest their hands equidistantly for maximum comfort during the task. Accuracy and reaction time were measured.
The procedure was as follows for all the trials in the Stroop task, approximately 3250 milliseconds (ms) per trial (1250 ms of fixed events +/- reaction time for button press). See Figure A1 for the trial sequence of the Stroop task.

![Diagram of the Stroop task trial sequence](image)

**Figure A1.** Trial sequence for the Stroop task

The results of the pretest show that the mean accuracy for the task was 59 out of 60 items ($M = .98$, $SD = .02$). Only correct answers were analyzed; the mean reaction time was 1802.4 ms ($SD = 114$). All reaction time data were treated in that any that fell below 200 ms or above 2.5 standard deviations was removed ($M_{items\_removed} = 1.2$, $SD_{items\_removed} = .86$).

Paired-sample $t$-tests revealed that the reaction time for congruent trials was faster than for incongruent trials ($t(14) = -5.84, p < .01$); and neutral trials were faster than incongruent trials ($t(14) = -5.29, p < .01$). As expected, the task seems to have
produced interference in its incongruent condition. Descriptive statistics for the mean RTs are listed in Table A2.

*Table A2.* Descriptive statistics for Stroop task by trial type. Mean reaction times (RTs) with standard deviations (SDs) in parentheses.

<table>
<thead>
<tr>
<th>Task</th>
<th>L1 English speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stroop Overall</td>
<td>1801 (140)</td>
</tr>
<tr>
<td>Neutral</td>
<td>1769 (117)</td>
</tr>
<tr>
<td>Congruent</td>
<td>1759 (122)</td>
</tr>
<tr>
<td>Incongruent</td>
<td>1894 (140)</td>
</tr>
<tr>
<td>N</td>
<td>15</td>
</tr>
</tbody>
</table>

There were no changes made to the Stroop task.

**Data Screening**

The accuracy and the reaction time data (RT) for the L1 and L2 groups largely follows the same pattern as the stand-alone cognitive tasks used in the dissertation. The accuracy data indicate that participants were more accurate than not on the experimental tasks; the data are negatively skewed and not normally distributed. To ensure that the accuracy scores were reliable, there were two raters; one rater determined whether the word was correctly repeated or not (for a score of 1 or 0) for 100% of the ENWR and 100% of the SNWR data and a second rater independently rated 25% of the ENWR and SNWR data. Interater reliability was calculated at 93% for the ENWR task and 92% for the SNWR task. See Table A3 for descriptive statistics and Table 4 for goodness-of-fit tests of the L1 and L2 accuracy data. The K-S test showed that the L1 group’s accuracy data were not normally distributed for the Stroop task, \( D(48) = 0.20, p < .05 \). The English NWR task appears to be normally distributed,
The K-S test showed that the L2 group’s accuracy data were not normally distributed for the Stroop task, \(D(22) = 0.24, p < .05\) and the Spanish NWR task, \(D(22) = 0.20, p < .05\). The English NWR task appears to be normally distributed, \(D(22) = 0.19, p = .05\). In addition, many of the converted skewness or kurtosis z-scores (calculated by dividing the skewness/kurtosis value by its standard error) were above or approaching 3.29 (see Field, 2009 for a description of measures of normality using z-scores). This is overwhelming evidence that the accuracy data violates normality. For this reason, nonparametric statistics are a more appropriate choice for inferential analyses.

Unlike the accuracy data, which were clustered at the higher end of the scale, the reaction time tended toward the lower (i.e., faster) end of the distribution; this shows that the majority of the participants responded fairly quickly to the stimulus. Reaction times were measured for all tasks, with the exception of the Non-Word Repetition task. See Table A5 for descriptive statistics and Table A6 for goodness-of-fit tests of the L1 and L2 reaction time data. The cognitive measures tend to have fairly symmetrical distribution, but in comparison to the other cognitive measures, the range of the Stroop task is positively skewed with higher (i.e., slower) overall reaction times. This divergence in the range of the reaction time is expected given the task design of the Stroop task. Since the task requires that the participant suppress the color word they have read and access the conceptual information about the color, the task typically yields a higher accuracy rate and slower reaction times. The K-S test showed that the L1 group’s reaction time data were normally distributed for the Stroop task, \(D(45) = \ldots\)
0.06, \( p > .05 \). The K-S test showed that the L2 group’s data were normally distributed for the Stroop task, \( D(22) = 0.10, p > .05 \).
Table A3. Descriptive statistics for the raw accuracy scores in the verbal cognitive tasks for the L1 and L2 group. Percentages in parentheses.

<table>
<thead>
<tr>
<th>Task</th>
<th>Mean (%)</th>
<th>Range</th>
<th>SD</th>
<th>SE</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td>L1 Stroop*</td>
<td>57.46 (96%)</td>
<td>8</td>
<td>1.99</td>
<td>0.28</td>
<td>57</td>
</tr>
<tr>
<td>L2 Stroop*</td>
<td>58.35 (97%)</td>
<td>4</td>
<td>1.2</td>
<td>0.256</td>
<td>57.8</td>
</tr>
<tr>
<td>L1 English NWR*</td>
<td>34.9 (87%)</td>
<td>11</td>
<td>2.60</td>
<td>0.37</td>
<td>34</td>
</tr>
<tr>
<td>L2 English NWR*</td>
<td>30.3 (76%)</td>
<td>17</td>
<td>3.9</td>
<td>0.821</td>
<td>28.96</td>
</tr>
<tr>
<td>L2 Spanish NWR*</td>
<td>25.7 (80%)</td>
<td>14</td>
<td>3.9</td>
<td>0.828</td>
<td>23.85</td>
</tr>
</tbody>
</table>

* Stand-alone task

Table A4. Goodness-of-fit tests to check for normality of distributions on accuracy of verbal cognitive tasks for L1 and L2 groups.

<table>
<thead>
<tr>
<th>Task</th>
<th>Skew</th>
<th>SE</th>
<th>Ratio</th>
<th>Kurtosis</th>
<th>SE</th>
<th>Ratio</th>
<th>Statistic</th>
<th>p-value</th>
<th>Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Kolmogorov-Smirnov</td>
<td></td>
<td>Shapiro-Wilk</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(K-S)</td>
<td></td>
<td>(S-W)</td>
<td></td>
</tr>
<tr>
<td>L1 Stroop*</td>
<td>-0.87</td>
<td>0.34</td>
<td>-2.57</td>
<td>0.35</td>
<td>0.66</td>
<td>0.53</td>
<td>0.20</td>
<td>0.00</td>
<td>0.91</td>
<td>0.00</td>
</tr>
<tr>
<td>L2 Stroop*</td>
<td>-0.10</td>
<td>0.48</td>
<td>-0.20</td>
<td>-0.51</td>
<td>0.94</td>
<td>-0.54</td>
<td>0.24</td>
<td>0.00</td>
<td>0.87</td>
<td>0.01</td>
</tr>
<tr>
<td>L1 English NWR*</td>
<td>-0.07</td>
<td>0.34</td>
<td>-0.20</td>
<td>-0.41</td>
<td>0.66</td>
<td>-0.62</td>
<td>0.12</td>
<td>0.10</td>
<td>0.97</td>
<td>0.16</td>
</tr>
<tr>
<td>L2 English NWR*</td>
<td>-0.50</td>
<td>0.48</td>
<td>-1.05</td>
<td>0.43</td>
<td>0.94</td>
<td>0.46</td>
<td>0.19</td>
<td>0.05</td>
<td>0.94</td>
<td>0.17</td>
</tr>
<tr>
<td>L2 Spanish NWR*</td>
<td>-0.91</td>
<td>0.48</td>
<td>-1.90</td>
<td>0.01</td>
<td>0.94</td>
<td>0.01</td>
<td>0.20</td>
<td>0.04</td>
<td>0.89</td>
<td>0.02</td>
</tr>
</tbody>
</table>

* Stand-alone + Performed with the Lilliefors correction

* Stand-alone task
**Table A5.** Descriptive statistics for the reaction times (in milliseconds) in the verbal cognitive tasks for the L1 and L2 groups

<table>
<thead>
<tr>
<th>Task</th>
<th>Mean</th>
<th>Range</th>
<th>SD</th>
<th>SE</th>
<th>95% Confidence Interval</th>
<th>% RT data removed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Lower</td>
<td></td>
<td></td>
<td>Upper</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bound</td>
<td></td>
<td></td>
<td>Bound</td>
<td></td>
</tr>
<tr>
<td>L1 Stroop*</td>
<td>1749</td>
<td>1723</td>
<td>1781</td>
<td>1.9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L2 Stroop*</td>
<td>1823</td>
<td>1776</td>
<td>1879</td>
<td>2.5%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Stand-alone task

**Table A6.** Goodness-of-fit tests to check for normality of distributions on RT of verbal cognitive tasks for L1 and L2 groups.

<table>
<thead>
<tr>
<th>Task</th>
<th>Skew</th>
<th>SE</th>
<th>Ratio</th>
<th>Kurtosis</th>
<th>SE</th>
<th>Ratio</th>
<th>Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Skewness</td>
<td></td>
<td>Kurtosis</td>
<td></td>
<td></td>
<td>Kolmogorov-Smirnov</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(K-S)</td>
<td></td>
<td>(K-S)</td>
<td></td>
<td></td>
<td>(K-S)</td>
<td></td>
</tr>
<tr>
<td>L1 Stroop*</td>
<td>0.22</td>
<td>0.34</td>
<td>0.64</td>
<td>-0.53</td>
<td>0.66</td>
<td>-0.80</td>
<td>0.06</td>
<td>0.20</td>
</tr>
<tr>
<td>L2 Stroop*</td>
<td>0.36</td>
<td>0.48</td>
<td>0.76</td>
<td>-0.37</td>
<td>0.94</td>
<td>-0.44</td>
<td>0.10</td>
<td><strong>0.20</strong></td>
</tr>
</tbody>
</table>

* Stand-alone task  + Performed with the Lilliefors correction
Correlations between verbal and non-verbal individual difference measures.

Independent measures were collected to assess each participant’s capacity to remember, retrieve, and manipulate stored information both verbally and non-verbally along cognitive control and short-term memory dimensions. The distraction paradigm did not permit the use of verbal tasks as distractors because participants had to complete two tasks within the same trial; participants needed to process and understand the two-sentence scenarios while completing a secondary distractor task. For this reason, a non-verbal cognitive control task and a non-verbal short-term memory task were used as distractors in the self-paced reading task. In order to gauge the cognitive control construct validity of the verbal (i.e., the Stroop task) and non-verbal (i.e., the N-Back task) measures and the short-term memory construct validity of the verbal (i.e., the NWR task) and non-verbal measures (i.e., the Visual Patterns task) for cognitive control capacity, correlations were run. The entire participant population \((N = 73)\) completed all the tasks, with the exception of the Spanish NWR task, which only the L2 participants completed. Therefore all the participants were used in the correlational analyses to increase the power of the test.

Bivariate correlations shed light on whether two variables are associated by investigating the degree that two continuous variables covary. The five assumptions of parametric correlations are that the data are independently collected (i.e., one person’s data does not influence the data of another), the data are assumed to be interval, the data are assumed to be normally distributed, the relationship between the variables is linear, and finally, the data should have homoscedasticity (i.e., the variances of their residuals are comparable). Since the accuracy and reaction time data were checked for the
assumptions of correlations in Chapter 5, it will not be reviewed here. The data do not adhere to the assumptions of parametric data; therefore Spearman’s Rho was used for the correlations. Table A7 shows the correlations between the verbal and non-verbal measures.

Table A7. Correlations (using Spearman’s rho) of verbal and non-verbal measures.

<table>
<thead>
<tr>
<th>Non-Verbal Measures</th>
<th>Stroop Accuracy</th>
<th>Stroop RT*</th>
<th>English NWR</th>
<th>Spanish NWR</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-Back Accuracy</td>
<td>0.11 (0.356)</td>
<td>-0.075 (0.533)</td>
<td>0.035 (0.772)</td>
<td>0.072 (0.548)</td>
</tr>
<tr>
<td>N-Back RT</td>
<td>.312 (0.007)</td>
<td>.405 (0)</td>
<td>-0.041 (0.732)</td>
<td>0.149 (0.207)</td>
</tr>
<tr>
<td>VPT Accuracy</td>
<td>.311 (0.008)</td>
<td>0.053 (0.662)</td>
<td>0.121 (0.315)</td>
<td>0.065 (0.592)</td>
</tr>
<tr>
<td>VPT RT</td>
<td>.252 (0.033)</td>
<td>.406 (0)</td>
<td>0.045 (0.705)</td>
<td>0.136 (0.256)</td>
</tr>
</tbody>
</table>

* The Stroop RT was represented by the overall RT, otherwise referred to as the Monitoring Effect.

The correlations were complex to interpret, although the verbal and non-verbal cognitive control measures were partially correlated in several ways; unexpectedly, there appear to be relationships between the verbal cognitive control measures and the non-verbal short-term memory measures. A verbal measure of cognitive control, the Stroop RT, and non-verbal measure of cognitive control, the N-Back RT, were positively correlated, \( r = .405, p < .05 \), but there was also a positive correlation between the verbal measure of cognitive control, the Stroop task RT, and the non-verbal measure of short-term memory, the VPT RT (non-verbal), \( r = .406, p < .05 \). In addition, the Stroop task accuracy, a measure of cognitive control was positively correlated with the N-Back task RT, a non-verbal cognitive control measure, \( r = .312, p < .05 \); this indicates that the more accurate participants were on the Stroop task, the slower they were to complete the N-Back task. Surprisingly, the Stroop task accuracy, a measure of cognitive control was also
positively correlated with the VPT task accuracy, a non-verbal short-term memory measure, $r = .311, p < .05$ and the VPT task RT, a non-verbal short-term memory measure, $r = .252, p < .05$. Taken together, it appears that the Stroop task possibly taps into both short-term memory and cognitive control. The verbal measures of short-term memory did not correlate with any other measures. The verbal stand-alone measures will not be used in the L1 or L2 analyses because there is only partial correlation between the verbal measures and the non-verbal measures.
Appendix E: Background Questionnaires

Appendix E1: L1 Background Questionnaire

SCREEN 1: Language Background Information

Welcome!

Thank you for participating in my dissertation research.

In the next few screens you will be asked about your language background. The language background questionnaire should take you about 5 minutes.

Thank you!

SCREEN 2: Introduction

Before you start the study, please fill out the following language background questionnaire.

- **RESEARCHER fills this in:** Which study is this?
  - Distractor Pilot
  - Sentence Pilot
  - Mini Pilot
  - L1
  - L2
  - Other (please specify)

- **RESEARCHER fills this in:** Participant number:
  - _____________

- Are you a native speaker of English?
  - Yes
  - No

SCREEN 3:

1. **Gender:**
   - Male
• Female

2. Age:
   • ________________

3. Please indicate which hand you WRITE with:
   • Right
   • Left

4. Please indicate which hand you CUT with:
   • Right
   • Left

5. Do you either have normal vision or do you wear glasses/contacts that correct your vision to normal?
   • Yes
   • No

6. Do you have any hearing problems that you are aware of?
   • Yes
   • No

7. Please write your native language(s) other than English below:
   • First language: __________________________
   • First language: __________________________
   • First language: __________________________
   • First language: __________________________

8. Was most of your schooling in English (i.e., elementary, middle and high school)?
   • Yes
   • No
   • Other (please specify)

9. Please write your second language(s) below:
   • Second language #1: __________________________
• Second language #2: ___________________________
• Second language #3: ___________________________
• Second language #4: ___________________________

10. Of your second languages (numbered above), please rate your fluency and usage.

**FLUENCY**: (i) beginner, (ii) intermediate, (iii) advanced, (iv) fluent

**USAGE**: (i) never, (ii) a few times a year, (iii) a few times a month, (iv) every day

• Second language #1: ___________________________
• Second language #2: ___________________________
• Second language #3: ___________________________
• Second language #4: ___________________________

**SCREEN 4**: Survey complete!

Thank you for filling out the language background survey. Please get the researcher and you can begin the study.
Appendix E2: L2 Background Questionnaire

SCREEN 1: Language Background Information

Welcome!

Thank you for participating in my dissertation research.

In the next few screens you will be asked about your language background. The language background questionnaire should take you about 5 minutes.

Thank you!

SCREEN 2: Introduction

Before you start the study, please fill out the following language background questionnaire.

❖ RESEARCHER fills this in: Which study is this?
  • Distractor Pilot
  • Sentence Pilot
  • Mini Pilot
  • L1
  • L2
  • Other (please specify)

❖ RESEARCHER fills this in: Participant number:
  • _____________

❖ Are you a native speaker of English?
  • Yes
  • No

SCREEN 3:

1. Gender:
  • Male
  • Female
2. Age:
   • __________________

3. Please indicate which hand you WRITE with:
   • Right
   • Left

4. Please indicate which hand you CUT with:
   • Right
   • Left

5. Do you either have normal vision or do you wear glasses/contacts that correct your vision to normal?
   • Yes
   • No

6. Do you have any hearing problems that you are aware of?
   • Yes
   • No

7. Please write your native language(s) below:
   • First language: ___________________________
   • First language: ___________________________
   • First language: ___________________________
   • First language: ___________________________

8. Please write your second language(s) other than English below:
   • Second language #1: ___________________________
   • Second language #2: ___________________________
   • Second language #3: ___________________________
   • Second language #4: ___________________________
9. Of your second languages (numbered above), please rate your fluency and usage (in drop-down menus).

**FLUENCY**: (i) beginner, (ii) intermediate, (iii) advanced, (iv) fluent

**USAGE**: (i) never, (ii) a few times a year, (iii) a few times a month, (iv) every day

- Second language #1: ___________________________
- Second language #2: ___________________________
- Second language #3: ___________________________
- Second language #4: ___________________________

***

The next 8 questions will ask you about your experience, usage and proficiency in English.

***

10. Please rate your skills in English below (a drop down menu)

**CHOICES**: (i) not at all confident, (ii) 2, (iii) 3, (iv) 4

- Speaking
- Reading
- Writing
- Listening

11. How old were you when you started to study English?
- ______________________

12. How many years have you studied English?
- ______________________
13. Where have you learned the majority of your English?
   • Formal
   • In informal settings
   • Mixed
   • Other (please specify)

14. Have you ever lived in an English-speaking country for more than one month?
   • Yes
   • No
   • If so, how long and what did you do there?

15. How important is it for you to learn English for the following reasons? (in drop-down menus).

   **CHOICES**: (i) not at all important, (ii) a little important, (iii) somewhat important, (iv) very important
   • For employment?
   • For social reasons?
   • For educational reasons?
   • Out of interest?

16. How much do you enjoy speaking English?
   • Not at all
   • Somewhat
   • Mostly
   • Completely

17. Have you taken an English proficiency exam? Please list your score and year taken below:
   • IELTS:
   • TOEFL:
   • Pearson’s English Exam:
   • Other:

**SCREEN 4**: Survey complete!

Thank you for filling out the language background survey. Please get the researcher and you can begin the study.
Appendix F: Scatterplots for Correlations (with Loess lines)

Appendix F1: Stand-Alone N-Back Task (by condition and scenario type)
Appendix F2: Stand-Alone VPT (by condition and scenario type)
Appendix G: L2 Statistical Analyses Without Possible Outliers

As we saw in the Discussion Chapter, three of the L2 participants had disproportionately larger reaction times when reading the scenarios in the control condition (S04, S21, and S26). This is a threat to the reliability of the L2 results, as the three L2 outliers were purported to have adopted different strategies throughout the experiment. All of the statistical analyses were rerun for the L2 group without the three participants to verify the results.

G1. Research question involving inferences versus non-inferences

Recall that the first research question in the L2 study looked at whether L2 readers process inferences differently than non-inferences. A repeated-measures analyses of variance (ANOVA) was carried out with the reaction times for the three item types (filler/no inference, bridging inference and predictive inference) in the control condition, where there was no interruption to the reading process. As we saw before, a reaction time that is significantly different between the inferences and the non-inferences shows that these stimuli are processed differently from one another.

G1.1. Repeated-Measures ANOVA. Revisiting the assumptions of the repeated-measures, the reaction time data were shown to be interval data that was normally distributed. See Table G1 for descriptive statistics of the three two-sentence scenario types: no inference (N), bridging (B) and predictive (P). Note that the reaction times were slower than those in the L1 study.
Table G1. Descriptive statistics for RQ1.

<table>
<thead>
<tr>
<th>Two-Sentence Scenario Type</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Inference (N)</td>
<td>4774</td>
<td>920</td>
</tr>
<tr>
<td>Predictive (P)</td>
<td>4639</td>
<td>761</td>
</tr>
<tr>
<td>Bridging (B)</td>
<td>3929</td>
<td>682</td>
</tr>
</tbody>
</table>

Next, the assumption of sphericity was checked. Mauchly’s test indicated that the assumption of sphericity was met, ($\chi^2(2) = 0.117, p > .01$) (see Table G2).

Table G2. Mauchly’s test of sphericity assumptions for RQ1.

<table>
<thead>
<tr>
<th>Within-Subjects Effect</th>
<th>Mauchly's W</th>
<th>Approx. Chi-Square</th>
<th>df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Story Type</td>
<td>0.777</td>
<td>4.288</td>
<td>2</td>
<td>0.117</td>
</tr>
</tbody>
</table>

The results of the repeated-measures ANOVA show that reaction time was significantly affected by the type of two-sentence scenario read, $F(2, 19) = 31.7, p < .05$. Partial $\eta^2$ was .638 and observed power was 1.0. Repeated-measures t-tests (using a Bonferroni adjustment, $\alpha = .05/3 = .017$) showed that subjects were significantly faster to read on the two-sentence scenarios with bridging inferences than they were for the two-sentence scenarios with predictive inferences or the two-sentence scenarios with no inference (B versus P: $t(18) = 8.57, p < 0.01$; B versus N: $t(18) = 6.63, p < 0.01$), but that there was no significant difference in reading time between the two-sentence scenarios with no inference and the two-sentence scenarios with predictive inferences (N versus P: $t(19) = 0.504, p = .620$, not significant). See Table G3 for the repeated-measures ANOVA of scenario type for the L2 group.
Table G3. RM ANOVA testing for scenario type in L2 group.

<table>
<thead>
<tr>
<th>Tests of Within-Subjects Effects*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Scenario Type</td>
</tr>
<tr>
<td>Error (Scenario Type)</td>
</tr>
</tbody>
</table>

* Performed with Greenhouse-Geisser corrections

Participants took longer to read the two-sentences scenarios with no-inference ($M = 4774$) and predictive inferences ($M = 4639$) than those with bridging inferences ($M = 3929$). These data reinforce that bridging inferences are processed more quickly, and therefore more automatically, than predictive inferences; this is the same result found in the larger L2 data set. Non-inferences were processed the most slowly, which is also seen in the overall L2 results.

**G2. Research question involving role cognitive resources**

The second research question in the L2 study looked at whether second language readers rely on cognitive resources to process inferences. Recall that the non-inferences were two-sentence scenarios that did not require readers to form an inference and the inferences were two-sentence scenarios that required readers to form a connection that was not explicitly provided, either a bridging inference or predictive inference. In order to address this research question, several correlations were carried out. The predictor variables were the type of distractor task used in each condition (i.e., control/none [CONT], cognitive control distractor [CC] and short-term memory distractor [STM]) and type of inference scenario (bridging or predictive), while the dependent variable was reaction time for reading. The co-variables were the independent measures of cognitive control and short-term memory. If there are significant correlations between the stand-
alone cognitive measures and the reaction times for the inferences in the conditions with distractor tasks, then this shows that the cognitive resources are being recruited to process those inferences.

**G2.1. Correlational Analyses.** As seen before, bivariate correlations shed light on whether two variables are associated by investigating the degree that two continuous variables covary. The L2 data were checked for all the assumptions in Chapters 5 and 6. Since the assumptions of normality, linearity, and homoscedasticity were not met for the L2 data, and as such Spearman’s rank order correlations were used for all analyses investigating the relationships of the variables. See Table G4 for the correlations. Since p-values have been included, the levels of statistical significance will not be denoted. As was previously reviewed in regard to correlations, the difference in the magnitudes of the correlations is most valuable, rather than which correlation is more or less significant.
Table G4. Correlations for the L2 group by scenario type (condition) and stand-alone cognitive measures. The p-value is represented (in parentheses).

<table>
<thead>
<tr>
<th>ID measure</th>
<th>Scenarios with bridging inferences</th>
<th>Scenarios with predictive inferences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(CONT) (STM) (CC)</td>
<td>(CONT) (STM) (CC)</td>
</tr>
<tr>
<td>N-Back Accuracy</td>
<td>-0.136 (0.295) -0.011 (0.483) 0.346 (0.074)</td>
<td>-0.236 (0.165) -0.225 (0.177) 0.364 (0.063)</td>
</tr>
<tr>
<td>N-Back RT</td>
<td>0.128 (0.301) 0.245 (0.149) 0.209 (0.188)</td>
<td>-0.095 (0.346) 0.227 (0.168) 0.29 (0.107)</td>
</tr>
<tr>
<td>VPT Accuracy</td>
<td>0.068 (0.394) 0.08 (0.372) 0.093 (0.353)</td>
<td>0.051 (0.418) -0.184 (0.225) 0.317 (0.093)</td>
</tr>
<tr>
<td>VPT RT</td>
<td>0.327 (0.093) 0.453 (0.026) 0.177 (0.234)</td>
<td>0.393 (0.048) .502 (0.014) 0.082 (0.369)</td>
</tr>
</tbody>
</table>

Spearman’s Rho, a non-parametric correlation statistic, was used to compute these statistics.

As mentioned previously, these correlations need to be interpreted with caution, as the r-squared values indicate the magnitude of the correlations. These results suggest that there is not a lot of shared variance between any two values reported, with the exception of the reaction time for the stand-alone short-term memory task (the Visual Patterns Task). Furthermore, the L2 data were less revealing than the L1 data, most likely due to a smaller sample size. There was a significant relationship between the reaction time on the Visual Patterns task, a measure of short-term memory, and the reaction time for the scenarios in the short-term memory condition for bridging inferences \((r = .453), p \text{ (one-tailed)} < .05\) and the predictive inferences \((r = .411), p \text{ (one-tailed)} < .05\). Additionally, there was a significant relationship for the scenarios with predictive
inferences in the control condition \((r = .393), p \text{ (one-tailed)} < .05\). Since the relationship was positive, the slower participants were to complete the visual patterns task, the slower they were to process scenarios with predictive and bridging inferences in the short-term memory condition. This mimics the pattern found in the L1 data for a link between short-term memory capacity for processing of both inference types.

**G3. Research question involving bridging inferences versus predictive inferences**

The third research question looked at how the participants processed the bridging inferences and predictive inferences in the different conditions; see Chapter 6 for predictions from the literature. The predictor variable is the type of distractor task used in each condition (i.e., control/none [CONT], cognitive control distractor [CC] and short-term memory distractor [STM]) and the dependent variable is reaction time. If participants processed predictive inferences with approximately the same speed in the control and short-term memory conditions, but were slower to read predictive inferences when distracted along the cognitive control dimension, this would indicate that cognitive control was more heavily relied upon to process this stimulus type. Similarly, if participants processed bridging inferences with approximately the same speed in the control and cognitive control conditions, but were slower to read bridging inferences when distracted along the short-term memory dimension, this would indicate that short-term memory was more heavily relied upon to process this stimulus type.

**G3.1. Processing of bridging inferences.** The first area investigated was the processing of bridging inferences in the different conditions and whether participants were slower to form bridging inferences when memory capacity is interfered with as compared to when it is not interfered with.
G3.1.1. Repeated-Measures ANOVA. The statistical test used to investigate this question was a repeated-measures ANOVA, which carries the assumptions of normal distribution of interval data and sphericity. The reaction time data were normally distributed continuous data. See Table G5 for descriptive statistics of the bridging inferences in the three conditions: control (CONT), short-term-memory (STM) and cognitive control CC).

Table G5. Descriptive statistics for bridging inferences in different conditions for L2 group.

<table>
<thead>
<tr>
<th>Two-Sentence Scenario Type</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control (CONT)</td>
<td>3929</td>
<td>682</td>
</tr>
<tr>
<td>Short-term memory (STM)</td>
<td>3949</td>
<td>740</td>
</tr>
<tr>
<td>Cognitive control (CC)</td>
<td>4999</td>
<td>1314</td>
</tr>
</tbody>
</table>

Next, the assumption of sphericity was checked. Mauchly’s test indicated that the assumption of sphericity was violated, ($\chi^2(2) = 15.42, p < .01$) (see Table G6).

Table G6. Mauchly’s test of sphericity assumptions for bridging inferences for L2 group.

<table>
<thead>
<tr>
<th>Mauchly's Test of Sphericity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within-Subjects Effect</td>
</tr>
<tr>
<td>Condition</td>
</tr>
</tbody>
</table>

Since the assumption of sphericity was violated, multivariate tests are reported ($\epsilon = .626$), using a Greenhouse-Geisser correction. The results of the repeated-measures ANOVA show that reaction time was significantly affected by the condition, $F(1.3, 12) = 15.93, p < .05$. Partial $\eta^2$ was .469 and observed power was 0.985. Repeated-measures t-tests (using a Bonferroni adjustment, $\alpha = .05/3 = .017$) showed that participants were significantly slower to read the two-sentence scenarios with bridging inferences in the
cognitive control condition (CC) condition than they were in the short-term memory condition (STM) or in the control condition (CONT) (CC versus STM: \( t(20) = 4.779, p < 0.01 \); CC versus CONT: \( t(18) = -3.923, p < 0.01 \)), but there was also not a significant difference in reading time between the two-sentence scenarios with bridging inferences in the short-term memory (STM) condition and the control condition (CONT) (STM versus CONT: \( t(18) = -1.77, p = 0.86 \), not significant). See Table G7 for the repeated measures ANOVA of bridging inference reaction times (in ms) by condition.

Table G7. RM ANOVA testing for bridging inferences by condition in the L2 group.

<table>
<thead>
<tr>
<th>Tests of Within-Subjects Effects*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Condition</td>
</tr>
<tr>
<td>Error (Condition)</td>
</tr>
</tbody>
</table>

*Performed with Greenhouse-Geisser corrections.

This also mirrors the results of the L2 analyses that incorporated data from the outliers and points to the fact that cognitive control is more influential in role the processing of bridging inferences than originally thought.

**G3.2. Processing of predictive inferences.** The next area investigated was the processing of predictive inferences in the different conditions. As we investigated above, this is a question of whether participants were slower to form predictive inferences when cognitive control capacity is interfered with as compared to when it is not interfered with. If participants are relying on their cognitive control to form predictive inferences, then a secondary task that taxes their cognitive control would result in slower reaction times when processing these inference types.
G3.2.1. Repeated-Measures ANOVA. The statistical analysis used to investigate this question is a repeated-measures ANOVA, which carries the assumptions of normal distribution of interval data and sphericity. As we saw above, the reaction time data were normally distributed continuous data. See Table G8 for descriptive statistics of the predictive inferences in the three conditions: control (CONT), short-term-memory (STM) and cognitive control (CC).

Table G8. Descriptive statistics for predictive inferences in different conditions.

<table>
<thead>
<tr>
<th>Two-Sentence Scenario Type</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control (CONT)</td>
<td>4785</td>
<td>987</td>
</tr>
<tr>
<td>Short-term memory (STM)</td>
<td>4462</td>
<td>959</td>
</tr>
<tr>
<td>Cognitive control (CC)</td>
<td>5646</td>
<td>1553</td>
</tr>
</tbody>
</table>

Next, the assumption of sphericity was checked. Mauchly’s test indicated that the assumption of sphericity was violated, ($\chi^2(2) = 0.521, p < .01$) (see Table G9).

Table G9. Mauchly’s test of sphericity assumptions for predictive inferences in L2 group.

<table>
<thead>
<tr>
<th>Mauchly's Test of Sphericity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within-Subjects Effect</td>
</tr>
<tr>
<td>Condition</td>
</tr>
</tbody>
</table>

Since the assumption of sphericity was violated, multivariate tests are reported ($\varepsilon = .676$), using a Greenhouse-Geisser correction. The results of the repeated-measures ANOVA show that reaction time was significantly affected by the condition, $F(1.4, 13.5) = 11.84$, $p < .05$. Partial $\eta^2$ was .38 and observed power was 0.96. Repeated-measures t-tests (using a Bonferroni adjustment, $\alpha = .05/3 = .017$) showed that participants were significantly slower to read the two-sentence scenarios with predictive inferences in cognitive control (CC) condition than they were in the short-term memory (STM) condition.
condition or the control (CONT) condition (CC versus STM: \( t(20) = -4.361, p < 0.01 \); CC versus CONT: \( t(20) = -2.794, p < 0.01 \)); the reading time for the two-sentence scenarios with predictive inferences in the control condition (CONT) was not significantly different than in the short-term memory (STM) condition (CONT versus STM: \( t(20) = 2.218, p = 0.039 \)), albeit just barely. See Table G10 for the repeated-measures ANOVA of predictive inferences by condition for the L2 group.

Table G10. RM ANOVA testing for predictive inferences by condition in the L2 group.

<table>
<thead>
<tr>
<th>Tests of Within-Subjects Effects*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Condition</td>
</tr>
<tr>
<td>Error (Condition)</td>
</tr>
</tbody>
</table>

*Performed with Greenhouse-Geisser corrections.

These results follow the same pattern as the L2 group that included the participants that were identified as possible outliers.

To summarize, I reran all of the analyses to verify the results from the L2 study without the smaller subset of L2 participants who were identified as outliers. These participants had longer reaction time latencies when compared the larger L2 set. Interestingly, the results of the L2 study were consistent in both sets of analyses – with and without the participants who were purported to have adopted different processing strategies. The only difference was more significant correlations for the reaction time of the Visual Patterns task in terms of reading time for both inference types regardless of condition. Implications for these results are explored in the discussion chapter.
References


Beeman, M. J., Bowden, E. M., & Gernsbacher, M. A. (2000). Right and left hemisphere


doi:http://dx.doi.org/10.1017/S0142716411000932


Language, 42, 526–544.


258


268


