

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Individual Cognitive Measures and Working Memory Accounts of Syntactic Island Phenomena /

Permalink

<https://escholarship.org/uc/item/3845q7jz>

Author

Michel, Daniel

Publication Date

2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Individual Cognitive Measures and Working Memory Accounts of Syntactic Island
Phenomena

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Linguistics and Cognitive Science

by

Daniel Michel

Committee in charge:

Professor Robert Kluender, Chair
Professor Grant Goodall, Co-Chair
Professor Ivano Caponigro
Professor Seana Coulson
Professor Marta Kutas

2014

Copyright

Daniel Michel, 2014

All rights reserved.

The dissertation of Daniel Michel is approved, and it is acceptable in quality and form for publication on microfilm:

Co-Chair

Chair

University of California, San Diego

2014

DEDICATION

For my parents

TABLE OF CONTENTS

Signature Page	iii
Dedication.....	iv
Table of Contents	v
List of Figures.....	xv
List of Tables	xxi
Acknowledgements	xxvi
Vita	xxx
Abstract of the Dissertation	xxxix
Chapter 1: Introduction.....	1
1.1 Overview	1
1.2 Islands as a processing phenomenon.....	3
1.3 Experimental approach.....	6
1.4 Dissertation overview	9
Chapter 2: Background	12
2.1 Introduction	12
2.2 Island phenomena.....	12
2.2.1 The basic data	13
2.2.2 Ameliorating effects on island violations.....	14
2.2.3 Experimental syntax	17
2.2.4 Satiation.....	20

2.2.5 On-line processing.....	21
2.2.5.1 Behavioral data (reading times).....	22
2.2.5.2 ERPs	26
2.2.5.2.1 LAN.....	28
2.2.5.2.2 P600.....	34
2.2.5.2.1 N400	45
2.3 Accounts of island phenomena.....	51
2.3.1 Grammatical	51
2.3.1.1 Syntactic	52
2.3.1.2 Semantic	56
2.3.2 Functional	57
2.3.3 Processing.....	58
2.3.3.1 Capacity-constrained	59
2.3.3.2 Similarity-interference.....	67
2.4 Current research agenda	73
Chapter 3: General Methods: Materials and Cognitive Measures.....	76
3.1 Introduction	76
3.2 Materials	76
3.3 Measures of Individual Differences	84
3.3.1 Reading Span.....	87
3.3.1.1 Task	88
3.3.1.2 Scoring.....	90

3.3.1.3 Results	91
3.3.2 N-back	91
3.3.2.1 Task	91
3.3.2.2 Scoring.....	94
3.3.2.3 Results	95
3.3.3 Flanker.....	95
3.3.3.1 Task	96
3.3.3.2 Scoring.....	97
3.3.3.3 Results	98
3.3.4 Memory interference	98
3.3.4.1 Task	99
3.3.4.2 Scoring.....	100
3.3.4.3 Results	101
3.3.5 Co-variation matrix	102
3.4 Conclusion	103
Chapter 4: Acceptability Judgment Experiment.....	105
4.1 Introduction	105
4.2 Background.....	107
4.2.1 Issues addressed by the current study.....	110
4.2.1.1 Choice of cognitive measures.....	110
4.2.1.2 The interpretation of null results	114
4.2.1.3 The interpretation of R^2	114

4.2.1.4 The reliance on DD scores	116
4.2.2 Cognitive Co-variation Intuition (CCI).....	120
4.2.2.1 The relationship between cognitive scores and sentence processing difficulty	122
4.2.2.2 The (potential lack of) transparency between processing and acceptability tasks: Rating task differences.....	128
4.2.3 Predictions and potential interpretations	134
4.3 Methods	136
4.3.1 Participants	136
4.3.2 Materials	136
4.3.3 Procedure.....	138
4.3.3.1 Cognitive measures	138
4.3.3.2 Acceptability ratings.....	138
4.3.4 Analysis	139
4.3.4.1 Linear mixed-effects model.....	140
4.3.4.2 Simple linear regression	141
4.3.4.3 Median split	142
4.4 Results and Discussion	143
4.4.1 Basic effects.....	143
4.4.1.1 Results	143
4.4.1.2 Discussion.....	147
4.4.2 Effects including cognitive measures	148

4.4.2.1 Linear mixed-effects model including cognitive measures.....	148
4.4.2.1.1 Results	148
4.4.2.1.2 Discussion.....	152
4.4.2.2 Pattern of results using simple linear regression score analysis.....	154
4.4.2.2.1 DD score	154
4.4.2.2.1.1 Results	154
4.4.2.2.1.2 Discussion.....	156
4.4.2.2.2 Other regressions	157
4.4.2.2.2.1 Results	158
4.4.2.2.2.2 Discussion.....	162
4.4.2.3 Median Splits.....	165
4.4.2.3.1 Results	165
4.4.2.3.2 Discussion.....	169
4.5 Summary.....	170
4.6 Conclusion	172
Chapter 5: Self-Paced Reading Experiment	175
5.1 Introduction	175
5.2 Predictions	177
5.3 Methods	179
5.3.1 Participants	179
5.3.2 Materials	180
5.3.3 Procedure.....	181

5.3.3.1 Cognitive measures	181
5.3.3.2 Self-paced reading	182
5.3.4 Analysis	184
5.4 Results and Discussion	186
5.4.1 Comprehension	186
5.4.1.1 Results	186
5.4.1.2 Discussion.....	187
5.4.2 Basic effects.....	187
5.4.2.1 Results	188
5.4.2.1.1 Position 5 (clause boundary)	189
5.4.2.1.2 Position 9 (<i>before</i>)	190
5.4.2.2 Discussion.....	191
5.4.2.2.1 Position 5 (clause boundary)	191
5.4.2.2.2 Position 9 (<i>before</i>)	195
5.4.3 Median splits	199
5.4.3.1 Results	199
5.4.3.1.1 Position 5 (clause boundary): Reading Span.....	200
5.4.3.1.2 Position 5 (clause boundary): Memory Lure.....	201
5.4.3.2 Discussion.....	202
5.4.3.2.1 Position 5 (clause boundary): Reading span	203
5.4.3.2.2 Position 5 (clause boundary): Form lure	205
5.5 Summary.....	209

5.6 Conclusion	210
Chapter 6: Event-Related Potentials Experiment	213
6.1 Introduction	213
6.2 Predictions	214
6.2.1 Lexical differences (positions 3 and 8, <i>_ openly/the sailor</i>)	215
6.2.2 Post-gap LAN (positions 4 and 9, <i>assumed/inquired</i>).....	216
6.2.3 Sustained LAN (position 4 or later)	217
6.2.4 Clause boundary N400 (position 5, <i>that/whether</i>)	218
6.2.5 Pre-gap P600 (position 7, <i>befriended</i>).....	219
6.2.6 Embedded gap position lexical differences and embedded post-gap LAN (positions 8 and 9).....	219
6.2.7 Sentence-final N400 (position 12, <i>hearing?</i>)	220
6.2.8 Processing cost of <i>whether</i> -island violation (multiple possible positions)	221
6.2.9 Cognitive measures co-variation (multiple possible positions).....	222
6.3 Methods	222
6.3.1 Participants	222
6.3.2 Materials	223
6.3.3 Procedure.....	225
6.3.3.1 Cognitive measures	225
6.3.3.2 Electrophysiological recording.....	226
6.3.3.3 Post-ERP acceptability judgments	227

6.3.4 EEG Analysis	228
6.4 Results and Discussion.....	231
6.4.1 Post-ERP acceptability judgments	231
6.4.1.1 Results	231
6.4.1.2 Discussion.....	233
6.4.2 Basic effects.....	234
6.4.2.1 Results	234
6.4.2.1.1 Position 2 (matrix pre-gap position: <i>had</i>).....	235
6.4.2.1.2 Position 3 (matrix gap position: <i>_ openly / the sailor</i>)	237
6.4.2.1.3 Position 4 (matrix post-gap position: <i>assumed / inquired</i>).....	241
6.4.2.1.4 Position 5 (clause boundary: <i>that/whether</i>).....	245
6.4.2.1.5 Position 7 (embedded pre-gap position: <i>befriended</i>).....	249
6.4.2.1.6 Position 8 (embedded gap position: <i>the sailor / _ openly</i>)	252
6.4.2.1.9 Position 9 (embedded post-gap position: <i>before</i>).....	264
6.4.2.1.8 Position 12 (sentence-final position: <i>hearing?</i>)	269
6.4.2.1.9 Slow wave: Sustained negativity.....	272
6.4.2.1.10 Summary.....	276
6.4.2.2 Discussion.....	277
6.4.2.2.1 P600.....	278
6.4.2.2.2 Lexical differences	285
6.4.2.2.3 Matrix verb N400	288
6.4.2.2.4 LAN.....	289

6.4.2.2.6 Sustained LAN	293
6.4.2.2.5 Clause boundary	296
6.4.2.2.7 N400 effects.....	301
6.4.2.2.8 Summary.....	312
6.4.3 Median splits	313
6.4.3.1 Results	313
6.4.3.1.1 Position 7 (<i>befriended</i>): Interaction of STRUCTURE x N-BACK group.....	314
6.4.3.1.2 Position 8 (<i>the sailor / _openly</i>)	315
6.4.3.1.2 Position 12 (<i>hearing?</i>) sentence final negativity	318
6.4.3.2 Discussion.....	322
6.4.3.2.1 Position 7 (<i>befriended</i>)	323
6.4.3.2.2 Position 8 (<i>the sailor / _openly</i>)	325
6.4.3.2.3 Position 12 (<i>hearing?</i>).....	330
6.5 Summary.....	331
6.6 Conclusion.....	333
Chapter 7: Discussion and Conclusion.....	335
7.1 Dissertation summary	336
7.1.1 Introduction of the similarity-interference account of islands	336
7.1.2 Development of a balanced factorial design	337
7.1.3 Advances in acceptability/processing frameworks: The Cognitive Co-variation Intuition and the Processing Benefit Schedule.....	338

7.1.4 Individual differences in the self-paced reading of islands	342
7.1.5 ERP responses to gap prediction and filler-gap association.....	343
7.1.6 Predictability in processing islands	344
7.2 New questions arising from the ERP findings	346
7.3 Updating working memory models	348
7.4 Processing accounts of islands	354
7.5 Gap predictability account of processing islands	358
7.6 Parser-grammar relationship	362
7.7 Future research	370
7.8 Concluding remarks.....	373
Appendix 1: Materials for acceptability judgment and self-paced reading experiments (Experiments 1 and 2, respectively).....	375
Appendix 2: Materials for ERP experiment (Experiment 3).....	381
References	412

LIST OF FIGURES

Figure 3.1: Right-facing arrow surrounded by incongruent flankers	97
Figure 4-1: Mean results (raw scores) for Experiment 1. Error bars indicate standard error	144
Figure 4-2: Mean results (z-cores) for Experiment 1. Error bars indicate standard error	145
Figure 4-3: Interaction of GAP and MEMORY LURE. Acceptability ratings of MATRIX GAP (black) and EMBEDDED GAP (red) sentences plotted against MEMORY LURE accuracy (higher accuracy indicates less susceptibility to similarity-based interference). Shaded area indicates standard error	149
Figure 4-4: Interaction of STRUCTURE and MEMORY LURE. Acceptability ratings of NON-ISLAND STRUCTURE (black) and ISLAND STRUCTURE (red) sentences plotted against MEMORY LURE accuracy (higher accuracy indicates less susceptibility to similarity-based interference)	150
Figure 4-5: Updated interaction of GAP and MEMORY LURE; scores 0.50 or greater. Acceptability ratings of MATRIX GAP (black) and EMBEDDED GAP (red) sentences plotted against MEMORY LURE accuracy (higher accuracy indicates less susceptibility to similarity-based interference)	151

Figure 4-6: Updated interaction of STRUCTURE and MEMORY LURE; scores 0.50 or greater. Acceptability ratings of NON-ISLAND STRUCTURE (black) and ISLAND STRUCTURE (red) sentences plotted against MEMORY LURE accuracy	152
Figure 4-7: Simple linear regression of n-back scores and DD scores. The regression line has an intercept of 1.47 and a slope of -1.59 , with $R^2 =$ 0.05. The data are marginally negatively correlated ($r = -0.22$, $p = 0.055$)	155
Figure 4-8: Simple linear regression of form lure scores and GAP position effect in NON-ISLANDS.....	159
Figure 4-9: Simple linear regression of form lure scores and GAP position effect in ISLANDS.....	160
Figure 4-10: Simple linear regression of form lure scores to island violation z-scores	161
Figure 4-11: Mean results (z-scores) for FORM LURE (high and low scorers) x GAP with standard error bars	167
Figure 4-12: Mean results (z-scores) for FORM LURE (high and low scorers) x STRUCTURE with standard error bars.....	168
Figure 5-1: Residual reading times.....	188
Figure 5-2: Position 5 residual reading times.....	190
Figure 5-3 Position 5 GAP x STRUCTURE (A) HIGH SPAN GROUP (B) LOW SPAN GROUP	201

Figure 5-4 Position 5 GAP x FORM LURE GROUP	202
Figure 6-1: Electrode locations	227
Figure 6-2: Post ERP acceptability scores.....	232
Figure 6-3: Position 2, pre-gap matrix clause (<i>had</i>) whole head ERPs.....	236
Figure 6-4: Position 2 (<i>had</i>) late positivity shown at CP4 (A) and in topographic isovoltage map showing MATRIX (preceding <i>_ openly</i>) - EMBEDDED (preceding <i>the sailor</i>) from 600-900 msec (B).....	237
Figure 6-5: Position 3, gap position matrix clause (<i>_ openly /the sailor</i>) whole head ERPs.....	240
Figure 6-6: Position 3 (<i>_ openly /the sailor</i>) negativities shown at F7 (A) and CPz (B) with topographic isovoltage map showing EMBEDDED (<i>the sailor</i>) - MATRIX (<i>_ openly</i>) from 300-600 msec (C).....	241
Figure 6-7: Position 4, post-gap matrix clause (<i>assumed / inquired</i>) whole head ERPs.....	244
Figure 6-8: Position 4 (<i>assumed/inquired</i>) negativites shown at F7 (A) and CPz (B) and topographic isovoltage map showing MATRIX (after <i>_ openly</i>) - EMBEDDED (after <i>the sailor</i>) from 300-600 msec (C)..	245
Figure 6-9: Position 5, clause boundary (<i>that / whether</i>) whole head ERPs	247
Figure 6-10: Position 5 (<i>that/whether</i>). Select electrodes shown with topographic isovoltage maps of ISLAND (<i>whether</i>) - NON-ISLAND (<i>that</i>) in time windows labeled above.....	248

Figure 6-11: Position 7, pre-gap embedded clause (<i>befriended</i>) whole head ERPs	250
Figure 6-12: Position 7 (<i>befriended</i>) late positivity shown at CP4 (A) and topographic isovoltage map showing EMBEDDED (before <i>_ openly</i>) – MATRIX (before <i>the sailor</i>) from 600-900 msec (B)	251
Figure 6-13 Comparison of positions 2 and 7. CP4 (A and B) and topographic isovoltage map showing [position immediately preceding the gap (<i>_ openly</i>)] – [position immediately preceding <i>the sailor</i>] from 600-900 msec (C, D)	252
Figure 6-14: Position 8, gap position embedded clause (<i>the sailor / _ openly</i>) whole head ERPs	254
Figure 6-15: Position 8 (<i>the sailor / _ openly</i>) negativities shown at F7 (A) and CPz (B) and topographic isovoltage map showing showing MATRIX (<i>the sailor</i>) – EMBEDDED (<i>_ openly</i>) from 300-600 msec (C)	256
Figure 6-16: Position 8 (<i>the sailor / _ openly</i>) main effect of GAP shown at F7 (A), interaction of GAP x STRUCTURE shown at CPz (B) and topographic isovoltage map showing EMBEDDED ISLAND (<i>_ openly</i>) – EMBEDDED NON-ISLAND (<i>_ openly</i>) from 300-600 msec (C)	259
Figure 6-17: Five word average starting at position 6 (<i>the captain</i>): Point of interest is interaction at { <i>_ openly / the sailor</i> }	260
Figure 6-18 Comparison of main effect of GAP (lexical difference of <i>the sailor</i> vs. <i>_ openly</i>) in positions 3 and 8: F7 (A and B) and CPz (C and D)	261

Figure 6-19 Comparison of positions 3 and 8. Topographic isovoltage map showing [the condition including the lexical item <i>the sailor</i>] – [the condition including the lexical item <i>_ openly</i>] from 600-900 msec.	263
Figure 6-20: Position 9, post-gap position embedded clause (<i>before</i>) whole head ERPs.....	266
Figure 6-21: Position 9 (<i>before</i>) negativity shown at F7 (A) and topographic isovoltage map showing EMBEDDED (following <i>the sailor</i>) – MATRIX (following <i>_ openly</i>) from 300-600 msec (B).....	267
Figure 6-22: Five word average starting at position 7 (<i>befriended</i>): Point of interest is reversal of more negative conditions from lexical LAN { <i>_ openly / the sailor</i> } to post-gap LAN (<i>before</i>)	268
Figure 6-23 Comparison of LAN responses at positions 4 and 9. F7 (A and B) and topographic isovoltage map showing [the condition after the gap (<i>_ openly</i>)] – [the condition after <i>the sailor</i>] from 300-600 msec (C, D).....	269
Figure 6-24: Position 12, sentence-final position (<i>hearing?</i>) whole head ERPs	271
Figure 6-25: Position 12 (<i>hearing?</i>) broad negativity shown at Pz (A) and topographic isovoltage map showing showing EMBEDDED - MATRIX from 300-600 msec (B)	272
Figure 6-26: Four word averages starting at post-gap positions. Position 4 through 7 (A), position 9 through 12 (B)	275

Figure 6-27: Matrix clause <i>_ openly</i> (black trace) compared to embedded clause <i>_ openly</i> (red trace) in a <i>whether</i> -island (A) and non-island <i>that</i> -clause (B)	303
Figure 6-28: Comparison of N400 amplitudes of <i>the sailor</i> at positions 3 and 8	308
Figure 6-29: Position 7 (<i>befriended</i>) STRUCTURE x N-BACK group mean scalp voltage (600-900 msec). Error bars denote standard error	315
Figure 6-30: Position 8: <i>the sailor / openly</i> GAP x STRUCTURE interaction at CPz in high (A) and low (C) span groups and topographic isovoltage map showing EMBEDDED ISLAND (<i>_ openly</i>) – EMBEDDED NON-ISLAND (<i>_ openly</i>) from 300-600 msec in high (B) and low (D) span groups	318
Figure 6-31: Position 12 (<i>hearing?</i>) potential N400 responses at Pz for high (A) and low (B) span groups	321
Figure 6-32: Position 12 (<i>hearing?</i>) topographic isovoltage map showing 300-600 msec. High (A,C) and low (B,D) span groups showing EMBEDDED – MATRIX (A,B) and ISLAND – NON-ISLAND (C,D)	322

LIST OF TABLES

Table 3-1: Sample stimuli set. Manipulation of STRUCTURE indicated in bold. Manipulation of <i>GAP</i> indicated by italics. No specific claims are intended by the placement of the gap, which is meant only to indicate the on-line point of disambiguation of the gap position	77
Table 3-2: Sample NON-ISLAND stimuli.....	78
Table 3-3: Position 3 & 8 controls for Experiments 1 and 2. Mean (Standard deviation).....	81
Table 3-4: Position 3 & 8 controls for Experiment 3. Mean (Standard deviation).....	81
Table 3-4: Position 4 matrix verb controls. Mean (Standard deviation)	83
Table 3-5 Reading span results across three experiments	91
Table 3-6 N-back (3-back) results across three experiments	95
Table 3-7 Flanker (incongruent - congruent) results across three experiments (msec)	98
Table 3-8 Memory Lure results across three experiments	101
Table 3-9 Form Lure results across three experiments	101
Table 3-10 Semantic Lure results across three experiments	102
Table 3-11 Correlation matrix: (Pearson's r), all experiments ($n = 160$).....	102

Table 4-1: Processing Benefits Schedule (PBS): Expectations of processing benefits for individuals with greater cognitive resources / higher cognitive scores (i.e. working memory, attention)	126
Table 4-2: Experiment 1 sample stimuli set. Manipulations of STRUCTURE indicated by bold. Manipulations of GAP indicated by italics. No specific claims are intended by the placement of the gap, which indicates the on-line point of disambiguation of the gap position.....	137
Table 4-3: Z-score transformed data. Means (standard deviation).....	146
Table 4-4: Significance testing of the basic model: linear mixed-effects model with no individual differences measures included.	146
Table 4-5: Significance testing of the memory-interference model: a linear mixed-effects model including MEMORY LURE as a factor	148
Table 4-6: Significance testing of updated memory-interference model: a linear mixed-effects model including MEMORY LURE as a factor, removing low-scorers (below 50%)	151
Table 4-7: Regressions of cognitive measures to DD scores	156
Table 4-8: Regressions of cognitive measures to MATRIX – EMBEDDED, NON-ISLANDS only	158
Table 4-9: Regressions of cognitive measures to MATRIX – EMBEDDED, ISLANDS only.....	160
Table 4-10: Regressions of cognitive measures to island violation z-scores	161
Table 4-11: ANOVAs including median split memory lure measures	166

Table 4-12: ANOVAs including median split cognitive measures (except memory lure)	168
Table 5-1 Predictions for the self-paced reading findings.....	178
Table 5-2: Experiment 2 sample stimuli set. Manipulations of STRUCTURE are indicated in bold while manipulations of GAP are indicated by italics. No specific claims are intended by the placement of the gap, which indicates the on-line point of disambiguation of the gap position.....	181
Table 5-3 Practice sentences	183
Table 5-4: Mean comprehension accuracy by condition	186
Table 5-5: Word positions 1-9.....	188
Table 5-6: Position 5 residual reading times. Mean (standard error).....	189
Table 6.1: Critical comparisons within the stimulus sentences, indicating both numbering and labels relative to the gap position in both the matrix and embedded clauses	214
Table 6-2: Experiment 3 sample stimuli set. Manipulations of STRUCTURE are indicated in bold while manipulations of GAP are indicated by italics. No specific claims are intended by the placement of the gap, which indicates the on-line point of disambiguation of the gap position.....	224
Table 6-3: Post ERP acceptability z-score transformed data. Means (standard deviation).....	232

Table 6-4: Critical comparisons within the stimulus sentences, indicating both numbering and labels relative to the gap position in both the matrix and embedded clauses	234
Table 6-5: Position 3 (<i>_ openly / the sailor</i>) 300-600 msec window	238
Table 6-6: Position 3 post-hoc analyses (<i>_ openly /the sailor</i>) 300-600 msec window	239
Table 6-7: Position 4 (<i>assumed / inquired</i>) 300-600 msec window	242
Table 6-8: Position 4 post-hoc (<i>assumed / inquired</i>) 300-600 msec window	243
Figure 6-7: Position 4, post-gap matrix clause (<i>assumed / inquired</i>) whole head ERPs.....	244
Figure 6-8: Position 4 (<i>assumed/inquired</i>) negativities shown at F7 (A) and CPz (B) and topographic isovoltage map showing MATRIX (after <i>_ openly</i>) - EMBEDDED (after <i>the sailor</i>) from 300-600 msec (C)	245
Table 6-9: Position 8 (<i>the sailor / _openly</i>) 300-600 msec window	255
Table 6-10: Position 8 post-hoc (<i>the sailor / _openly</i>) 300-600 msec window...	257
Table 6-11: Position 8 post-hoc (<i>the sailor / _openly</i>) 300-600 msec window paired comparisons	258
Table 6-12: Position 9 (<i>before</i>) 300-600 msec window	265
Table 6-13: Critical comparisons within the stimulus sentences, indicating both numbering and labels relative to the gap position in both the matrix and embedded clauses	273

Table 6-14: Location of (non-lexical) N400 effects shaded in gray. Critical indicators of condition are underlined. Note: the same experimental conditions are not represented by the matrix clause (MATRIX GAP) and embedded clause (EMBEDDED ISLAND)	303
Table 6-15: Position 8: <i>the sailor / _openly</i> (300-600)	316
Table 6-16: Position 8 post-hoc (<i>the sailor / _openly</i>) 300-600 msec window...	317
Table 6-17: Position 12 (<i>hearing?</i>) 300-600 msec window	319
Table 6-18: Position 12 post-hoc (<i>hearing?</i>) 300-600 msec window	320
Table 7-1: Sample stimulus set. Manipulation of STRUCTURE indicated in bold. Manipulation of GAP indicated by italics. No specific claims are intended by the placement of the gap, which is meant only to indicate the on-line point of disambiguation of the gap position	338
Table 7-2: Processing Benefit Schedule (PBS): Expectations of processing benefits for individuals with greater cognitive resources / higher cognitive scores (i.e. working memory, attention)	341

ACKNOWLEDGEMENTS

While one person's name appears front and center on a dissertation, the contributions of others is the structural core around which the dissertation is written. Without the guidance, assistance and support of many others, this dissertation would not resemble anything close to its current form, and would likely not be able to be raised off the ground. I am thankful for the many individuals who have helped to shape this work.

First and foremost, I express my gratitude to my committee members, especially my co-chairs and co-advisors: Grant Goodall and Robert Kluender. I benefited greatly from the balanced team that Grant and Robert formed. Both were willing and able to provide their insights and support, each in their own style and specialty. I am especially grateful for Grant's (often understated) guidance in developing syntactic acceptability studies and Robert's freeform brainstorming sessions and careful examination of experimental materials. I thank Ivano Caponigro for his insightful questions and guidance, especially in terms of packaging and delivery of my findings. I thank Seana Coulson for graciously welcoming me into her lab to run my ERP experiment, including providing a rich range of technical guidance as well as guidance in writing multiple abstracts based on the results. And I thank Marta Kutas for generously sharing her expertise when that ERP experiment produced an unexpected result. To all of my committee members, I have immense gratitude for your time, reflection and feedback on this work. I know it is better thanks to your involvement.

I would also like to specifically thank some other faculty members at UC San Diego: Gabriela Caballero for giving me the opportunity to work on her Choguita Raramuri project; and for the many stimulating interdisciplinary classes that make UC San Diego such a unique place to be learning about and researching language, I thank Farrell Ackerman, Vic Ferreira, Terry Jernigan, Roger Levy, Rachel Mayberry, and Keith Rayner (as well as classes taken with Grant, Robert and Marta). I want to acknowledge and thank the Linguistics Department staff for all their day-to-day efforts that allow the research and teaching done here to proceed smoothly: Allen Allison, Gris Arellano, Dennis Fink, Corie Gochicoa, Alycia Randol, Marc Silver, Rachel Pekras, Cheri Radke, Ezra Van Everbroeck and Lucie Wiseman.

I am thankful for the faculty at the University of Florida: Caroline Wiltshire, who gave me my introduction to linguistics, Eric Potsdam who advised me through my M.A., and Wind Cowles and Edith Kaan for introducing me to the world of psycholinguistics and neurolinguistics.

At various conferences and talks, I have benefited from carefully considered feedback on this research. I thank Collin Phillips, Jon Sprouse, Laura Staum-Casasanto and Matt Wagers for their reflections and insights. To all those researchers whose work has preceded and laid the groundwork for this dissertation, I am grateful.

I thank my colleges in Robert Kluender's Language and Brain Lab. I am grateful to members past and present whose stimulating discussions have directly and indirectly influenced my research: Simone Gieselmann, Dave Hall, Nayoung Kwon,

Lisa Rosenfelt, Mieko Ueno, and especially Chris Barkley for insightful comments, dedicated feedback, and for keeping me awake on the drive from Florida.

I also want to thank those in Grant Goodall's Experimental Syntax Lab, where I have embedded myself on campus to work amongst an array of decorative embellishments (thanks Rudolfo). I thank the numerous 199 research assistants in the lab that I have worked with: Camille Asaro, Taylor Braxel, Karly Fasth, Nikolai Gutierrez, Adrienne Le Fevre, Danielle Marom, and Anju Shimura. I am grateful to the 'regulars' of the lab meetings over the years for thoughtful conversations and feedback: Henry Beecher, Kate Davidson, Shin Fukuda, Boyoung Kim, Bethany Keffala, Leslie Lee, Ryan Lopic, Emily Morgan, Savi Namboodiripad, Amanda Richart, and Alex Stiller-Shulman. I want to especially thank Gabe Doyle for useful and relevant discussions (for example modeling and statistics) as well as helpful and distracting conversations and activities (examples too numerous to name).

I want to express my deep gratitude for the members of Seana Coulson's Brain and Cognition Lab, who guided me through the process of running my ERP experiment in their space: Tristan Davenport for sample scripts, guidance on programs, procedures and troubleshooting; Megan Bardolph for guidance on data processing and analysis; and Josh Davis for letting me get basic experience by helping to run his participants. Thanks is also due to the 199 research assistants that helped me run participants: Karla Barranco Marquez, Isabella Jones, Seerat Jammu, Michael Belcher, Lisset Berrios, Tracy Hoang, and especially Pat Samermit for her invaluable

experience in running participants- more than anyone else, she gave me hands on training until I was ready to run participants on my own.

Thanks also for support from classmates and friends who were not associated with the above labs: Lucien Carroll, Naja Ferjan Ramirez and Hope Morgan. I'd also like to thank some of the department's more senior students (and now alumni) for taking the time to welcome and assist newer students acclimate: Rebecca Colavin, Laura Kertz, Cindy Kilpatrick and Hannah Rohde. And for extracurricular outings, board games, dinners and just general sanity-restoring activities, I am grateful to Lucien Carroll, Alex Del Giudice, Kate Davidson, Gabe Doyle, Boyoung Kim, Naja Ferjan Ramirez, Rudolfo Mata and Hope Morgan. For other sanity-restoring reasons, I thank God and gravitational tidal forces for the calming waves, and Roger Revelle for founding UCSD so close to them.

Finally, I owe an incalculable debt to my family. Many thanks to my parents, Dan and Grace Michel, for their unwavering support and faith over the years; to my siblings, Christie, Brian and Lisa; and to my nephews and nieces: Zach, Layla, Nathaniel, Jillian, Jonah and Bethany (your names are now in a dissertation!). To my wonderful wife Zoe, thank you for your support, encouragement and love. I could not have done this without you.

VITA

Education

- 2014 Ph.D. in Linguistics and Cognitive Science
University of California, San Diego
- 2007 M.A. in Linguistics
University of Florida
- 2000 B.A. in Religion
University of Florida

Publications

- 2013 Michel, D. and Goodall, G. **Finiteness and the nature of island constraints.** In Nobu Goto, Koichi Otaki, Atsushi Sato, and Kensuke Takita (eds.), *Proceedings of GLOW in Asia IX 2012: The main session*. 187-197. Mie University, Japan.
- Michel, D. **Individual on-line processing differences are not necessarily reflected in off-line acceptability judgments.** LSA 2013 Annual Meeting Extended Abstracts. Accessible at:
<http://www.linguisticsociety.org/>
- Michel, D. **Agreement and pronoun incorporation in ASL verbs from a Bantu perspective: Reporting of pilot data.** In Lucien Carroll, Bethany Keffala, and Dan Michel (eds.), *San Diego Linguistic Papers, Issue 4*. Department of Linguistics, UCSD, UC San Diego.
- 2012 Michel, D. **Positional Transparency in c’Lela.** In Ryan Bochnak, Nassira Nicola, Peet Klecha, Jasmin Urban, Alice Lemieux and Christina Weaver, (eds.), *Proceedings from the Annual Meeting of the Chicago Linguistic Society* 45:1, pgs 455-469.
- Fukuda, S., Goodall, G., Michel, D. & Beecher, H. **Is magnitude estimation worth the trouble?** In Jaehoon Choi, E. Alan Hogue, Jeffery Punske, Deniz Tat, Jessamyn Schertz and Alex Truman (eds.), *Proceedings of the 29th West Coast Conference on Formal Linguistics*. 328-336. Somerville, MA. Cascadilla Proceedings Project.

ABSTRACT OF THE DISSERTATION

Individual Cognitive Measures and Working Memory Accounts of Syntactic Island

Phenomena

by

Daniel Michel

Doctor of Philosophy in Linguistics and Cognitive Science

University of California, San Diego, 2014

Professor Grant Goodall, Co-Chair

Professor Robert Kluender, Chair

This dissertation examines the on-line processing and off-line acceptability judgments of *whether*-islands using an individual differences approach in order to test processing accounts of island phenomena. Processing accounts of islands propose that the unacceptability of an island violation can be attributed to difficulties in on-line

processing, but accounts differ in how this difficulty is characterized, based on the view of working memory adopted.

The three experiments reported here (acceptability judgments, Chapter 4; self-paced reading, Chapter 5; event-related potentials- ERPs, Chapter 6) test the capacity-constrained account of islands (e.g. Kluender 1991), based on working memory as capacity-constrained (Just & Carpenter 1992), as well as a novel ‘similarity-interference’ account of islands (Chapter 2) based on working memory as subject to similarity-based interference (e.g. Gordon, Hendrick & Johnson 2001; Lewis, Vasishth & Van Dyke 2006).

I introduce two frameworks (Chapter 4) - the Cognitive Co-variation Intuition (CCI) and the Processing Benefits Schedule (PBS) - to clarify the relationship between off-line acceptability, on-line processing and individual differences (i.e. reading span, memory interference). Ultimately, the data reported here do not support a view where processing factors directly and transparently predict the unacceptability of island violations (neither do they directly support a grammatical account of islands). However, the ERP data indicate the importance of real time prediction for the on-line processing of islands. This is formalized as the gap predictability account of processing islands (Chapter 7).

Specifically, high span readers are better able to adjust their predictions for a gap online (evidenced by an N400 response at the embedded gap, suggesting lowered expectation for a gap in an island context), but both high and low span readers show evidence of filler-gap association (evidenced by post-gap LANs). There was no

evidence of a failed parse or reanalysis in any condition or in any group of participants, as predicted by both processing and grammatical accounts, yet these same participants still rated island violations as the least acceptable sentences. There is no apparent ERP evidence of a large on-line processing cost that would account for this difference in acceptability. These island violations appear to be unacceptable, but not unparseable.

Chapter 1 Introduction

1.1 Overview

This dissertation examines the on-line processing and off-line acceptability of long-distance filler-gap dependencies and certain configurations, termed ‘islands,’ that disrupt these dependencies. Specifically, bi-clausal *wh*-questions, such as in (1.1) will be examined.

(1.1) **Who** had Mary thought [that John saw ___?]

In psycholinguistics, *who*, in (1.1), is referred to as a ‘filler’ and the syntactic position it must associate with in order to be interpreted is indicated by the underscore and referred to as a gap (Fodor 1978). The square brackets indicate an embedded clause. The question in (1.1) is asking about a person (that Mary thought) John saw. Note that if the gap is elsewhere, as in (1.2), the interpretation of the filler *who* changes. The question is no longer about who was seen, but who was thinking.

(1.2) **Who** had ___ thought [that John saw Bill?]

Other types of embedded clauses are possible. The embedded clauses in (1.1) and (1.2) are declarative clauses (even though the entire sentence is an interrogative, the clause in square brackets is itself declarative). The sentence in (1.3) shows an embedded interrogative clause.

(1.3) **Who** had ___ wondered [whether John saw Bill?]

The interpretation of *who* in (1.3) is not very different from in (1.2): instead of asking about the person doing the thinking, this sentence asks about the person doing the wondering. However, (1.4) does not appear to be as similar to its counterpart in (1.1).

(1.4) * **Who** had Mary wondered [whether John saw ___?]

In (1.4), *who* should be just as interpretable as (1.1) in that the sentences are asking about the person seen by John. However, many people report that examples like (1.4) are unacceptable or ungrammatical to them as sentences of English. The asterisk before the sentence indicates that most native speakers judge the sentence to be unacceptable. Other researchers use the asterisk to indicate ungrammaticality, but I intend no claim on the grammatical status of such a judgment. The sentence in (1.4) is an example of a violation of a so-called *whether*-island, one of a number of syntactic configurations that disrupt the dependency between a filler and its gap in similar ways.

The term ‘island’ was coined by Ross (1967) as a metaphor indicating that a particular part of the sentence is isolated, like an island, from other parts of the sentence. A gap inside an island is isolated from a filler that is outside of the island, resulting in an island violation (that is, the filler-gap dependency is unacceptable). The original cataloging of syntactic islands organized them as a series of constraints: each island type was given its own specific constraint. Subsequent theoretical analyses of island phenomena, however, attempted to provide a unified analysis for these various structures (e.g. Subjacency: Chomsky 1973, 1977, 1981; Barriers: Chomsky 1986; Relativized Minimality: Rizzi 1990, Cinque 1990). Views of unacceptability/

ungrammaticality also changed over time. Ross (1987) saw ungrammaticality as due to cumulative small deviations from a prototype reaching a certain threshold, at which point an individual perceives the utterance as ungrammatical. Fodor (1983) characterizes ungrammaticality similarly, but in terms of a build-up of markedness. This incremental view of the unacceptability of sentences like those in (1.4) opened the way for a processing account of islands.

1.2 Islands as a processing phenomenon

Kluender (1991) presented a different approach to islands, arguing that the unacceptability of islands need not be accounted for by a theoretical syntactic constraint or analysis but could instead be captured by the interaction of independently motivated difficulties in sentence processing. Thus sentences like (1.4) are less acceptable than ones like (1.1) not because a constraint or particular syntactic configuration rules them out, but because (1.4) is more difficult to parse in real time. This account, and others that followed from it (e.g. Kluender 1998; Kluender and Kutas 1993a,b; Hofmeister 2007; Sag et. al. 2007) rely on a specific view of working memory to explain the processing difficulties. This view of working memory is the Just and Carpenter (1992) Capacity Constrained Comprehension Theory.

Working memory is a cognitive construct that involves both computational processes and memory storage, distinguishing it from short-term memory, which includes only storage (e.g. Cowan 2004, 2008). In the Just and Carpenter model of working memory, the system has a limited resource capacity that both storage and

processing must draw from. If a task requires many items to be stored, then there is less capacity available for processing. If complex processing is required, then there is less capacity for storage. Kluender (1991) argues that the storing of the filler (*who* in examples 1.1 - 1.4) combines with processing complexities, such as those present in the interrogative *wonder* [*whether* clause boundary (that are not present in the declarative *think* [*that* clause boundary), to overload the capacity of the working memory system. This overload results in the sentence being deemed unacceptable. The combination of independent factors proposed by Kluender (1991) is more like the Ross (1987) view of island violations as an accumulation of small deviations than it is like the Ross (1967) view of island violations each requiring specific global constraints to explain their unacceptability.

I will refer to this general approach towards island phenomena as a capacity-constrained account, since it relies so heavily on the capacity-constrained view of working memory. It was this account that first seeded the idea for this dissertation. In what is now, with hindsight, a somewhat naïve assumption, I thought of a way to test the capacity-constrained account of islands. Since (i) there is a task, the reading span task of Daneman & Carpenter (1980), that is purported to be a cognitive measure of working memory capacity, and (ii) the capacity-constrained account of islands claims that the unacceptability of island violations is due to an overload of working memory capacity, then individuals with a measurably higher capacity should be able to process island violations easier and thus rate them higher. I call this the Cognitive Co-variation Intuition (CCI). Complications to this apparently straightforward idea are discussed in

detail in Chapter 4 (section 4.2.2). Sprouse, Wagers and Phillips (2012) conducted an acceptability judgment study based on the same basic intuition. Chapter 4 (section 4.2.1) details the benefits that the acceptability study in this dissertation has over that study.

In the years since Kluender first proposed the capacity-constrained account of islands, the views and understanding of working memory in the sentence processing literature have changed. The idea of working memory having a capacity limit on a common pool of resources that both storage and processes must draw from has fallen out of favor. In its place is a view of working memory as a system that uses a content-addressable retrieval process to retrieve items/words from recent memory (e.g. Gordon, Hendrik and Johnson 2001; Gordon, Hendrick and Levine 2002; Lewis and Vasishth 2005; Gordon et al. 2006; Lewis, Vasishth and Van Dyke 2006; Van Dyke and McElree 2006). That is, there is no specific cost for storage, since there is no ‘active’ storage. The retrieval system is, however, susceptible to interference from items/words in memory that have similar features to the target being retrieved. In the case of a filler-gap dependency, the filler is not actively stored, but is instead retrieved when a cue that it is needed (the gap) is encountered. If there is something in recent memory that overlaps in features with that filler (such as the interrogative *whether*, having a [+wh] feature, just like the [+wh] filler *who*), then similarity-based interference is predicted and the retrieval process is rendered more difficult. Unfortunately, our understanding of what features are relevant for similarity-based interference is still in its infancy. To date, similarity-based interference has not been

explicitly proposed as an explanation for island phenomena in the literature, even though it appears that such an account would be plausible. I present this plausible account as an alternative to the capacity-constrained account of islands.

In what I will call the similarity-interference account of islands, the main difficulty in processing is located not at the clause boundary, as in the capacity-constrained account, but at the gap site. Since there is no active storage cost in the similarity-interference view of working memory, when the clause boundary is encountered, only the inherent processing difficulty of the clause boundary should be observed (not an overload of capacity as in the capacity-constrained account). However, when the gap position is encountered, the retrieval process is cued to retrieve a filler with certain features. If the island structure overlaps with the filler in some of those features, the retrieval process should be more difficult in these cases compared to non-overlapping controls (i.e. a declarative clause boundary with the complementizer *that*). This extra difficulty should be observed at or after the embedded gap position. Thus, while these two accounts of islands are similar in that they both claim that difficulties in processing result in sentences being deemed unacceptable, they differ in the locus at which they predict those difficulties to occur, and what the underlying processes responsible for those difficulties are.

1.3 Experimental approach

Three different experimental methodologies, acceptability judgments (Chapter 4), self-paced reading (Chapter 5) and event-related potentials (Chapter 6), were used

to examine *whether*-islands and closely related control sentences. *Whether*-islands were chosen, in part, because they allow for a balanced factorial design including both matrix clause and embedded clause gaps (Chapter 3). Using the same sentence types across experiments allowed for more direct comparisons to be made across these various methodologies. Additionally, an individual differences approach was adopted, in which participants were tested on a number of cognitive measures and the linguistic data they provided (whether acceptability ratings, reading times or elicited brain responses) were checked for co-variation with those cognitive measures. This provided another commonality and point of comparison across different methodologies.

The cognitive measures were chosen in an attempt to tap into the various cognitive skills assumed by the capacity-constrained and similarity-interference views of working memory. Examination of these data and of the locus of processing difficulty in the sentence (i.e. focused on the clause boundary or the embedded gap) was designed to help decide between these two accounts of island phenomena.

Ultimately, neither account was fully supported by the data presented here (though the capacity-constrained account finds partial support from the self-paced reading results, Chapter 5). To foreshadow the conclusions of the dissertation, I present the gap predictability account of processing islands in (1.5), annotated with the sections that discuss each part in detail.

- (1.5) The gap predictability account of processing islands
- a) If there is an unresolved filler-gap dependency in a sentence, upon encountering an island boundary (5.4.2.2.1), the parser revises its predictions that a gap will be forthcoming (6.4.2.2.4).
 - b) High span readers are better able to revise/modulate this prediction (6.4.3.2.2).
 - c) If evidence for a gap is encountered within an island, it is straightforwardly identified (6.4.2.2.2) and associated (6.4.2.2.1) with a filler.
 - d) Neither of these processes (b or c) directly influences the acceptability ratings assigned to an island violation (4.4.1; 6.4.1.2).

As can be seen from (1.5), islands differ from non-islands in how predictable a gap is within an island. While there is reading-time evidence of processing difficulty at the clause boundary, there is no brain response evidence of a failed parse or reanalysis in the island violations. No difference is observed for the process/cost of filler-gap association, which occurs in both island and non-island clauses. Even so, island violations are rated as unacceptable. This dissociation between on-line processing cost patterns and acceptability judgment patterns makes it unlikely that a processing account of island phenomena is a viable explanation for their unacceptability. However, this does not allow us to conclude that a grammatical account for islands (Chapter 2, section 2.3.1) is necessarily preferred. While grammatical accounts do not typically make predictions about processing data, one must expect, barring clear

evidence to the contrary, that the brain would be aware of such constraints and show a response when they are violated. The lack of such an effect is just as problematic for the grammatical accounts of islands as it is for the processing accounts.

1.4 Dissertation overview

The organization of this dissertation is as follows:

Chapter 2 presents background information relevant to the rest of the dissertation. The basic island data are presented, including prior experimental work. An overview of event-related potential (ERP) components is provided in preparation for the ERP experiment in Chapter 6. Accounts of island phenomena are reviewed, with a focus on processing accounts.

Chapter 3 presents the methodologies that are common to all three experiments in the dissertation. The design of the materials used throughout the dissertation is explained. Each of four cognitive measures (reading span, n-back, flanker and memory interference) is explained. The results of these cognitive measures are compared across the participants of each experiment, and a co-variation matrix for these measures is provided.

Chapter 4 presents the first of three experiments that examine linguistic data for co-variation with cognitive measures. Experiment 1 is an acceptability judgement study. The chapter first reviews a similar study done by Sprouse, Wagers and Phillips (2012), highlighting the advantages that Experiment 1 has over that study. Two frameworks acting as conceptual aids are presented and explained, namely the

Cognitive Co-variation Intuition (CCI), and the Processing Benefits Schedule (PBS). Through multiple analyses and guided by the frameworks above, the acceptability judgment chapter concludes in agreement with Sprouse, Wagers and Phillips' (2012) assessment that these results do not support a capacity-constrained account of islands. However, because of the advantages of the current study, we are better able to understand the complexities of looking for co-variation of cognitive measures with acceptability judgments.

Chapter 5 presents a self-paced reading study using the same materials used in Chapter 4. The basic findings (not including cognitive measures) appear to support the capacity-constrained account of islands, as the processing difficulty occurs at the clause boundary. However, the picture becomes more complex when cognitive measures are considered. Low span readers show a graded pattern of difficulty, while high span readers show processing difficulty only for the island violation condition.

Chapter 6 presents the final experiment of the dissertation, an ERP examination of the same types of sentences from Chapters 4 and 5. The ERP data suggest that readers identify and fill gaps embedded in *whether*-islands just as readily as they do for control sentences. The only difference appears in the modulation of gap predictability inside the island. Gaps are less predicted in an island domain and when evidence for a gap is encountered there, an N400 response is elicited. This effect is significant in high but not low span readers, and does not influence the parser's ability to associate the filler and gap, nor does it influence the pattern of acceptability judgments given to these sentences. Additionally, the results reported here raise

questions about the standard interpretation of pre-gap P600 effects and sustained LAN effects.

Finally, Chapter 7 concludes the dissertation. This chapter summarizes the key findings from the prior chapters and discusses how their results combine to further our understanding of island phenomena and working memory.

Chapter 2: Background

2.1 Introduction

In this chapter I present a brief overview of island phenomena (section 2.2). I review the processing findings of both behavioral (section 2.2.5.1) and electrophysiological (section 2.2.5.2) studies that will be relevant to the experiments of this dissertation. Section 2.3 presents a sketch of different accounts of island phenomena, with a focus on the capacity-constrained (section 2.3.3.1) and similarity-interference (section 2.3.3.2) accounts that are the focus of inquiry here. Finally, section 2.4 concludes with a summary of the current research agenda.

2.2 Island phenomena

Throughout the dissertation I will use the term *island* to indicate a certain structure type, and *island violation* to indicate when the addition of a filler-gap dependency to that structure results in it being deemed unacceptable. Specifically, an *island violation* occurs when a filler occurs outside an island domain and the gap associated with that filler occurs inside the island domain. An asterisk (*) before an example sentence indicates that the sentence is judged as unacceptable. A question mark (?) before an example sentence indicates that the sentence is judged as somewhat/borderline unacceptable.

2.2.1 The basic data

Long-distance filler-gap dependencies have long been a subject of inquiry among linguists. Comparing the declarative statement in (2.1) to the related question in (2.2), we observe that the object of *see* has been replaced with an interrogative *who* which no longer occurs in its canonical object position.

(2.1) John will see Mary.

(2.2) Who will John see _?

In sentence (2.2) the word *who* (referred to as a *filler*) forms a long-distance dependency with its empty canonical position (referred to as a *gap*). It is crucial that the gap position be empty, otherwise the result is a sentence with an ungrammatical ‘filled gap,’ shown in (2.3).

(2.3) * Who will John see Mary?

The nature of the dependency between filler and gap is highlighted here. If the gap is filled, the dependency fails. Similarly, the filler cannot be removed as in (2.4) and have the meaning intended in (2.2; the lack of intended reading is marked by the ‘#’).

(2.4) # Will John see _?

In (2.5) we can see that a filler-gap dependency can potentially be arbitrarily long and cross many clause boundaries.

(2.5) Who will John report [that he thought [that he saw _?]]

Early in the examination of these types of long-distance dependencies, Ross (1967) reported a number of structures that disrupted their acceptability. Ross termed these structures *islands*, and the number of islands reported in the literature has increased since his original findings were presented. As shown in the example of a *whether* island in (2.6) below, if the filler is not displaced, but instead remains *in situ* as is appropriate for a so-called echo question, then there is no apparent unacceptability.

(2.6 a) Bill wondered whether John saw who?

(2.6 b) * Who did Bill wonder whether John saw _?

This indicates that the cause of the unacceptability in the island examples is not purely semantic or interpretational, but crucially involves the syntactic displacement present in (2.6 b).

2.2.2 Ameliorating effects on island violations

The basic pattern that fillers outside of islands cannot be associated with gaps inside of islands has a number of exceptions. These exceptions indicate that the

patterns of islands are neither absolute nor purely syntactic. For example, if the semantics of the filler are altered to be more specific (alternatively characterized as individuated, Szabolcsi and Zwartz 1990, 1993; or d(iscourse)-linked, Pesetsky 1987) as in (2.7), the effect of the island is ameliorated.

(2.7) ? Which man did Bill wonder whether John saw _? >

(2.6 b) * Who did Bill wonder whether John saw _?

The difference between (2.6 b) and (2.7) is subtle, and it appears that not all native speakers make this distinction (Michel 2010). The core intuition is that *which man* invokes a more specific, definite, or limited set of possible referents, namely those who are both a human and an adult male, while *who* invokes a less well-defined set of referents, namely those who are a human.

Pesetsky (1987) suggested that *which-N* is discourse-linked (d-linked) while other *wh*-phrases are not. He proposed that for the question *Which book did you read?* the set of possible felicitous answers is limited to the set of books present in the common ground of both the speaker and hearer. This is not the case for *What did you read?*, for which the set of possible felicitous answers is limited only by those things that can be read. Similar to the notion of d-linking is that of a *wh*-phrase being ‘referential’ (Cinque 1990; Chung 1994). Cinque (1990) defines referentiality as a quality held by arguments that are either currently in the discourse or “refer to specific members of a set in the mind of the speaker (pg 16).” With either characterization, the

common point of interest for current purposes is that neither d-linking nor referentiality can be defined in purely syntactic terms.

While recent work has reported sensitivity to the d-linking effect (Hofmeister and Sag 2010), other recent work has either not found the effect (Kim 2010) or has found it to be robust only in certain populations (Michel 2010). Michel (2010) found that high working memory capacity individuals were sensitive to the difference between sentences like (2.6 b) and (2.7) while low working memory capacity individuals were not. This type of individual differences approach will be pursued throughout this dissertation.

Another factor claimed to have an ameliorating effect on island phenomena is the notion of finiteness. Finiteness is claimed to act as an ‘overlay,’ strengthening islands (Szabolcsi and Zwartz 1990). Compare the non-finite examples in (2.8 a) with the finite versions in (2.8 b). The island effect in (2.8 a) is greatly reduced if not eliminated entirely.

(2.8 a) {Who/ which man} did Bill wonder whether to invite _? >

(2.8 b) {*Who/ ?which man} did Bill wonder whether John invited _?

Note that the finiteness manipulation between (2.8 a) and (2.8 b) actually involves two differences: the finiteness of the verb and the presence of an additional noun phrase referent (*John* in 2.8 b). Michel and Goodall (2013) separated out these effects in a series of acceptability judgment manipulations. They found that the

finiteness of the verb itself resulted in lower acceptability only in island violations, but the effect of an additional noun phrase was more global, lowering the acceptability of the sentence in both islands and non-islands. Michel and Goodall interpreted the finiteness effect itself as more compatible with a grammatical view of island phenomena (section 2.3.1), since the effect was limited to island structures. On the other hand, the additional referent was interpreted as being more compatible with a processing view of islands (section 2.3.3), since the effect was found to contribute to islands as well as non-islands. Thus, it may be that both grammatical and processing factors are implicated in island phenomena (see section 2.3 for more discussion of accounts of island phenomena).

2.2.3 Experimental syntax

Experimental syntax, or an experimental acceptability judgment task, is the systematic gathering of acceptability ratings for sentences of interest (e.g. Cowart 1997; Snyder 2000; Sprouse 2007). These types of studies aim to quantify sentences more finely than the grammatical / ungrammatical distinction often presented in the syntactic literature, and based on introspective judgments and self-report by a limited number of native speakers. There are a number of measurements used in such acceptability experiments, such as binary yes/no responses, Likert scales, and magnitude estimation. The consensus thus far is that these different response measures all produce generally similar patterns of results (Bader & Häussler 2010; Fukuda et al. 2012). However, subjects do not follow necessarily the assumptions underlying

magnitude estimation (Sprouse 2011) and magnitude estimation introduces additional spurious variance (Wescott & Fanselow 2011). therefore all acceptability judgments in the current study are done using 7-point Likert scales (see Chapter 4, section 4.3.3.2).

Syntacticians have recognized the need for a way to indicate more gradient judgments in their research, and have used a variety of symbols (?, #, *?) for these marginal sentences. While this is an improvement over strict grammaticality / ungrammaticality, it still does not capture the range of possible differences between sentences that an acceptability experiment can capture. By having a number of participants rate sentences for acceptability, an acceptability score can be generated. Rather than relying on the notion of grammaticality, it is often more useful to refer to sentences as more or less acceptable (this is the approach taken in this dissertation). This is in part due to the recognition that factors other than the status of a sentence being grammatical in a person's grammar come into play when that person is tasked with rating that sentence.

Factors such as processing difficulty are known to influence acceptability ratings given to different sentences, even when both sentences would be considered fully grammatical in syntactic theory. For example, a much replicated finding in the sentence processing literature is that object relative clauses in sentences like (2.9 a) are more difficult to process than subject relative clauses in (2.9 b) (King & Just 1991).

(2.9 a) Object relative:

The reporter who the senator harshly attacked _ admitted the error.

(2.9 b) Subject relative:

The reporter who _ harshly attacked the senator admitted the error.

Modified from King and Just (1991; 581)

While both (2.9 a) and (2.9 b) are grammatical in English, sentences like (2.9 a) are nonetheless rated as less acceptable than sentences like (2.9 b) in experimental studies (e.g. Keffala 2013). Thus we see how processing costs can modulate acceptability scores.

While acceptability judgment tasks have indicated effects of processing difficulty in general, there have been differing results with respect to whether this methodology is sensitive to measures of individual cognitive differences. Michel (2010) reported an interaction of working memory scores and acceptability scores in the rating of d-linked sentences. However, Sprouse, Wagers and Phillips (2012) failed to find robust interactions with island phenomena. This study is discussed in detail in Chapter 4, section 4.2. Hofmeister, Staum-Casasanto and Sag (2014) grapple with the issue of individual variation in acceptability judgments, and in Chapter 4, I lay out a number of frameworks in order to advance the discussion of this complex topic (the Cognitive co-variation Intuition (CCI), section 4.2.2; the Processing Benefit Schedule (PBS), section 4.2.2.1; and rating task differences, section 4.2.2.2).

It is worth noting at this point that despite their status as a ‘weak’ island, and the intuition that they are easily ameliorated compared to other islands, acceptability experiments on *whether*-islands have clearly and consistently found the *whether*-island violations to be judged as the least acceptable compared to relevant controls (Sprouse, Wagers and Phillips 2012; this dissertation Chapter 4, Chapter 6).

2.2.4 Satiation

Repeated exposure to unacceptable sentences has been anecdotally reported among syntacticians for some time.¹ Snyder (2000) reported the first experimental results showing syntactic satiation, using an acceptability judgment paradigm. A number of satiation studies have been undertaken since then, many focusing on island phenomena (e.g. Snyder 2000, Hiramatsu 2000, Francom 2009, Sprouse 2009, Goodall 2011, Crawford 2012). Results are inconsistent between studies, with certain sentence types showing satiation patterns in some studies but not others. *Whether*-islands are one of the more consistent structures investigated however, showing a satiation pattern in most studies (but not Sprouse 2009).

This is of concern to the current set of experiments as participants in the ERP experiment (Chapter 6) were exposed to 40 examples of each sentence type. It could thus be that the ERP results reflect (at least partially) the responses to structures that participants have satiated on. That is, if participants satiate on the *whether*-island violation sentences then the participants may no longer make a clear acceptability

¹ To add to the anecdotes, I have had enough exposure to the unacceptable sentences reported on in this dissertation that they feel only slightly (if at all) degraded to me.

distinction between these sentences and the control sentences. In order to test for this possibility, participants in the ERP study completed an acceptability judgment after the ERP session (see Chapter 6, section 6.4.1). These results did not differ substantially² from the results of the full acceptability study (Chapter 4), indicating that such concerns are unwarranted in the current dataset.

2.2.5 On-line processing

Both behavioral measures and brain measures have been used to study the processing of filler-gap dependencies and island phenomena. Behavioral measures are most frequently realized as reading time studies, either self-paced or using eye-tracking. Other behavioral measures include sentence matching (e.g. Freedman & Forster 1985) and cross modal priming (e.g. Nicol & Swinney 1989). Behavioral measures are characterized by the dependent measure being a physical response that the participant makes, such as a button press or eye fixation. Brain response measures, on the other hand, require no task other than reading the sentences (though some task is often included to engage the participant). The dependent measure here is not a physical action that the participant does, but rather the response of the brain to the stimulus. In the case of event-related potentials (ERP), this is an electrical signal measured at the scalp. This signal comes from the summed activity of post-synaptic potentials from a large population of synchronously firing pyramidal neurons in the cortex (e.g. Peterson et al. 1995). In the following sections I review behavioral and

² An interaction that was significant in the full acceptability study was only marginal in the short acceptability study, see section 6.4.1.2 for discussion.

ERP findings relevant to this dissertation and introduce the ERP components that will be discussed in Chapter 6.

2.2.5.1 Behavioral data (reading times)

Behavioral reading times are measured either in a self-paced reading experiment, where a participant advances through the sentence incrementally via a button press, or an eye-tracking experiment, where a participant can read freely, but the fixations of their eyes are recorded and timed. In both case, slower reading times are widely accepted to indicate processing difficulty compared to a control condition. Inferences about what these difficulties represent depend on the linguistic manipulation and experimental design.

The on-line processing of filler-gap dependencies (Fodor 1978) is largely thought to involve some version of the Active Filler Strategy (Frazier & Clifton 1989; 2.10) rather than a ‘last resort’ strategy (e.g. Jackendoff & Culicover 1971; Wanner & Maratsos 1978). That is, the parser does not wait for all elements of a sentence to be encountered before attempting to assign a filler to a gap.

(2.10) Active Filler Strategy (AFS)

When a filler has been identified, rank the option of assigning it to a gap above all other options.

(Frazier & Clifton 1989, pg 95)

Evidence for this strategy comes from the examination of sentences like (2.11), in which a gap could be interpreted at either position (1) or (2).

(2.11) Who did Fred tell (_1_) Mary (_2_) left the country?

Frazier and Clifton (1989) reported that the preferred reading of (2.11) was the one where Fred told someone (*who*) that Mary left the country, consistent with a gap at position (1). The reading where Fred told Mary that someone else (*who*) left the country, consistent with a gap at position (2), was less preferred.

On-line evidence largely comes from the filled-gap effect (e.g. Crain & Fodor 1985; Stowe 1986; Frazier & Clifton 1989; Bourdages 1992; Pickering, Barton & Shillcock 1994; Boland, Tanenhaus, Garnsey & Carlson 1995) and plausibility manipulations (e.g. Traxler & Pickering 1996, Phillips 2006). Other methods, such as visual-world paradigm (Sussman & Sedivy 2003), and cross-modal priming (i.e. trace-reactivation: Bever & McElree 1988; Nicol & Swinney 1989; MacDonald 1989) have also provided converging evidence for an Active Filler Strategy. The filled-gap effect occurs in an unresolved filler-gap dependency and is measured by a slowdown in reading times at a position where a gap could have been located, but is instead ‘filled’ with some other lexical item. In (2.12 a) *us* was read more slowly when there is a filler *who* that could associate with the gap position (the object of *bring*) that *us* is ‘filling.’

(2.12 a) My brother wanted to know

who Ruth will bring **us** home to _ at Christmas.

(2.12 b) My brother wanted to know

if Ruth will bring **us** home to at Christmas.

Modified from Stowe 1986

The actual gap position for *who* occurs a few words later, but the parser attempts to associate the filler with the earlier possible gap. Since these effects occur immediately at a possible gap position, and do not wait until other viable gap positions occur, they are taken as evidence for the parser trying to assign the filler “to a gap above all other options” (Frazier & Clifton 1989).

The plausibility manipulation is similar, but instead of having the gap position be filled, the filler is paired with an implausible verb to associate with, again resulting in slower reading times. In (2.13 b) for example, readers immediately slowed down upon reading *shot* when *the garage* was an implausible antecedent (compared to *the pistol*).

(2.13 a) That’s the pistol

with which the heartless killer **shot** the man yesterday afternoon.

(2.13 b) That’s the garage

with which the heartless killer **shot** the man yesterday afternoon.

Modified from Traxler & Pickering 1996

Crucially, these filled-gap effects and plausibility manipulation effects were not found when the potential gap site was inside an island domain. Stowe (1986) did not obtain a filled-gap effect inside subject islands. Traxler and Pickering (1996) did not obtain a plausibility effect inside subject islands.

These findings have led to “[t]he prevailing opinion in psycholinguistics [being] that the evidence supports the position that island constraints are immediately effective in parsing, and that contrary findings may be due to flaws in experimentation” (Phillips 2006, pg 800)³. These contrary findings include the aforementioned sentence matching task of Freedman and Forster (1985), criticized on methodological grounds by Crain and Fodor (1987) and Stowe (1992). Thus far, experiments using event-related potentials have not conclusively weighed in on the issue of the immediate application of island constraints.

ERP experiments have demonstrated that the brain is sensitive to island boundaries (i.e. N400 response of Kluender and Kutas 1993b; P600 response of McKinnon and Osterhout 1996, but see discussion below), but due to how the materials were designed and what comparisons could be made, these experiments have not been informative as to whether the brain response also indicates that gaps are not posited within islands. “[S]ince they indicate only that the start of the island domain is detected, they do not provide clear information on whether gaps are posited at potential gap sites inside islands” (Phillips 2006, pg 800). In the current ERP study (Chapter 6), comparisons at the gap are possible. The results of Experiment 3 (Chapter

³ Although, in the same study, Phillips presents evidence that the parser does posit gaps in subject islands if such a gap would be allowed in parasitic gap constructions (Phillips 2006).

6) indicated that while readers (especially high span readers) did not appear to predict a gap inside the island, they still associated a filler with a gap inside that island. The next section lays the foundation for interpreting those results.

2.2.5.2 ERPs

The recording of event related brain potentials, or ERPs, to linguistic stimuli has a number of advantages over the behavioral measure of reading times.⁴ For example it is not necessary for participants to respond to each word (as in self-paced reading). Nor is it even necessary to have a specific task for the participants except to read (or listen) passively. Rather than generate a single (in the case of self-paced reading) or small set of (in the case of eye-tracking) reading time measure(s) for each word, ERPs allow the researcher to examine the time-course of reactions in more detail. As the ERPs are not dependent on a specific participant response (i.e. a button press), but instead unfold in real time, we are able to examine effects at different latencies time-locked to the same stimulus. For example, a word might elicit an earlier N400 response followed by a later P600 response in one condition, but only an N400 response in another. This allows for more specific inferences to be drawn than, for

⁴ On the other hand, there are at least two (related) major disadvantages to using ERPs rather than reading times. First, a much larger time commitment is needed for the experiment, both in terms of (i) participant screening, as certain participant profiles are not ideal ERP participants unless such participants are being specifically studied (i.e. individuals with a history of head trauma, or on medication designed to alter brain chemistry) and (ii) individual sessions, which require a more complex setup and more experimental materials. The need for more experimental materials arises because the electrical activity recorded at the scalp is a very small signal. In order to obtain a favorable signal-to-noise ratio, many (usually around 40 for syntactic manipulations) trials of each condition must be recorded. This limits the number of conditions that can be tested while keeping the overall experimental running time manageable. This also raises concerns of syntactic satiation (section 2.2.4), for Experiment 3, but this did not turn out to be an issue (Chapter 6, section 6.4.2.2.5).

example, the measuring of reading times being slower in one condition than another. The ERP responses vary not only in timing, but also amplitude, polarity and distribution across the scalp. A certain number of combinations of these characteristics have become considered ‘components’ in the literature, meaning that they are generally reliable responses to certain kinds of stimuli (though this does not mean that the interpretations of these responses are a settled matter). The more specific inferences one can draw from ERPs depends, in part, on a solid understanding of what kinds of stimuli these components are elicited by, both in the language domain and in other cognitive domains. I discuss the components most often associated with language in the following sections: the LAN (section 2.2.5.2.1; including the sustained left-anterior negativity and eLAN), P600 (section 2.2.5.2.2) and N400 (section 2.2.5.2.3).

The LAN and P600 components have both been previously reported in studies of filler-gap dependencies, though due to differences in experimental stimuli design or other factors, it is not uncommon that a study can only report an effect of either a LAN or a P600. A recent unpublished meta-analysis of these studies (p.c. Chris Barkley) indicates a consistent pattern across studies when looking at the second item in a filler-gap dependency. This second item is usually a gap (as in the current experiments, so I will continue to refer to this second element as a gap for consistency and ease of exposition; see Chapter 3, section 3.2), but can also be a filler in languages where the gap position can be encountered first (e.g. Japanese; Ueno & Garnsey 2008), or the subcategorizing verb if the gap is separated from it (e.g. Gouvea et al. 2010). The

generalization is that a P600 is elicited in the pre-gap position (Kaan et al. 2000; Fiebach et al. 2002; Phillips et al. 2005 and Gouvea et al. 2010) and a LAN is elicited in the post-gap position (Kluender & Kutas 1993a,b; King & Kutas 1995; Fiebach et al. 2002; Felser et al. 2003; Ueno & Kluender 2003; Phillips et al. 2005; Kwon 2008; Ueno & Garnsey 2008). Again, the apparent asymmetry in the literature in that some of these studies report a pre-gap P600 and some report a post-gap LAN is due to differences in experimental design. Differences in materials lead to differences in which sentence positions the researchers have been able to reliably measure. The current experimental materials are designed such that both the pre- and post-gap positions are lexically matched, allowing for the potential to measure both the P600 and LAN effects (see Chapter 3, section 3.2 for materials design). Both effects are found in Experiment 3 (Chapter 6).

2.2.5.2.1 LAN

The Left Anterior Negativity (LAN) is a negative going deflection starting around 300 msec post-stimulus onset, originally reported with a left anterior scalp distribution (Kluender & Kutas 1993a,b). Some subsequent studies have reported a bilateral, but still anterior scalp distribution (Fiebach et al. 2002; King & Kutas 1995; Phillips et al. 2005). LANs are often considered to be either phasic (i.e. non-continuing) or sustained. Additionally, the early Left Anterior Negativity (eLAN) is discussed briefly below.

Three ‘flavors’ of the phasic LAN have been reported in the literature: a morphosyntactic violation LAN, a definiteness LAN and a working memory LAN. Phasic LANs have been elicited by various morphosyntactic violations, frequently preceding a P600 response, such as verbal agreement (e.g. Kutas & Hillyard 1983; Münte et al. 1993), case violations (e.g. Coulson et al. 1998a) and phrase structure violations (e.g. Münte et al. 1993; Neville et al. 1991). Additionally, increased LAN responses have been reported to nouns that follow a definite (compared to indefinite) determiner (Anderson & Holcomb 2005; possibly related to the ‘Nref’ e.g. Van Berkum et al. 2007; Barkley, Kluender & Kutas 2011). This ‘definiteness LAN’ will be relevant for certain lexical differences in the current materials (Chapter 6, section 6.4.2.2.2). However, it is the sensitivity of the LAN to working memory processes that is most important to the current research.

In filler-gap dependencies, LANs have been reported following both the filler and following the gap. As the experiments in this dissertation focus on the gap positions rather than the fillers, the post-gap LAN is discussed first. Kluender and Kutas (1993a,b) reported the first post-gap LAN effects in sentences like those in (2.14 b).

(2.14 a) Can't you tell

[if she intends to drum this stuff **into** you by the end of the quarter?]

(2.14 b) Can't you tell

[what she intends to drum __ **into** you by the end of the quarter?]

Modified from Kluender & Kutas (1993a, Figure 5)

In (2.14 b), at the post-gap position (*into*), Kluender and Kutas (1993a) reported a LAN compared to the same lexical item when it doesn't follow a gap (2.14 a). They interpreted this LAN as reflecting the retrieval of the filler (*what*) from working memory (see section 2.3.3.1 below) so that the filler and gap can be integrated. Numerous other studies have adopted this working memory retrieval/integration view of the post-gap LAN (e.g. King & Kutas 1995; Müller et al. 1997; Weckerly & Kutas 1999; Matzke et al. 2002; Felser et al. 2003; Ueno & Garnsey 2008; Kwon et al. 2013). As mentioned above, A LAN is elicited after both the filler and the gap. Kluender and Kutas interpreted these LANs as a unified working memory process involved in "the storage of a filler in working memory and its subsequent retrieval" (Kluender and Kutas 1993a, pg 205)

The first post-filler LAN was originally reported as a phasic effect as the sentence materials prohibited measuring a longer epoch (Kluender & Kutas 1993a,b) but now the post-gap LAN is often characterized as continuing with a sustained anterior negativity (e.g., King & Kutas 1995; Fiebach et al. 2002; Phillips et al. 2005). That is, the initial negativity difference starting 300 msec post-stimulus onset

continues throughout the epoch and into the following words. This post-filler LAN has been associated with entering the filler into working memory, and the sustained negativity following it has been associated with the active maintenance cost of holding that filler in memory (e.g. Kluender & Kutas 1993a,b; King & Kutas 1995; Fiebach et al. 2002; Phillips et al. 2005)

The amplitude of this post-filler sustained anterior negativity has been reported to co-vary with working memory scores (Fiebach et al. 2002) and reading comprehension scores (King & Kutas 1995). A sustained anterior negativity in non-filler-gap sentences has also been reported to co-vary with working memory scores (Münte et al. 1998, see 2.15 below). The direction of co-variations with individual differences is not consistent between the studies, however. Fiebach et al. (2002) report that “the sustained negativity was stronger, more broadly distributed, and present earlier for individuals with low working memory capacity” (pg. 262). However, in Münte et al. (1998) and King & Kutas (1995), it is the high scoring group that shows the larger effect (this type of potentially counter-intuitive pattern is discussed in more detail in Chapter 4, section 4.2.2.1).

Results have also varied as to whether this sustained anterior negativity amplitude increases over time. Fiebach et al. (2002) reported an increase over time in German filler-gap dependencies, which they interpreted as reflecting the ongoing cost of storing the filler as more sentential material is encountered. This cumulative storage cost is reminiscent of Gibson’s Dependency Locality Theory (e.g. Gibson 2000). However, other researchers did not find this cumulative pattern in the sustained

anterior negativity (King and Kutas 1995; Phillips et al. 2005). In particular, Phillips et al. (2005) presented a clear demonstration in English that the sustained nature of the negativity following the filler disappeared when subsequent words were re-baselined. Furthermore, some studies have not been able to detect this ongoing negativity at all.

McKinnon & Osterhout (1996) looked for, and failed to find, the sustained negativity. Kaan et al. (2000) noted a possible phasic response, but reported that it was not statistically significant. Both of these studies used a heavier, d-linked filler (*which of his staff members* and *which popstar*, in the respective example sentences). It could be that these multi-word fillers either reduce the sustained negativity, or spread its effect across the multiple words of the filler, thus making it more difficult to detect and measure. In comparison, both King and Kutas (1995) and Fiebach et al. (2002) observed the sustained LAN following the bare filler *who*. However, the materials in Kaan et al. (2000) also contained a condition with *who*, and this condition patterned similarly to *which popstar* (a non-significant phasic, but not sustained LAN). Phillips et al. (2005) also used a d-linked filler, *which accomplice*, and reported a sustained negativity⁵. To further illustrate that it is not clear exactly under which conditions a sustained negativity will be elicited, Münte et al. (1998) reported a sustained LAN in sentences with no filler-gap dependencies (2.15).

⁵ The ‘weight of the filler’ hypothesis sketched here still seems like a plausible explanation for why McKinnon & Osterhout (1996) do not detect a LAN. The lighter, but still d-linked fillers used by Kaan et al. (2000) and Phillips et al. (2005) patterned more closely to the bare filler *who* (especially evident in Kaan et al. 2000, where both fillers are used). The outstanding question then is why Kaan et al. (2000) don’t observe a sustained negativity.

(2.15 a) After the scientist submitted the paper,
the journal changed its policy.

(2.15 b) Before the scientist submitted the paper,
the journal changed its policy.

Modified from Münte et al. (1998)

Münte et al. (1998) reported a sustained LAN in the *before* sentences, where word order does not match the temporal order described. They interpreted these results as readers having an increased memory load when the clause order must be rearranged to match temporal order. That is, the *before* clause creates a working memory burden. This negativity was again greater in the high span participants than in low span participants, matching the pattern reported above for King & Kutas (1995), although not in a filler-gap dependency.

The current experimental materials were designed to highlight the gap position, rather than the filler position, so we are not able to report on a sustained LAN effect here. Thus on two issues for which there are conflicting data in the literature, namely (i) the potential increase of the negativity over time and (ii) the potential pattern that the negativity has with respect to individual differences, we can make no direct comment based on the results of the current ERP experiment. However, we did find that the post-gap LAN is a sustained effect in the current experiment. There does not seem to be a reason for this that would invoke memory load, as has been provided for the sustained LAN following the filler. The presence of a sustained post-gap LAN

calls into question the interpretation of the sustained post-filler LAN as an index of storage cost for the filler. This is discussed in Chapter 6, section 6.4.2.2.5.

Finally, a response known as the ‘early Left Anterior Negativity’ (eLAN) appears with the same scalp distribution (left anterior) as the LAN, but much earlier, onsetting in a 100-300 msec window. This response was originally thought to be elicited by word-category violations (Neville et al. 1991). In Friederici’s (2002) model of sentence processing, much theoretical weight has been balanced on the eLAN as a representation of modular syntax-first parsing. However, recent findings have greatly undermined this association, instead supporting a theory where the eLAN reflects early form-based processing of a word based on predictions made by the parser (e.g. Lau et al. 2006; Rosenfelt et al. 2009; Steinhauer & Drury 2012). This response is not apparent in the current experiment’s data and so is not discussed further here. However, it is noteworthy to highlight the eLAN response as an example of the parser making predictions about upcoming words. The importance of predictions for ERP responses is relevant for the discussion of the N400 effect discussed in Chapter 6, section 6.4.2.2.7

2.2.5.2.2 P600

The P600 (Osterhout & Holcomb 1992) is so named because it is a positive-going deflection often peaking around 600 msec post-stimulus onset, but unlike the N400 described below, its latency can vary considerably between experimental manipulations. This component is often broadly elicited across the scalp, but has

frequently been reported with a centro-posterior distribution. The component has also been referred to in the literature as the ‘late positive component’ and the ‘syntactic positive shift’ (LPC and SPS, respectively; Hagoort et al. 1993). This latter label was invoked because whereas the N400 had been associated with semantic aspects of language, the P600 originally appeared to be elicited by syntactic violations.

Early P600 research indicated that a P600 could be elicited by (syntactic) phrase structure and subcategorization violations (Neville et al. 1991; Osterhout & Holcomb 1992; Hagoort et al. 1993; Osterhout et al 1994) as well as (morpho-syntactic) agreement violations (Hagoort et al. 1993; Osterhout & Mobley 1995; Münte et al. 1997; Coulson et al. 1998a; Hagoort & Brown 1999). Additionally, it was reported that the P600 was elicited by violations of syntactic movement constraints (Neville et al. 1991; McKinnon & Osterhout 1996). These movement constraints are of particular interest with respect to the present study, and so they will be examined in more detail.

McKinnon and Osterhout (1996) reported on two linguistic manipulations involving a filler and gap. However, they did not measure near the gap site in either case. Because their results occurred in intermediate positions, it is unclear to what extent the effects should be directly attributed to the presence of a filler-gap dependency in the sentence rather than other ungrammatical structures that arise within the sentences. The first comparison was termed a ‘subjacency violation’ by McKinnon and Osterhout, and represents a *wh*-island in (2.16 b) compared to an embedded interrogative clause control in (2.16 a).

(2.16) Subjacency violation

- a. I wonder whether the candidate was annoyed

when his son was questioned by his staff member.

- b. * I wonder which of his staff members the candidate was annoyed

when his son was questioned by _.

Modified from McKinnon & Osterhout (1996, Table 1)

A P600 was reported at the island boundary (*when*) in (2.16 b). However, the fact that *when* serves as an island boundary is not the only difference between (2.16 a) and (b). In the (b) sentence, the presence of *when* also preempts a preferred gap position for the filler *which of his staff members*, as the indirect object of *annoyed* (i.e. *annoyed with* _). Thus, at *when*, readers may be attempting a reanalysis of a garden path or attempting to repair a violation of an expected subcategorization frame. Both garden path sentences (Osterhout & Holcomb 1992; Osterhout et al. 1994) and subcategorization violations (e.g. Neville et al. 1991; Osterhout & Holcomb 1992; Hagoort et al. 1993; Osterhout et al. 1994) have been reported to elicit P600s. So there is no need to consider this effect as directly due to a subjacency violation. Note that this effect was measured neither at the filler nor at the gap, which is also true of the ‘ECP violation’ reported by McKinnon & Osterhout (1996). McKinnon & Osterhout reported a P600 effect, measured at *that*, in the ungrammatical (2.17 b) compared to (2.17 a).

(2.17) ECP violation

- a. It seems that it is likely that the man will win
- b. * The man seems that it is likely _ to win

Modified from McKinnon & Osterhout (1996, Table 2)

The Empty-Category Principle (ECP) (e.g. Chomsky 1981, 1986; Huang 1982; Lasnik & Saito 1992) states that a trace (i.e. a gap) must be properly governed (in this case by its filler). The presence of *that* in (2.17 b) prevents this proper governing relationship and the sentence is ungrammatical. Note that the ‘base generated’⁶ sentence (without a displaced *the man*) is also ungrammatical: *It seems that it is likely *the man to win*. Thus the comparison in (2.17) does not represent a minimal pair. Considering this, it is unsurprising that the violation in (2.17 b) can be explained long before the gap position is encountered. Unaccusative verbs like *seem* do not subcategorize for both an NP (*the man*) and a CP-*that* clause. It is possible for *seem* to subcategorize for the NP (e.g. *The man seems tired*) and a CP-*that* clause (as in 2.17 a), but not both simultaneously. Thus, the P600 elicited by *that* in (2.17 b) is most straightforwardly explained as a subcategorization violation. Again, while McKinnon and Osterhout labeled this as a violation of a movement phenomenon (i.e. a filler-gap dependency), and there is a filler and gap present in the materials, the violation has a more direct antecedent than the filler-gap dependency.

⁶ Expletive insertion is included here for readability.

Neville et al. (1991) reported a P600 response to what they call the ‘subjacency constraint violation,’ (i.e. a subject island) presented in (2.18 b). Here the ungrammatical filler-gap dependency elicited a P600 measured at the main verb *admired*.

(2.18) Subjacency constraint violation (Subject island)

- a. Was [a sketch of the landscape] admired by the man?
- b. * What was [a sketch of _] admired by the man?

Modified from Neville et al. 1991

While Neville et al. labeled this comparison a subjacency violation, it is not clear what the P600 effect at *admired* (2.18 b) was in response to. This response could be due to the incomplete noun phrase *a sketch of* which could trigger reanalysis or repair processes, or it could be due to the parser attempting a number of integration processes: integration of the filler *what* with the gap position, or *what* with the verb *admired*, or the entire noun phrase *what was a sketch of* with the verb *admired*. As we will see below, P600 responses have been reported at the pre-gap position in a number of studies (Kaan et al. 2000; Fiebach et al. 2002; Phillips et al. 2005 and Gouvea et al. 2009), making it unclear how to reconcile the post-gap effect in (2.18 b) with other P600 responses to gaps.

These early studies associating a P600 response with island structures (Neville et al. 1991; McKinnon & Osterhout 1996) do not convincingly hold up under further

examination. As such, I do not predict similar effects in the ERP experiment presented in Chapter 6. However, as was alluded to above, there have been multiple interpretations of what the P600 is a response to. Some of these studies, comparing grammatical filler-gap dependencies (rather than using a ‘violation’ paradigm as above), have illustrated that there is a pre-gap response P600, which will be relevant for the experiment in Chapter 6.

Studies on long-distance filler-gap dependencies have reported a P600 response to the verb preceding the gap, when measured (Kaan et al. 2000; Fiebach et al. 2002; Phillips et al. 2005 and Gouvea et al. 2009). Kaan et al. (2000) interpreted this finding as a reflection of the process of integrating the filler with the gap position. Kaan et al. (2000, Experiment 1) compared three sentence types, shown in (2.19 a-c).

(2.19 a) Emily wondered who the performer in the concert
had **imitated** _ for the audience’s amusement.

(2.19 b) Emily wondered whether the performer in the concert
had **imitated** a pop star for the audience’s amusement.

(2.19 c) Emily wondered which pop star the performer in the concert
had **imitated** _ for the audience’s amusement.

(modified from Kaan et al. 2000, 2a-c)

Kaan et al. (2000) reported a larger P600 response at *imitated* in (2.19 a, c), which both have a gap following *imitated*, compared to (2.19 b), which does not have a gap.

Since a filler needs to be integrated with its gap only in conditions (2.19 a, c), Kann et al. interpreted this as the index of that integration process. The current ERP study (Chapter 6) reveals not only a pre-gap P600 in the embedded position as previously reported, but also a pre-gap P600 in the matrix clause, where the conditions do not yet differ (except for the following word; see Chapter 3, section 3.2 for materials). This finding is problematic for integration interpretation of the P600. The findings of these prior studies with respect to the results of the current ERP experiment are examined in further detail in Chapter 6, section 6.4.2.2.1

Additionally, the P600 has been elicited by so-called ‘garden path’ sentences (Osterhout & Holcomb 1992; Osterhout et al. 1994). In a garden path sentence (2.20 b), the parser has a preferred structure for the encountered words, but then that preferred and predicted structure is shown to be incorrect by new input.

(2.20 a) The doctor charged the patient (for the operation).

(2.20 b) The doctor charged the patient was lying.

(2.20 c) The doctor believed the patient was lying.

Modified from Osterhout et al. (1994)

In the example above, *charge* can (among other possibilities) take either a direct object or a sentential complement. A direct object (2.20 a) is preferred over the sentential complement (2.20 b). However, in other verbs, such as *believe*, a sentential complement is preferred (2.20 c). Osterhout and colleagues found that (2.20 b) elicited

a larger P600 response compared to (2.20 c) following *was*. Like the filler-gap studies above, these garden path studies indicate that a violation is not a necessary condition for a P600. In the case of fillers and gaps, it has been claimed that the process of integrating the filler and gap results in a P600. In the case of garden path sentences, the P600 is a response to a dispreferred parse. Osterhout and colleagues interpreted this P600 response as reflecting a process of syntactic reanalysis (accommodating a sentential complement in (2.20 b) when a direct object was expected/preferred). Thus there are two fairly disparate views of the P600 represented here; a reflection of difficulty in an early stage of syntactic integration (Kaan et al. 2000), or a reflection of effort in a late stage of syntactic reanalysis (Osterhout & Holcomb 1992; Osterhout et al. 1994). Added to this picture is the view of Münte et al. (1997) that the P600 reflects a repair process that occurs in response to ungrammatical structures in semantically contentful sentences (i.e. no P600s were elicited when Münte et al. used violations to pseudo-words). This repair process view does not account for P600s elicited when there is no ungrammatical structure present, however.

The complex set of conditions that have been reported to elicit a P600 grow more complex when we consider that non-syntactic violations also elicit a P600. Orthographic violations have elicited a P600 (Münte et al. 1998, Vissers et al. 2006), as well as pragmatic violations (Kuperberg et al. 2003, 2007, Hoeks et al. 2004; Kim & Osterhout 2005; van Herten et al. 2006). These results in particular have generated a wide range of theories as to what the P600 is responding to, some of which are

compatible with one of the theories mentioned above (i.e. integration, reanalysis or repair related explanations).

The theoretical picture of the P600 remains complex, and many of the theories discussed above appear to be mutually exclusive (i.e. how does the P600 reflect both early integration and late reanalysis?) and a common ground account for the P600 remains elusive in the linguistic domain. However some researchers have reached beyond the linguistic domain to associate the P600 response in language with a set of domain-general responses known as the P300 family (e.g. Chapman & Bragdon 1964; Sutton et al 1965; Donchin et al. 1978; Pritchard 1981).

Gunter, Stowe and Mulder (1997) and Coulson, King and Kutas (1998a,b) associate the P600 response with the P300 response (specifically the P3b) that is elicited in a variety of ‘odd-ball’ paradigms. The P300 has a similar scalp distribution to the P600 and its post-stimulus latency is known to vary based on the difficulty of the experimental task (Polich 1987; Picton 1992; Kok 2001). The P300 response is the most robust when a monitoring task is the focus of the subject’s attention (as opposed to a passive monitoring)⁷ and is elicited by low-probability target items in a group of non-target/standard items (see Polich 2007 for a review).

By manipulating the local probability of syntactic violation stimuli, Coulson et al. (1998a,b) were able to reverse the pattern of what conditions elicited a P600. Violations (pronominal case violations and verb agreement violations) elicited P600s as in prior experiments. However, when violations were plentiful and grammatical

⁷ This parallels a general trend in the literature where experiments that involve participants making an explicit judgment task tend to elicit more robust P600 effects than those that don’t- including acceptability judgment tasks in sentence processing experiments.

sentences were rare, it was the grammatical sentences that elicited the P600, albeit smaller in amplitude. That is, both the grammaticality and the probability of encountering a grammatical / ungrammatical sentence modulated the P600. Coulson et al. take this to be evidence that the P600 effect is not domain specific for language⁸ and is a member of the P300 family of brain responses (though see Osterhout & Hagoort 1999 for counterarguments).

If the P600 is taken to be a special case of the P300 family such that it is a response to unexpected events, then many of the disparate findings appear to have a common explanation. Experimental participants certainly encounter grammatical sentences much more frequently than ungrammatical sentences in their day-to-day lives. It is unsurprising that there is a neural response to such improbable sentences in an experimental setting then. Similarly, garden path sentences are an improbable continuation given the subcategorization preferences of the verb. Orthographic violations, specifically in highly predictable contexts (Vissers et al. 2006) are similarly compatible with a probabilistic account of the P300/P600 response. The pragmatic violations often involve improbable animacy conditions (*javelins* throwing *athletes*, Hoeks et al. 2004) and/or subcategorization violations (*for breakfast the eggs would bury*, Kuperberg et al 2003). Under this view of the P600, the more striking question

⁸ Additionally, the P600 has also been elicited in musical (e.g. Patel et al. 1998) and arithmetic (e.g. Núñez-Peña & Honrubia-Serrano 2004) experiments, further indicating that it is not language specific.

becomes not why do all of these disparate constructions elicit a P600, but why don't the sentences that elicit an N400 elicit a P600 instead?⁹

Kaan et al. (2000) reported in a footnote that their integration cost view of the P600 is in principle “compatible with both a language specific and a language non-specific interpretation of the P600, assuming that integration and structural predictions also occur in domains other than language” (pg 161). However, we did not have to invoke ideas of integration in order to accommodate the other linguistic data discussed above into the domain general view of the P300/P600.¹⁰ Can the same be done for filler-gap constructions? That is, can we rely only on ‘structural predictions?’ Are filler-gap constructions less probable than non-filler gap constructions? In day-to-day life, it would seem so. Additionally, syntactic acceptability studies (section 2.2.3) demonstrate that experimental participants disprefer long-distance dependencies even when fully grammatical. Thus, if a dispreferred subcategorization of a verb (i.e. a garden path, Osterhout & Holcomb 1992; Osterhout et al. 1994) is surprising to the parser, so too a dispreferred complex filler-gap construction may be surprising to the parser. In this case, the elicitation of a P600 to both sentence types can be given a unified explanation.

While the domain general picture just described for the P600 may appear promising, it is by no means widely accepted among researchers in sentence

⁹ Though it remains an outstanding question why cloze-probability violations elicit an N400 and not a P600, one option is that the parser presumably encounters unexpected or even novel lexical items in sentences as a regular part of communicating and thus at a much higher rate than ungrammatical structures, leading to different patterns of brain response.

¹⁰ Though it may yet prove useful to do so in order to distinguish it from the N400 response. In language, this may essentially be a rephrasing of the P600 ~ syntax, N400 ~ semantics association.

processing, and the interpretation of the P600 remains an unsettled issue. For present purposes, the most relevant interpretation of the P600 is the integration cost hypothesis put forth by Kaan et al. (2000). Like Kaan et al. and following studies, we find a pre-gap P600 response, but the design of the current materials allowed us to additionally observe a pre-gap P600 response when the gap position could not yet be predicted. This led us to interpret the P600 as reflecting gap recognition rather than syntactic (filler-gap) integration difficulty (Chapter 6, section 6.4.2.2.2). As a gap recognition response, our view of the P600 is compatible with the domain general view of the P600 being sensitive to less probable stimuli and thus related to the P300.

2.2.5.2.1 N400

The first ERP component to be identified with language processes was the N400 (Kutas & Hillyard 1980a,b,c), a negative going voltage peaking about 400 msec post-stimulus onset often with a right centro-posterior scalp distribution. While not usually associated with filler-gap dependencies, the N400 is highly relevant to the findings of Experiment 3 (Chapter 6, section 6.4.2.2.7) and is presented here as preparation.

The original N400 findings reported that the amplitude of the N400 was larger to a word that was more incongruous in a sentence (e.g. *He took a sip from the transmitter/waterfall*, Kutas & Hillyard 1980a) but it was not sensitive to physical changes of a word (i.e. font size, Kutas & Hillyard 1980c). Some have thus characterized the N400 as an index of semantic incongruity, or in other words, a

'semantic violation.' The N400 is more than a simple incongruity detector, however. Kutas and Hillyard (1984) reported that the N400 amplitude was negatively correlated with how predictable a word was. The more predictable a word was (determined by cloze probability), the smaller the N400's amplitude. Thus in (2.21 b), where *ladder* has a low cloze probability and is less predictable, a larger N400 response is elicited than in (2.21 a), where *safe* has a high cloze probability and is more predictable.

(2.21 a) Hi cloze: She locked the valuables in the safe.

(2.21 b) Lo cloze: The dog chased our cat up the ladder.

Modified from Kutas & Hillyard (1984, Fig 1 A)

While cloze probability correlates strongly with the N400 response, contextual constraint does not. Where the cloze probability is the probability that a particular word completes/continues an utterance, contextual constraint is a reflection of how many or how few possible completions/continuations are provided for a given sentence.

(2.22 a) Hi constraint & cloze: He mailed the letter without a stamp.

(2.22 b) Lo constraint, hi cloze: There was nothing wrong with the car.

Modified from Kutas & Hillyard (1984, Fig 1 A)

In (2.22 a) *stamp* represents both a high cloze probability (it was the most frequently provided completion in a cloze task) and a high contextual constraint (few other completions were provided). In (2.22 b), on the other hand, *car* also represents a high cloze probability, but it has low contextual constraint. Participants provided a wide range of completions for this sentence, but the most frequent (highest cloze) was *car*.

This dissociation indicates that the N400 is not reflecting how a semantic prediction is ‘violated,’ because then we would expect contextual constraint to influence the response (i.e. a higher constraint means a stronger expectation and should result in a larger response if violated). Instead, the N400 reflects how context has prepared the processor for the current word and its properties (Kutas & Federmeier 2009). As we see below, numerous properties of a word have been shown to influence the N400 in addition to cloze probability.

In addition to the robust predictability effect discussed above, there are a number of lexical manipulations that have been shown to modulate the amplitude of the N400. Open class words elicit larger amplitude N400s than closed class words (Van Petten & Kutas 1991; Neville, Mills, & Lawson 1992; Münte et al. 2001). Low frequency words elicit larger amplitude N400s than high frequency words (Van Petten & Kutas 1990, 1991; Van Petten 1993; Münte et al. 2001; Allen et al. 2003; Barber et al., 2004). More concrete words elicit larger amplitude N400s than less concrete, more abstract words (Paller et al. 1987; Kounios & Holcomb 1994; Holcomb et al. 1999; West & Holcomb 2000). Words with a higher orthographic neighborhood density elicit larger amplitude N400s than words with a smaller orthographic neighborhood

density (Holcomb, Grainger, & O'Rourke 2002). The fact that there are larger amplitude N400 responses to open class, low frequency, concrete words and words with a high neighborhood density are not immediately explicable purely by a predictability view of the N400. While it could be claimed that open class words are less predictable than closed class words, or that low frequency words are less predictable than high frequency words, it is counterintuitive that concrete words should then be less predictable than abstract words. Similarly, it is unclear how orthographic neighborhood density interacts with predictability.

Orthographic neighborhood density could straightforwardly interact with lexical access, however. As a word has more neighbors, there could be (i) more competitors with the target word or (ii) more spreading activation around the target word. Under either view, a larger N400 response is not unreasonable as it could reflect either (i) more effort required for lexical access, or (ii) more (incidental) activity surrounding that access.

In order to accommodate the data from orthographic neighbors and frequency effects, a notion of lexical access must be included in our understanding of the N400. As Van Petten and Luka (2006, pg 281) state, "data suggest that N400 amplitude is a general index of the ease or difficulty of retrieving stored conceptual knowledge associated with a word, which is dependent on both the stored representation itself, and the retrieval cues provided by the preceding context." Frequency and neighborhood density effects are part of 'the stored representation itself,' while

predictability (e.g. cloze probability) serves as ‘retrieval cues provided by the preceding context.’

Note that while the N400 is sensitive to manipulations requiring lexical access (i.e. frequency, orthographic neighborhood, concreteness) this does not necessarily mean that the N400 reflects semantic integration, though some researchers do hold this position. In order to account for pragmatic and discourse manipulations of the N400, some researchers have claimed that the N400 reflects a process of connecting the meaning from the sentential context with the semantic information retrieved from the current word; that is, a semantic integration process (Brown & Hagoort 1993, 1999; Chwilla, Brown, & Hagoort 1995; Hagoort et al. 2009). Thus, under this view, the N400 reflects a late, post-lexical process for these researchers (see Kutas and Federmeier 2011 for further discussion on this view, and Chapter 6, section 6.4.2.2.4 for the relevance of this view to the current ERP experiment).

The elicitation of N400s is not limited to sentential, or even linguistic, environments. N400s have been reported in word list and priming paradigms (Bentin, McCarthy, & Wood, 1985; Kutas and Hillyard 1989; Kutas, Neville, & Holcomb 1987; Holcomb 1988; Holcomb & Neville 1990, 1991; Neville, Mills, & Lawson 1992). N400s have been elicited by pseudo-words (but not ill-formed non-words). Pseudo-words are not actual words, but are possible words in that they follow the phonotactics and orthographic conventions of the language (Bentin et al. 1985; Bentin 1987; Rugg & Nagy 1987; Smith & Halgren 1987; Holcomb 1988, 1993; Holcomb & Neville 1990; Bentin, Mouchetant-Rostaing, Giard, Echallier & Pernier 1999). N400s

have been reported for other meaningful, visual stimuli such as movies (Sitnikova et al. 2008), drawings (Nigam, Hoffman, & Simons 1992; Ganis et al. 1996; Ganis & Kutas 2003), faces (Barrett & Rugg 1989; Bobes, Valdes-Sosa, & Olivares 1994) and gestures (Kelly, Kravitz, & Hopkins 2004; Wu & Coulson 2005) as well as non-linguistic but meaningful sounds (Chao, Nielsen-Bohlman, & Knight 1995; Plante, Van Petten, & Senkfor 2000; Van Petten & Rheinfelder 1995) and music (Koelsch et al. 2004; Daltrozz & Schön 2009).

In summary, the N400 is a response to a wide range of potentially meaningful stimuli. In the specific domain of language, the N400 can potentially be found to all words in a sentence¹¹ and represents “part of the brain’s normal response” to those words (Kutas & Federmeier 2009). This normal response is modulated by the interaction of predictability and lexical features, which appears to drive many of the N400 effects reported in the literature. Context influences the predictability of a given word which then prepares the parser for the features of that word. To the extent those predicted word’s features are encountered, the N400 response is reduced. In the current ERP experiment (Chapter 6) we see a similar pattern, albeit with a new twist. Section 6.4.2.1.6 presents an N400 effect to a lexical item that indicates the presence of a syntactic gap. This is discussed in section 6.4.2.2.7 in terms of the prior sentence context (whether the parser encountered (i) an interrogative *whether*-island clause boundary or (ii) a declarative *that* clause boundary, both while there was an unresolved filler-gap dependency) influencing the predictability of the gap. The

¹¹ Depending on experimental design.

pattern of results in Chapter 6 indicated that gaps are less predicted inside *whether*-islands.

2.3 Accounts of island phenomena

Having introduced the basic phenomenon, including processing methods and data, I turn now to the proposed accounts for island phenomena. While these accounts are numerous, I will group them into three broad categories here: grammatical, functional and cognitive.

2.3.1 Grammatical

The grammatical accounts for islands include both syntactic and semantic accounts. Syntactic constraints have moved from the relatively ad hoc nature of Ross' original island constraints (1967) to broad theoretical constructs such as Subjacency (Chomsky 1973, 1977, 1981), Barriers (Chomsky 1986) or Relativized Minimality (Rizzi 1990; Cinque 1990) that have been proposed as efforts to unify the various individual stipulations that island constraints reflect. While these accounts have had success in reducing the ad hoc nature of the constraints, they encounter difficulties in accounting for the 'd-linking' phenomena (Pesetsky 1987) in (2.6 b) and (2.7) without recourse to concepts most would concede are outside the domain of theoretical syntax. In order to address d-linking, theoreticians need to invoke pragmatic (d-linking) or semantic (specificity, individuation in a set) notions. While the syntactic theory cannot

account for all the facts within its own domain, neither can semantics, as extraction is an integral component of island violations in *wh*-movement languages.¹²

2.3.1.1 Syntactic

Chomsky (1964) proposed the A-over-A Condition, which states that ‘an element of category A cannot be extracted out of a phrase of category A.’ This represents the first attempt to capture the idea that ‘syntactic movement’ was not permissible out of certain domains. Ross (1967) noted both over-generations and under-generations of the A-over-A condition, cataloging structures from which a *wh*-phrase could not be extracted to form *wh*-questions or relative clauses. These structures, called ‘islands,’ have not only become the focus of much work in theoretical syntax themselves, but have served as diagnostic tests for examining whether a syntactic structure involves movement.

Ross (1967) noted a number of separate domains that blocked extraction, as well as ameliorations of these effects (e.g. finiteness). From the beginning of the research into island phenomena, we see competing drives to (i) capture all of these extraction effects uniformly (e.g. the A-over-A condition) and (ii) capture the facts that the linguistic data are variable (e.g. the ameliorations observed by Ross 1967). The *wh*-islands that this dissertation focuses on were not one of the original island constraints introduced by Ross (1967), who was focused on the strongest/most clear

¹² For languages that are *wh*-in-situ, island sensitivity varies. Huang (1982) reported island effects in Mandarin Chinese for *wh*-adjuncts, but not arguments. Other in-situ languages may not show any apparent island effects (Quechua, Cole & Hermon 1994) or may show effects for both *wh*-adjuncts and arguments (Hindi, Malhotra 2009).

cases preventing extraction. While experimental acceptability studies have since demonstrated that *whether* island effects are just as clear and consistent as in other islands (e.g. Sprouse, Wagers, Phillips 2012), there has been a sense that *whether* island violations were not as robust as other types of island violations. Subsequent influential theoretical proposals to address island phenomena, such as Subjacency (Chomsky 1973, 1977, 1981) and Barriers (Chomsky, 1986), included *wh*-islands with the Complex NP, Coordinate Structure and Subject islands of Ross (1967), though a distinction of ‘strong islands’ versus ‘weak islands’ (for example *whether*-islands), was later introduced to account for the fact that islands differ in the types of phrases that may be extracted from them (Huang 1982, Chomsky 1986).

The ‘strong/weak’ distinction represents an attempt at compromise between the competing factors of providing a universal explanation and accounting for diverse data, mentioned above. According to Cinque (1990) a ‘weak’ island allows a PP-gap while strong islands can only allow a DP-gap (if at all). Crucially, the ‘strong/weak’ distinction is not in reference to how unacceptable the island violation is deemed, although this later usage often appears to creep into discussions about island phenomena.

The theoretical desire to be explanatory, and to avoid stipulation, provides a goal for subsequent treatments of islands of having this strong/weak distinction fall out from other theoretical constructs. In Rizzi’s (1990) Relativized Minimality (see also Cinque 1990) these ‘classes’ of island phenomena are accounted for by the interaction of the Empty Category Principle (ECP, Chomsky 1981) and Subjacency

(Chomsky 1973). However, even in this compromise between looking for a universal account of islands and the reality of differences in the data, the notions of ‘strong’ versus ‘weak’ islands do not represent a strict dichotomy and proposed diagnostics for them differ (Cinque 1990; Postal 1998, but see Stepanov 2007 for an attempt to unify ‘strong/weak’ islands).

While the theoretical machinery applied to islands has changed over the years as the dominant syntactic theory has changed, the basic approach of syntactic accounts of island phenomena share the same goal: to illustrate how the structure of the language prevents extraction out of islands, but not out of non-islands. Comparing (2.23 b) to (2.23 a), for example, *who* is analyzed as being extracted out of the post-verbal position, but it does not move immediately to the beginning of the sentence, where we see it realized.

(2.23 a) Bill thought [that Mary insulted who?]

(2.23 b) Who did Bill think [(_) that Mary insulted _ ?]

First, the filler stops at the specifier position of the embedded clause (Spec-CP, indicated by (_) in 2.23 b), and only then continues to its matrix clause position. That is, movement does not occur in one fell swoop, but rather operates successive-cyclicly. This basic picture remains the same regardless of whether this style of movement is motivated by avoiding Subjacency violations (Chomsky 1973, 1977, 1981), jumping

over Barriers (Chomsky, 1986), or skirting the Phase Impenetrability Condition (Chomsky 2004) when a Derivation by Phase (Chomsky 2001) analysis is used.

However, in the island violation (2.24 b), this successive-cyclic movement must be blocked in order to explain the unacceptability of the sentence.

(2.24 a) Bill wondered [whether Mary insulted who?]

(2.24 b) *Who did Bill wonder [whether Mary insulted _ ?]

In order to do this, it is claimed that the ‘landing site’ at Spec-CP of the embedded clause is not available. The presence of *whether* precludes a moved constituent from landing here. The most common claim is that *whether* is itself already residing in Spec-CP (unlike *that* in (2.23 b), which is in C), so the moved constituent cannot make use of this structural location. Thus, the only movement available to *who* is the long ‘one fell swoop’ movement, which is ruled out by the architecture of the grammar. This movement either violates Subjacency, crosses too many Barriers or is not available at the phase’s edge, depending on the theoretical architecture used. The exact theoretical machinery invoked matters less for the current purposes than the unifying intuition that the long-distance relationship is blocked, either by the unavailability of well-formed successive-cyclic movement or another disruptor is present (such as a possible antecedent-governor in Relativized Minimality).

2.3.1.2 Semantic

Semantic accounts of islands (e.g. De Swart 1992; Honcoop 1998; Szabolsci & Zwarts 1993; Szabolsci & Den Dikken 2002; Truswell 2007) focus not on preventing a filler from moving out of an island domain, but rather on semantic notions such as scope or event structure disrupting the filler-gap dependency. For example, de Swart 1992 proposes that a quantifier (Q1) can only separate another quantifier (Q2) from its restrictive clause if Q1 has wide scope over Q2. So we see an intervention effect, much like in a syntactic account, but movement need not be a part of this explanation.¹³ Szabolsci (2006) provides an overview of both Szabolsci & Zwarts' (1993) algebraic approach to scopal intervention and Honcoop's (1998) Dynamic Semantic approach. In this same overview, Szabolsci points out that the *wh*-island phenomenon that is the focus of this dissertation can be captured by a range of accounts, both syntactic (Subjacency and Relativized Minimality) and semantic (Monotonicity: Szabolsci & Zwarts 1990; and the aforementioned scopal approach). As the focus of the dissertation is not to distinguish between these grammatical accounts, but rather between processing accounts of islands, I will not delve into the technical details of these grammatical accounts further. The reader is referred to the above references for discussions. The relevant point for the current research is to note that there is not a singular 'grammatical account' of islands, although some version of a syntactic account is generally more widely assumed in the literature.

¹³ Although for most semantic theorists, movement will be required to either provide the surface structure of the sentence, or to obtain the proper scopal relationships at LF.

2.3.2 Functional

I will refer to the second type of account for islands as functional accounts. While these accounts vary in the terminology that they use - Erteschik-Shir and Lappin speak of *dominance* (1979), Kuno *topics* (1976), Takami *focus* or ‘*important information*’ (1989), and Goldberg (2006) *background*, they all converge on a similar idea. The basic intuition is that sentences are *about* something (but may also contain additional, less crucial information), and only that ‘something’ which the utterance is *about* is salient enough to be extracted. From this point of view islands are simply constructions that may modify what the utterance is about, but no individual element within the island represents what the utterance itself is *about* and thus extraction from within that island is disallowed. Questioning an item in a sentence marks it as important information. In English this includes having it at the beginning of the sentence, where it can be extremely salient. If this salient filler is associated with a gap in a non-salient, or backgrounded, part of the sentence, then there is a clash of information structure and the sentence is deemed to be unacceptable.

Under a functional account, the ‘light’ bridge verb *think* does not create an island because they “are generally used to introduce a complement clause containing the foregrounded information” of an utterance (Ambridge & Goldberg 2008). *Who* highlights/focuses the person insulted (in 2.23 b), which is in the foregrounded clause. No information structure clash occurs and the sentence is acceptable. In order to differentiate the conditions, we must assume that *wonder whether* does not typically indicate foregrounded information. *Who* would again highlight/focus on the person

insulted, but it is not clear in (2.24 b) what is foreground and what is background. Is the topic of the sentence Bill's wondering or what he wondered? In order to best account for the unacceptability of (2.24 b), it would be preferable if it could be shown that Bill's wondering is foregrounded, and the remainder is backgrounded, resulting in an information structure clash. However, it is difficult to convincingly demonstrate this in experimental findings as 'foregrounded information' (for example) is difficult to operationalize when it is not explicitly being manipulated by a factor such as (auditory) stress. So while the functional account may present an intuitively satisfying explanation for island phenomena, studies need to be constructed around the careful manipulation of information structure to convincingly demonstrate that participants are not applying alternate readings/assumptions about information structure to experimental sentences.

2.3.3 Processing

The basic claim of processing accounts for islands is that the unacceptability of the island violation is due to the difficulty in some part of the on-line parsing of the sentence. Where processing accounts differ from each other is in how they characterize this on-line parsing difficulty. I will discuss these processing accounts of islands based on what view of working memory they hold. The capacity-constrained based account relies on the Just and Carpenter (1992) model of working memory. The similarity-interference account on the other hand views working memory in terms of content-addressable retrieval processes that are susceptible to similarity-based

interference (e.g. Lewis and Vasishth 2005). Both of these accounts share the same underlying position: difficulties in processing island violation sentences lead to those sentences being deemed unacceptable. While there are other approaches (e.g. Deane's 1991 attention-based account, or a mixed grammar/processing account suggested by Michel & Goodall 2013), I will focus on the capacity-constrained account, as it is the most prevalent processing account of islands (e.g. Kluender 1991, 1998; Kluender and Kutas 1993a,b; Hofmeister 2007; Sag et. al. 2007) and a newer working memory model that has been supplanting the capacity-constrained view of working memory in recent years: similarity-interference (e.g. Gordon, Hendrik and Johnson 2001; Gordon, Hendrick and Levine 2002; Lewis and Vasishth 2005; Gordon et al. 2006; Lewis, Vasishth and Van Dyke 2006; Van Dyke and McElree 2006). As this newer account has not yet been applied specifically to island phenomena, I introduce the similarity-interference account of islands below.

2.3.3.1 Capacity-constrained

Linguistic theory and theories of language processing have long assumed that some form of working memory influences the parser (e.g. Chomsky & Miller 1963; Wanner & Maratsos 1978; Daneman & Carpenter 1980; Just & Carpenter 1992; Caplan & Waters 1999). Working memory (and how it applies to language) has been described in a number of ways, from the multiple sub-components of the phonological and articulatory loops and acoustic store (Baddeley and Hitch 1974; Baddeley 1983, 2002), to working memory being the 'focus of attention' on four (Cowan 1995, 2001)

or even just one (Oberauer 2002) representation(s) in long-term memory. The relevant theory of ‘working memory’ for the capacity-constrained account of islands is the Capacity Constrained Comprehension Theory (Just and Carpenter 1992).

Kluender (1991) adopted the Capacity Constrained Comprehension Theory (Just and Carpenter 1992) in order to develop a processing account of island phenomena. In this working memory model, there is a common pool of resources used for two distinct tasks: (i) computation and (ii) storage. Because these two tasks share the same common pool of resources, as the demands on one task increase, the resources available for the other decrease. If demands are near capacity, certain items held in memory may be expunged or certain processes slowed or abandoned in order to free up this mental resource. The key point of comparison with other models (e.g. Waters and Caplan 1996) is that stresses on the capacity limit are not domain specific, but can also be due to other unrelated memory requirements or computational processes.

Kluender (1998) presented the foundation for the case for a sentence processing explanation of island constraints from converging ideas from Fodor (1978, 1983) and Ross (1987), ironically the originator of the island constraints. Ross (1987) posited that ungrammaticality is due to cumulative small deviations from a prototype reaching a certain threshold, at which a speaker perceives the utterance as ungrammatical. Fodor (1983) characterized ungrammaticality similarly, but in terms of a build-up of markedness. Kluender (1998) treated processing strains as causing the kind of small deviations from prototypes that Ross (1987) and Fodor (1983) invoke to

account for ungrammaticality. As such, the ‘certain threshold’ that these deviations must together exceed is understood as the capacity of a person’s working memory.

The capacity-constrained account of islands draws on a collocation of factors independently known to add to the processing difficulty across different sentence types: (i) the storing of a filler in memory, (ii) having this filler stored in memory when a clause boundary is encountered and (iii) processing differences in clause boundary types (Kluender, 1991, 1998; Kluender & Kutas, 1993a).

The first strain is induced by the presence of a filler in the sentence, specifically, storing this filler in memory. Fodor (1978) noted that filler-gap dependencies are more difficult to process than sentences without filler-gap dependencies, even when they do not violate any proposed syntactic constraints. Gibson’s (2000) Dependency Locality Theory addresses why this might be in its Storage-Based Resource Theory. In this theory, each syntactic head imposes a storage cost on the processor. These costs accumulate as more syntactic heads are encountered, and the costs are reduced as each head is successfully integrated into the meaning of the sentence. A filler is an additional syntactic head, and thus comes with an additional processing (specifically memory storage) cost. There is no way to interpret the referent of the *wh*-filler in the discourse when it is first encountered, and thus the referential ‘space’ for the *wh*-filler must be held until a referent is found. That is, a filler is often not immediately integrated, so the cost of storing it lingers.

Kluender and Kutas (1993a) proposed that entering a filler into storage is reflected in an event related potential (ERP) signal obtained at the scalp that they dub

the Left Anterior Negativity (LAN). In sentences with *wh*-extraction, they found that the LAN was elicited 300 to 500 msec after a subject had encountered the filler (interrogative *wh*-pronoun) as compared to a baseline where no filler was present. In the sentences in (2.25), a LAN was present at *she* in (2.25 a) when compared to (2.25 b).

(2.25 a) What has she forgotten that he dragged her to ___ on Christmas Eve?

(2.25 b) Has she forgotten that he dragged her to a movie on Christmas Eve?

Kluender (1998)

This LAN effect was also evident in a subordinate clause at the gap position associated with the filler (e.g. 2.14 b). Thus the LAN is elicited by both the storing of the filler, and its retrieval from memory for integration. In this sense the LAN appears to index the ‘lid’ to the storage mechanism. The lid opens when the filler is deposited, and a LAN is observed. When the lid is opened again so the filler can be withdrawn, a LAN is again observed.

The second strain on the working memory system is caused by storing a filler across a clause boundary, which is more taxing than storing a filler within a clause of the same length. Frazier and Clifton (1989) matched sentences for length and found that *wh*-extraction out of a clause resulted in an additional processing cost (measured by an increase in reading times) compared to *wh*-extractions that occurred within a single clause. This processing difficulty was paired with a drop in grammaticality

ratings for extraction out of a clause boundary. It is likely that this reflects a general processing cost of initiating a new clause.

The third and final proposed strain on working memory resources that is relevant to island phenomena is the amount of additional referential processing needed at the clause boundary. This is roughly equivalent conceptually to the complexity of the lexical item that indicates the subordinate clause. Kluender (1998) leaves this concept underspecified, but contrasts three options of *that*, *if* and *who* at the clause boundary of yes/no questions as in (2.26).

(2.26 a) Has she forgotten [**that** he dragged her to a movie on Christmas Eve?]

(2.26 b) Has she forgotten [**if** he dragged her to a movie on Christmas Eve?]

(2.26 c) Has she forgotten [**who** he dragged ___ to a movie on Christmas Eve?]

Kluender (1998)

Kluender links the relative amount of referential processing needed for each lexical item to the peak amplitude of the N400 component (Kutas and Hillyard 1980a, 1980b, 1984, section 2.4.1.2.3), with *that* being understood as the least complex and eliciting the smallest N400, *if* occupying the middle ground and *who* being the most complex item at the clause boundary and eliciting the largest N400.

The results of two acceptability judgment tasks (an offline scalar judgment and an online forced choice task) indicated an effect of extraction with *wh*-extraction questions (2.27 d-f) rated lower than yes/no questions (2.27 a-c).

- (2.27 a) Has she forgotten
 [that he dragged her to a movie on Christmas Eve?] >
- (2.27 b) Has she forgotten
 [if he dragged her to a movie on Christmas Eve?] >
- (2.27 c) Has she forgotten
 [who he dragged ___ to a movie on Christmas Eve?] >
- (2.27 d) What has **she** forgotten
 [that he dragged her to ___ **on** Christmas Eve?] >
- (2.27 e) What has **she** forgotten
 [if he dragged her to ___ **on** Christmas Eve?] >
- (2.27 f) What has **she** forgotten
 [who he dragged ___ to ___ **on** Christmas Eve?]

Modified from Kluender (1998)

Additionally, an effect of complement type was found (*that* was rated highest, followed by *if* and then *who*) as well as an interaction between extraction and complement type. The interaction resulted in the worst judgments being assigned to extraction out of a *wh*-clause (2.27 f; a *wh*-island violation). This indicated that the unacceptability of *wh*-island extraction was due to an interaction of extraction and complement type, as has been replicated for multiple island types (e.g. Sprouse, Wagers and Phillips 2012; this dissertation Chapter 4). Kluender concludes that this

eliminates the need for positing a separate grammatical island constraint to account for the unacceptability of extraction out of a *wh*-island.

This interaction was also present in the neurophysiological response to the same stimuli. The main effect of complement type was demonstrated by an increasing N400 to the less preferred complement in the yes/no questions (2.27 c > b > a, underlined words). The explanation given for why this effect is not present in the *wh*-questions is that the referential processing that would have normally occurred at the complementizer (as in the yes/no questions) is postponed. The reason for this delay is that the *wh*-extraction has caused a temporary referential ambiguity, pushing the working memory system to capacity. Unfortunately, in this dissertation, we do not observe a similar N400 effect at the clause boundary, likely for methodological reasons (see Chapter 6, section 6.4.2.2.6 for discussion).

The main effect of extraction is shown by the LAN, indexing the working memory costs of storing a *wh*-filler and associating it with its gap site (2.27 d-f, bolded words). The effect of double extraction is seen in a second LAN following the embedded *who* in the *wh*-island sentence (2.27 f, bolded underlined word), as the *wh*-complement *who* creates an additional dependency. This results in an additional processing cost above and beyond the one indexed by the matrix *wh*-questions. Note that this extraction in the subordinate clause is also present in the (grammatical) yes/no question in (2.27 c). This is expected based on the LAN's association with filler storage. However, this additional process does not itself result in (2.27 c) being considered unacceptable. This is further evidence that the processing strains can be

isolated and that it is only their accumulation beyond a certain threshold that results in unacceptability.

In sum, the Kluender (1991, 1998) and Kluender and Kutas (1993a,b) studies have argued that the acceptability patterns of *wh*-islands can be generated by the interaction of known strains on processing. Further, two of these three strains (holding a filler in memory and clausal referential processing) have been shown to have separate neurophysiological correlates, lending empirical evidence to the claim that there are independent memory and integration costs to online sentence processing.

In the experiments presented in this dissertation, working memory capacity was measured with the reading span task (Just and Carpenter 1992) the cognitive measure of choice in studies that examine the processing of filler-gap dependencies (e.g. King & Just 1991; Fiebach, Schlesewky & Friederici 2002). Additionally, the n-back task (Kirchner 1958) was used to assess more general working memory. The reading span task is described in chapter 3, section 3.3.1 and the n-back follows in section 3.3.2. If the capacity-constrained account of islands is correct, then it is expected that these working memory measures will co-vary with the experimental measures of the following experiments. Additionally, while one could expect processing difficulties to occur at a gap located within an island, the key claim of the capacity-constrained account is that the confluence of processing difficulties reaches its high point at the clause boundary. This is the opposite pattern as what is expected under a similarity-interference account (see section 2.3.3.2, below). Thus, if the main processing difficulty of the island violation sentence occurs at the clause boundary,

this will favor the capacity-constrained account over the similarity-interference account.

2.3.3.2 Similarity-interference

A more recent view of working memory is implemented in similarity-based interference accounts of sentence processing (e.g. Gordon, Hendrik and Johnson 2001; Gordon, Hendrick and Levine 2002; Lewis and Vasishth 2005; Gordon et al. 2006; Lewis, Vasishth and Van Dyke 2006; Van Dyke and McElree 2006). The treatment of long-distance dependencies differs from the capacity-constrained view in that no special storage is required. All words that the parser encounters are stored, and a filler does not enter a special storage that must be actively maintained. Thus, there is no storage ‘cost.’ Instead, processing costs can be observed in the retrieval process due to similarity-based interference.

Gordon, Hendrik and Johnson (2001) demonstrated the importance of similarity-interference in relative clause filler-gap dependencies. The object relative clause in (2.28 a) was more difficult to process (slower reading times) than the subject relative clause in (2.28 b) at (and immediately before) *climbed* only when *the lawyer* is in the relative clause. When *Joe* is in the relative clause, the subject/object relative clause asymmetry disappears.

(2.28 a) The barber that

[{ the lawyer / Joe } admired _] **climbed** the mountain.

(2.28 b) The barber that

[_ admired { the lawyer / Joe }] **climbed** the mountain.

Modified from Gordon, Hendrik and Johnson (2001, 6)

Because *the lawyer* is similar to *the barber*, retrieval of *the barber* is more difficult. However, since *Joe* is dissimilar to *the barber*, no interference is observed. The similarity between *the barber* and *the lawyer*, but not *Joe*, could be interpreted as form-based (*the x*) or semantically based (definite description vs. individual name). In either case, the difficulty of object relatives compared to subject relatives is not seen here as inherently due to linear or structural distance, or the amount of time a word has to be held in memory. The difficulty is due to (potentially) similar words intervening between and interfering with the filler-gap dependency.

While similarity-interference views of working memory have presented explanations for relative clause asymmetries in terms of retrieval interference rather than active storage, they have not explicitly done so for island phenomena. It is not difficult to see how such an account would be stated, however. The presence of the island introduces features that are similar to those of the filler, meaning that when the retrieval cue (the gap) is encountered, similarity-based interference would occur. This interference creates difficulty for the parser and in turn, this difficulty results in the

sentence being deemed unacceptable. This approach can be summarized succinctly in (2.29).

(2.29) Similarity-interference account of islands: Island boundaries contain features that interfere with the retrieval of fillers.

The broad claim in (2.29) recalls the A-over-A condition (Chomsky 1964) where a similar intervener disrupts a dependency. In order to narrow this broad claim, I focus specifically on a recent view of similarity-interference (Lewis and Vasishth 2005; Lewis, Vasishth and Van Dyke 2006) that has been developed in a specific cognitive architecture.

Based on the ACT-R model (Adaptive Control of Thought-Rational, Anderson et al. 2004), Lewis and Vasishth (2005) and Lewis, Vasishth and Van Dyke (2006) proposed that the parser has (i) a limited focus of attention, with (ii) rapid, content-addressable access to items in memory, which are (iii) stored as bundles of features, which are in turn (iv) subject to similarity-based retrieval interference; but (v) the parser does not have fast access to serial order information (i.e. which item in memory was encountered first, second, etc.) and (vi) has fluctuating activation of items as a function of decay and retrieval history. While much of the focus of this model has been on demonstrating (ii) and (v), point (iii) is still fairly underspecified. If a word in memory is represented by a bundle of features, it is possible that any number of those features could be contributing to similarity-based interference.

We can get a sense of what features may be of import by examining an example of content-addressable memory and cue-based parsing (2.30, summarized from Lewis, Vasishth and Van Dyke 2006).

(2.30) Melissa knew that **the toy** from her uncle in Bogotá **arrived** today.

Modified from Lewis, Vasishth and Van Dyke (2006, Figure 1)

Lewis, Vasishth and VanDyke (2006) explain that the NP *the toy* is the subject of the verb *arrived*, but these words are separated from each other by an adjunct phrase¹⁴. *The toy* is in an embedded clause, and it is encoded as a subject based on its positioning. At this point, however, *the toy* has no verb to associate with, so a prediction for such a verb is generated. Later, the lexical item *arrived* is encountered and requires a constituent to be the ‘arriver,’ that is, a subject for that verb. In this sentence, that ‘arriver’ is *the toy*, but this distant NP must be retrieved from memory. The lack of subject for *arrived* is a cue that triggers a search for a likely candidate. This candidate should be (i) an NP, (ii) in a subject position, that (iii) is predicting an upcoming verb (that is, it has not been associated as a subject of another verb yet). *The toy* fits these requirements and is successfully retrieved. Other words in the sentence (and otherwise in recent memory) represent potential distracters for this retrieval process. For example, *Melissa* is also (i) an NP, (ii) in a subject position, but it is no longer predicting a verb- it has already been associated with *knew* as its subject. Thus

¹⁴ Note, however, that the entire phrase *the toy from her uncle in Bogotá* could be analyzed as the subject of *arrived*, raising questions about how a ‘subject feature’ or ‘subject position’ is defined and assigned in this account.

there is partial overlap between the features of *Melissa* and *the toy*, which may have slowed down the retrieval of *the toy* at *arrived*, but not as much as if the phrase matched on all three example criteria.

It is important to highlight two things from this example. First, predictions are made based on what words are encountered.¹⁵ This makes the similarity-interference approach compatible with the increasing evidence for the importance of prediction in sentence processing (e.g. Altmann and Kamide, 1999; 2009; Kamide, Altmann & Haywood 2003; Federmeier 2007; Pickering and Garrod 2007; Levy 2008). Second, the features relevant for similarity-based interference include structural notions (i.e. subject) as well as relational notions (i.e. whether a NP is still predicting / has associated with a verb). The possibility of structural features causing interference makes this type of approach straightforwardly applicable to *whether*-islands.

In the *whether*-island sentences that are the focus of this dissertation, *wh*-fillers have certain features attached to them, such as being an interrogative [+WH] word in a high structural position in its clause (i.e. Spec-CP). When the parser reaches the retrieval cue (gap), the filler must be retrieved and associated with the gap. In a normal situation, this is not problematic. The parser retrieves the lexical item that had the [+WH] feature and makes the association. However, if there are other [+WH] items that have been recently encountered in similar structural configurations, such as *whether* in a *wh*-island, the parser has difficulty resolving the conflict from this

¹⁵ In fact, the ACT-R architecture includes a specific component, the ‘control buffer,’ that updates syntactic predictions (Lewis and Vasishth 2005, pg 383)

interference, and processing difficulty ensues.¹⁶ If this is the case, then rather than high and low working memory capacity individuals behaving differently, I predict that individuals who are better at suppressing distracters would be better able to process these interfering island violation sentences.

While the examples in (2.28) and (2.30) suggest possible features that could interfere with retrieval, there is not yet a consensus in the field for what features are most likely to be disruptive. The Gordon, Hendrik and Johnson (2001) example above indicates that either form-based similarity or the type of description of a noun phrase may be disruptive (or both), while the Lewis, Vasishth and Van Dyke (2006) example focuses on syntactic category, position and relationships.

In the experiments presented in this dissertation, the ability to suppress distracters in real time was measured with the Eriksen flanker task (Eriksen & Eriksen 1974; Eriksen & Schultz 1979) while the ability to suppress distracters in memory was measured by the memory interference task that includes semantic (Deese 1959, Roediger & McDermott 1995) and form (orthographical/phonological) lures (Reinitz, Lammers & Cochran 1992). This represents a first attempt using an individual differences approach to narrow down the range of possibly relevant features for similarity-based interference. The flanker task is described in chapter 3, section 3.3.3 and the memory interference task follows in section 3.3.4. Additionally, while one could expect some processing difficulties to occur at the clause boundary (based on

¹⁶ Note that while this example makes use of a [+WH] feature, not all islands will have this as the interfering item. At first glance, it would seem that the best candidates for interference in CNPC and subject islands are the other DPs present in the sentence, though this may make untenable predictions with respect to non-island questions that include multiple DPs.

predictive processing), the key claim of the similarity-interference account is that the processing difficulty is in the retrieval process, which is cued by encountering the gap. Thus the largest processing difficulty is expected at or around the gap embedded in an island. This is a distinct pattern from what is expected under a capacity-constrained account (see section 2.3.3.1, above). Thus, if the main processing difficulty of the island violation sentence occurs at the embedded gap, this will favor the similarity-interference account over the capacity-constrained account.

2.4 Current research agenda

This dissertation uses an individual difference approach with three different methodologies: acceptability judgments (section 2.2.3; Chapter 4), self-paced reading (section 2.2.5.1; chapter 5) and ERPs (section 2.2.5.2; Chapter 6) in order to test whether a capacity-constrained (section 2.3.3.1) or similarity-interference (section 2.3.3.2) account of island phenomena better accounts for the data. The same materials design was used for each experiment (Chapter 3, section 3.2), and participants in all three experiments were scored on the same cognitive measures (Chapter 3, section 3.3), allowing for comparisons across experimental methods. The materials were designed to match lexical items before, at and after the gap positions across conditions, allowing on-line measurements to be made at each of these positions with minimal risk of artifacts. These experimental design considerations lead to a focus on *whether*-islands, which can be manipulated to meet the above requirements while maintaining a factorial design. The use of a factorial design allowed for the

examination of separate aspects of the *whether*-island: the nature of the clause boundary, whether the filler gap dependency extends into the embedded domain, as well as the interaction of the two.

The experiments presented in the remainder of the dissertation are focused on the processing accounts of islands. The reasons for this focus, rather than attempting to compare and contrast these approaches with grammatical or functional accounts of islands are threefold. First, in making use of measures of individual differences, it must be noted that these are cognitive measures, and thus finding co-variation with language data that is approached from a processing/cognitive standpoint is most directly interpretable. Second, while the debate between the grammatical, functional and cognitive accounts of islands is a significant intellectual pursuit, it is not the focus of this dissertation. It is crucial to be comparing the best example of a given account when each claims to have explanatory arguments for the data. Thus, rather than attempting to decide between them, this dissertation focuses on either strengthening or updating the processing account(s), as guided by the data. Thus, while this dissertation is not focused on the ‘cause of island phenomena’ debate per se, the results will be of interest to research that is, and relevant findings will be highlighted throughout. Additionally, regardless of which approach one deems best in explaining island phenomena, the fact remains that these sentences have generated processing data, and these data must be explained if we are to better understand the human language processing faculty, even if these processing data are not the ‘cause’ of the overall island effects. Thirdly, a focus on comparing working-memory-based approaches to

language leads to a comparison of underlying working memory models. These models of verbal working memory are of interest to researchers outside of the study of language as well as within. Note that this focus on the processing accounts is not to preclude the importance of the contributions of other accounts, but instead to be able to focus more intensely on how research at the intersection of language and cognition can help inform both sciences.

Chapter 3: General Methods: Materials and Cognitive Measures

3.1 Introduction

This chapter presents the methodologies that are common across all three experiments. Methodology specific to each experiment can be found in their respective chapters: Experiment 1, acceptability judgments in Chapter 4; Experiment 2, self-paced reading in Chapter 5; Experiment 3, event-related potentials in chapter 6. In section 3.2 I discuss the design of the linguistic stimuli that are used in all three experiments. In section 3.3 I discuss the design, procedure, and results for four measures of individual cognitive differences: reading span (section 3.3.1), n-back (section 3.3.2), flanker (section 3.3.3) and memory interference (section 3.3.4). Section 3.3.5 presents the co-variation matrix for these measures. Section 3.4 concludes the chapter.

3.2 Materials

The trio of experiments in the following chapters focus on using different methodologies (acceptability judgments, self-paced reading, and event-related potentials in Chapters 4, 5, and 6, respectively) to examine the same linguistic phenomena: *whether*-islands. Two sets of materials were developed. The first set, available in Appendix 1, was used for the acceptability judgment and self-paced reading experiments and contains 32 sets of sentences like the one found in Table 3-1, below. Since ERP studies require many more trials to be run, a second, larger set of

materials was developed that contains 160 sets of sentences. These sentences are available in Appendix 2. Both sets of sentences follow the same design outlined below.

As discussed in Chapter 2, an island violation can arise when two conditions are met. First, an island structure is present, and second, the filler is outside of the island and the gap is inside of the island. As such, the island violation for the *whether* island is operationalized here as having the island structure (as opposed to a non-island structure) and having the gap be embedded within that structure (as opposed to being outside of that island domain, i.e. in the matrix clause). This forms a 2 x 2 factorial design with the factor STRUCTURE having two levels: NON-ISLAND, ISLAND, and the factor GAP having two levels: MATRIX, EMBEDDED. An example stimuli set is presented in Table 3-1.

Table 3-1: Sample stimuli set. Manipulation of STRUCTURE indicated in bold. Manipulation of GAP indicated by italics. No specific claims are intended by the placement of the gap, which is meant only to indicate the on-line point of disambiguation of the gap position.

		STRUCTURE	
		NON-ISLAND	ISLAND
GAP	MATRIX	Condition 1: Who had <i>_ openly</i> assumed [that the captain befriended <i>the sailor</i> before the final mutiny hearing?]	Condition 2: Who had <i>_ openly</i> inquired [whether the captain befriended <i>the sailor</i> before the final mutiny hearing?]
	EMBEDDED	Condition 3: Who had <i>the sailor</i> assumed [that the captain befriended <i>_ openly</i> before the final mutiny hearing?]	Condition 4: Who had <i>the sailor</i> inquired [whether the captain befriended <i>_ openly</i> before the final mutiny hearing?]

Because these stimuli will be used for online measures (self-paced reading/event-related potentials), many more aspects of the sentences have been controlled for than is typical for sentences used only in an off-line acceptability judgment study (like the one implemented in Chapter 4). In order to explain the various constraints involved in constructing these stimuli sets, I focus now on the NON-ISLAND conditions (condition 1 and 3), which are presented again in Table 3-2.

Table 3-2: Sample NON-ISLAND stimuli

1	2	3	4	5	6	7	8	9
MATRIX GAP (condition 1):								
Who	had	<i>_ openly</i>	assumed	[that	the captain	befriended	<i>the sailor</i>	before ...
EMBEDDED GAP (condition 3):								
Who	had	<i>the sailor</i>	assumed	[that	the captain	befriended	<i>_ openly</i>	before ...

Table 3-2 shows the example stimulus sentence divided into 9 positions. These positions represent the presentation of word(s) in the self-paced reading and ERP experiments. Sentences were presented in their entirety for the acceptability judgment experiment. One core concern addressed in the building of these stimuli was keeping the length of all the conditions equal and consistent throughout a trial. Van Petten and Kutas (1990) reported a decrease in N400 amplitude as word position increases (becomes later) in a sentence. As the N400 response is one of the possible responses to be elicited in Experiment 3 (Chapter 6), it is critical that comparisons occur at the same ‘depth’ into the sentence.

In all sentences, including fillers, the first two positions were always *who* and *had*. Using an auxiliary that undergoes subject-auxiliary inversion is required for the condition in Table 3-2 where the gap is EMBEDDED (i.e. not in the matrix subject position). Compare the sentences in (3.1).

(3.1 a) Who did the sailor see _?

(3.1 b) *Who the sailor (did) see _?

When the gap is in the matrix subject position it is possible to form a *wh*-question without subject-auxiliary inversion (*Who _ saw the sailor?*) but then the MATRIX GAP sentences would be one word shorter than the EMBEDDED GAP sentences. Therefore both sentence types included the auxiliary *had*.

Position 3 is the matrix clause gap position. If the sentence has a MATRIX GAP, this is the position when that is known. That is, this position indicates the on-line point of disambiguation where the reader knows whether the sentence has a matrix clause gap or not. The gap, represented in Table 3-2 as an underscore (_) is not marked for the participants in any way (they do not see an underscore). As can be seen in position 3 (Table 3-2), this gap co-occurs with an adverb, *openly*. While the gap is marked in the sample materials as adjacent to and before *openly*, this should not be interpreted as a claim that the gap is necessarily in a pre-adverbial position. This representation of ‘ _ *openly*’ is used as a notation of convenience in order to parallel the embedded clause gap position. Theoretical arguments could be made for the location of the matrix

clause gap being immediately after the filler, or after the adverb. Neither of these options would alter the predictions or inferences of this dissertation. What is critical, however, is that this is the position of the sentence where it is clear that there is a gap in the matrix clause. That is, the sentence cannot grammatically continue from this point with an overt NP following this adverb (see 3.2).

(3.2) * Who had openly the sailor assumed that the captain befriended _ ?

Use of the adverb in position 3 crucially keeps the MATRIX sentences from being one position shorter than the EMBEDDED sentences until they would ‘catch up’ at position 8. For the EMBEDDED conditions in position 3, Table 3-2 shows two words: *the sailor*. For both the self-paced reading and ERP experiments, full determiner-noun noun phrases were presented simultaneously at position 3. This presentation could have been reduced to one word, either by employing bare plurals (e.g. *sailors*) or proper names (e.g. *James*), but the former sounded less natural and the latter would be more difficult to control for frequency.

As two different sets of lexical material would be compared in position 3 (*openly* and *the sailor*), these were controlled for length and frequency (using log HAL frequency, Balota et. al. 2007), not including *the* in the noun phrases. As *the* is the most frequent word in English, it would be nearly impossible to balance frequencies between the adverbs and nouns with *the* included in the calculations. The frequency and length values are shown in Tables 3-3 and 3-4. Since the number of

sentences used for the ERP experiment is much larger than the number used for acceptability judgments and self-paced reading, the frequencies and lengths involved differ for this material set. However, in no case was there a statistically significant difference of mean length or frequency within a set of materials.

Table 3-3: Position 3 & 8 controls for Experiments 1 and 2. Mean (Standard deviation)

	Log HAL Frequency	Length
Adverbs	7.59 (2.66)	7 (1.11)
NPs	7.69 (1.59)	7.03 (1.69)

Table 3-4: Position 3 & 8 controls for Experiment 3. Mean (Standard deviation)

	Log HAL Frequency	Length
Adverbs	5.58 (2.99)	9.52 (2.28)
NPs	5.72 (2.49)	8.45 (2.20)

Care was taken to choose adverbs that not only matched their position 3 nouns in frequency and length, but also to choose adverbs that were compatible with both the matrix verb (*assumed*) and the subordinate verb (*befriended*; each of which can be done *openly*). This is because the adverbs used in the matrix gap position for MATRIX GAP sentences would also be used in position 8, the embedded gap position, for EMBEDDED GAP sentences.

The ability to use these adverbs in multiple sentence positions in this way was one of the determining factors in choosing *whether*-islands among all the island types to examine. Other island types proved to be much more difficult to construct a balanced factorial design for. Constructing sentences where the gap could be either embedded within the island or outside of it while still controlling for frequency as best

as possible required extremely complex sentences that would have likely introduced more confounds than they might mitigate.

In order to further ensure that the presence of these adverbs would not introduce additional confounds to the *whether*-islands, a pilot study was conducted with 172 native English speakers. The study was a 2 x 2 x 2 design, comparing the conditions in Table 3-1 (STRUCTRE and GAP) with the presence or absence of these adverbs. There was no main effect of presence/absence of adjuncts, nor were there any interactions involving the presence/absence of adjuncts (all $F_s < 0.64$). This indicates that the addition of these adverbs to the GAP and STRUCTURE manipulations does not alter the pattern of acceptability of these sentences.

To summarize, the use of these adverbs serve to keep the two different GAP conditions aligned, position for position. They are controlled for frequency and length, and they are pragmatically compatible with both the matrix and embedded verbs. Finally, a pilot study indicates that sentences like those in Table 3-1 are equally acceptable with or without these adverbs.

The matrix verb is encountered at position 4. Only verbs that resisted an interpretation of having an immediately post-verbal gap were used. This was to avoid participants attempting to posit and/or fill a gap in the matrix clause instead of the embedded clause. Compare (3.3) where no immediate post-verbal gap is possible and (3.4) where it is. Verbs that patterned like (3.4) were avoided.

(3.3) Who had the sailor inquired (* _) whether the captain ...

(3.4) Who had the sailor asked (_) whether the captain ...

Some of the verbs used allow a (linearly) post-verbal gap as the head of a complement clause (3.5), but these options were precluded if *that* or *whether* followed the matrix verb (3.6), which it always did.

(3.5) Who had the sailor declared _ was the winner?

(3.6) *Who had the sailor declared _ that was the winner?

When possible, verbs that could take both a declarative embedded clause (NON-ISLAND conditions) and an interrogative embedded clause (ISLAND conditions) were used (*deduced, said*). Otherwise, declarative complement verbs (*assumed, contended, declared*) were frequency and length matched with interrogative complement verbs (*inquired, speculated, wondered*). This is shown in Table 3-5. There was not a statistically significant difference of mean length or frequency for these verbs.

Table 3-4: Position 4 matrix verb controls. Mean (Standard deviation)

Verb type	Log HAL Frequency	Length
Declarative complement	8.73 (2.77)	7 (1.87)
Interrogative complement	8.38 (2.68)	7.4 (2.19)

Position 5 is the clause boundary position. This position was always filled with either the declarative complementizer *that* (NON-ISLAND conditions) or the

interrogative complementizer *whether* (ISLAND conditions). In addition to the obvious length difference, *that* is more frequent than *whether* (log HAL 15.48 vs. 11.63). As these are individual words, no further methodological controls can be used here. Any differences observed between these words will be discussed in the appropriate analyses. Note also that the clause boundary marker '[' is used in this presentation to aid the reader, but this bracketing was not visible to participants.

In position 6, *the captain* is again a pair of words presented simultaneously. This is done so that it is not only at the critical gap sites (MATRIX: position 3; EMBEDDED: position 8) that two-word presentations occur. One additional difference between the stimuli used for the self-paced reading experiment and those used for the ERP experiment occur with respect to two word presentation. Word position 10 (not illustrated in Table 3-2) is *the final* in the ERP experiment, but only *the* in the self-paced reading. As such, the self-paced reading materials are 13 positions long, while the ERP materials are 12 positions long. This was a result of trying to shorten the amount of time participants in the ERP experiment had to keep from blinking. This position is not critical to either the self-paced reading or ERP analyses.

Position 8 is the inverse of position 3, and all the same controls apply. All other sentence positions are identical across conditions.

3.3 Measures of Individual Differences

The following sections detail the methods of collecting the individual differences measures. The order of the sections below corresponds to the order in

which the measures were administered: reading span, n-back, flanker and memory interference. All four of these tasks were completed using the e-prime software program (Schneider, Eschman, & Zuccolotto 2002) on an HP laptop with a 14” diagonal screen running Windows XP.

Participants used an ‘X-box’ style video game controller to respond to each task, but only used three of the many buttons present on the controller. The ‘A’ button, reached by the thumb of the right hand was used most frequently to advance through the tasks. The ‘Left’ and ‘Right’ shoulder buttons, reached by the left and right index fingers, respectively, were used when the participant needed to make one of two alternative responses. The buttons used for the correct response were counter-balanced across participants where appropriate. The experimenter indicated which buttons would be used before the tasks began. No participants had difficulty using the correct buttons. Video game controllers are designed to have fast and accurate timing responses and can easily be configured to work with the e-prime software.

In addition to the game controller, a USB microphone was used for the reading span task to record the participants’ responses for later verification of the responses if needed. Participants sat 18”-24” from the laptop screen, with the microphone in front of the laptop and pointed towards them. Participants were free to hold the controller above or below the desk and were free to adjust the chair and angle of the laptop screen for ease of viewing and comfort. In all cases text was presented as black on a white background in Courier New 18 point font unless specifically noted otherwise. The experimenter remained in the room with the participant for the reading span task,

to ensure that the task was completed promptly and properly, to record responses, and to respond to any questions posed by the participants. Upon completion of the reading span task, the experimenter left the room (in order to, for example, set up the ERP capping station) and the participant completed the remainder of the tasks on their own. Participants were instructed to seek out the experimenter if they had questions.

There were many opportunities for participants to take breaks or ask questions. Participants were told that whenever they were at a screen that stated 'Press A to continue,' that they could take a break or approach the experimenter with a question. Before the tasks were begun, participants were asked to turn off their cell phones or similar devices. All tasks below used the exact same stimuli in the exact same order for each participant. The entire set of measures took between fifteen and twenty-five minutes for each participant. Informed consent was obtained for all of the following tasks.

For each of the following tasks, the procedure and instructions will be presented, followed by the scoring method and results. Results are organized into a table showing the mean, standard deviation and median scores for participants in each of the three experiments in the dissertation as well as a grand aggregate of all 160 participants for all three experiments. In addition to the median, a low/high count is provided. This is to indicate how many participants were assigned to the low and high scoring groups for the purposes of median splits submitted to repeated measures ANOVA analyses for each experiment. This analysis was chosen as a 'common denominator' analysis that has previously been used in all three methodologies

employed in this dissertation.¹ Some analyses that are straightforward to do in the more simple acceptability judgment procedure such as a linear mixed-effects model, for example, are difficult to implement in an event-related potential study where one must contend with a noisier signals, reliance on averaging across more trials per participant and complexities of the distributional analysis of electrodes. In order to keep comparisons across experiments as straightforward as possible, this same median split approach was used for each of the three experiments.

3.3.1 Reading Span

The reading span task (Daneman & Carpenter 1980) is a complex span task, wherein both storage and processing ability are required to complete the task successfully. Participants must read sentences aloud, thus requiring the engagement of all the normal processes required to do so (the processing component). Additionally, they are tasked with remembering the final word of each of the sentences they read and recalling them in order (the memory component). This is in contrast to simple span tasks such as digit span or serial recall tasks wherein the task can be completed successfully by memory alone, without the need to engage any additional processing

¹ For example: in ERP studies, subjects have been grouped by median splits of sentence comprehension (King & Kutas 1995, Müller, King, & Kutas 1997), reaction times (Reinhart, Carlisle, Kang, & Woodman 2012) and most importantly for the current research, cognitive measures (e.g. Hampton Wray & Weber-Fox 2013; Nakano, Saron, & Swaab 2010). Median splits on cognitive scores used recently in self-paced reading include (e.g. Borovsky, Elman, & Fernald 2012; Soederberg Miller, Cohen, & Wingfield 2006), though some groups prefer to organize participants into a three-way high/medium/low distinction (e.g. Bornkessel, Fiebach, & Friederici 2004; King & Just 1991; Waters & Caplan 1996).

of the stimulus beyond what might be automatic and required for the memory trace to be established.

3.3.1.1 Task

Participants were given the following instructions on screen:

‘In this task you will read sentences into the microphone.
After each sentence you will press A for the next sentence.
Be careful not to press A too early or the sentence you are reading will
vanish!
After reading the sentences, you will be asked to recall the last word of
each sentence IN ORDER.
Try your best to get the words IN THE CORRECT ORDER, but if you
can’t, just say the ones you can recall.’

Participants were then given the following example:

“If you see the sentences
“Bob saw Mary.”
“The dog is in the car.”
“The apples are bright red.”
You will read each of these out loud.
Then when you see the slide that says:
“Repeat the last word of each sentence.”
You will say: “Mary, car, red” ”

If participants had no questions, they began the task. The task was broken into three sections. In the first section, three sentences were presented before the “Repeat the last word of each sentence” prompt occurred. This was thus measuring a reading span of three. After the first section, participants saw a slide which read: “Now you will do the same thing, except there will be four sentences. Remember to read each sentence out

loud.” In the second section, four sentences were presented before the recall prompt, measuring a span of four. After the second section participants saw a slide which read: “Now you will do the same thing, except there will be five sentences. Remember to read each sentence out loud. This is the last group of sentences for this task.” In the third section, five sentences were presented before the recall prompt, measuring a span of five. Each of the sections had five trials, each trial consisting of a group of sentences to read and a recall prompt. Thus each participant read $(3 \times 5 =)$ 15 sentences at the 3 span level, $(4 \times 5 =)$ 20 sentences at the 4 span level and $(5 \times 5 =)$ 25 sentences at the 5 span level for a total of 60 sentences. Participants were not stopped from continuing if they could not complete a majority of trials from a span level.

Participants’ responses and the order of the responses were marked on a sheet containing all the correct final words for the sentences by the experimenter, who sat behind the participant while the task was completed. Additionally, responses were recorded via a microphone and the e-prime software in case any results needed to be double-checked at a later time.

At the end of the task participants saw a slide that read: “Great. Feel free to take a little break now.” At this point, the experimenter excused himself from the room, telling the participant to come out of the room when they were done, and additionally to feel free to come out and ask questions if any of the tasks or their instructions were confusing. Participants were then allowed to proceed through the rest of the individual differences tasks on their own.

Participants frequently expressed surprise at how difficult the reading span task was. Additionally, once all the individual tasks were completed, many participants expressed that they ‘did better’ at the other tasks.

3.3.1.2 Scoring

Participants were scored using the partial credit method (Conway et al. 2005), gaining one point for each correct sentence final word that they recalled out of a possible score of 60. This method of scoring was chosen over the original method of scoring for two reasons. In the original method of scoring, participants would earn a span score equal to the highest level at which they could answer a certain portion of trials with complete accuracy. Once this criterion could not be met for a certain span level (due to too many incorrect responses) the task was halted. I have found that this procedure results in the potential loss of data as some participants can perform poorly on the low span levels but then increase their performance greatly at later span levels. Because of this pattern, presumably due to understanding the task better after the initial trials, I did not want to prematurely stop the task. Secondly, the partial credit method generates a wider range of scores (0-60 in this case) compared to the original scoring method, which ranges up to 6. This is useful for forming balanced median split groups for analysis.

3.3.1.3 Results

The results for the reading span measure are presented in Table 3-5. Higher values indicate a higher working memory score.

Table 3-5: Reading span results across three experiments

	Mean (SD)	Median	low/high
Acceptability judgment (n = 80) (Experiment 1, Chapter 4)	38.45 (6.45)	38	40/40
Self-paced reading (n = 48) (Experiment 2, Chapter 5)	38.74 (7.70)	40	22/26
Event-related potentials (n = 32) (Experiment 3, Chapter 6)	40.66 (5.31)	41	15/17
All participants combined (n=160)	38.99 (6.65)	39	

3.3.2 N-back

The n-back measure (Kirchner 1958) is a commonly used working memory task where participants must not only store representations for recall, but constantly update these representations. The basic task involves indicating whether a stimulus that is currently being observed is the same as a stimulus seen n stimuli ago. When n is 1, this is a fairly simple task, but as n increases to 2 or higher, multiple representations need to be held, attended to, and updated.

3.3.2.1 Task

The n-back task has the most complex instructions of any of the individual differences tasks that participants completed for this dissertation. As such, the

instructions encourage them to seek out the experimenter for clarification if the instructions are not clear to them. The task began with the following information:

“This next task has the most complicated directions of any of the tasks.
Please ask the experimenter for assistance if you have difficulty
understanding what you are supposed to do.”

This was followed by the initial instructions:

“In this part of the experiment you will need to remember what letter
you just saw and compare it to what you currently see.
If the letter that you see matches a letter you just saw, you will press the
A button.
If the letter that you see does NOT match a letter you just saw, you will
NOT press any buttons.
Try to respond as quickly and accurately as possible.”

As indicated by the instructions above, all of the stimuli presented for the n-back task were letters, specifically: F H K L T V X Z in Courier New Bold 36 point font. These eight letters were chosen because participants would be highly familiar with them, but they also have overlapping visual features (vertical lines: F, H, K, L, T; horizontal lines: F, H, L, T, Z; and diagonal lines: K, V, X, Z) so it would not be too simple to distinguish between these letters in memory.

Three levels of n were tested for the n-back task: 1, 2 and 3. The 1-back served as a familiarization for the participants before the more difficult 2- and 3-back levels. Separate instructions were given before each level. First, the 1-back:

“You will see a list of letters, one letter at a time.
If the letter that you are currently looking at matches the letter you saw
EXACTLY ONE LETTER AGO, you will press the A button.
Otherwise, do not press anything.
For example, if you saw: O E E
You would do nothing for the first two letters, but press the A button for the
third, since it matches the letter exactly one space before it.
Do you have any questions?
Press the A button to begin”

Note that participants are again reminded that the experimenter can be asked for clarification if the task is unclear. Participants only had to respond (via button press) if they thought the letter they were currently looking at was the same as a letter n letters ago. The trial began with a fixation cross for 500 ms, followed by a letter for 1000 ms. This was repeated for a series of 15 letters. Each letter was followed by a fixation cross so it would be clear when the letters were being updated. Of the 15 letters, 5 of them should have elicited a response from the participant (they matched the letter one before them- which I will refer to as a ‘match’ condition) while the other 10 should have not. The letters and correct responses were psuedo-randomized so that correct hits were distributed throughout the trial, including back-to-back correct hits. The software recorded whether the participants pressed the button for each letter.

Following the 1-back trial, the instructions for the 2-back were presented:

“Great, Now you are going to do the same thing, EXCEPT, you will only press the A button when the letter you see matches what you saw EXACTLY 2 LETTERS AGO.

Otherwise, do not press anything.

For example, if you saw: A O E O E

You would press the A button at the second 'O' since exactly two letters ago there was an 'O'. You would also press the A button at the second 'E'.

You would not press A for any of the other letters.

Press the A button to begin”

The procedure for the 2-back was the same as the 1-back except 30, rather than 15, letters were presented. Of these 30, 10 of them (the same proportion as in the 1-back: 1/3) were matches and should have elicited responses from the participants.

Finally, after the 2-back task was completed, the instructions for the 3-back task were given. The procedure was the same as the 2-back.

“Great, Now you are going to do the same thing, EXCEPT, you will only press the A button when the letter you see matches what you saw EXACTLY 3 LETTERS AGO.

Otherwise, do not press anything.

For example, if you saw: O O E O E

You would press the A button at the third 'O' since exactly three letters ago there was an 'O'.

You would not press A for any of the other letters.

Press the A button to begin

(This is the last one of this task)”

3.3.2.2 Scoring

Accuracy was recorded separately for each level of the n-back ($n = 1, 2$ and 3). For each level a total accuracy score was obtained counting the number of correct responses to match conditions as well as correct lack of responses to non-match conditions (that is when participants did not press the button when they shouldn't

have). This generates an accuracy figure with the highest possible score being 15/15 for the 1-back and 30/30 for the 2- and 3-back levels. Only the 3-back accuracy was used in the analyses for Experiments 1, 2 and 3. The 1-back task was extremely easy and there was little variance in how well participants did. There was slightly more variance in the 2-back, but the most differentiation in scores was in the 3-back task.

3.3.2.3 Results

The results for the n-back measure are presented in Table 3-6. Higher values indicate a higher working memory score.

Table 3-6: N-back (3-back) results across three experiments

	Mean (SD)	Median	low/high
Acceptability judgment (n = 80) (Experiment 1, Chapter 4)	0.80 (0.08)	0.80	37/43
Self-paced reading (n = 48) (Experiment 2, Chapter 5)	0.74 (0.09)	0.73	20/28
Event-related potentials (n = 32) (Experiment 3, Chapter 6)	0.81 (0.06)	0.80	15/17
All participants combined (n=160)	0.78 (0.08)	0.78	

3.3.3 Flanker

The Eriksen flanker task (Eriksen & Eriksen 1974; Eriksen & Schultz 1979) using arrows (Kopp, Mattler, & Rist 1994) as stimuli has been used as a measure of processing speed and selective attention. A target stimulus in the center of an array of stimuli is responded to. The stimuli surrounding the target stimulus would either generate the same response as the target (these are congruent flankers) or they would generate the opposite response (incongruent flankers). Participants are tasked with

responding only to the target stimulus while ignoring the flankers. Responses to the target with congruent stimuli give a measure of reaction speed while the amount that a participant's reaction slows when the target is surrounded by incongruent stimuli gives a measure of how susceptible they are to interference.

3.3.3.1 Task

Participants were first instructed that: "In this next task, you will pay attention to the direction of the center Arrow while ignoring the other Arrows." They were then shown a fixation cross which was replaced by a right-facing arrow. In order to proceed, they needed to press the corresponding button on the control (right button on the controller). Then participants were presented another fixation cross, which was replaced by a left-facing arrow flanked by a right-facing arrow on each side. They were instructed to respond only to the center arrow (by pressing the left button). The instructions were clear to distinguish the right and left *buttons* on the controller, which were used in this task, from the right and left *triggers* on the controller, which were not used and would not advance the participant through the instructions if pressed.

After this short training, participants completed 32 trials, split 50/50 between right-facing and left-facing targets, and split 50/50 between congruent and incongruent flankers (example in Figure 3.1). These were counterbalanced such that there was a 50/50 split in whether these generated 'match' or 'no match' conditions.



Figure 3.1: Right-facing arrow surrounded by incongruent flankers

Arrows were bitmap images that were mirror images of each other, as shown above. In each trial, the fixation cross appeared for 500 ms, followed by the array of arrows for 1000 ms. The target arrow was always presented in the same location as the fixation cross.

3.3.3.2 Scoring

Two scores were recorded for each participant from the flanker task. Only correct responses were analyzed. First, average reaction time to congruent trials formed the ‘reaction time’ measure. The ‘interference’ measure was obtained by subtracting the average reaction time of congruent trials from the average reaction time of incongruent trials. Thus a higher flanker interference score indicates more susceptibility to interference. As susceptibility to interference is one of the focuses of the current studies, but reaction time is not, only the ‘interference’ measure was analyzed in Experiments 1, 2 and 3.

3.3.3.3 Results

The results for the flanker measure are presented in Table 3-7. Higher values indicate a larger reaction time penalty in the presence of distractors. Thus lower values indicate less susceptibility to interference.

Table 3-7: Flanker (incongruent - congruent) results across three experiments (msec)

	Mean (SD)	Median	low/high
Acceptability judgment (n = 80) (Experiment 1, Chapter 4)	32.78 (30.05)	30.94	40/40
Self-paced reading (n = 48) (Experiment 2, Chapter 5)	40.22 (35.23)	40.29	24/24
Event-related potentials (n = 32) (Experiment 3, Chapter 6)	34.40 (46.17)	32.82	16/16
All participants combined (n=160)	35.26 (35.21)	34.94	

3.3.4 Memory interference

The memory interference task is an old-new recognition task for words (Warrington 1984) that includes semantic (Deese 1959, Roediger & McDermott 1995) and form (orthographical/phonological) lures ('feature' lures in Reinitz, Lammers & Cochran 1992). In an old-new recognition task participants are given a list of items to study during the study phase. During the test phase, participants are presented with items that were either on the study list (old items) or were not (new items). Participants indicate whether the test items are old or new. The current task follows this same pattern except that some of the new items are similar to some of the old

items in a particular feature: semantic (i.e. *cheetah* ~ *jaguar*) or form (orthographic/phonological; i.e. *grass* ~ *glass*).

3.3.4.1 Task

Before starting this task, participants were trained to use the left and right buttons to indicate a ‘yes’ or ‘no’ response. The buttons being used for each answer were counterbalanced across participants such that half of the participants would use the left button to indicate ‘yes’ and the other half would use the left button to indicate ‘no.’ Participants needed to correctly associate the buttons with the ‘yes/no’ responses to progress through the instructions and complete six test prompts (‘press the yes button’).

After completing the ‘yes/no’ button training, participants saw the following instructions:

“In this task you will memorize a short list of ten words.
Then, you will be presented with ten more words, some of which you
memorized, and some of which you didn't.
You will press 'yes' if you see a word you memorized.
Press 'no' if you see a word you didn't memorize.
Each word will only be on the screen for a short time, so be sure to pay close
attention.
There will be three lists overall.
Press A to continue”

Each of the three study trials started with a 500 ms long fixation cross followed by a study word for 1500 ms. This process (fixation cross then word) repeated until all ten

study words had been presented. After all ten words were presented, participants saw the following:

“Now there will be ten more words.
Press 'yes' if you memorized the word.
Otherwise press 'no'
Each word will be on the screen for about a second.
Press A to continue”

The ten test words were presented in the same format as the study list (500 ms fixation, 1500 ms word) and participants' button presses (indicating 'yes' and 'no') were recorded.

This procedure was repeated two more times for a total of 30 study words in three lists. Half of the test words in each list were study words (old) and the other half were new. The new words were one of three types: unrelated to old words, semantically related to the old words (semantic lures), or orthographically/phonologically related to the old words (form lures). Of the fifteen total new words, five were in each of these categories. The new words were distributed across the three test blocks such that there were three representatives of a given category and one of each other category in a test list. For example, one test list consisted of 5 old words, 1 new unrelated word, 3 semantic lures and 1 form lure.

3.3.4.2 Scoring

The memory interference task generated four scores. The memory score indicated how many of the 15 old words were correctly identified as such during the

test phase. The memory lure score is a count of how many of the 10 lure conditions a participant gave a correct response to (they were not lured). This is the key score used in the following three experiments. However, this can further be broken down into the semantic lure and the form lure scores. When possible in the analyses for Experiments 1, 2, and 3 the separate semantic lure and form lure scores are examined. As shown in section 3.3.5, below, these scores are only marginally correlated with each other.

3.3.4.3 Results

The results for the memory lure are presented in Table 3-8. The form lure is in Table 3-9 and the semantic lure in Table 3-10. In all cases higher values indicate better accuracy in the face of memory lures. Thus higher scores indicate less susceptibility to similarity-based interference in memory.

Table 3-8: Memory Lure results across three experiments

	Mean (SD)	Median	low/high
Acceptability judgment (n = 80) (Experiment 1, Chapter 4)	0.77 (0.15)	0.80	40/40
Self-paced reading (n = 48) (Experiment 2, Chapter 5)	0.82 (0.13)	0.80	25/23
Event-related potentials (n = 32) (Experiment 3, Chapter 6)	0.83 (0.13)	0.85	15/17
All participants combined (n=160)	0.79 (0.14)	0.80	

Table 3-9: Form Lure results across three experiments

	Mean (SD)	Median	low/high
Acceptability judgment (n = 80) (Experiment 1, Chapter 4)	0.75 (0.22)	0.80	40/40
Self-paced reading (n = 48) (Experiment 2, Chapter 5)	0.79 (0.23)	0.80	26/22
Event-related potentials (n = 32) (Experiment 3, Chapter 6)	0.69 (0.25)	0.70	17/15
All participants combined (n=160)	0.75 (0.23)	0.80	

Table 3-10: Semantic Lure results across three experiments

	Mean (SD)	Median	low/high
Acceptability judgment (n = 80) (Experiment 1, Chapter 4)	0.79 (0.19)	0.80	38/42
Self-paced reading (n = 48) (Experiment 2, Chapter 5)	0.85 (0.16)	0.80	26/22
Event-related potentials (n = 32) (Experiment 3, Chapter 6)	0.81 (0.21)	0.80	15/17
All participants combined (n=160)	0.81 (0.19)	0.80	

3.3.5 Co-variation matrix

Scores from the four measures presented above, reading span, n-back, flanker and memory lure will be used to test for co-variation with the acceptability judgments, reading time and brain responses of/to linguistic stimuli (Experiments 1, 2, and 3 respectively). As such, it is important to know to what degree, if any, these measures are correlated with each other.

Table 3-11 provides the Pearson's r correlation values for these measures based on all 160 participants from all three experiments. As we can see, only one pair of measures reaches statistical significance: reading span and memory lure ($r = 0.31$, $p < 0.001$).

Table 3-11: Correlation matrix: (Pearson's r), all experiments (n = 160)

	Flanker	N-back	Reading span	Memory lure
Flanker	--			
N-back	-0.03	--		
Reading span	0.04	0.15	--	
Memory lure	0.03	< 0.01	0.31 ***	--

It is unsurprising that there is some correlation between reading span and memory lure. Reading span is intended to be a measure of memory and processing

while memory lure is intended to be a measure of memory and (lack of) susceptibility to interference. To the extent that the same memory process is involved in each of these tasks, a modest correlation is expected. However, while this correlation is significant, it is modest. There is still a substantial amount of variance that each task is capturing that the other does not. Some of this variance is presumably the processing (sentences) and interference components of the reading span and memory lure tasks, respectively. As such, they should still prove to be useful as largely independent measures. In fact, in the experiments that follow we see numerous dissociations between these measures in terms of which measure co-varies with a linguistic manipulation.

The memory lure score is formed by responses to both semantic and form lures. When these scores are checked for correlation, we find that the semantic and form lures are only marginally correlated ($r = 0.14$, $p = 0.08$). Additionally, the semantic lure is not significantly correlated with reading span ($r = 0.12$, $p = 0.13$), while the form lure is ($r = 0.26$, $p < 0.001$). We have the opportunity then to see the two aspects of the memory lure task as separate but related measures. Thus, when possible, the memory lure analyses in Experiment 1, 2, and 3 will include additional analyses indicating the patterns of the semantic and form lure.

3.4 Conclusion

The material design and cognitive measures outlined here represent the common methodology used in Experiments 1, 2, and 3. Any methodology that is unique to these specific experiments is discussed in its relevant chapter (acceptability

judgments, Chapter 4; self-paced reading, Chapter 5; event-related potentials, Chapter 6).

Chapter 4: Acceptability Judgment Experiment

4.1 Introduction

This chapter presents an acceptability judgment study of *whether*-islands that examines co-variation of those judgments with measures of individual cognitive differences. Ultimately, this experiment does not support a processing account of islands. The data do not support either a capacity-constrained account of islands or a similarity-interference account of islands. However, this lack of direct support should not be taken as counter-evidence. Issues surrounding the apparent simple intuition connecting acceptability and working memory are discussed in detail (section 4.2.2.1) that make the lack of support for these processing views unsurprising.

The basis for the acceptability judgment experiment presented in this chapter is an intuition, which I call the Cognitive Co-variation Intuition (CCI, Michel 2013). The intuition is simply that if island phenomena are due to working memory related processing costs (e.g. Kluender & Kutas 1993b) then individuals who have greater working memory capacities should be able to process the island violation sentences better and thus rate them as more acceptable. This intuition is outlined in more detail in (4.1).

- (4.1) Cognitive Co-variation Intuition (CCI) applied to island phenomena
- a) If the unacceptability of a sentence (here specifically an island violation) is due to processing difficulties
 - b) And these processing difficulties arise from constraints on measurable cognitive resources (such as WM)
 - c) Then those individuals with a measurably greater cognitive score are expected to process the sentence (island violation) more easily
 - d) And this will result in these high-scoring individuals rating these difficult to process sentences as more acceptable

In the discussion that follows, it will become more clear that there are a number of assumptions made that are inherent in the CCI, some of which are not borne out by the data from the current experiment. One of the goals of this chapter is to examine this intuition more closely and to determine what our expectations about the co-variation of cognitive scores and acceptability judgments should reasonably be.

Sprouse, Wagers and Phillips (2012, henceforth SWP) conducted an independent study based on this same intuition. The remainder of this chapter is organized as follows. In section 4.2 I review the work by SWP and discuss how the current study can be seen to extended and improve it. I also present a framework of expectations for how cognitive measures and acceptability judgments might interact (section 4.2.2). Section 4.3 presents the methods of the current experiment, though for details about the measures of individual differences or materials design see Chapter 3.

Section 4.4 presents the results of three different analyses, with discussion after each. Section 4.5 summarizes the findings from these analyses and section 4.6 concludes the chapter.

4.2 Background

Sprouse, Wagers and Phillips (2012, henceforth SWP) conducted a study based on their own version of the Cognitive Co-variation Intuition (CCI, though not formulated the same way as in 4.1). SWP tested four types of island phenomena (*whether*-islands, subject islands, adjunct islands and complex noun phrase violations) while the current study focuses on *whether*-islands (for reasons discussed in Chapter 3). However, the research done by SWP has been criticized by Hofmeister, Staum-Casasanto and Sag (2012a,b). The current research, though independently conducted, addresses many of Hofmeister, Staum-Casasanto and Sag's concerns, as discussed below (section 4.2.1).

The research agenda of SWP is clear and specific: by looking for co-variation of the judgments of island phenomena with working memory, they set out to test Kluender's processing account of islands (Kluender 1991, 1998; Kluender & Kutas 1993a,b). Kluender argues that the degradation of island violation sentences can be accounted for by processing costs, specifically, a too-high burden on the working memory system (see Chapter 2 for discussion.) The types of sentences SWP tested are like those in (4.2):

(4.2) A factorial design for island effects:	STRUCTURE × GAP POSITION
a. Who ___ thinks [that John bought a car?]	NONISLAND MATRIX
b. What do you think [that John bought ___?]	NONISLAND EMBEDDED
c. Who ___ wonders [whether John bought a car?]	ISLAND MATRIX
d. *What do you wonder [whether John bought ___?]	ISLAND EMBEDDED

Modified from SWP

Like the materials used in this dissertation (see Chapter 3), the four sentences in (4.2) are arranged into a set of 2 x 2 comparisons. The factor STRUCTURE has two levels, NON-ISLAND (4.2a,b) and ISLAND (4.2c,d). The factor GAP POSITION also has two levels, a MATRIX gap (4.2a,c) and an EMBEDDED gap (4.2b,d). It is only with a certain combination of factors, the EMBEDDED ISLAND condition in (4.2d), that the sentence is deemed to be unacceptable. Neither the STRUCTURE itself nor the GAP POSITION are enough to generate an island violation. The ‘island effect’ then, is a combination of factors resulting in the EMBEDDED ISLAND condition being deemed the least acceptable of the four. This characterization is true whether one approaches the issue from a theory of grammar or a theory of processing. It is only when the filler is outside, and the gap is inside an island structure that a violation occurs.

In the aggregate, acceptability ratings given to island violations like (1d) can be characterized as superadditive since the rating for the island violation condition is less than the sum of any ‘penalties’ given for STRUCTURE or GAP POSITION (c.p.,

Fukuda, Goodall, Michel & Beecher 2012; Michel & Goodall 2013). SWP operationalize this superadditivity in acceptability ratings with a differences-in-differences score (DD). The DD score represents how much lower (presumably) a person rates the island violation in (4.2d) than could be expected from the independent effects of STRUCTURE and GAP POSITION. The DD score is thus a measure of an interaction effect: a higher DD score represents a larger superadditive effect, and a lower score represents a smaller superadditive effect.

SWP then fit the DD score to a simple linear regression with working memory measures that the participants had taken (serial recall and n-back). If the capacity-constrained account of islands is correct, then the individuals scoring higher on the working memory tasks were predicted to have a smaller DD score. These individuals will have more working memory resources and will thus be less troubled by the processing difficulty, resulting in less of a superadditive penalty being applied to the island violation condition. This correlation was not found, however, and SWP concluded that the capacity-constrained account of islands was not supported.

The approach that SWP took, as well as the interpretation of their results, has not been without criticism. Specifically, in a pair of replies, Hofmeister, Staum-Casasanto and Sag (2012a,b; henceforth HSCS) raise a number of issues concerning SWP's study, including (1) questioning whether we have reason to think that offline acceptability judgments will or should co-vary with cognitive measures of online performance, (2) questioning conclusions based on null results, (3) questioning SWP's

assessment and interpretation of the R^2 goodness of fit metric and (4) questioning the choice of WM measures employed by SWP.

The current experiment focuses on one of the four island effects tested in SWP (*whether-islands*, as in (4.2)) but expands upon SWP by providing additional analyses, including testing additional cognitive measures. Through these additional analyses, SWP's general conclusion is supported, namely that there is a lack of co-variation between cognitive measures and judgments on island violations. However, positive results in the current experiment put these null results into perspective and assist in their interpretation. The inclusion of the memory lure task (see chapter 3) highlights the importance of similarity-based interference in the judgments of long-distance dependencies (though not specifically island violations).

4.2.1 Issues addressed by the current study

4.2.1.1 Choice of cognitive measures

The cognitive measures used in SWP were the serial recall task and the n-back task. The n-back is a measure of general working memory, also used in the current experiment, and is discussed in Chapter 3. Serial recall is a simple span task that requires no complex computation. Participants are given an increasing list of stimuli to remember and they have to repeat the stimuli back to the experimenter in order. The highest number of stimuli that a participant can consistently recall is the serial recall score. As is evident from this description, this is a simple memory task.

HSCS were unsatisfied with SWP's choice of cognitive measures, a reasonable criticism. SWP set out to specifically test Kluender's processing account of islands. Kluender's account is very clear that it builds on the Just and Carpenter capacity-constrained model of working memory, which posits a single cognitive resource for both memory and computation (Just & Carpenter 1992). The serial recall task used by SWP appeared to measure only memory and not computation. It is unclear that serial recall should then be expected to co-vary with linguistic judgments of island phenomena as the computation component is absent.

The n-back task used by SWP is also frequently used in fMRI imaging studies to help researchers locate areas of the brain engaged in working memory, and has additionally been demonstrated to show co-variation with linguistic stimuli (Michel 2010). As such it is a better candidate to find a co-variational relationship with acceptability scores of islands than the serial recall task, though HSCS argue that the n-back is more of a short-term memory task than a working memory task. The task requires constant updating of representations in memory, but it is arguable whether this qualifies as a computational component.

While the n-back task is a better fit than serial recall to a capacity-constrained model of working memory (and is used in the current dissertation), it is the reading span task that is most associated with Just and Carpenter's capacity-constrained view. Because of this connection, the reading span task is the most obvious measure to use when attempting to test the capacity-constrained account of islands with a co-variational approach. The reading span task requires remembering the last word of a

series of sentences while performing whatever processes are normally used in reading those sentences out loud (see Chapter 3 for further discussion). While SWP did not make use of reading span, this dissertation does.

The current study uses the reading span task, as well as the n-back and two other cognitive measures, each motivated from a specific view of the interaction between cognitive factors and sentence processing (see Chapter 3). The additional cognitive task that will prove to be crucial in the current study is one that tests susceptibility to similarity-based interference in memory: the memory lure task.

The memory lure task, like the serial recall task, is a simple memory task. Participants are tasked with recalling a list of words, though it is not a free recall task. Participants are given a new list of words and they must indicate whether each word on the new list was one that they were tasked with remembering (from the old list). Crucially, some of the new words are similar in either form or meaning to words that they had to remember (i.e. lures). Thus, if a participant is susceptible to similarity-based interference, they may respond positively to a lure (for example ‘jaguar’) when it was not on the study list (but the related word ‘panther’ was). For further discussion, see Chapter 3.

This task is designed to test the similarity-based interference view of working memory (e.g. Gordon, Hendrik and Johnson, 2001; Gordon, Hendrick and Levine, 2002; Lewis and Vasishth, 2005; Gordon et al. 2006; Lewis, Vasishth and Van Dyke, 2006; Van Dyke and McElree, 2006) which unlike the Just and Carpenter model, does not place an emphasis on *actively* holding words (in the current study, fillers) in

memory. Instead, when the gap is encountered, the filler is retrieved from recent memory. If there are other items in recent memory that could interfere with this process (such as *whether* in a *wh*-island), then the parser has difficulty resolving the conflict from this interference, and processing difficulty ensues (see Chapter 2 for further discussion). If this similarity-interference view of working memory is correct, then we would not expect to see co-variation with tasks that focus on participants' ability to actively store items (such as serial recall or verbal span), but instead we would expect to see co-variation with tasks that can measure how successful individuals are at suppressing distractors. If a participant can successfully ignore the interference present from similar items, then they should be able to process a complex sentence more easily. In the current study, the Eriksen flanker attention task (Eriksen & Eriksen 1974; Eriksen & Schultz 1979) provides a measure of how well participants can suppress simultaneous distractors as they are encountered while the memory lure task provides a measure of how well participants can suppress distractors that compete with items in recent memory.

The use of a variety of cognitive measures in the current experiment does much to address the concerns HSCS raise with SWP's choice of measures. As will be seen below, the choice of measures is crucial for the results of this experiment, which demonstrate variability of the acceptability judgment ratings with the memory lure task scores.

4.2.1.2 The interpretation of null results

When looking for a relationship between cognitive measures and acceptability scores, SWP ultimately reported that they found none. That is, they reported a null result. HSCS were concerned about the interpretability of these null results. It is not known if these null results were due to there being no relationship between the cognitive measures and acceptability scores, as SWP claimed, or if the results did not reach significance for some methodological reason and/or lack of statistical power. HSCS mention, for example, that the wrong choice of cognitive measures (see section 4.2.1.1) could result in these null effects, where the proper measure would not. The current study addresses the choice of cognitive effects, and finds statistically significant results with the memory lure task. In the current experiment, while there are null results with some cognitive tasks, others (i.e. memory interference) did provide statistically significant results (section 4.4.2). By virtue of the fact that statistically significant results were found with one measure, the interpretation of the null results becomes less problematic. Finding significant effects indicates the experiment has sufficient statistical power to detect these effects, and that it is possible to obtain these co-variational effects in this type of study.

4.2.1.3 The interpretation of R^2

I have previously stated that SWP ultimately concluded that they found no evidence for a relationship between acceptability scores and the cognitive measures that they used. This conclusion is based on a series of simple linear regressions, some

of which, however, do reach statistical significance. For example, in subject islands, SWP reported that the best-fit regression line's slope had a p -value of 0.02. However, the R^2 value of the line was only 0.04. R^2 is a measure of the goodness of fit of the regression line and represents the percentage of variance in the data that is explained by the linear model. In the simple linear models SWP used, this is equivalent to how much the correlation between the DD score and the cognitive measure accounts for the variance in the data.

The R^2 metric is different from p -values, which are used to measure statistical significance. In 3 out of the 12 comparisons that SWP made across two experiments¹ SWP obtained a statistically significant p -value, (Experiment 1 subject islands $p = 0.02$; Experiment 2 adjunct islands $p = 0.04, 0.01$ for serial recall and n-back, respectively). However, the R^2 scores for these comparisons were 0.04, 0.02 and 0.04, respectively, leading SWP to conclude that they did not account for a meaningful percentage of the variance in the models. As HSCS pointed out, how to interpret R^2 values is an open question, as is the question of how much variance one should expect the model to account for in this situation. HSCS argued that while there is not consensus in the field as to how to interpret R^2 values, there is consensus that $p < 0.05$ is taken to be statistically significant. HSCS argued that SWP's statistically significant findings should be taken as evidence in support of the capacity-constrained processing account of islands.

¹ I am not counting the separate analyses where SWP included only the participants with a positive DD score (see SWP 2012). It should be expected that responses from some individual participants would not necessarily pattern the same way that the aggregate data does (see 4.2.1.4).

The current study also reports simple linear regressions, including p -values and R^2 values. However, since the analysis for the current experiment tests more than only the relationship between DD and cognitive measures (see section 4.2.1.4), the current study is in a position to *compare* R^2 scores and say which approach can account for more of the variance in the data, rather than attempt to interpret such figures in isolation.

4.2.1.4 The reliance on DD scores

One concern with SWP's analysis, which is not brought up by HCSC, is the exclusive reliance on the DD scores for the co-variational analysis. There are two issues with focusing solely on this measure. First, the DD score obscures any effects that might occur in only the STRUCTURE manipulation or the GAP position manipulation. This is problematic both for (i) the interpretation of the aggregate response and (ii) how it limits the ways individuals can be observed to differ from each other. Since the DD score is a derived score, it reduces variation in its four component parts to a single measurement. Second, using DD scores requires assumptions about scale uniformity that does not appear to hold based on recent research (Michel, in prep).

SWP reported a lack of co-variation between their cognitive measures and acceptability DD. In order to calculate the DD score, the difference between the two MATRIX GAP conditions was subtracted from the difference between the two EMBEDDED GAP conditions (4.3).

$$(4.3) \quad \text{DD score} = \quad D1 ([\text{EMBEDDED NON-ISLAND}] - [\text{EMBEDDED ISLAND}]) \\ - D2 ([\text{MATRIX NON-ISLAND}] - [\text{MATRIX ISLAND}])$$

This metric gives a good measure of a superadditive effect, but the focus on this measure leads SWP away from analyzing other useful contrasts. Consider that SWP were attempting to use co-variation of cognitive scores with the DD score to look for support of the capacity-constrained account of islands. This is an open and vigorously debated claim, and so it was quite reasonable to test it. But when they reported no results that supported this account, HSCS criticized that they were only reporting null results, which are difficult to interpret. Imagine instead if SWP had also looked for co-variation of cognitive scores with the effect of GAP POSITION in the factorial design. Distance effects such as this are widely accepted as having a processing explanation. If this effect showed co-variation, but the superadditivity didn't, then the lack of effect for the latter would immediately be more interpretable, strengthening SWP's case. Or, if the distance effect did not show co-variation, then HSCS's concerns about the proper choice of cognitive measures and whether this co-variational approach is an appropriate test of the capacity-constrained account of islands would be further supported. The current experiment provides these analyses and reports on co-variation with the well-accepted GAP POSITION (i.e. distance) effect (which is found²) in addition to those with the more contentious superadditive island violation effect (which is not found).

² Though in an unexpected pattern; see section 4.4.2.2.2.

Consider also that, since a DD score is a set of subtractions, there are multiple ways to arrive at the same DD score. A higher DD score indicates a larger superadditive effect. It is generally assumed that this effect is the result of the acceptability score for the EMBEDDED ISLAND in (4.8) being lower, thus making the derived DD score higher. This need not be the case, however. If the ratings for the EMBEDDED NON-ISLAND or MATRIX ISLAND conditions are higher than related conditions, a high DD score can also be obtained without the island violation condition (EMBEDDED ISLAND) being rated lower than all other conditions. It would not be appropriate to conclude that such a pattern showed an 'island effect.' This demonstrates the importance of additional analysis beyond the DD score. Plotting the pattern of results or following up with an analysis of paired comparisons can clarify what a reasonable interpretation of a DD score should be.

SWP provided this type of clarification for the aggregate data, but not for individual DD scores. Thus the use of DD scores for individuals may obscure differences between those individuals in how those DD scores were obtained. The issue is that we should not expect individuals to necessarily pattern like the aggregate. SWP expected that individuals could have different DD scores (it was their dependent measure), but they did not allow for individual differences in how one could arrive at that DD score for an individual. In this way SWP's focus on the DD score forced interpretations of the data where individuals can differ from each other only in terms how much of a superadditive effect they show. The current study provides analysis of the component parts of this DD score.

Another concern with the use of DD scores is that it assumes participants successfully use the rating scale uniformly when rating sentences. Since DD scores are a measure of superadditivity, their use assumes that simple additivity should be observable. Michel (in prep) presents data from a 7-point scale acceptability task showing that adding a second grammatical error of the same type (e.g. over-regularization of irregular verbs) within a sentence does not result in a simple additive effect, but a sub-additive one. For example, a single error was rated as a 2.5 out of 7 (about a 3-point penalty from the no error control sentence), but an additional error of the same type was rated a 2 out of 7 (only a 0.5-point penalty for the same type of error). To my knowledge, no researchers have claimed that such an effect should be additive, but the fact that it is not should give us pause when focusing on a requirement of superadditivity. Whether this pattern is the fault of the size of the scale (a floor effect) or representative of a genuine sub-additive pattern for errors, participants are not using the scale uniformly.³ If scale uniformity is in doubt in simple cases that result in subadditivity, they also need to be addressed in more complex combinations of errors (combining two different types of error/difficulty; see discussion in HSCS and Hofmeister, Staum-Casasanto and Sag 2010). It is further unclear if individuals differ in how they treat these issues of scale and additivity.

In order to address these issues, the current study uses a variety of ways to measure participants' acceptability responses in addition to the DD score (e.g. independent effects of GAP POSITION and STRUCTURE as well as the ratings specifically

³ Similarly, Sprouse (2011) argued that assumptions about participants' use of scale in magnitude estimation studies (a method used in SWP Experiment 2) do not hold.

given to the island violation condition). These additional measurements are not as obscuring as the DD score, involving only two conditions at a time. Furthermore, the use of multiple such comparisons allows a richer understanding on the effects in a way that the DD score does not. Since these measurements are not dependent on a statistical interaction, they are less influenced by participants' potential lack of scale uniformity. Again, the use of multiple such measures allows for the easier identification of an issue that could arise from a lack of scale uniformity. The DD score is still used here for comparison with SWP, but the use of other analyses allows for more clear examination of comparisons.

4.2.2 Cognitive Co-variation Intuition (CCI)

At the beginning of this chapter, I introduced the intuition that represents the basis for both this study and SWP. The Cognitive Co-variation Intuition, or CCI, is repeated in (4.4). While SWP were careful to lay out their reasoning as to why this co-variation should be expected, they did not break down the reasoning into these exact terms. I will refer to SWP's intuitions as essentially parallel to the CCI in (4.4), but we will see that modifications will need to be made to (4.4 c) and (d).

(4.4) Cognitive Co-variation Intuition (CCI) applied to island phenomena

- a) If the unacceptability of a sentence (here specifically an island violation) is due to processing difficulties
- b) And these processing difficulties arise from constraints on measurable cognitive resources (such as WM)
- c) Then those individuals with a measurably greater cognitive score are expected to process the sentence (island violation) more easily
- d) And this will result in these high-scoring individuals rating these difficult to process sentences as more acceptable

One of HSCS's most fundamental criticisms of SWP was that it is unclear that we should expect to find differences in an off-line measure (acceptability judgments) modulated by cognitive scores that are associated with on-line processing. That is, it has not been demonstrated that something like the CCI holds for the unacceptability of sentences that are uncontroversially thought to have a processing explanation. As will be seen below, the CCI as it stands in (4.4) will need to be modified in order to account for prior data. Specifically, the idea that higher cognitive scores correlate with ease in processing more difficult sentences does not hold (4.4 c; section 4.2.2.1). Additionally, the assumption that ease in processing a sentence will result in a higher acceptability score being assigned to that sentence does not always hold (4.4 d; section 4.2.2.2).

4.2.2.1 The relationship between cognitive scores and sentence processing difficulty

Since Kluender and Kutas (1993b) do not explicitly predict co-variation, SWP were careful to lay out the components of the capacity-constrained processing account of islands, as well as the necessary extensions to it in order to be able to test the theory with a co-variational approach. One such extension is the linking hypothesis that “processing costs are reflected in acceptability judgments” (SWP, pg 89). This ‘linking hypothesis’ is akin to the final clause of the CCI (4.4 d), which will be addressed below, but first we must examine what these ‘processing costs’ are expected to be. What processing pattern is it that is being ‘reflected’ in the acceptability judgments?

In the CCI (4.4 c) there is an expectation that individuals with higher cognitive scores will be able to process difficult (island violation) sentences better. This expectation is in line with Just and Carpenter’s capacity-constrained theory. It is only when the capacity limits are reached that a processing bottleneck occurs. If the capacity is less constrained in people with higher cognitive scores, then they should have more capacity available in order to process sentences of greater complexity. I consider this view of individual differences a ‘push the limits’ scenario, where an increase of working memory capacity means that one can sustain more complex storage and processing before one’s limit is reached. This general view is also compatible with a similarity-interference view of working memory, as the less susceptible to interference one is, the more able one is to process complex sentences

without confusing cues needed for retrieval. This is not the only view of how individual differences interact with processing complexity, however, and it does not appear to be the view that is supported by the data in the literature.

If we compare the processing difficulty of (4.5 a) and (4.5 b), below, (4.5 b) is the more difficult to process sentence as it contains a longer distance dependency.

(4.5 a) Who ___ thinks that John bought a car? >

(4.5 b) What do you think that John bought ___?

The more difficult sentence should be rated less acceptable, since it was harder to process. This is simple, straightforward, and what has been reported in the literature (including in SWP). But if we want to look for co-variation of cognitive scores with acceptability judgments (via processing ability), this simple picture becomes much more complicated.

SWP assume a ‘push the limits’ view of the CCI assumption in (4.4 c), but prior research does not support this view; at least for the processing of dependency length. In an ERP study, King and Kutas (1995) compared long-distance dependencies (object relatives (4.6 a)) and short-distance dependencies (subject relatives (4.6 b)).

(4.6 a) The reporter who [the senator harshly attacked _] admitted the error.

(4.6 b) The reporter who [_ harshly attacked the senator] admitted the error.

King and Kutas (1995)

The long-distance dependency in (4.6 a) is expected to be more difficult and it elicited a sustained anterior negativity when compared to (4.6 b). A sustained anterior negativity is thus associated with processing difficulty. When the participants were split into high and low performing groups based on comprehension question accuracy, however, only the high scoring group showed the effect. In the high scoring group, a clear distinction was made between the difficult (4.5a), which elicited the negativity, and the easier (4.5b), which elicited a more positive waveform. In the low scorers, however, the sustained negativity was elicited for both the difficult (4.5a) and easy (4.5b) sentences. Thus, instead of the high group getting a boost on the difficult condition (a ‘push the limit’ pattern), they showed a benefit in processing the easy condition. A similar pattern was found with a working memory span split in Münte, Schiltz and Kutas (1998) which compared sentences with initial *before* (more difficult) and *after* (less difficult) clauses.

We thus see a disconnect then in the ‘push the limits’ view of cognitive ability that SWP build their analysis on (CCI assumption in (4.4c)) and the actual pattern attested in the processing literature. We see a need to update the CCI to reflect the possibility that high scorers may find the less difficult sentences (rather than the more difficult sentences) easier to process (as suggested by the data above). However, we do not want to assume that all processing difficulties will pattern like these long-distance dependencies. (4.4c) has been updated in (4.7c) to allow for multiple relationships between cognitive scores and the processing of easier/more difficult sentences.

(4.7) Cognitive Co-variation Intuition (CCI) applied to island phenomena

(first updated version)

- a) If the unacceptability of a sentence (here specifically an island violation) is due to processing difficulties
- b) And these processing difficulties arise from constraints on measurable cognitive resources (such as WM)
- c) Then those individuals with a measurably greater cognitive score are expected to process the sentences in question *differently than lower scorers*
- d) And this will result in these high-scoring individuals rating these difficult to process sentences as more acceptable

What forms could the differences alluded to in (4.7 c) take? The logical possibilities are presented as the Processing Benefits Schedule (PBS) in Table 4-1. If a person or group demonstrates a higher cognitive score, we assume that they will have some kind of processing benefit. But there are various ways that processing benefits can manifest. If difficulties in processing are viewed as an individual being pushed to their individual limits, then a higher cognitive score could represent an extension of those limits. This would result in the ability to more easily process complex sentences. It could also be the case that such an expansion would also benefit the individual in the processing of simpler sentences, creating a situation where the high scorer has a global processing benefit over the low scorer. However, if processing limits represent a 'hard cap' that is more or less even across the population, then there is no room at

that upper limit for increased performance; all participants will have roughly the same ceiling for processing complex sentences. In this case, the only room for a high scorer's processing benefit is in the easier, less complex sentences. Finally, it is logically possible that scoring highly on a given cognitive measure provides no processing benefits for either difficult or easy sentences.

Table 4-1: Processing Benefits Schedule (PBS): Expectations of processing benefits for individuals with greater cognitive resources / higher cognitive scores (i.e. working memory, attention)

Higher cognitive resources benefit...	Does apply to difficult to process sentences (a 'push the limits' view)	Does not apply to difficult to process sentences
Does apply to easy to process sentences	(A) Global benefits: All sentences become easier to process	(B) Simple (only) benefit: Difficult sentences are at ceiling for everyone: no benefit available. Room available for benefit only in simple sentences.
Does not apply to easy to process sentences	(C) Complex (only) benefit: Difficult sentences require more resources that, if present, allow faster resolution of difficulties. Simple sentences do not need nor can they benefit from these extra resources.	(D) No benefits: Cognitive co-variation is irrelevant to processing

I have given these cells descriptive labels so that they can be referred back to easily. If high scorers get a processing benefit for both the simple and complex sentences, this is a pattern of 'global benefits' (A). It could be, however, that simple sentences are easy enough that additional cognitive resources aren't beneficial; these extra resources are only engaged in the difficult sentences and we only see the benefit for them there. In this case we have a 'complex (only) benefit' (C). However, if

complex sentences are equally difficult for everyone, but high scorers can gain a processing benefit with the simpler sentences, there is a ‘simple (only) benefit’ pattern (B). Finally, we could have ‘no benefits’ for higher cognitive scores for sentence processing (D).

While they did not express it in these terms, SWP assumed that higher cognitive resources will lead to a person being able to process the more difficult sentence (the island violation condition) better. No specific claims are made about the less difficult sentences, so we cannot definitively distinguish between a ‘global benefits’ or a ‘complex (only) benefit’ view. However, SWP assume that if there are differences based on cognitive scores, that these will be measurable with the DD score. Based on this, we can intuit that SWP do not expect the high scorers to have such a large benefit on the easier sentences (those that are not island violations) that the resulting DD scores would wash out any effects in the island violation sentence (i.e. leaning more towards a ‘global benefits’ view). Checking for the possibility that cognitive scores are influencing the GAP and STRUCTURE manipulations is yet another reason why it is important to examine not just the superadditive DD score, but the more simple comparisons as well.

It should be clear that the Processing Benefits Schedule (PBS) in Table 4-1 represents a certain level of abstraction in characterizing ‘difficulty’ and ‘complexity.’ I do not intend that there are only two levels of complexity relevant to processing. I simply wish to use this table as a point of reference to illustrate the complexities that are added to simple assumptions of “processing costs are reflected in acceptability

judgments” when individual differences are introduced to this claim. Still the above terms and comparisons are useful for being more explicit with our assumptions and the pattern of findings revealed in the data below.

In summary, SWP assume either a ‘global benefits’ or a ‘complex (only) benefits’ view of the relationship between variation in cognitive scores and processing sentences. Both of these views are consistent with a ‘push the limits’ view of working memory, where a higher cognitive score is assumed to result in less difficulty processing complex sentences. On the other hand, when the processing literature is examined, we see a pattern more consistent with the ‘simple (only) benefit’ view of cognitive scores and processing difficulty.

4.2.2.2 The (potential lack of) transparency between processing and acceptability tasks: Rating task differences

The Cognitive Co-variation Intuition (CCI) has another assumption that does not appear to hold universally, namely that while participants are expected to vary in how they process a sentence, it is assumed that they do not vary in how they approach the task of rating a sentence for acceptability. However, there are at least two (non-exclusive) ways in which high and low scorers could be approaching the rating task differently. First, a group may not be transparently transferring their processing ease/difficulty onto acceptability scores. Second, the groups may differ in how they treat the scale (i.e. the upper and lower bounds, mean, etc.).

Acting transparently on the rating task simply means that if a sentence is more difficult to process then this will result in lower (and crucially *not* higher) acceptability

scores.⁴ The CCI assumes that processing difficulty transparently maps onto lower ratings in acceptability. In the aggregate, this often appears to be the case. For example, as previously discussed, long-distance dependencies are more difficult to process, on average, and are rated lower than short-distance dependencies, on average. This represents a *transparent* relationship. As discussed below, a Processing Discernment Penalty (PDP, Michel 2010) represents a *non-transparent* relationship.

Even if all participants are acting transparently, it is possible that they are approaching the scale differently from each other. It is known that individuals differ in how they assign acceptability ratings to a scale (e.g. some may favor using extreme values, while others keep towards the middle— though that ‘middle’ can also differ by participant). Typically z-score transformations of the raw responses are used to account for these differences. However, even after normalizing the data in this way, it may be that high scorers and low scorers are using the scale differently (in terms of upper and lower bounds, mean response, etc.). Both of these interpretations (PDP and scale use differences) can be applied to the same pattern of data, as shown below.

Michel (2010) reported a possible example of this latter situation in an acceptability judgment manipulation of d-linked (*which man*) vs. bare (*who*) fillers in *wh*-islands. Ratings were made on a large, unmarked, 1-36 point scale and normalized as a percentage of the actual range used by each participant.⁵ Participants were split into high and low working memory groups based on median split n-back scores. The

⁴ This is independent of which processing pattern from the Processing Benefits Schedule (PBS) (Table 4-1) may be found to hold in the data.

⁵ For example, if a participant regularly used the entire scale, a rating of 18 would represent a 50% rating.

d-linked sentences were expected to be rated higher than the bare sentences (see Chapter 2, section 2.2.2), but only the high working memory group made this distinction. The low working memory group rated both the bare (41%) and d-linked sentences (44%) statistically on par with the high working memory group's rating of the d-linked sentences (41%). That is, all the sentences were rated equally except for the high working memory group's rating of bare sentences, which were rated lower than all the rest (30%).

The basic claim of d-linking is that a more d-linked/individuated/specific filler restricts the set of referents that the filler could possibly refer to, resulting in the sentence being more acceptable (Chapter 2, section 2.2.2). If some participants (such as the low scorers here) did not show a distinction between bare and d-linked fillers, this would not in and of itself be surprising. It would simply mean that they do not notice and/or benefit from this distinction. However, we would expect that this lack of distinction would appear as a lack of benefit. That is, both conditions should be rated at the (lower) 'bare' filler level of acceptability. This is because it is more likely that a reader fails to notice and/or benefit from the filler having a restricted set (and it is intuitive to attribute such a failure to low scoring individuals) than the alternative. The alternative is the unlikely scenario in which the reader restricts the set of both the d-linked and bare fillers, making both relatively more acceptable (and this seems even less likely considering that it is the low scorers that are involved). If this latter unlikely scenario is discarded, then we must assume that the low scorers are rating both the bare and d-linked sentences at the bare filler level of acceptability (41-44%). Again,

this is not problematic in and of itself. The complication arises when we see that the high scorers are rating the d-linked sentences at the same level of acceptability (41%). There are two ways to try to account for this pattern of results, but both lead to the understanding that the groups differ in how they are using the rating scale.

Originally, this pattern of results was interpreted as a Processing Discernment Penalty (PDP; Michel 2010), meaning that the group with more cognitive resources was able to notice a distinction (the d-linking effect) that the low group did not, but instead of processing benefits from the easier, d-linked condition being transparently applied to the acceptability judgments (4.7 d), it appears a penalty was assessed on the more difficult condition. That is, in order to differentiate these conditions, the bare sentences were penalized, resulting in the high working memory group actually having a *lower average rating* on the sentences than the low working memory group. None of the processing patterns in the Processing Benefits Schedule (PBS, Table 4-1) predicts this pattern. That is, the high scorers did not act transparently.

This pattern of results could also be interpreted differently, however. Rather than assuming that the high working memory group is engaging in a different task related behavior (i.e. rating the more difficult to process sentence lower rather than the easier to process sentence higher, as above), it may be that the high working memory group is using the scale differently than the low working memory group. To the low working memory group, perhaps 41% acceptable represents an extremely unacceptable sentence. However, an extremely unacceptable sentence for the high group is 30% acceptable. Under this view, the high group is showing the predicted d-

linking effect (amelioration for d-linked fillers), but they are showing it in a different (lower) part of the scale compared to the low group's responses. In this case, a comparison between the groups would be obscured by this difference in use of the scale.

A similar pattern can be found for center-embeddings. Both Sprouse (2009) and HSCS report having acceptability judgment data on center embedding sentences, long taken to be the prototypical example of unacceptability judgments being due to processing considerations, rather than grammatical ones (Chomsky & Miller 1963). When participants are split into high and low working memory groups, the high working memory groups actually rate the difficult to process center embedded sentences *lower* than the low working memory group did. The high working memory group was expected to be able to process the complex sentence better and thus rate it higher (a 'push the limits' view of processing on the Processing Benefits Schedule, see Table 4-1, above), but this did not occur for either group of experimenters. This unexpected data pattern *does* make sense from both of the options presented above, however. From a PDP perspective, the high working memory group is able to better recognize just how difficult the center embedded sentences are and so, in the task of assigning scores to sentences, they rate it lower. Alternatively, the high working memory group could be using the scale differently than the low working memory group. It is not the goal of the current discussion to decide between these two interpretations, but to highlight that in either case, a rating task difference is present between the groups.

We do not yet have a clear understanding of when we should predict these types of rating task differences, but we should be aware of the possibility of their existence and take steps to check for them. The mere possibility that (i) the transfer of processing difficulty to acceptability scores is not transparent and/or (ii) different cognitive groups are using the scale in different ways, preventing transparent comparisons, represents a further complication to SWP's linking hypothesis, and requires another update to the CCI (4.8 d).

(4.8) Cognitive Co-variation Intuition (CCI) applied to island phenomena

(final version)

- a) If the unacceptability of a sentence (here specifically an island violation) is due to processing difficulties
- b) And these processing difficulties arise from constraints on measurable cognitive resources (such as WM)
- c) Then those individuals with a measurably greater cognitive score are expected to process the sentences in question *differently than lower scorers*
- d) And this will result in these high-scoring individuals as rating these difficult to process sentences as more acceptable, *assuming there are no rating task differences between scorers*

The fairly straightforward original intuition of the CCI has become somewhat burdened with caveats, but these are all concerns that must be considered when i)

moving from an aggregate response to an individual differences approach and ii) moving from processing measurements to acceptability measurements. We have seen that SWP's assumptions differ from the more cautious formulation of the updated CCI (4.8 c, 4.8 d). While rephrased in different terms, this discussion has addressed core issues of HSCS' criticisms regarding the uncertainty of the relationship between processing and acceptability data (see also Hofmeister, Staum-Casasanto and Sag 2010 for further discussion). In addition to covering HSCS's general concerns, by articulating the CCI carefully and examining its assumptions, by identifying logical possibilities (PBS, Table 4-1), and by associating these possibilities directly with the processing data, we are better equipped to address and discuss these issues in the current study.

4.2.3 Predictions and potential interpretations

It is important to note what this type of endeavor can and can't show. If transparent co-variation of cognitive scores and the island effect are found, this would be support for a processing account of islands (assuming an extension of it via the CCI). Finding this same pattern, however, does not itself constitute an argument against a grammatical approach. While Kluender's capacity-constrained processing account does not explicitly predict cognitive co-variation, it is a reasonable extension, as the account is focused on working memory capacity limits. The grammatical account makes no predictions and has no obvious extensions that would connect to an expectation (or lack thereof) for cognitive co-variation. It is possible that any co-

variation found would simply be a reflection of the processing of grammatical constraints, much like processing effects can be observed for grammatical errors.⁶ Similarly, the lack of finding co-variation with the island effect does not constitute direct evidence for the grammatical approach or direct evidence against the processing approach. A number of issues were outlined above that could contribute to a lack of finding an effect. In short, SWP and the current study represent a check on one possible prediction attributed to the capacity-constrained processing account of islands.

While direct support may be difficult to obtain from these data, much indirect and suggestive data will be presented below and in the experiments in Chapters 5 and 6 that bear not only on the debate over the grammatical and/or processing origins of island effects, but also on which view of working memory is more relevant to the processing and judging of these sentences, and how we can proceed in looking at cognitive co-variation with acceptability ratings.

At the very least, it is predicted that the basic pattern of the *whether*-island effects will be replicated (see section 4.4.1 for this replication). Additionally, if any of the cognitive measures co-vary with the acceptability judgments, this will be taken as evidence for the importance of the process(es) associated with the measure to the judgments (and presumably processing, though this is better tested in Experiment 2, Chapter 5). For example, co-variation with reading span would implicate the

⁶ For example, processing effects of grammatical agreement violations. The fact that there is an observable processing cost to reading an agreement violation does not constitute evidence that the agreement violation is not part of the grammar. If there were also co-variation with cognitive scores for this effect, it could simply be co-variation with the processing cost, and not the grammatical nature of the violation.

importance of active storage and computation, while co-variation with memory lure would implicate the importance of similarity-based interference (see sections 4.4.2 for these results).⁷ Finally, any findings that portions of the 2 x 2 manipulation of GAP and STRUCTURE co-vary with cognitive score will be taken as evidence for the ability of this approach to capture patterns of co-variation and will aid in the interpretation of any null results.

4.3 Methods

4.3.1 Participants

80 undergraduate students from UC San Diego participated in this experiment (44 female, mean age: 20.4). All were native English speakers and gave informed consent. Participants received course credit for their participation.

4.3.2 Materials

The design of the experimental sentences is detailed in Chapter 3 (section 3.2), but is briefly summarized here for convenience. Full materials can be found in Appendix 1.

The experimental sentences manipulated two factors of *whether*-islands. The factor GAP (two levels: EMBEDDED, MATRIX), indicating which clause the gap was located in, was crossed with the factor STRUCTURE (ISLAND, NON-ISLAND), indicating

⁷ As discussed below, the importance of the similarity-interference view of working memory is at least in part reflected by a task related (rating) process and not clearly due to the online processing of the sentence. For arguments on the online importance of similarity-interference, see section 5.4.3.2.2.

the nature of the embedded clause boundary. There were eight items for each of these four conditions. These were arranged in a Latin square design, forming four lists. Four additional lists of reverse order were also generated. 168 fillers were included in each list, for a total of 200 sentences in the experiment. The stimuli were pseudo-randomized such that no individual level of a factor (ex. EMBEDDED) was presented more than twice in a row. Additionally, the 200 sentences were split into eight blocks of 25 sentences each. No experimental condition (ex. EMBEDDED ISLAND) was presented more than once in a block. See Table 4-2 for sample sentences.

Table 4-2: Experiment 1 sample stimuli set. Manipulations of STRUCTURE indicated by bold. Manipulations of GAP indicated by italics. No specific claims are intended by the placement of the gap, which is meant only to indicate the on-line point of disambiguation of the gap position.

		STRUCTURE	
		NON-ISLAND	ISLAND
GAP	MATRIX	Condition 1: Who had <i>_ openly</i> assumed [that the captain befriended <i>the sailor</i> before the final mutiny hearing?]	Condition 2: Who had <i>_ openly</i> inquired [whether the captain befriended <i>the sailor</i> before the final mutiny hearing?]
	EMBEDDED	Condition 3: Who had <i>the sailor</i> assumed [that the captain befriended <i>_ openly</i> before the final mutiny hearing?]	Condition 4: Who had <i>the sailor</i> inquired [whether the captain befriended <i>_ openly</i> before the final mutiny hearing?]

The stimuli used for this experiment differ from SWP's in that here we have held the filler constant as animate, while SWP used an inanimate filler for the EMBEDDED cases (3b,d). Additionally, because the current stimuli are also used in a

self-paced reading (Chapter 5) and ERP experiment (Chapter 6), adverbs at the gap position have been included (*quickly* in the example in Table 2). On average, the adverbs used were controlled for frequency with the alternating nouns used (*carpenter*). The inclusion of these adverbs allowed us to control for word position and to have consistent comparisons across conditions. A pilot study indicated that the presence of these adverbs did not alter the pattern of acceptability judgments of these sentences (see Chapter 3, section 3.2 for discussion).

4.3.3 Procedure

4.3.3.1 Cognitive measures

Prior to the acceptability rating task, the e-prime software program (Schneider, Eschman, and Zuccolotto 2002) was used to administer four cognitive individual differences measures to the participants in the following order: reading span, n-back, flanker and memory-interference (see section 3.3 for details).

4.3.3.2 Acceptability ratings

Following the completion of the individual cognitive differences measures, participants completed the acceptability judgment experiment with paper and pen. Participants rated the sentences on a scale of 1 (least acceptable) to 7 (most acceptable).

4.3.4 Analysis

All statistical analyses presented below were done on the z-score transformation of participants' responses on the 7-point scale. Z-scores are useful because participants may not make use of the 7-point scale in the same way as each other (e.g. one subject might tend to give only extreme ratings of 1 and 7, while another rarely makes use of the most extreme ratings). Z-scores were calculated separately for each participant, taking into account their responses on all 200 sentences, including fillers. A z-score of zero represents the mean rating that was given by that participant for all sentences. Each full point of z-score represents one standard deviation from that personal mean, which can be either positive or negative.

A linear mixed-effects model was constructed with PARTICIPANTS and ITEMS as random factors. This will be referred to as the 'basic model.' The linguistic factors GAP and STRUCTURE were included as fixed effects. Markov chain Monte Carlo sampling was used to estimate p-values in the languageR package for R (Baayen 2007, Baayen et al. 2008, R Development Core Team 2009, see also SWP).

Three types of analyses were used to test for effects of the individual difference measures. First, the 'basic' linear mixed-effects model (above) was extended to include the individual difference scores in the model. This allows for testing of the interactions of the individual differences measures with each of the linguistic manipulations (GAP and STRUCTURE) without having to group the participants into high and low-scorers (as in the median-split analysis, below). Second, to provide for the most direct comparisons with SWP, simple linear regressions were

fit for the cognitive measures scores and difference-in-differences (DD) scores. In addition to the DD score, these simple linear regressions were fit to a variety of other measures including the z-scores of the island violation (EMBEDDED ISLAND) condition. Finally, the data was submitted to an ANOVA including median splits on each individual difference measure (as discussed in Chapter 3).

4.3.4.1 Linear mixed-effects model

In order to test for the significance of the individual difference measures in the linear mixed-effects model, scores from all individual difference measures (flanker score, n-back score, reading span score and memory lure score) were added as fixed effects to the basic model (in addition to PARTICIPANTS and ITEMS as random factors and GAP and STRUCTURE as fixed effects). Following a backward selection procedure, individual difference measures were removed from the largest model (the ‘parent model’) one at a time and this larger parent model was compared to the resulting reduced model (the ‘daughter model’) using a Chi square test. If the Chi square test indicated a significant difference between the two models then the removed individual difference measure had greater explanatory power than could be expected from just the added degrees of freedom in the model. Thus, if the Chi square test was significant, the individual difference measure was kept in the model, but if the Chi square test was not significant, the individual difference measure was removed from the model. This newly reduced daughter model became the new parent model and the process was repeated until no element could be removed from the model by this

method. Markov chain Monte Carlo sampling was again used to estimate p-values in the languageR package for R (Baayen 2007, Baayen et al. 2008, R Development Core Team 2009).

4.3.4.2 Simple linear regression

Simple linear regressions were fit between one individual difference measure and one rating measure at a time. The rating measures were: DD score (following SWP), the difference in scores to NON-ISLAND and ISLAND sentences in each of the two GAP conditions (equivalent to the D1 and D2 measures used to form the DD score, see below), the difference in scores to the MATRIX and EMBEDDED conditions in each of the two STRUCTURE conditions, and finally the z-score to the island violation condition (EMBEDDED ISLAND).

To measure the DD score, the mean MATRIX ISLAND condition was subtracted from the mean MATRIX NON-ISLAND condition to obtain a D2 value for each participant. This was then subtracted from D1, which is the mean EMBEDDED ISLAND condition subtracted from the mean EMBEDDED NON-ISLAND condition, obtained for each participant. This equation is shown in (4.9). A larger DD score represents a larger superadditive effect of GAP and STRUCTURE.

$$(4.9) \quad \text{DD score} = \quad \text{D1 ([EMBEDDED NON-ISLAND]} - \text{[EMBEDDED ISLAND]}) \\ - \text{D2 ([MATRIX NON-ISLAND]} - \text{[MATRIX ISLAND]})$$

Simple linear regression lines were fit for each of the four different measures of individual differences (n-back, reading span, form lure, flanker and verbal fluency) with the DD score. As discussed previously, there should be no expectation that comparative scores, such as the DD score, should have positive values for every subject. Experimental noise and individual variation can result in some participants exhibiting a pattern that does not support or even contradicts the aggregate pattern.⁸ As such, unlike SWP, multiple analyses where individuals who exhibit a sub-additive effect are removed from analysis were not run.

The five other scores that were fit with simple linear regression lines were as follows. D1 and D2, as defined in (4.9), which represent the effect of STRUCTURE in EMBEDDED and MATRIX conditions respectively. Similarly, the effect of GAP was examined in both ISLAND ([EMBEDDED ISLAND] - [MATRIX ISLAND]) and NON-ISLAND ([EMBEDDED NON-ISLAND] - [MATRIX ISLAND]) conditions. Finally, the z-scores to the island violation condition (EMBEDDED ISLAND) were used to give a measure more reflective of a threshold ‘island effect,’ following Ross (1987), as compared to the ‘interaction effect’ represented by the DD score.

4.3.4.3 Median split

The data were submitted to a series of (2 x 2 x 2) repeated measures ANOVAs with the within subject factors GAP (two levels: EMBEDDED and MATRIX) and STRUCTURE (two levels: ISLAND and NON-ISLAND) and between subject factor of

⁸ This could be due to differences between individuals themselves, or differences in how some individuals are responding to the specific items on a certain experimental list (since lexicalizations are balanced across the experiment, not the individual).

cognitive measure (either flanker score, n-back score, reading span score or memory lure score, each with two levels: HIGH and LOW). Cognitive measure groups were formed by median split. Where possible, the memory lure scores were tested separately between scores on the form lures and scores on the semantic lures. When these show different patterns from the general memory lure scores it is reported below.

4.4 Results and Discussion

In the following sections, I present the results of the basic effects (section 4.4.1) and the three analyses that consider the individual cognitive measures (4.4.2) separately. A separate discussion follows each individual presentation of results.

4.4.1 Basic effects

This section focuses on the basic effects in the data, without the inclusion of measures of individual differences.

4.4.1.1 Results

The mean acceptability rating for each condition on the 7-point scale is shown in Figure 4-1. Z-score transformations of these results are shown in Figure 4-2. As expected, the island violation condition (EMBEDDED ISLAND) was rated the lowest of the four conditions. MATRIX GAPS were rated more highly than EMBEDDED GAPS, and the NON-ISLAND STRUCTURE was rated more highly than the ISLAND STRUCTURE.

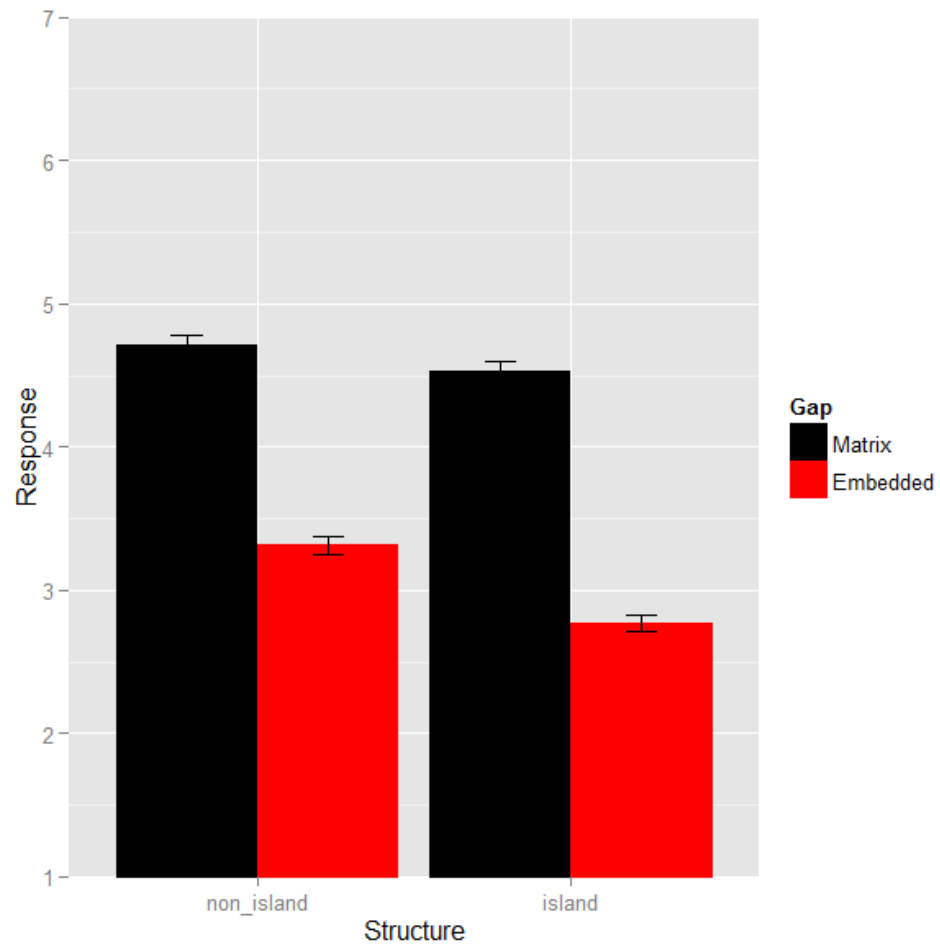


Figure 4-1: Mean results (raw scores) for Experiment 1. Error bars indicate standard error

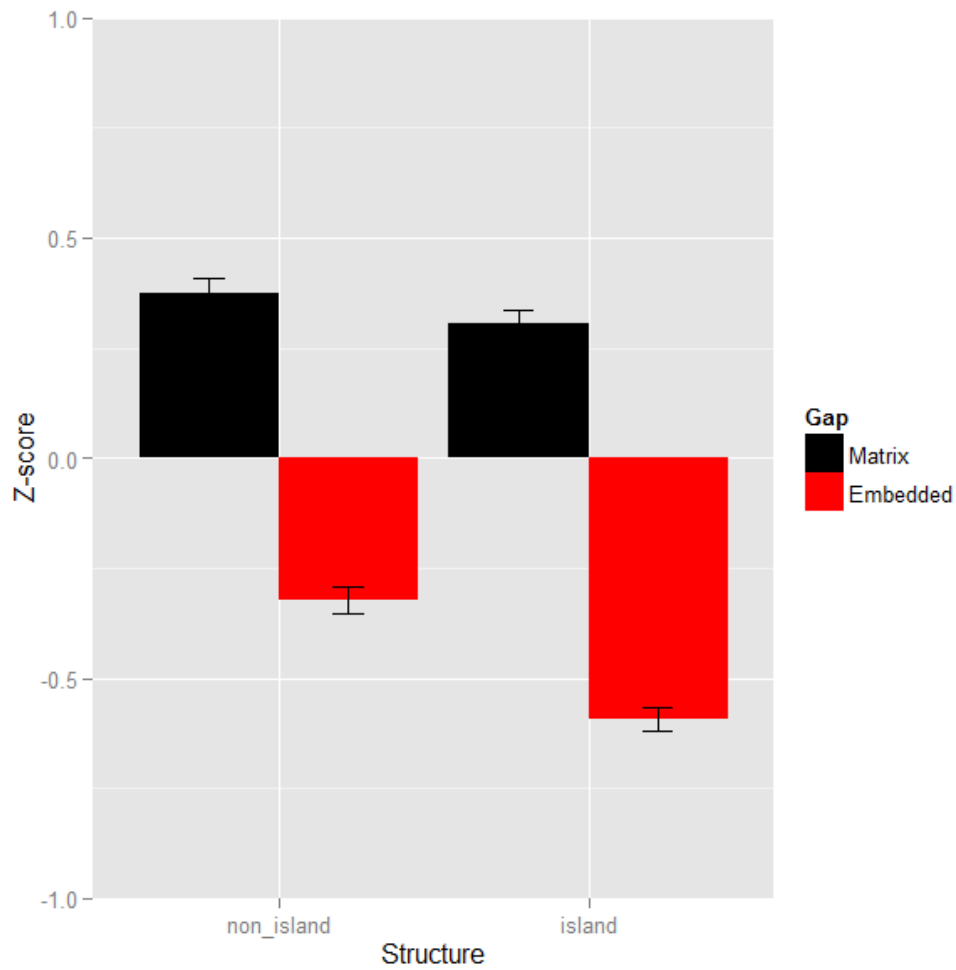


Figure 4-2: Mean results (z-cores) for Experiment 1. Error bars indicate standard error

The means and standard deviations for these data, as well as the means and standard deviations for the conditions overall, are presented in Table 4-3.

Table 4-3: Z-score transformed data. Means (standard deviation)

		STRUCTURE		
		NON-ISLAND	ISLAND	
GAP	MATRIX	0.376 (0.425)	0.304 (0.371)	0.34 (0.4)
	EMBEDDED	-0.322 (0.317)	-0.593 (0.316)	-0.458 (0.344)
		0.027 (0.51)	-0.145 (0.566)	

The results of the basic linear mixed-effect model reveal significant main effects of STRUCTURE and GAP as well as an interaction of STRUCTURE and DISTANCE. The significance values for the main effects, interaction and pairwise condition comparisons are given in Table 4-4. The only pairwise comparison that did not reach statistical significance was that of STRUCTURE (ISLAND: 0.304 vs. NON-ISLAND: 0.376) when the GAP factor was MATRIX.

Table 4-4: Significance testing of the basic model: linear mixed-effects model with no individual differences measures included.

Full 2 x 2 model		<i>p</i> -value
Main effect of STRUCTURE		<i>p</i> < 0.001 ***
Main effect of GAP		<i>p</i> < 0.001 ***
Interaction of STRUCTURE x GAP		<i>p</i> = 0.002 **
Pairwise comparisons (conditions from Table 4-2)	<i>t</i> (639) =	<i>p</i> -value
MATRIX NON-ISLAND vs. MATRIX ISLAND (1 vs. 2)	-1.60	<i>p</i> = 0.11
EMBEDDED NON-ISLAND vs. EMBEDDED ISLAND (3 vs. 4)	-6.80	<i>p</i> < 0.001 ***
MATRIX NON-ISLAND vs. EMBEDDED NON-ISLAND (1 vs. 3)	-15.22	<i>p</i> < 0.001 ***
MATRIX ISLAND vs. EMBEDDED ISLAND (2 vs. 4)	-21.57	<i>p</i> < 0.001 ***

4.4.1.2 Discussion

The basic pattern of results were as expected, with the ISLAND condition being rated lower than the NON-ISLAND condition and the EMBEDDED GAP being rated lower than the MATRIX GAP. Additionally, we see an interaction between the factors of GAP and STRUCTURE, with the island violation condition (EMBEDDED GAP in an ISLAND) rated the lowest.

However, pairwise comparisons of the four conditions reveal that there is not a statistically significant difference between the MATRIX ISLAND and MATRIX NON-ISLAND conditions.⁹ This is in contrast to SWP, who did find this manipulation of STRUCTURE to be significant in *whether* constructions, though they did not find a significant effect of STRUCTURE in complex noun phrase constructions ($p = 0.57$) and only marginal significance in adjunct constructions ($p = 0.06$, SWP 2012). This manipulation of ISLAND/NON-ISLAND STRUCTURE, without a concurrent long-distance dependency crossing into that STRUCTURE appears to be rather subtle. The lack of an effect here presents a complication for a capacity-constrained account of islands, as there is not a clear cost of clause boundary complexity represented in acceptability judgments.

⁹ This pattern for *whether*-island remains the same in the present study if raw, rather than z-scores are analyzed.

4.4.2 Effects including cognitive measures

The following sections present the results and discussion of three analyses that include the cognitive measures: linear mixed-effects modeling (section 4.4.2.1), simple linear regressions (section 4.4.2.2) and median split ANOVAs (section 4.4.2.3).

4.4.2.1 Linear mixed-effects model including cognitive measures

4.4.2.1.1 Results

The results of the backward selection procedure resulted in a fairly simple model that included fixed effects of DISTANCE, STRUCTURE and MEMORY LURE and random factors of PARTICIPANTS and ITEMS. That is, of the four individual differences measures administered, only the memory lure task added additional explanatory power to the basic model when considering the additional degrees of freedom that would be added to the model as a result of its inclusion. This new model will be referred to as the ‘memory-interference model.’ Markov chain Monte Carlo sampling was again used to estimate p-values, reported in Table 4-5.

Table 4-5: Significance testing of the memory-interference model: a linear mixed-effects model including MEMORY LURE as a factor

Effects:	<i>p</i> -value
Main effect of STRUCTURE	$p = 0.647$
Main effect of GAP	$p = 0.339$
Main effect of MEMORY LURE	$p = 0.588$
Interaction of STRUCTURE x GAP	$p = 0.749$
Interaction of STRUCTURE x MEMORY LURE	$p = 0.022$ *
Interaction of GAP x MEMORY LURE	$p < 0.001$ ***
Interaction of STRUCTURE x GAP x FORM LURE	$p = 0.351$

In the memory-interference model we see an interaction of MEMORY LURE with STRUCTURE ($p = 0.02$) and an interaction of MEMORY LURE with GAP ($p < 0.001$). The interaction of MEMORY LURE with GAP results from high scorers on the lure task (those least susceptible to similarity-based interference) making a greater differentiation in acceptability of the EMBEDDED and MATRIX conditions than the low scorers do. This is illustrated in Figure 4-3.

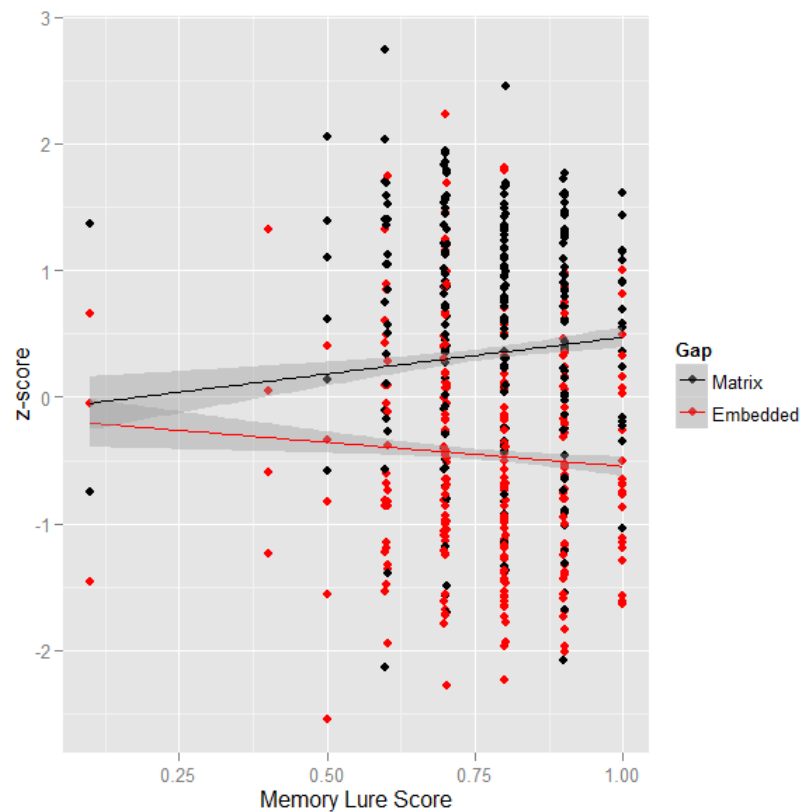


Figure 4-3: Interaction of GAP and MEMORY LURE. Acceptability ratings of MATRIX GAP (black) and EMBEDDED GAP (red) sentences plotted against MEMORY LURE accuracy (higher accuracy indicates less susceptibility to similarity-based interference). Shaded area indicates standard error.

A similar, but smaller, pattern appears for the interaction of MEMORY LURE and STRUCTURE, shown in Figure 4-4, where high scorers made a (slight) differentiation between the ISLAND and NON-ISLAND sentences but low scorers did not.

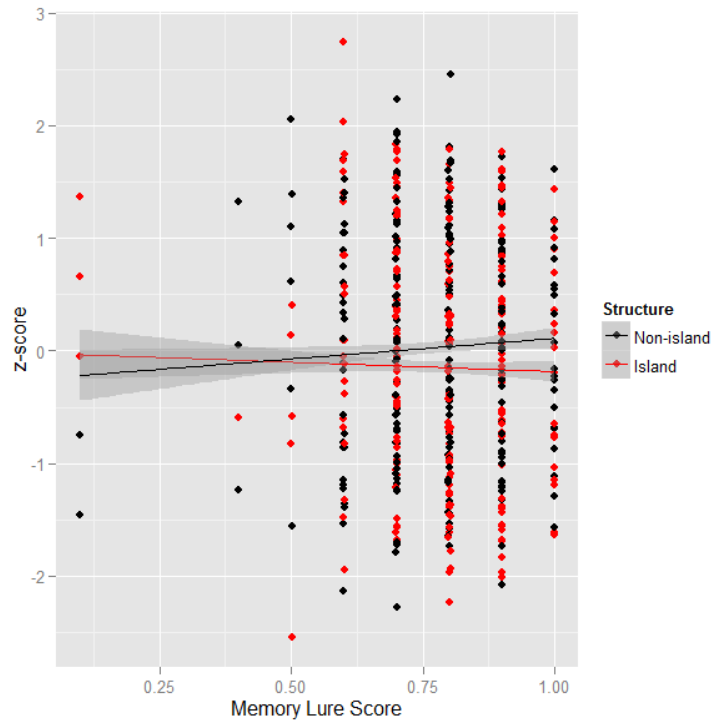


Figure 4-4: Interaction of STRUCTURE and MEMORY LURE. Acceptability ratings of NON-ISLAND STRUCTURE (black) and ISLAND STRUCTURE (red) sentences plotted against MEMORY LURE accuracy (higher accuracy indicates less susceptibility to similarity-based interference). Shaded area indicates standard error.

In both Figures 4-3 and 4-4, visual inspection indicated that there may be some low-scoring outliers for the memory lure task. To ensure that these low scorers weren't responsible for the effects reported above, the analysis was repeated while excluding the participants who scored less than 0.50 on the memory lure task. As can be seen in Table 4-6, the interaction of GAP and MEMORY LURE remains significant ($p < 0.001$), but the interaction of STRUCTURE and MEMORY LURE does not. Figures 4-3 and

4-4 are re-plotted as Figures 4-5 and 4-6, excluding participants who scored less than 0.50 on the memory lure task.

Table 4-6: Significance testing of updated memory-interference model: a linear mixed-effects model including MEMORY LURE as a factor, removing low-scorers (below 50%)

Effects:	<i>p</i> -value
Main effect of STRUCTURE	$p = 0.714$
Main effect of GAP	$p = 0.665$
Main effect of MEMORY LURE	$p = 0.244$
Interaction of STRUCTURE x GAP	$p = 0.645$
Interaction of STRUCTURE x MEMORY LURE	$p = 0.146$
Interaction of GAP x MEMORY LURE	$p < 0.001$ ***
Interaction of STRUCTURE x GAP x FORM LURE	$p = 0.337$

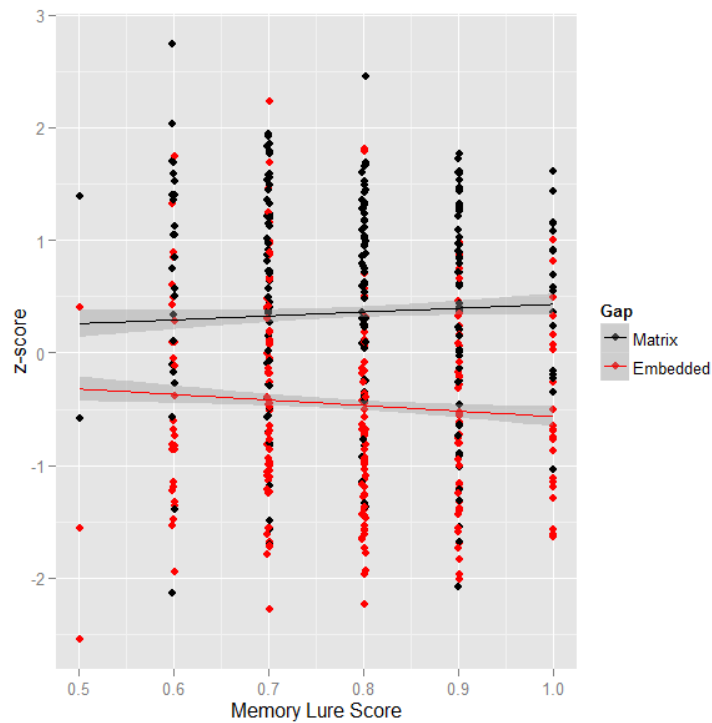


Figure 4-5: Updated interaction of GAP and MEMORY LURE; scores 0.50 or greater. Acceptability ratings of MATRIX GAP (black) and EMBEDDED GAP (red) sentences plotted against MEMORY LURE accuracy (higher accuracy indicates less susceptibility to similarity-based interference). Shaded area indicates standard error.

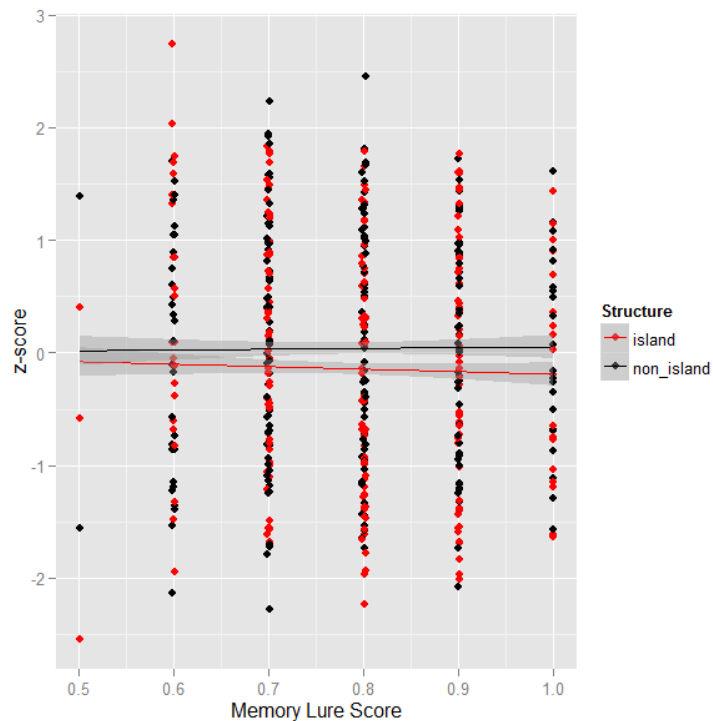


Figure 4-6: Updated interaction of STRUCTURE and MEMORY LURE; scores 0.50 or greater. Acceptability ratings of NON-ISLAND STRUCTURE (black) and ISLAND STRUCTURE (red) sentences plotted against MEMORY LURE accuracy (higher accuracy indicates less susceptibility to similarity-based interference). Shaded area indicates standard error.

4.4.2.1.2 Discussion

Only the memory lure scores were found to contribute significantly to the linear mixed effects model. This supports the idea that similarity-based interference is involved in the rating of the GAP manipulation. This is a consistent finding throughout all three analyses (see below). There is also an interaction found between the STRUCTURE manipulation and the memory lure scores, but unlike the GAP interaction, this does not remain when low-scoring outliers are removed (Table 4-6). As such, the

focus here will be predominantly on the GAP manipulation. Crucially, there is no three-way interaction of GAP x STRUCTURE x MEMORY LURE.

This lack of a three-way interaction indicates that the cognitive measure (memory lure) is not interacting with the superadditive island effect (similar to the DD score that SWP examined), but only the independent factors of GAP and (more weakly) STRUCTURE.

Looking at the interaction of GAP and MEMORY LURE, we can ask what type of processing pattern this represents. Where do the high scorers show a processing benefit? Figures 4-3 and 4-5 (low scoring outliers removed) indicate that as the memory lure score increases, the z-score ratings for the MATRIX condition sharply (4-3) or slightly (4-5) increase, while the z-score ratings for the EMBEDDED condition decrease (both figures). The MATRIX condition is the shorter dependency between filler and gap, and is the easier to process condition. That the MATRIX condition shows increasing z-score ratings with higher MEMORY LURE scores could be indicative of a 'simple (only) benefit' where the more difficult to process sentences do not benefit from the increased cognitive score; only the easier sentences do so (Table 4-1). However, it is not clear that this is the best explanation of the data. While the more difficult condition does not show increasing z-score ratings with increasing cognitive scores (contra a 'push the limits' type view assumed by SWP), it is more striking that these scores are actually decreasing with increasing cognitive scores.

This decrease in z-score ratings is not predicted by any of the views linking cognitive scores and processing difficulty (Processing Benefits Schedule, Table 4-1).

The decrease of acceptability rating as cognitive score increases recalls the pattern discussed in section 4.2.2.2 for the Processing Discernment Penalty (PDP) reported for d-linking (Michel 2010) and center embedding sentences (Sprouse 2009; HSCS). It should also be considered then that the high scorers on the memory lure task are taking note of the distinction between EMBEDDED and MATRIX GAPS in a way that the lower scorers are not. Either interpretation represents a rating task difference, contra the CCI (4.8 d). As MEMORY LURE score increases, so does the amount of differentiation between the GAP conditions. This interpretation of the data receives additional support from the self-paced reading data (Chapter 5, section 5.4.3.2.2). The data here do not support a processing account of islands since increasing cognitive scores do not map on to increasing acceptability scores.

This pattern also demonstrates a possible reason why the interaction effect (and DD score in other analyses) does not significantly co-vary with MEMORY LURE. As the form lure score increases, the ratings for MATRIX GAP sentences improve, but the ratings for EMBEDDED GAP sentences decline. These effects pull in opposite directions, effectively ‘washing out’ in the DD score. Thus, the concern that the DD score could be obscuring results appears to be justified (section 4.2.1.4).

4.4.2.2 Pattern of results using simple linear regression score analysis

4.4.2.2.1 DD score

4.4.2.2.1.1 Results

The correlation of DD score and n-back score just missed statistical significance ($r = -0.22$, $p = 0.055$). The negative correlation indicates that the higher individuals scored on the n-back task (on average), the less of a superadditive effect they would have (see Figure 4-7). This is the pattern predicted by SWP if the capacity-constrained account of islands is correct. No other comparison approached statistical significance with the DD score (see Table 4-7).

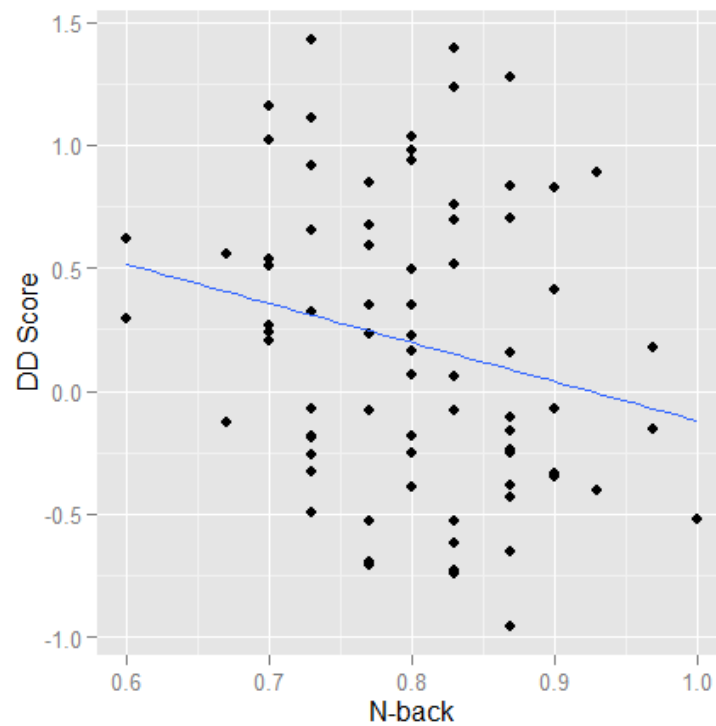


Figure 4-7: Simple linear regression of n-back scores and DD scores. The regression line has an intercept of 1.47 and a slope of -1.59 , with $R^2 = 0.05$. The data are marginally negatively correlated ($r = -0.22$, $p = 0.055$).

Table 4-7: Regressions of cognitive measures to DD scores

	slope	intercept	R ²	r	t	p
Flanker	< 0.01	0.13	0.01	0.11	0.96	0.341
N-back	-1.59	1.47	0.05	-0.22	-1.95	0.055 .
Reading Span	0.01	-0.05	< 0.01	0.07	0.64	0.527
Mem Lure ¹⁰	0.22	0.03	< 0.01	0.06	0.49	0.622

4.4.2.2.1.2 Discussion

The current study finds a marginal ($p = 0.055$) effect of n-back scores with DD scores. As n-back score increases, the DD score (i.e. how superadditive the island effect is) decreases. This pattern is the one predicted if the capacity-constrained processing account is correct. SWP also tested n-back scores for correlations with DD scores in *whether*-islands, but failed to find a significant effect ($p = 0.24$). This difference could be for a number of reasons. First, this small effect may not be particularly robust. Additionally, there are methodological differences between the two experiments. For example, the current experiment controlled for animacy in the *whether*-islands, while SWP did not. Differences in fillers could result in different influences on participants' z-scores. The magnitude estimation task used by SWP could introduce more spurious variance (Weskott and Fanselow 2008, Fukuda et al. 2010) compared to the 7-point scale used here. Finally, SWP used a D' measurement for their n-back scores, where simple accuracy was used here. However, reexamining the current data using a D' measurement generates the same pattern of data.¹¹ But the real issue is how we should interpret this type of small but (nearly) statistically

¹⁰ Neither subset of the memory lure task was significant (Form Lure $p = 0.469$, Semantic Lure $p = 0.94$)

¹¹ Simple linear regression of n-back scored using D' with DD scores: $R^2 = 0.01$, $r = -0.11$, $p = 0.05$

significant result. Should this be taken as support for a processing account of islands? Or is this a spurious result?

We obtained a small R^2 value (0.05) for this regression, indicating that it accounts for a minimal amount of the variance of the data, though this is larger than the value reported in SWP (0.01). As HSCS pointed out, unlike statistical significance for p -values, there is not a generally accepted value for R^2 that is taken to indicate a meaningful result. A similar situation arises if we choose to instead look at Pearson's r . Judgments vary in what is a small, medium or large correlation, but $r = -0.22$ would often be judged to be a small effect size (e.g. Cohen 1988). However, rather than attempt to divine whether a small value for r or R^2 is suitable to support a capacity-constrained account of islands, it is more informative to compare this to other measures and see what appears to account for the data better. These other regressions are reported below.

4.4.2.2 Other regressions

In addition to the simple linear regressions fit to DD scores, regression lines were fit to the size of four pairwise effects and the z -score assigned to the island violation condition. The four pairwise effects are: the difference between the ISLAND and NON-ISLAND STRUCTURES in both the EMBEDDED and MATRIX cases (equivalent to D1 and D2, respectively; see (4.8)) and the difference between the EMBEDDED and MATRIX GAPS in both the ISLAND and NON-ISLAND cases.

4.4.2.2.1 Results

No significant results were found for the slopes of any regression lines between any of the four cognitive measures and the D1 ([EMBEDDED NON-ISLAND] – [EMBEDDED ISLAND]) or D2 ([MATRIX NON-ISLAND] – [MATRIX ISLAND]) scores.

The results of regressions to the difference in GAP conditions did generate significant results both in NON-ISLANDS (see Table 4-8) and ISLANDS (Table 4-9). In the NON-ISLAND case, a marginal effect was found between the difference between the GAP conditions and n-back scores. A significant effect was found with the memory lure scores; specifically the form lure sub-section, which had a positive slope (slope: 0.2, $p = 0.015$) with an R^2 of 0.07. This upward slope, shown in Figure 4-8, indicates that as the form lure scores increase, the z-score differentiation between the EMBEDDED and MATRIX GAPS also increase (in NON-ISLANDS).

Table 4-8: Regressions of cognitive measures to MATRIX – EMBEDDED, NON-ISLANDS only

	slope	intercept	R^2	r	t	p
Flanker	< 0.01	0.70	< 0.01	< 0.01	< 0.01	0.998
N-back	-1.3	-0.34	0.04	0.19	1.67	0.097 .
Reading Span	0.01	0.20	0.02	0.15	1.34	0.184
Mem Lure	0.85	0.05	0.05	0.23	2.07	0.042 *
Form Lure	0.67	0.20	0.07	0.27	2.48	0.015 *
Semantic Lure	0.11	0.61	< 0.01	0.04	0.35	0.728

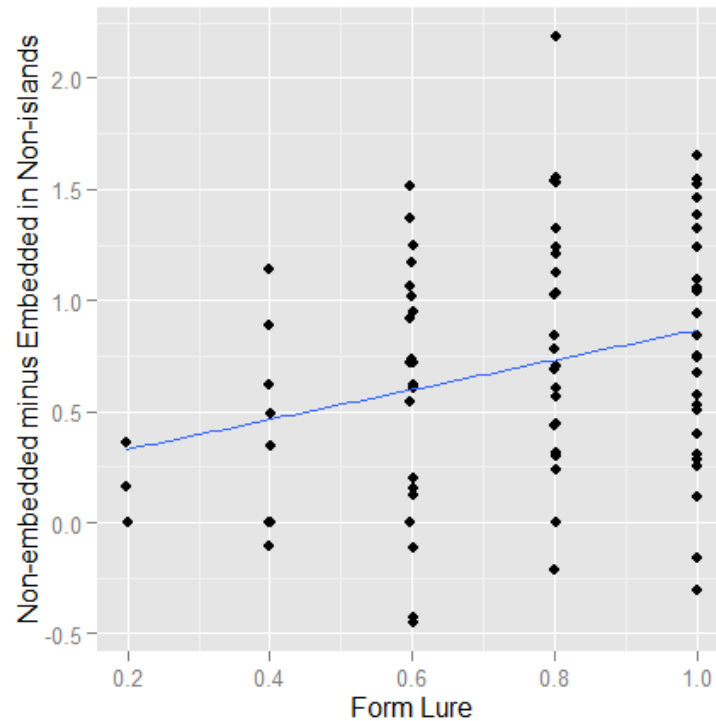
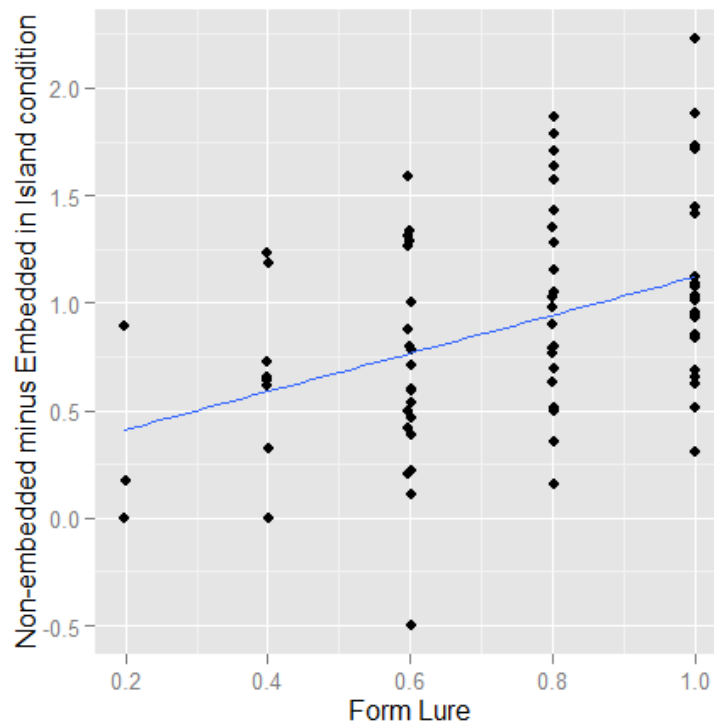


Figure 4-8: Simple linear regression of form lure scores and GAP position effect in NON-ISLANDS

In the ISLAND case, a significant effect was found between the difference between the GAP conditions and form lure scores, which had a significant positive slope (slope: 0.23, $p < 0.001$) with an R^2 of 0.16. This upward slope, shown in Figure 4-9, indicates that as the form lure scores increase, the z-score differentiation between the EMBEDDED and MATRIX GAPS also increase (in ISLANDS). A similar, but smaller effect was found for the reading span measure (Table 4-8).

Table 4-9: Regressions of cognitive measures to MATRIX – EMBEDDED, ISLANDS only

	slope	intercept	R ²	r	t	p
Flanker	< 0.01	0.83	0.02	0.13	1.13	0.261
N-back	-0.27	1.13	< 0.01	-0.05	-0.40	0.687
Reading Span	0.02	0.15	0.06	0.25	2.29	0.024 *
Mem Lure	1.06	0.08	0.10	0.32	2.98	0.004 **
Form Lure	0.87	0.23	0.16	0.40	3.83	< 0.001 ***
Semantic Lure	0.09	0.83	< 0.01	0.03	0.30	0.765

**Figure 4-9: Simple linear regression of form lure scores and GAP position effect in ISLANDS**

Finally, simple linear regressions were fit to the z-scores assigned to the island violation condition (EMBEDDED ISLAND). Significant slopes were found for form lure scores (slope: -0.21, $p < 0.001$) with an R² of 0.14. This is plotted in Figure 4-10. A similar, but smaller effect was found for memory lure and reading span scores (Table

4-10). Note that these z-scores are not differences, so a negative slope indicates a decrease in actual scores, not a decrease in the difference between two scores.

Table 4-10: Regressions of cognitive measures to island violation z-scores

	slope	intercept	R^2	r	t	p
Flanker	-0.001	-0.56	0.01	0.10	-0.85	0.398
N-back	0.08	-0.66	< 0.01	0.02	0.17	0.853
Reading Span	-0.02	-0.01	0.10	-0.31	-2.87	0.005 **
Mem Lure	-0.70	-0.06	0.11	-0.33	-3.12	0.003 **
Form Lure	-0.52	-0.21	0.14	-0.37	-3.49	< 0.001 ***
Semantic Lure	-0.14	-0.48	0.01	-0.09	-0.79	0.433

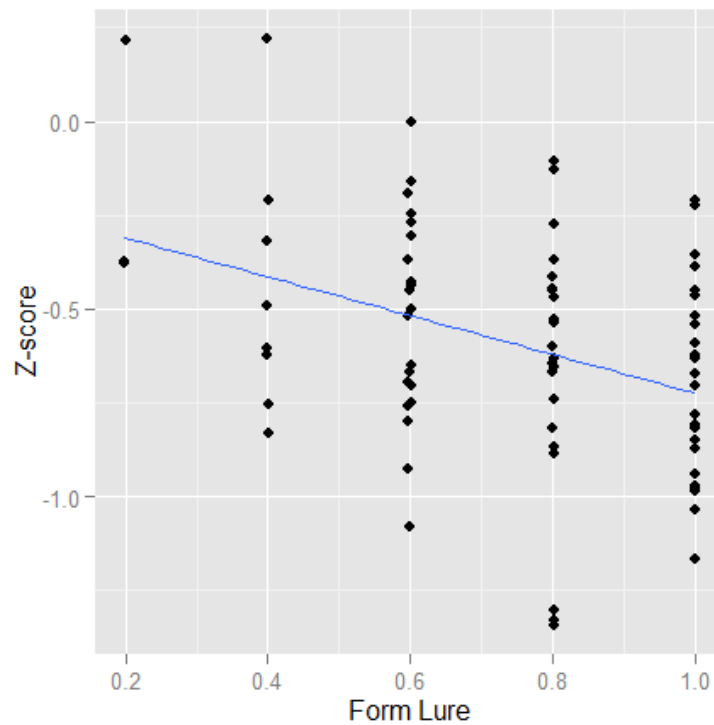


Figure 4-10: Simple linear regression of form lure scores to island violation z-scores

4.4.2.2.2 Discussion

As mentioned previously, looking only at DD scores is problematic due to the information that is not being used (i.e. patterns of the GAP and STRUCTURE effects with respect to cognitive differences). The patterns seen in the linear mixed-effects model analysis in section 4.4.2.1, are replicated here using the same simple linear regressions method used for DD scores.

We see co-variation effects with the GAP condition in both NON-ISLAND (Table 4-8) and ISLAND (Table 4-9) comparisons. The largest effects in each of these were from the form lure scores. Memory lure scores are also significant, but in these simple linear regressions we can easily separate out the form and semantic sub-portions of the memory lure task, and it is clear that the form lure scores are co-varying with the acceptability scores and the semantic lure scores are not. In both cases, as the cognitive score increases, the differentiation between the MATRIX and EMBEDDED conditions also increase. This mirrors the findings from the linear mixed-effects model (section 4.4.2.1), and is unexpected under a ‘push the limits’ view of processing difficult and acceptability (Processing Benefits Schedule, Table 4-1).

We see no effects in the STRUCTURE comparisons (neither with EMBEDDED nor MATRIX GAPS). This is consistent with the data in Table 4-6, where the linear mixed effects model was run excluding the low-scoring outliers. In both these cases the effects of STRUCTURE failed to interact with the cognitive measures.

As an alternate to the DD score, we examined the z-score ratings given for the island violation condition (EMBEDDED ISLAND). As there are no assumptions of

superadditivity involved, this avoids many of the concerns for relying on a DD score (section 4.2.1.4). By looking at just these z-scores, we are simply asking how low/unacceptable participants rate the island-violating sentences. Note that this is not a ‘raw’ score, as we are still using z-scores, which were determined for each participant based on how they rated all 200 sentences in the experiment. So these scores are still ‘relative’ to the other sentences, but there is no superadditivity assumption or requirement. When we do this, we see that it is again the form lure scores that correlated with the judgment data.

The form lure task generated a statistically significant negative correlation with z-scores of the island violation condition (slope: -0.52 , $p < 0.001$, $R^2 = 0.14$). Note that this negative slope should not be interpreted the same way as with difference measures, such as the GAP POSITION effect (Figures 4-8, 4-9). Here the negative slope indicates that as form lure scores increase, the z-score rating given to the island violation condition decreases. That is, the higher scoring individuals are rating these difficult sentences lower, not higher (see Figure 4-10), contra the expectations of a ‘push the limit’ view of processing difficulty on the Processing Benefits Schedule (section 4.2.2.1).¹² These data do not directly support a processing account of islands then, as the increasing cognitive scores do not map on to increasing acceptability scores.

Comparing the pattern of form lure scores and z-score ratings (Figure 4-10) with the form lure and GAP POSITION effects (Figures 4-8, 4-9), we see that part of the

¹² While the island violation z-score does not show a comparison between an easier and more difficult condition, the island violation condition is assumed to be (and argued to be by processing accounts of islands) the most difficult condition of the four.

increased differentiation between the MATRIX and EMBEDDED conditions in Figures 4-8 and 4-9 is attributable to the high form lure group assigning lower acceptability scores to the EMBEDDED ISLAND (island violation) condition. The GAP POSITION effects do not appear to be exclusively the result of lower acceptability being assigned to the EMBEDDED ISLAND condition, however.

The higher form lure group is rating the island violation condition's z-scores lower than the low form lure group, indicating the possibility of (i) a Processing Discernment Penalty (PDP; Michel 2010) pattern, or (ii) that the high form lure group is using the scale differently than the low form lure group. Under either of these interpretations of a rating task difference, a processing account of islands is not supported, since the increasing cognitive scores do not map on to increasing acceptability scores.

We now have a small number of statistically significant results from regression analyses between cognitive measures and various permutations of the acceptability scores. Taking the regression slopes that had the highest R^2 value for each comparison in order from highest R^2 to lowest, these are: form lure and the GAP effect in ISLANDS ($R^2 = 0.16$), form lure and the z-score response to the island violation condition ($R^2 = 0.14$), form lure and the GAP effect in NON-ISLANDS ($R^2 = 0.07$), and finally the n-back and DD scores ($R^2 = 0.05$). These values are from statistically significant (or nearly significant in the case of n-back and DD scores) comparisons, but it is clear that there are a range of R^2 values present. While we are still not in a position to claim that R^2 values above a certain numeric threshold should be considered to be experimentally

informative while others are not (to paraphrase HSCS' critique of SWP), we can at a glance see which comparisons are better able than others to account for variance in the data.

The form lure scores consistently account for more variance than the n-back scores do. Unlike the memory/form lure scores, the n-back scores were not included in the linear mixed effects model (section 4.4.2.1) and do not reach significance in the median split ANOVA analysis below (section 4.4.2.3) (and were not significant in SWP's analysis). Taking these findings into consideration, the results from the form lure scores should be given more weight than results from the n-back and DD score regression. That is not to say that the n-back and DD score regression results are unimportant; they may still prove to be informative. But for the present discussion the results of the form lure scores are more informative. In light of this comparison, and earlier concerns about the DD score, the small albeit statistically significant correlation of DD score with n-back score will not be treated as evidence in favor of the capacity-constrained processing account of islands.

4.4.2.3 Median Splits

4.4.2.3.1 Results

Repeated measure ANOVA analyses were conducted separately for each cognitive score, with subjects divided by median split. The interaction of MEMORY LURE group x GAP just missed significance by subjects but was not significant by items (see Table 4-11). The interaction of MEMORY LURE x STRUCTURE was significant by

items, but not by subjects. Examining the sub-sections of the memory lure scores, we find the interaction of FORM LURE x GAP was significant by subjects and items ($F_1(1,304) = 16.01, p < 0.001, F_2(1,1072) = 4.89, p = 0.03$). This is shown in Figure 4-11. The interaction of FORM LURE x STRUCTURE was not significant by subjects ($F_1(1,304) = 2.09, p = 0.149$), but was significant by items ($F_2(1,1072) = 8.64, p = 0.003$). This is shown in Figure 4-12. The semantic lure scores did not interact with either GAP or STRUCTURE. There were no three-way interactions between cognitive measure, GAP and STRUCTURE.

Table 4-11: ANOVAs including median split memory lure measures

Measure	F_1		F_2	
Memory Lure				
x GAP	$F_1(1,304) = 3.86$	$p = 0.051 .$	$F_2(1296) = 0.27$	$p = 0.605$
x STRUCTURE	$F_1(1,304) = 2.80$	$p = 0.100$	$F_2(1296) = 5.64$	$p = 0.018 *$
Form Lure				
x GAP	$F_1(1,304) = 16.01$	$p < 0.001 ***$	$F_2(1072) = 4.89$	$p = 0.030 *$
x STRUCTURE	$F_1(1,304) = 2.09$	$p = 0.149$	$F_2(1072) = 8.64$	$p = 0.003 **$
Semantic Lure				
x GAP Lure	$F_1(1,304) = 0.19$	$p = 0.666$	$F_2(656) = 0.47$	$p = 0.494$
x STRUCTURE	$F_1(1,304) = 0.73$	$p = 0.392$	$F_2(656) = 0.16$	$p = 0.690$

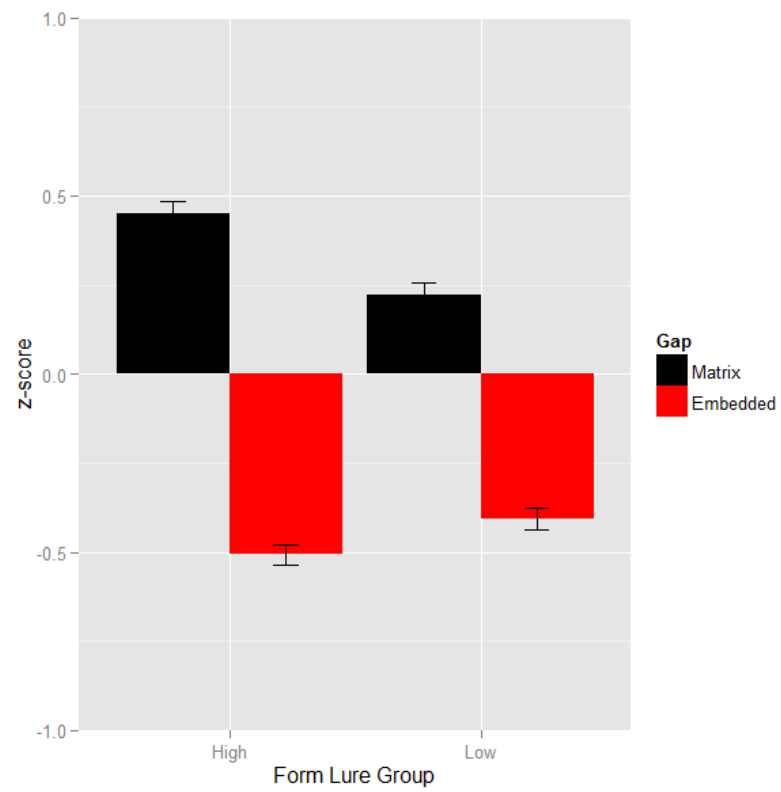


Figure 4-11: Mean results (z-scores) for FORM LURE (high and low scorers) x GAP with standard error bars

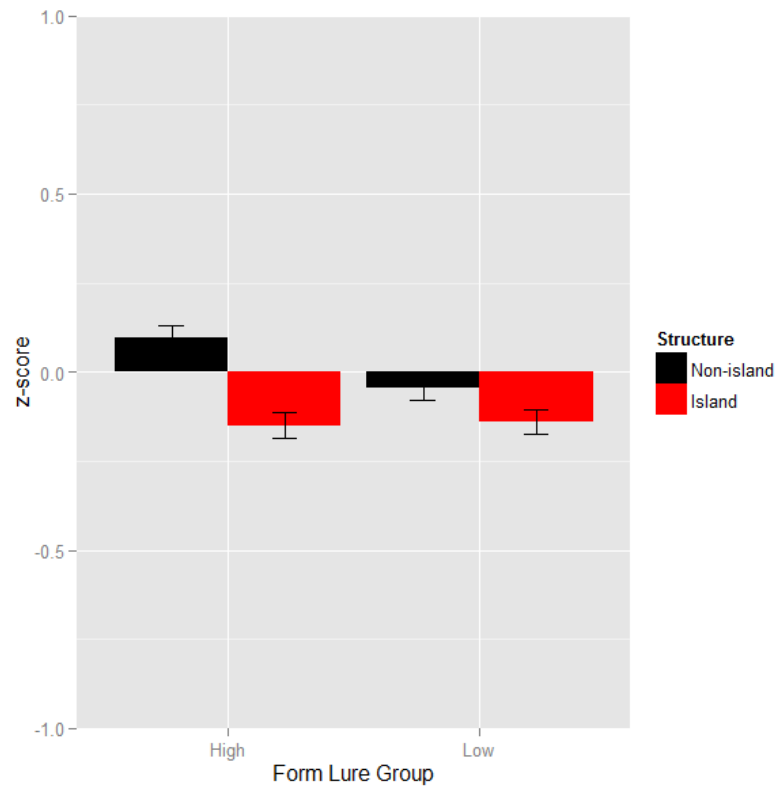


Figure 4-12: Mean results (z-scores) for FORM LURE (high and low scorers) x STRUCTURE with standard error bars

None of the other cognitive scores resulted in significant interactions with the linguistic manipulations (see Table 4-12).

Table 4-12: ANOVAs including median split cognitive measures (except memory lure)

Measure	F_1		F_2	
Reading span				
x GAP	$F_1(1,304) = 1.52$	$p = 0.218$	$F_2(1616) = 0.80$	$p = 0.371$
x STRUCTURE	$F_1(1,304) = 0.186$	$p = 0.667$	$F_2(1616) = 0.41$	$p = 0.521$
N-back				
x GAP	$F_1(1,304) = 1.61$	$p = 0.206$	$F_2(1072) = 1.72$	$p = 0.190$
x STRUCTURE	$F_1(1,304) = 0.087$	$p = 0.768$	$F_2(1072) = 0.45$	$p = 0.502$
Flanker				
x GAP	$F_1(1,304) = 2.39$	$p = 0.123$	$F_2(1,944) = 0.02$	$p = 0.896$
x STRUCTURE	$F_1(1,304) = 0.17$	$p = 0.684$	$F_2(1,944) = 0.43$	$p = 0.512$

4.4.2.3.2 Discussion

Similar to the linear mixed effects model results (section 4.4.2.1) and the simple linear regressions (section 4.4.2.2.2), we see that memory lure scores, specifically form lure scores, interacted with the acceptability judgment data. In the FORM LURE x GAP interaction, we again see that the high scorers (those who are least susceptible to similarity-based interference in memory) rated the difference between the GAP positions larger than low scorers. The high scorers both rated the easier to process MATRIX GAPS more highly (which could be interpreted as a ‘simple (only) benefit’ pattern on the Processing Benefits Schedule, Table 4-1) and rated the more difficult to process EMBEDDED GAPS lower (a non-transparent pattern, section 4.2.2.2).

The interaction of FORM LURE x STRUCTURE was only significant by items, but the general pattern indicates that the high scorers rated the easier to process NON-ISLANDS higher than the low scorers did, while both rated the more difficult to process ISLANDS equally. This also reflects a ‘simple (only) benefit’ pattern on the Processing Benefits Schedule (PBS, Table 4-1).

It is clear in both cases that the ‘push the limits’ view of working memory assumed by SWP is not reflected by the data reported here. That is, higher form lure scores did not pattern with higher ratings for the more difficult to process sentence. It is also clear that there is no three-way interaction of FORM LURE, GAP and STRUCTURE that would be akin to an interaction of FORM LURE and DD score. This lack of interaction between the cognitive measures and the superadditive effects of GAP and STRUCTURE persist throughout this experiment.

4.5 Summary

In the above data I have argued, like SWP, that the co-variation of DD scores and cognitive measures should not be weighted heavily as evidence for the capacity-constrained account of islands (section 4.4.2.2.2). However, there were a number of concerns underlying this co-variation approach such that we should not be surprised that this endeavor did not produce convincing evidence for this account (cp. HSCS). These concerns include the choice of cognitive measure (section 4.2.1.1), the reliance on null results (section 4.2.1.2), the interpretation of R^2 values (section 4.2.1.3) as well as questions about the expectations we should have with respect to the mapping of processing difficulty and cognitive measures onto offline acceptability judgments (section 4.2.2). The current experiment, analysis and discussion represent gains in all of these areas of concern. Through these improvements we have seen that cognitive co-variation *does* occur with the GAP manipulation. This is to be expected, based on the consensus that processing larger distances between fillers and gaps results in decreases in acceptability. This gives us confirmation that the current acceptability methodology is sensitive to cognitive co-variation.

The consistent significant co-variation of the memory lure task, especially form lure, indicates the importance of similarity-based interference in the judgments of filler-gap dependencies. The pattern of co-variation is not the one originally assumed in the Cognitive Co-variation Intuition (CCI), where a high score is thought to increase performance on the processing of more difficult sentences. Instead, it is the less difficult sentences that benefit with a higher form lure score, and the more

difficult sentences are actually rated lower. This can be taken as new evidence favoring the view that working memory is best thought of as a system of content-addressable memory with similarity-based interference (Gordon, Hendrik and Johnson, 2001; Gordon, Hendrick and Levine, 2002; Lewis and Vasishth, 2005; Gordon et al. 2006; Lewis, Vasishth and Van Dyke, 2006; Van Dyke and McElree, 2006) rather than a system focused on active storage costs, as would be supported if the reading span task was the cognitive measure consistently co-varying with the linguistic data. However, as the co-variation of the form lure scores and acceptability judgments were focused on the GAP position, and not the interaction of GAP and STRUCTURE we do not find support for a view of island violations themselves being due to similarity-interference processing difficulties.

In summary then, we did not find direct evidence to support either the capacity-constrained account of the similarity-interference account of islands. Instead, we found that similarity-interference is important to the rating, and possibly the processing of, long-distance dependencies. The experiments in Chapters 5 and 6 will further assess the importance of similarity-interference to the online processing of these sentences. While the self-paced reading experiment in Chapter 5 does reveal co-variation with the memory lure task (section 5.4.3.2.2), the ERP experiment in Chapter 6 does not. We also have a better understanding of the complexities involved in (i) predicting how differences in cognitive scores will manifest processing benefits and (ii) mapping those benefits onto off-line acceptability scores.

4.6 Conclusion

The data presented here, like that in SWP, did not ultimately provide support for a capacity-constrained account of islands. Specifically, there was no support for a view wherein individuals with greater cognitive resources will be able to process island violations more easily and thus rate them as having comparatively higher acceptability. This is not the same as finding evidence against processing factors having explanatory power for island phenomena, however. As has been shown, an understanding of how processing benefits could manifest for high scorers on cognitive tasks is needed before being able to make clear predictions about how cognitive variation should influence acceptability (i.e. the PBS, Table 4-1). SWP assume that processing benefits would potentially be seen in the more difficult to process sentences, while the data here showed benefits surfacing on the less difficult to process sentences. Additionally, we saw further complications coming from apparent rating task differences (section 4.2.2.2). Different cognitive groups are either using the scale differently from each other or showing non-transparent effects of processing on acceptability scores, such as the Processing Discernment Penalty (PDP, Michel 2010), wherein participants with higher cognitive scores appear to be assigning a (larger) penalty to the difficult sentences.

The results here provide evidence that, when individual differences are taken into account, operationalizing an ‘island effect’ as a statistically significant superadditive island effect (such as a DD score) in a factorial design is problematic. Mixed-effect models, ANOVA analysis and simple linear regression have repeatedly

failed to find an interaction of cognitive scores and DD scores, while these analyses (and simple linear regressions) indicate that cognitive scores did vary with sub-components of the DD score, but in opposite directions. Specifically, as form lure scores increase, ratings for (short-distance, easier to process) matrix gap sentences improved, while the (long-distance, difficult to process) embedded gap sentences declined. Thus it appears that relying on a combination of comparisons between factors, rather than a composite score, is best. As we have seen, the DD score conflates too much other data.

Examination of the component parts of the islands, especially the GAP manipulation, which co-varied with the form lure task, not only help us to better interpret the results as they apply to island phenomena, but also give us insights into the characterization of working memory. The co-variation of this particular cognitive measure supports the importance of similarity-based interference with the processing and rating of these sentences in particular as well as the growing shift in the field from a capacity-constrained view of working memory to one based on content-addressable memory and similarity-based interference. However, it is important to note that this co-variation pattern was found for both ISLAND and NON-ISLAND conditions; no discrimination was made between them. Thus, while these results may indicate the importance of similarity-interference for judging the difference between long and short dependencies (the GAP POSITION manipulation), they do not support a similarity-interference account of islands. The form lure task and why it co-varies with this acceptability manipulation is discussed further in Chapter 5, section 5.4.3.2.2, which

also discusses co-variation of this measure with reading times measured at the clause boundary.

The results of the current study indicate that there is much more complexity in the relationship between processing data and acceptability judgments than may have been originally thought. However, the current work presents more clearly defined ways to think about these complexities in our intuitions about cognitive co-variation and acceptability (i.e. the revised Cognitive Co-variation Intuition – CCI), and expectations of where to find and how to identify processing benefits (i.e. the Processing Benefits Schedule – PBS, Table 4-1). Being able to check for these patterns will enable us to move forward in this endeavor and better understand this area of research.

Chapter 5: Self-Paced Reading Experiment

5.1 Introduction

This chapter presents a self-paced reading time study of *whether*-islands that examines co-variation of reading times with measures of individual cognitive differences. Unlike the acceptability judgment study (Chapter 3), which was concerned with the interface between off-line judgments, conceptualizations of online processing costs, and co-variation with cognitive measures, this chapter focuses directly on the on-line processing costs of *whether*-islands and how those costs are modulated by differences in cognitive measures.

It is widely accepted that island violation sentences are more difficult to process than non-violating controls. This characterization is equally unproblematic for proponents of either a processing account of islands or a grammatical account of islands. As such, that debate is not addressed further here. At issue is whether the capacity-constrained account of islands (e.g. Kluender 1991) is supported by co-variation with cognitive measures, or whether the conception of working memory used by this approach should instead be updated to one of similarity-based interference (e.g. Lewis & Vasishth 2005).

There are two main lines of inquiry for this chapter. First, where in the sentence is the greatest cost for the island violation condition found? The capacity-constrained view predicts that the greatest cost should be observed at the clause boundary. The similarity-interference view predicts that the greatest cost should be

most apparent at the gap position. Second, once the locus (or loci) of the processing cost is found, does that cost vary with cognitive scores, and what does the nature of that variation reveal about the processor?

The self-paced reading results reported below find that the processing cost occurs at the clause boundary, consistent with a capacity-constrained account (section 5.4.2.2.1). This cost is modulated by reading span scores in a pattern that indicates that while both high and low span readers show the same overall processing cost for the island violation condition, the low span readers additionally have a processing cost for the *whether* clause boundary when there is no incomplete filler-gap dependency. While this incremental processing cost for the low span readers is consistent with a capacity-constrained account, the fact that both groups show the same overall processing cost for the island violation condition is problematic for such an account (section 5.4.3.2.1).

The remainder of this chapter is organized as follows. In section 5.2 I briefly review the predictions made for the capacity-constrained and similarity-interference views of processing islands (see Chapter 2 for more detail). Section 5.3 presents the methods of the current experiment, though for details about the measures of individual differences or materials design see Chapter 3. Section 5.4 presents results and discussion of the basic data (section 5.4.2) as well as the co-variation analysis (5.4.3). Section 5.5 summarizes these findings and section 5.6 concludes the chapter.

5.2 Predictions

Chapter 4 demonstrated that *whether*-island violations are rated as less acceptable than closely related control sentences. It is not unreasonable to assume that *whether*-island violations will therefore also be more difficult to process. This expectation holds independently of any claims about whether those processing difficulties are the cause of lower acceptability ratings (see Chapter 2, section 2.3.3 for discussion). In this chapter, I use the exact same sentences as were used in the acceptability study in a self-paced reading experiment to determine where exactly in the sentence the processing difficulty occurs.

If the difficulty in processing *whether*-island violations is due to the combination of small processing costs such as (i) holding a filler in memory, (ii) crossing a clause boundary while that filler is held in memory, and (iii) the greater complexity of an island (*whether*) clause boundary compared to a non-island (*that*), as predicted by the capacity-constrained processing account of islands (e.g. Kluender 1991, 1998; Kluender & Kutas 1993a,b), then the greatest processing cost in the sentence is predicted to be observed at the clause boundary (either directly at the clause boundary or in a spillover region).

If, however, the difficulty in processing *whether*-island violations is due to difficulties in retrieving the filler from memory once the retrieval cue (i.e. the gap) is encountered, as expected under a similarity-interference account of processing islands, then the greatest processing cost in the sentence is predicted to occur at the embedded

gap site. Note that sentence processing based on cue-based retrieval and similarity-based interference (e.g. Gordon, Hendrik and Johnson, 2001; Gordon, Hendrick and Levine, 2002; Lewis and Vasishth, 2005; Gordon et al. 2006; Lewis, Vasishth and Van Dyke, 2006; Van Dyke and McElree, 2006) does not predict that there will not be a processing cost at the clause boundary. Even though these types of processing models focus on retrieval costs, they must also allow for the effects of predictive processing, which are supported by an abundance of evidence, starting with findings from the visual world paradigm (e.g. Altman & Kamide 1999; Tanenhaus et al. 1995). Similarly, the capacity-constrained account does not preclude the possibility of processing difficulty at the embedded gap position. The difference lies in the relative severity of the processing costs predicted by the two accounts. The predictions of these two approaches are summarized in Table 5-1.

Table 5-1: Predictions for the self-paced reading findings

Working memory theory	Sentence position	
	Clause boundary	Embedded gap
Capacity-constrained	focus of slowdown	possible slowdown
Similarity-interference	possible slowdown	focus of slowdown

If a slowdown in reading times for the island violation condition is found *only* at the clause boundary, this will be taken as evidence that the processing difficulty of *whether*-islands is due to an accumulation of costs that combine at that clause boundary. Finding a slowdown in reading times *only* at the embedded gap, however, will be taken as evidence in favor of the processing difficulty being due to complexity in the retrieval of the filler from recent memory. If slowdowns in reading times are

found at both locations, then the slowdowns will need to be compared to determine if it can be concluded that one of the two indicates greater difficulty for the parser. Such a comparison was not needed in the current experiment as the slowdown was localized to the clause boundary (section 5.4.2.1.1).

In addition to examining where in the sentence a processing cost occurs, we used co-variation with cognitive measures to further test these working memory theories. If the locus of processing cost is at the embedded gap, then it would not be surprising if this co-varies with the memory lure task. Similarly, if the processing cost occurs at the clause boundary, it would not be surprising if this cost co-varied with the reading span task. This second finding, that the processing cost occurs at the clause boundary and co-varies with reading span, is what was found in the experiment reported below, though not in a pattern that supports the capacity-constrained account of islands (section 5.4.3.2.1).

5.3 Methods

5.3.1 Participants

48 undergraduate students from UC San Diego participated in this experiment (26 female, mean age: 20.8). All were native English speakers and gave informed consent. Participants received course credit for their participation.

5.3.2 Materials

The design of the experimental sentences is detailed in Chapter 3 (section 3.2), but is briefly summarized here for convenience. Full materials can be found in Appendix 1. The experimental sentences manipulated two factors of *whether*-islands. The factor GAP (two levels: EMBEDDED, MATRIX), indicating in which clause the gap was located, was crossed with the factor STRUCTURE (ISLAND, NON-ISLAND), indicating the nature of the embedded clause boundary. There were eight items for each of these four conditions, as well as 168 fillers, for a total of 200 sentences in the experiment. These were arranged in a Latin square design, forming four lists. Four additional lists in reverse order were also generated. The stimuli were pseudo-randomized such that no individual level of a factor (ex. EMBEDDED) was presented more than twice in a row. Additionally, the 200 sentences were split into eight blocks of 25 sentences each. No experimental condition (ex. EMBEDDED ISLAND) was presented more than once in a block. See Table 5-2 for sample sentences.

Table 5-2: Experiment 2 sample stimuli set. Manipulations of STRUCTURE are indicated in bold while manipulations of GAP are indicated by italics. No specific claims are intended by the placement of the gap, which is meant only to indicate the on-line point of disambiguation of the gap position.

		STRUCTURE	
		NON-ISLAND	ISLAND
GAP	MATRIX	Condition 1: Who had <i>_ openly</i> assumed [that the captain befriended <i>the sailor</i> before the final mutiny hearing?]	Condition 2: Who had <i>_ openly</i> inquired [whether the captain befriended <i>the sailor</i> before the final mutiny hearing?]
	EMBEDDED	Condition 3: Who had <i>the sailor</i> assumed [that the captain befriended <i>_ openly</i> before the final mutiny hearing?]	Condition 4: Who had <i>the sailor</i> inquired [whether the captain befriended <i>_ openly</i> before the final mutiny hearing?]

5.3.3 Procedure

5.3.3.1 Cognitive measures

Prior to the acceptability rating task, the e-prime software program (Schneider, Eschman, and Zuccolotto 2002) was used to administer four cognitive individual differences measures to the participants in the following order: reading span, n-back, flanker and memory interference (see Chapter 3, section 3.3 for details).

5.3.3.2 Self-paced reading

Following the completion of the individual cognitive differences measures, participants completed the self-paced reading experiment, administered with the e-prime software program (Schneider, Eschman and Zuccolotto 2002).

Trials began with a black fixation cross that appeared in the center of a white screen for 1000 msec. The first word (always *who*) then appeared in black 18 point Courier New font on a white background. The word remained until the 'A' button on an 'X-box' style controller was pressed. The thumb was used for this button. Words continued to be presented centrally. The central presentation was chosen rather than a moving window style of self-paced reading so that the current experiment's method of presentation would be most similar to the RSVP presentation used for the ERP experiment (Chapter 6). After each sentence a 'yes/no' comprehension statement was presented. As all the sentences read were questions, in order to judge comprehension, participants were given a statement that they had to judge for compatibility with the question. A compatible statement represented a possible situation where the question that they read could be asked. Participants were given three practice sentences, with an explanation on how they should have answered the comprehension checks. The practice sentences are provided in Table 5-3 and the explanations of the correct responses, as presented to the participants, in (5.1).

Table 5-3: Practice sentences

Practice sentences:	Comprehension statements:
Who / took / the dog / for / a walk?	The owner took the dog for a walk.
Who / had / the dog / been / biting?	The cat is a fantastic animal.
Who / had / followed / the dog / home?	The dog followed the mailman home.

(5.1) For the sentence: ‘Who took the dog for a walk?’ You should respond that, YES, ‘The owner took the dog for a walk’ is a possible situation.

But, for the sentence: ‘Who had the dog been biting?’ You should respond NO, ‘The cat is a fantastic creature’ is not a possible situation for that question as it is totally unrelated.

Finally, for: ‘Who had followed the dog home?’ You should respond NO, ‘The dog followed the mailman home’ is not a possible situation because the dog should be followed, not following someone.

Participants responded to the comprehension checks using the left or right index finger buttons on the game controller. These buttons were counter-balanced for their ‘yes/no’ mapping and matched the mapping used in the memory lure task (Chapter 3, section 3.3.4). The right thumb was used to advance through the self-paced

reading sentences, but the index fingers were used to respond to the comprehension checks.

After each comprehension check, participants were presented with a screen that prompted them to press the 'A' (thumb) button to continue. In this way, they were able to take a break after any sentence before the next trial commenced, starting again with the fixation cross.

5.3.4 Analysis

Raw reading times were examined for outliers. First, responses greater than 2500 msec were trimmed from the data. Outliers were defined over the remaining data as values more than 3 standard deviations from the mean reading time for each word position. Outliers were treated by replacing them with the value 3 standard deviations from the mean. This procedure affected less than 3% of the data.

Basic analyses (those not including cognitive measure scores) were done on residual reading times in order to control for length effects and individual differences in overall reading speed. Residual reading times were calculated separately for each participant by first calculating a linear regression equation between reading times and word/position length (in number of characters). This linear regression provides a slope and intercept for each participant. With this information, a predicted reading time was calculated for each word/position by multiplying the length of that word/position by the slope and adding this to the intercept value. This predicted value was subtracted

from the observed value in order to generate the residual reading time. This procedure allows for each participant to have independent slopes and intercepts, modeling how long each participant takes (on average) to read a word of a given length. The residual reading times represent deviations from that average. If the residual reading time is positive, then a word was read slower than predicted. If the residual reading time is negative, then a word was read faster than predicted.

To test for the basic effects of the linguistic manipulation, the residual reading times for each position in the sentence were submitted to a 2x2 repeated measures ANOVA with within subject factors GAP (two levels: EMBEDDED and MATRIX) and STRUCTURE (two levels: ISLAND and NON-ISLAND).

To test for the effects of the cognitive measures, the raw reading time data (treated for outliers) were submitted to a series of (2 x 2 x 2) ANOVAs with the within subject factors GAP (two levels: EMBEDDED and MATRIX) and STRUCTURE (two levels: ISLAND and NON-ISLAND) and a between subject factor of individual difference measure (median split groups of either flanker score, n-back score, reading span score or memory lure score, each with two levels: HIGH and LOW). Additionally, the memory lure scores were tested separately between scores on the form lures and scores on the semantic lures (see Chapter 3, section 3.3.4). When these showed different patterns from the general memory lure scores, it is reported below. Residual reading times were not used to examine the individual cognitive differences data because residual reading times are designed to control for variance between individuals. As the interest in the individual cognitive measure analysis was to examine the data for differences

between individuals, using the residual reading times would have potentially obscured some of those differences.

5.4 Results and Discussion

5.4.1 Comprehension

The comprehension checks after each trial were not planned to be analyzed but are presented here briefly for completeness.

5.4.1.1 Results

Overall, including fillers, participants averaged only 67% accuracy on the comprehension checks. This was slightly higher (67.8%) for the experimental sentences, with the MATRIX NON-ISLAND condition having the highest mean accuracy (72.8%). This was marginally higher than the EMBEDDED NON-ISLAND condition (62.9%, $p = 0.077$). No other pairwise comparisons were statistically significant. The average accuracy for each experimental condition can be found in Table 5-4.

Table 5-4: Mean comprehension accuracy by condition

		STRUCTURE	
		NON-ISLAND	ISLAND
GAP	MATRIX	72.8%	65.7%
	EMBEDDED	62.8%	69.8%

5.4.1.2 Discussion

The motivation for the comprehension checks was to keep participants attending to the task of reading the sentences. Based on debriefing interviews, this goal was reached. Also based on those interviews, however, it was found that the comprehension check used here was rather subjective.

Consider that the participants were first reading a question and then given a statement, the opposite order of most other comprehension testing that student participants are accustomed to. Participants were asked to decide if the statement was consistent with a situation where the question could be asked. The judgments of this consistency depend on how one interprets the discourse context of both the question and the statement. Participants reported that they were often unsure whether they answered correctly.¹ Based on these debriefings, no further analyses are conducted using the comprehension check results.

5.4.2 Basic effects

In this section, I present and then discuss the findings of the self-paced reading experiment before consideration of the cognitive measures is included.

¹ No feedback was given during the experiment.

5.4.2.1 Results

Residual reading times for all sentence positions are shown in Figure 5-1.

Table 5-5 shows example stimuli for sentence positions 1-9 for reference.

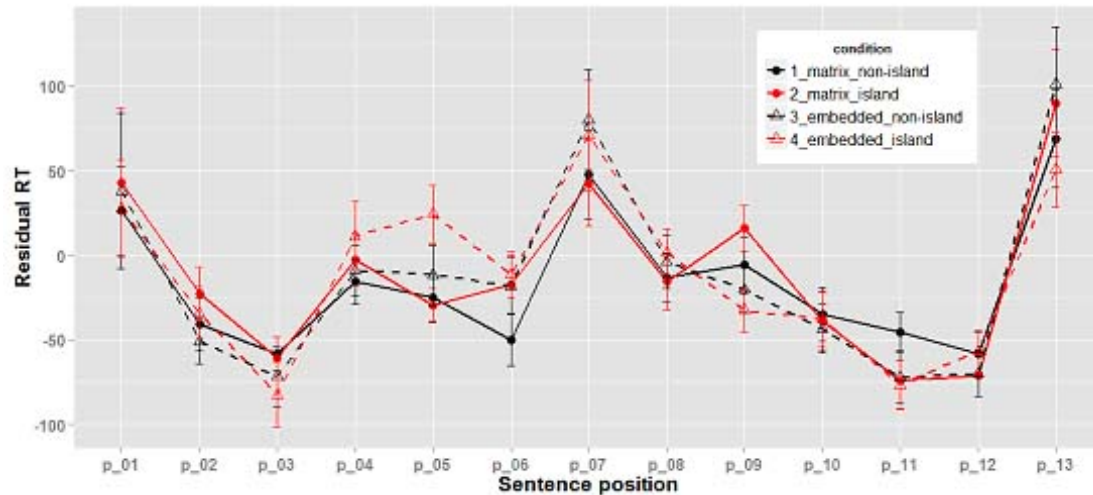


Figure 5-1: Residual reading times

Table 5-5: Word positions 1-9

1	2	3	4	5	6	7	8	9
MATRIX GAP, NON-ISLAND:								
Who	had	– <i>openly</i>	assumed	[that	the captain	befriended	<i>the</i> <i>sailor</i>	before ...
MATRIX GAP, ISLAND:								
Who	had	– <i>openly</i>	inquired	[whether	the captain	befriended	<i>the</i> <i>sailor</i>	before ...
EMBEDDED GAP, NON-ISLAND:								
Who	had	<i>the</i> <i>sailor</i>	assumed	[that	the captain	befriended	– <i>openly</i>	before ...
EMBEDDED GAP, ISLAND:								
Who	had	<i>the</i> <i>sailor</i>	inquired	[whether	the captain	befriended	– <i>openly</i>	before ...

The results of the ANOVA revealed only two sentence positions with significant effects (positions 5 and 9).²

5.4.2.1.1 Position 5 (clause boundary)

At sentence position 5, the clause boundary (*that* vs. *whether*), there was a main effect of GAP [$F_1(1,47) = 5.63, p = 0.022, F_2(1,31) = 4.48, p = 0.043$] and a marginal interaction of GAP with STRUCTURE by items [$F_1(1,47) = 2.52, p = 0.12, F_2(1,31) = 3.05, p = 0.091$]. Paired comparisons revealed that the EMBEDDED ISLAND condition was read more slowly than the MATRIX ISLAND condition ($t(667.2) = 2.49, p = 0.013$) and marginally slower than the EMBEDDED NON-ISLAND condition ($t(688.6) = 1.72, p = 0.084$). There was no statistical difference found between either the MATRIX and EMBEDDED NON-ISLANDS (both containing *that*-clauses; $t(684.7) = 0.7, p = 0.483$) or the MATRIX NON-ISLAND and MATRIX ISLAND conditions (both containing matrix subject gaps; $t(695.37) = -0.15, p = 0.882$). The means at position 5 are shown in Table 5-6 and plotted in Figure 5-2.

Table 5-6: Position 5 residual reading times. Mean (standard error)

		STRUCTURE	
		NON-ISLAND	ISLAND
GAP	MATRIX	-25.16 (13.91)	-29.7 (10.11)
	EMBEDDED	-11.73 (17.89)	24.32 (17)

² All positions were analyzed, but as inspection of Figure 5-1 may induce some curiosity, it is worth explicitly stating that there were no significant differences at positions 6, 7, or 11.

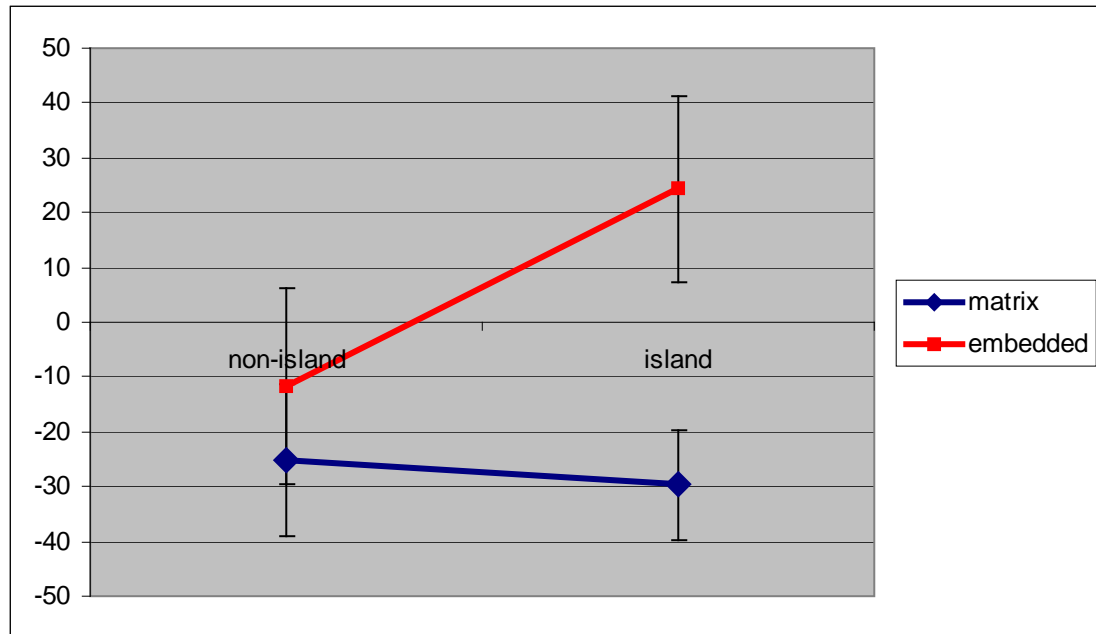


Figure 5-2: Position 5 residual reading times

5.4.2.1.2 Position 9 (*before*)

At sentence position 9, immediately after the embedded gap (*before*), there was a main effect of GAP ($F_1(1,47) = 8.49, p = 0.006, F_2(1,31) = 5.63, p = 0.024$), but no statistical interactions. The MATRIX condition (immediately following *the sailor*) was read more slowly (5.24 residual RT) than the EMBEDDED condition (immediately following *_ openly*, -26.42 residual RT).

5.4.2.2 Discussion

5.4.2.2.1 Position 5 (clause boundary)

The island violation condition was read the most slowly at the clause boundary, as predicted by the capacity-constrained account of islands (and not necessarily by the similarity-based interference account). If the filler is being actively held in working memory, then the additional cost of crossing the more complex (ISLAND) clause boundary (position 5) should be evident. While the current data appear to support the possibility of this compound penalty, we do not see independent evidence of costs of either (i) holding a filler across the clause boundary or (ii) clause boundary complexity.

The cost of holding a filler in memory while crossing the clause boundary should be evident when comparing the two NON-ISLAND conditions. With the same lexical item (*that*) at the clause boundary, the capacity-constrained account predicts that the EMBEDDED GAP condition should incur a cost compared to the MATRIX GAP condition as it contains a long-distance dependency that ranges across the clause boundary. While there was a numeric trend in this direction, it is not statistically significant (EMBEDDED -11.73 residual reading time, slower than MATRIX -25.16 residual reading time, $p = 0.33$). This is contrary to the findings in Frazier & Clifton (1989) who reported a processing cost for carrying a filler across a (non-island) clause boundary.

Frazier and Clifton compared single (5.2 a,b) and bi-clausal sentences (5.2 c,d) while manipulating whether the gap occurred earlier (5.2 a,c) or later (5.2 b,d) in the sentence.

(5.2 a) One-clause, early gap

What did / the cautious old man / whisper _ / to his fiancée /
during the movie / last night?

(5.2 b) One clause, late gap

What did / the cautious old man / whisper / to his fiancée about _ /
during the movie / last night?

(5.2 c) Two clauses, early gap

What did / you think the man / whispered _ / to his fiancée /
during the movie / last night?

(5.2 d) Two clauses, late gap

What did / you think the man / whispered / to his fiancée about _ /
during the movie / last night?

Modified from Frazier and Clifton (1989)

Frazier and Clifton reported that the two-clause sentences (5.2 c,d) were read more slowly than the one-clause sentences. This slowdown did not occur prior to encountering the verb that governs the gap (*whispered*), which Frazier and Clifton took as suggestive evidence “that carrying a filler across a clause boundary and

assigning it to a gap, rather than some other aspect of two-clause sentences, is the source of the difficulty...” (pg 104).

The position in the current experiment equivalent to Frazier and Clifton’s gap-governing verb (*whispered*) is position 7 (*befriended*). At position 7, the numeric trend was that that the EMBEDDED conditions (in which a filler would have had to be held in working memory while crossing a clause boundary) were read more slowly (95.96 residual reading times) than the MATRIX conditions (49.25 residual reading times), as predicted by the Frazier and Clifton results. However, there was no statistical difference found at this position.³

The comparisons made in the current experiment differ in a number of respects from those in Frazier and Clifton. First, Frazier and Clifton’s manipulation of gap location did not correspond to the MATRIX vs. EMBEDDED gap comparison in the current experiment. All conditions in the current experiment were bi-clausal and differed in whether the filler-gap dependency crossed into the second clause or not. Frazier and Clifton compared filler-gap dependencies of similar lengths (both temporally and structurally) but which differed in whether a second clause was introduced or not. Second, the materials for the current experiment used an overt complementizer (*that/whether*), while the Frazier and Clifton materials did not. Third, Fraizer and Clifton’s single clause conditions contained a larger NP (*the cautious old*

³ Examining the paired comparisons, we can confirm that the individual effects were not statistically significant. In the NON-ISLAND conditions, the EMBEDDED condition was read more slowly (96.74 residual reading time) than the MATRIX condition (53.58 residual reading time), but this was not statistically significant ($p = 0.21$). In the ISLAND conditions, the EMBEDDED condition was read more slowly (94.48 residual reading time) than the MATRIX condition (41.04 residual reading time), but this was also not statistically significant ($p = 0.10$). The raw reading times were also not statistically significant.

man) than did the bi-clausal sentences (*the man*), whereas the current experiment's sentences differed in whether there was a gap/adverb or NP before the embedded clause. Any or all of these differences could be contributing factors to why the pattern in Frazier and Clifton (1989) is not replicated here.

It is also possible that 'some other aspect of two-clause sentences is the source of the difficulty' (Frazier and Clifton 1989, pg 104). It could be that for Frazier and Clifton's observed slowdown it is sufficient merely to have a filler and a clause boundary and not crucial to have a filler-gap dependency cross a clause boundary. Then no difference would be predicted for the current experiment. Another option is that it is not the clause boundary itself, but rather the second verb of the sentence that causes the slowdown for Frazier and Clifton's bi-clausal sentences. This would be compatible with the slowdown not starting until *whispered* is encountered. In either case, one of these reinterpretations of Frazier and Clifton's reading time penalty would still be a complication for the capacity-constrained account of islands, which relies on Frazier and Clifton's 'clause difficulty' interpretation for one of the three processing difficulties that contribute to an island-violation (see Chapter 2, section 2.3.3.1).

The complexity of the clause boundary is also expected to generate a processing cost (Kluender and Kutas 1993b). This should be evident in the current results when comparing the two MATRIX GAP conditions. When the *wh*-dependency has already been resolved in the matrix clause, the only difference that remains at the clause boundary position is the lexical difference in the clause boundary itself (and the previous verb that selects it). The comparison of *that* and *whether* was again not

statistically significant, and the residual reading times actually trended in the opposite direction, with the more complex ISLAND *whether* boundary read faster. Since residual reading times were used here, the difference in length was already factored out.

It is possible that using the residuals here also unintentionally factored out other lexical or lexically dependent syntactic differences. To assess this possibility, this sentence position was also examined using raw reading times rather than residuals. The paired comparison of ISLAND vs. NON-ISLAND in the MATRIX GAP conditions still did not reach significance ($p = 0.145$), though the numerical trend was now in the predicted direction, with ISLANDS read more slowly (446.48 msec) than NON-ISLANDS (416.63 msec). However, as discussed in section 5.4.3.2.1, the lack of such a clause boundary effect can be explained when the cognitive measures are taken into account. The low reading span group did show the expected processing cost of the clause boundary, but the high reading span group did not.

5.4.2.2.2 Position 9 (*before*)

In the first spillover region after the embedded gap, there was a slowdown in the MATRIX conditions compared to the EMBEDDED conditions. All conditions at position 9 contained the same lexical item, but the preceding position differed according to the GAP manipulation. The MATRIX condition at position 8 contained an NP (*the sailor*) while the EMBEDDED condition at position 8 contained an adverb (*_openly*). If the slowdown were due to the process of gap-filling, we would expect the

EMBEDDED condition to be slower than the MATRIX condition, but this was not the case.⁴

Instead, the word following the definite referent *the sailor* was read more slowly than the adverb *openly*. Recall that residual reading times were measured here, so length differences between the NP and adverb were already controlled for. Additionally, these items were frequency matched (with the exception of *the*). If length, frequency, or word class differences between an NP and an adverb were directly responsible for this difference, then there should have been a similar difference at position 4 as well (immediately following a position where the same two lexical items were compared). However, there was no indication of such a lexical difference in the matrix clause. Assuming that the embedded gap was successfully processed in the EMBEDDED NON-ISLAND condition, it is impossible to claim that participants failed to process the gap in the EMBEDDED ISLAND condition, as reading times did not differ in these conditions at or around the embedded gap.⁵ There has to be a reason why the difference at position 9 (bold and underlined in 5.3) did not also emerge at position 4 (underlined in 5.3) following the same lexical items, and what differs between these two positions is the surrounding sentence (5.3).

⁴ Additionally, note that there is no reading time evidence of a cost of gap filling in the matrix clause in the current experiment (position 3 is the gap position and position 4 would be the first spillover region).

⁵ The fact that no processing cost was observed at the embedded gap position is evidence against a similarity-based interference view of islands, as there is no apparent difference in processing costs when the filler should be retrieved.

(5.3 a) MATRIX GAP: read more slowly at *before*

Matrix clause: Who had openly assumed

Embedded clause: that the captain befriended the sailor **before** ...

(5.3 b) EMBEDDED GAP:

Matrix clause: Who had the sailor assumed

Embedded clause: that the captain befriended openly **before** ...

If we are observing the processing cost of introducing a definite discourse referent to the sentence at position 9, we can plausibly entertain reasons as to why this is more difficult than it is in the matrix clause. At position 3, the EMBEDDED conditions introduce a definite discourse referent (*the sailor*; 5.3 b), which must be processed as the subject of a main verb. At position 8 however, while the MATRIX conditions (5.3 a) introduce this same definite discourse referent (*the sailor*), it must now be processed as the object of a verb that has as its subject another similarly definite NP (*the captain*). That is, the processing of *the sailor* in the embedded (but not the matrix) clause involves a more complex clausal integration with both another definite NP and a verb. This could be considered an instance of similarity-based interference, since two similar NPs are being integrated with a single verb.

In order to explain why this effect is not present when a gap (indicated by *openly*) is present in the embedded clause, we only have to observe that the filler *who* is not as similar to *the sailor* as *the captain* is. If this explanation is on the right track, then we predict that the difference between (5.3 a) and (5.3 b) at *before* would be

greatly reduced or even disappear if more similar (definite) fillers were used. Testing this prediction is not within the current scope of this experiment, however.

A second possibility is that the slowdown to *the sailor* in the MATRIX GAP conditions (but in the embedded clause, 5.3 a) reflects an end of clause wrap-up effect. This has the benefit of explaining why there are no effects to *the sailor* in the matrix clause, since *the sailor* is not sufficient to complete this clause. On the other hand, all of the arguments for the embedded verb (*befriended*) have been directly encountered (*the captain* and *the sailor*) when *the sailor* is read, and so the parser can consider this clause complete. One would have to argue, then, that when a gap is encountered in that embedded clause, the parser has to complete filler-gap integration before the clause is complete. Presumably, this filler-gap integration would take effort and have a cost, but such a cost was not found in the current experiment. So while a clause wrap-up effect may explain slower reading times following the embedded *the sailor*, the lack of a filler-gap integration cost is still puzzling. While a reading-time cost of filler-gap integration was not found here, the ERP experiment successfully demonstrates a post-gap LAN, which is interpreted as indexing this process (Chapter 6, section 6.4.2.2.4)

While the reason why the spillover region following the embedded gap position was read more slowly in the MATRIX conditions (i.e. reading times after *the sailor* were slower than after *_ openly*) cannot be tested here, the crucial finding for our present purposes is that the gap (*_ openly*) was not read more slowly in the ISLAND condition than in the NON-ISLAND condition. In other words, there is no evidence that

the difficulty in processing an island violation occurs when the cue for retrieval of the filler (i.e. the gap) is encountered. Thus we find no evidence that the processing difficulty of the island violation is due to content-addressable similarity-based interference, contra the predictions of the similarity-interference account of islands.

5.4.3 Median splits

In this section, I present and then discuss the findings of the self-paced reading experiment including the scores from the cognitive measures.

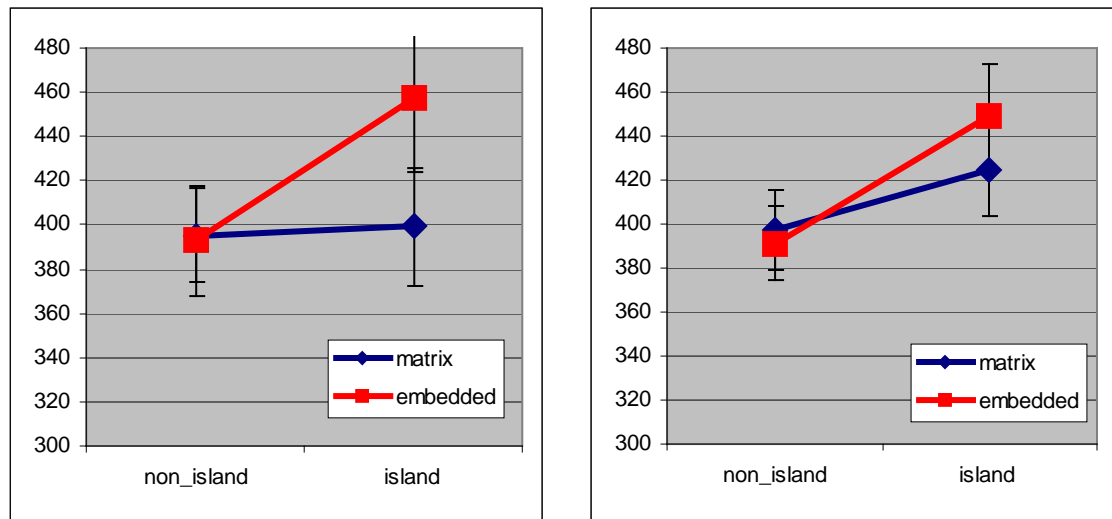
5.4.3.1 Results

In the following sections, I present only the results in which the median split groups interact with at least one of the linguistic manipulations (GAP and/or STRUCTURE). Main effects of the high and low groups for the n-back and flanker tasks occurred fairly regularly, with the high scorers reading 20-30 msec faster overall at a word position. As these are main effects of the cognitive measure and do not interact with the linguistic manipulations, they are not of interest here. Only position 5, the clause boundary, yielded significant interactions of the linguistic manipulations and cognitive measures and it did so with both reading span and memory interference, albeit in different ways.

5.4.3.1.1 Position 5 (clause boundary): Reading Span

At position 5, the clause boundary, the effect of READING SPAN GROUP was marginal by subjects and significant by items [$F_1(1,164) = 3.41, p = 0.067, F_2(1,239) = 6.24, p = 0.013$], indicating that the low span group read more slowly overall (415.54 msec) than the high span group (411.32 msec). The low span group (but not the high span group) read the clause boundary position more slowly in the MATRIX ISLAND condition than the MATRIX NON-ISLAND condition (424.56 vs. 399.25 msec) as indicated by a marginal three-way interaction of READING SPAN GROUP with STRUCTURE and GAP by items [$F_1(1,164) = 0.28, p = 0.60, F_2(1,239) = 3.36, p = 0.068$]. Paired comparisons indicated that in both span groups, the EMBEDDED ISLAND condition was read more slowly than the EMBEDDED NON-ISLAND condition (all $ps < 0.05$). However, only in the high span group was the EMBEDDED ISLAND condition (457.60 msec) read more slowly than the MATRIX ISLAND condition (399.25 msec, $t(198.2) = 2.025, p = 0.044$).

Additionally, the low span group showed a distinction that the high span group did not. The low span group read the MATRIX ISLAND condition (i.e. the *whether* complementizer: 424.56 msec) marginally more slowly than the MATRIX NON-ISLAND condition (i.e. the *that* complementizer: 397.56 msec, $t(472) = 1.83, p = 0.067$). That is, the low span group showed a difference of clause boundary (*that* vs. *whether*) even when the filler-gap dependency had already been resolved in the matrix clause. This is shown in Figure 5-3 B.



A: HIGH SPAN GROUP

B: LOW SPAN GROUP

Figure 5-3: Position 5 GAP x STRUCTURE (A) HIGH SPAN GROUP (B) LOW SPAN GROUP

5.4.3.1.2 Position 5 (clause boundary): Memory Lure

At the clause boundary the overall memory lure group analysis showed only a marginal main effect of MEMORY LURE GROUP by subjects [$F_1(1,168) = 2.76, p = 0.09, F_2(1,240) = 0.41, p = 0.52$], with high scorers reading the complementizer (*that/whether*) more slowly (422.46 msec) than low scorers (406.83 msec). Significant differences were found for the FORM LURE GROUP but not for the SEMANTIC LURE GROUP. There was a main effect of FORM LURE GROUP by items only [$F_1(1,160) = 0.26, p = 0.77, F_2(1,240) = 9.88, p < 0.001$], with the high group reading the complementizer more slowly (437.46 msec) than the low group (382.76 msec). There was also a marginal interaction of GAP with FORM LURE GROUP by items [$F_1(1,160) = 0.21, p = 0.81, F_2(1,240) = 2.48, p = 0.08$], while both groups read the EMBEDDED GAP

conditions slower than MATRIX GAP conditions, this slowdown was larger for the high lure group than the low group. Pairwise comparisons revealed that the HIGH FORM LURE GROUP read more slowly than the LOW FORM LURE GROUP both for the MATRIX GAP condition ($t(67.8) = 2.07, p = 0.04$) and in the EMBEDDED GAP condition ($t(67.9) = 2.33, p = 0.02$). These findings are shown in Figure 5-4.

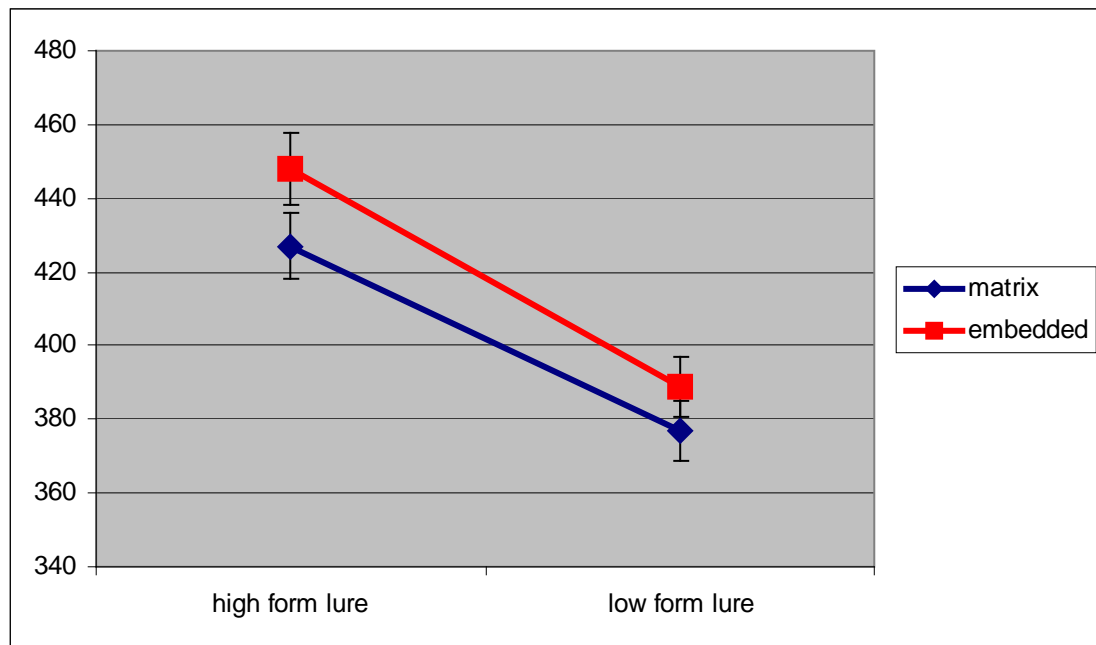


Figure 5-4: Position 5 GAP x FORM LURE GROUP

5.4.3.2 Discussion

Overall there was relatively little interaction of cognitive measures with the linguistic manipulations of GAP and STRUCTURE. From a certain point of view, this is reassuring, as the experimental stimuli used here were carefully controlled and

matched each other word for word as much as possible (Chapter 3, section 3.2). The interactions between the cognitive measures and linguistic manipulations were found only at the clause boundary (*that/whether*). While only one sentence position interacted with the cognitive measures, two different measures- reading span and memory lure- interacted in different ways with the linguistic data. The following sections discuss these effects in turn.

5.4.3.2.1 Position 5 (clause boundary): Reading span

To facilitate the discussion of the reading span effects, (5.4) presents the experimental conditions up to the current point of comparison (the clause boundary)

- | | |
|------------------------------|--|
| (5.4 a) MATRIX NON-ISLAND: | Who had openly assumed <u>that</u> ... ? |
| (5.4 b) EMBEDDED NON-ISLAND: | Who had the sailor assumed <u>that</u> ... ? |
| (5.4 c) MATRIX ISLAND: | Who had openly inquired <u>whether</u> ... ? |
| (5.4 d) EMBEDDED ISLAND: | Who had the sailor inquired <u>whether</u> ... ? |

In both the high and low span groups the island violation condition (EMBEDDED ISLAND; 5.4 d) was read more slowly at the clause boundary than the EMBEDDED NON-ISLAND condition (5.4 b; i.e. the acceptable long-distance filler-gap dependency into an embedded *that*-clause; see Figure 5-3), indicating a cost of encountering the *whether*-island clause boundary while the filler-gap dependency was still unresolved. These

results are consistent with the results discussed in section 5.4.2.2.1. That this processing burden occurred at the clause boundary was consistent with the capacity-constrained view of working memory.

Additionally, the low span group (and not the high span group) showed evidence of a processing cost for the more complex interrogative *whether* clause boundary compared to the declarative *that* clause boundary. This effect was observed in the MATRIX conditions with (5.4 c) read significantly more slowly than (5.4 a), indicating that the low span group had difficulty with the *whether* clause boundary even when there was no incomplete filler-gap dependency present. The high span group showed no such slowdown, only reading the island violation (EMBEDDED ISLAND; 5.4 d) more slowly than the other three conditions (see Figure 5-3).

Thus, the high span group made a two-way distinction and the low span group made a three-way distinction (see Figure 5-3). In the high span group, there was a slowdown in reading times for the island violation condition and everything else was read equally quickly. In the low span reading group, in addition to the slowdown for the island violation condition, there was also a slowdown for just the island clause boundary. This latter pattern is the pattern predicted by the capacity-constrained view.

As discussed in section 5.4.2.2.1, if there is a cost for processing a more complex clause boundary in and of itself, without a filler-gap dependency crossing it, it is reasonable to expect this to be visible in the reading times. The fact that only the LOW SPAN GROUP showed this pattern while the HIGH SPAN GROUP showed no cost for processing the clause boundary by itself is evidence for the importance of working

memory in processing that complex clause boundary. The LOW SPAN GROUP had some difficulty with the complex clause boundary and slowed down for it. In contrast, the HIGH SPAN GROUP, having more cognitive resources available to them, was not vexed by the complexity of the clause boundary and showed no slowdown in reading time.

Both groups showed the processing penalty for the combination of having an EMBEDDED GAP within an ISLAND. While the HIGH SPAN GROUP was better able to process the complex clause boundary when the filler-gap dependency did not cross it, they had no such benefit in the island-violation condition when it did. Here we see yet another pattern where the high scoring cognitive group demonstrates a processing benefit for an easier condition, but not for the more complex one (i.e. a Processing Benefits Schedule, PBS ‘simply (only) benefit;’ see discussion in Chapter 4, section 4.2.2.1).

5.4.3.2.2 Position 5 (clause boundary): Form lure

At the clause boundary, form lure effects did not interact with the STRUCTURE (ISLAND/NON-ISLAND) manipulation. To facilitate discussion, (5.5) collapses the relevant experimental manipulation (GAP POSITION) up to the current point of comparison (the clause boundary *that/whether*)

(5.4 a) MATRIX: Who had openly {assumed that / inquired whether } ... ?

(5.4 b) EMBEDDED: Who had the sailor {assumed that / inquired whether } ... ?

At the clause boundary, the HIGH FORM LURE GROUP slowed down overall compared to the low scoring group, and slowed down more in the EMBEDDED GAP conditions (5.4 b) than in the MATRIX GAP conditions (5.4 a; see Figure 5-4). Unlike the READING SPAN GROUP differences (section 5.4.3.2.1), there was no interaction with STRUCTURE, indicating that this was only a distance effect (i.e. whether the GAP was located in the MATRIX or EMBEDDED clause). In other words, clause boundary type did not matter for this comparison. Most striking, however, was that the high scoring group slowed down rather than speeded up compared to the low scoring group. This pattern was thus the opposite of a processing benefit for the high scorers, and is inconsistent with both capacity-constrained and similarity-interference views of working memory in sentence processing.

When we compare this pattern of responses to the pattern of responses in the acceptability judgment experiment (Chapter 4, section 4.4.2.3.2), the overall picture becomes more clear. In the acceptability study, the high scoring FORM LURE GROUP rated sentences differently than the low scoring group did. Like the effect here, the acceptability effect was specific to the GAP manipulation (and independent of the STRUCTURE manipulation). The high scorers rated MATRIX gaps (5.4 a) higher than, and EMBEDDED gaps (5.4 b) lower than the low scorers did; effectively distinguishing between the short-distance and long-distance filler-gap dependencies more clearly. This was interpreted as the high group being more aware of this distinction and assigning acceptability scores accordingly. Here, we see the on-line reflection of that off-line judgment. At the clause boundary, the high group appears to be more aware of

the difference between the MATRIX and EMBEDDED conditions and is more affected by it. The low group, which is less aware of the distinction between the gap positions at this point in the sentence, reads through the clause boundary more quickly.

An outstanding question is why either of these effects (the on-line reading time and/or off-line acceptability distinctions) should co-vary specifically with the form lure task. One possibility is that high form lure scores require a fine-grained attention to detail. In order to score highly on the form lure, a participant must avoid lures based on sublexical differences (phonological or orthographical differences, compared to more general semantic categories in the semantic lure task). This heightened attention to detail would then manifest here as increased awareness of the MATRIX/EMBEDDED GAP distinction. It is interesting that this occurred at the clause boundary and not at the gap positions (matrix or embedded). It seems these readers were aware in particular of when a filler-gap dependency was crossing a clause boundary (though it apparently did not matter whether this clause boundary was an island or not).

An alternative possibility is that the differences in the acceptability judgment task reflect off-line, post-processing rating task differences (Chapter 4, section 4.4.2.2). While this may be a contributing factor, it cannot be the entire story. If the differences in the acceptability judgments were due to differences in the way in which the high and low scorers approached the rating task, then we would not expect to see complementary patterns in the on-line reading task – because these on-line effects occurred before any potential post-processing could have taken place, and also because there was no rating task for the self-paced reading experiment. Since this

possibility appears untenable, it is more reasonable to conclude (as stated above) that the differences in FORM LURE GROUP, both reading times and acceptability judgments, reflect a difference in diligence of processing.

The increased diligence of processing by the HIGH FORM LURE GROUP patterns in a way that suggests a speed-accuracy tradeoff. At the clause boundary, these readers slow down overall compared to low scorers. The high scorers additionally slow down more than low scorers do for the EMBEDDED conditions (5.4 b). This slowdown in processing (for only the high scorers) mirrors a greater acceptability distinction made between these same conditions (again, only for high scorers; Chapter 4, section 4.4.2.3.2). It appears that the high scorers are trading processing speed for sensitivity.

Based on this discussion, it is unlikely that the differences found at the clause boundary in the FORM LURE GROUP were due to differences in how the groups employed a content-addressable retrieval mechanism (as proposed in the similarity-interference view of working memory; Chapter 2, section 2.3.3.2). This is because at this sentence position (clause boundary) the gap has either already been filled (5.4 a) or has not yet been encountered (5.4 b). This effect does not interact with the STRUCTURE manipulation, meaning that it does not directly bear on island-violations (and interaction of GAP and STRUCTURE). Thus, even though co-variation with a memory lure measure was found, this reading time data does not constitute evidence in favor of a similarity-interference account of islands.

5.5 Summary

The key self-paced reading time results from this chapter are that (i) the processing cost of *whether*-island violations occurred at the clause boundary and not at the embedded gap position, (ii) at this clause boundary the high span readers did not show a cost of the type of clause boundary by itself, while low span readers did, and (iii) the form lure scores co-varied with reading times for the GAP manipulation in a way that suggests a type of speed-accuracy tradeoff in attending to details of gap position.

In the acceptability judgment study reported in Chapter 4 (section 4.4.2.3.2), the high scorers on the form lure task rated the difference between the MATRIX and EMBEDDED GAP conditions as greater (the MATRIX GAP rated higher and the EMBEDDED GAP rated lower) than did low scorers. In the present reading time study, high scorers read both the MATRIX and EMBEDDED GAP conditions more slowly at the clause boundary than did low scorers. This is interpreted as high scorers being on some level more cognizant than the low scorers of the difference between a shorter and longer filler-gap dependency when a clause boundary is crossed (section 5.4.3.2.2).

Variations in the reading time patterns of high and low span participants indicated a difference in how the clause boundary was processed. Low span participants had difficulty processing the interrogative *whether* clause boundary compared to the declarative *that* clause boundary even when it was not crossed by a filler-gap dependency. The high span participants showed no such penalty when the

clause boundary was not crossed by a filler-gap dependency (section 5.4.3.2.1), suggesting that they have a processing benefit for this easier to process condition (i.e. a Processing Benefits Schedule ‘simply (only) benefit;’ see discussion in Chapter 4, section 4.2.2.1).

However, both the high and low span readers showed a processing penalty at the clause boundary for the island violation condition (5.4.2.2.1), exactly where processing difficulty is predicted under the capacity-constrained view of working memory. While the similarity-interference view is also compatible with a processing cost at the clause boundary (because it must allow for effects of predictive processing), there is no processing penalty found at the embedded gap position, precisely where the similarity-interference view predicts that it should occur.

5.6 Conclusion

The results of this study were more consistent with a capacity-constrained view of working memory than with a similarity-interference view. As predicted by the capacity-constrained account, an interaction of GAP and STRUCTURE at the clause boundary, which was furthermore modulated by reading span, was found. This is not to say that content-addressable memory and similarity-based interference are unimportant in filler-gap dependencies, but merely that this model of working memory does not successfully capture the differentiation between processing costs of *whether*-island violations and closely related controls observed here.

Not every aspect of the current data aligns with a capacity-constrained processing account of islands, however. Not all readers showed a cost for a *whether* clause boundary compared to a *that* clause boundary, independent of a filler-gap dependency, as the capacity-constrained account predicts. The basic idea of the capacity-constrained account was supported, however, at least in low span readers. Processing difficulties in *whether*-islands can be explained as the combination of separate processing difficulties, including the cost of having a longer filler-gap dependency and the cost of encountering a more complex clause boundary before it is resolved. But for high span readers, no cost of the complex clause boundary was observed. These high span readers were able to handle this lexical semantic complexity with no slowdown in reading time. However, when this lexical semantic complexity was combined with an unresolved filler-gap dependency (EMBEDDED ISLAND condition), the high span readers slowed down just as much as the low span readers did. This indicates that both high and low span readers had an equal amount of difficulty processing the clause boundary in an island violation.

This pattern reflects a Processing Benefits Schedule (PBS) ‘simple (only) processing benefit’ for the high span readers (see Chapter 4, section 4.2.2.1). The high span readers had a processing benefit for the *whether* clause boundary itself compared to low span readers, but did not have a benefit in the more difficult condition- when this clause boundary was combined with an unresolved filler-gap dependency. This lack of benefit for the high span readers suggests the possibility that the combined processing cost present in the island violation condition may represent a ceiling effect

for what the human parser can handle simultaneously. Whether this processing ceiling is the reflection of a grammatical constraint or is itself a cause of the unacceptability of island violations remains an unresolved issue, however (see Chapter 4, section 4.2 for discussion). The next chapter examines these processing costs using event-related potentials, uncovering processing patterns not apparent in the reading time measures reported here.

Chapter 6: Event-Related Potentials Experiment

6.1 Introduction

In Chapter 5, we saw behavioral evidence for a processing cost at the clause boundary for *whether*-island violations, which were also rated as the least acceptable in Chapter 4. We now turn our attention to the brain responses to islands by using Event-Related Potentials (ERPs). Key points of interest are: how the brain responds in real time to sentences rated as unacceptable off line (Chapter 4), what this can tell us about the processing of these sentences in addition to reading time data (Chapter 5), and how these responses vary with cognitive measures.

The results from the ERP experiment reported below find consistent brain responses to gaps, namely a Left Anterior Negativity (LAN) elicited in the position following the gap. This occurs in both matrix and embedded gaps, and even in gaps embedded within a *whether*-island. These LAN effects are statistically indistinguishable from one another. The ERP response elicited from *whether*-island violations that distinguishes it from other conditions is an N400 effect at the embedded gap position (see Table 6-1). This effect is argued to reflect the low predictability of a gap inside a *whether*-island.

The remainder of this chapter is organized as follows. Section 6.2 presents the predictions for this experiment. Section 6.3 presents the methods of the current experiment - though for details about the measures of individual differences or

materials design see Chapter 3. Section 6.4 presents results and discussion of the basic data (section 6.4.2) as well as the co-variation analysis (6.4.3). Section 6.5 briefly summarizes these findings and section 6.6 concludes the chapter.

6.2 Predictions

The design of the materials for this experiment is detailed in Chapter 3, section 3.2. The materials are again briefly discussed below in section 6.3.2, but for the purpose of presenting the predictions of the current experiment, it is useful to refer to specific sentence positions. Critical positions are presented in Table 6-1 for convenience.

Table 6.1: Critical comparisons within the stimulus sentences, indicating both numbering and labels relative to the gap position in both the matrix and embedded clauses

		<i>pre-gap position</i>	<i>gap position</i>	<i>post-gap position</i>			
Matrix clause:							
Position:	1	2	3	4			
	Who	had	_ openly/ the sailor	assumed/ inquired			
Embedded clause:							
Position:	5	6	7	8	9	10/11	12
	that/ whether	the captain	befriended	the sailor/ _ openly	before	the final/ mutiny	hearing?

6.2.1 Lexical differences (positions 3 and 8, _ *openly/the sailor*)

I will proceed through these predictions, for the most part, chronologically. However, since the design of the materials matching positions of interest in both the matrix and embedded clauses (Table 6-1), I will refer to related positions as appropriate. Recall from Chapter 3 (section 3.2) that the ‘gap position’ refers to the position in the sentence when a reader knows whether the gap is in the matrix or embedded clause (i.e. it is not necessary that there is theoretical agreement about where the gap should be assumed; the gap position is a disambiguation point). The gap positions (3 and 8 in Table 6-1) both compared words of different grammatical classes (*the sailor* vs. *openly*). While the words chosen were controlled for length and frequency (Chapter 3, section 3.2), it seemed highly unlikely that there would be no observable differences between *the sailor* and *openly*. Crucially, the design of the experiment balances the occurrence of these items across conditions. *Openly* is present in the matrix clause in the MATRIX GAP conditions, and in the embedded clause in the EMBEDDED GAP conditions. When *openly* does not appear, *the sailor* does (see Table 6-1). Thus, any lexical differences between *the sailor* and *openly* should be visible at both locations (though the effects should be ‘flipped’ between the conditions at the two locations). Any effects not evident at both sentence positions must be assumed to be more than a simple lexical difference between *the sailor* and *openly*.

These differences, namely an increased N400 and LAN to *the sailor* compared to *openly*, are compared in section 6.4.2.1.6 and discussed in section 6.4.2.2.4.

6.2.2 Post-gap LAN (positions 4 and 9, *assumed/inquired*)

A response of Left-Anterior Negativity (LAN) in filler-gap dependencies has been reported following both fillers and gaps. This has been interpreted as “the storage of a filler in working memory and its subsequent retrieval” (Kluender and Kutas 1993a, pg 205). Since the current materials uniformly begin with the filler *who* (position 1, Table 6-1), it is not possible to observe the filler-related LAN response. However, based on the pattern of results reported in the literature (see Chapter 2, section 2.2.5.2.1 for discussion) we expected to observe a gap-related LAN response after both the matrix and the embedded gap positions (4 and 9, Table 6-1). While we predicted that the post-gap LAN would be visible in the matrix clause, Kluender and Kutas (1993a,b) did not observe this effect. The current materials differ from Kluender and Kutas’ in that the materials they used did not separate the filler from the gap position (i.e. point of gap position disambiguation).¹ Since we do separate the filler from the gap position (point of disambiguation), we expected to be able to see this effect if the same process reported in the literature for longer distance filler-gap dependencies also occurs for shorter dependencies. In the embedded clause (position 9, Table 6-1) LAN effects should be observed after the gap position both when embedded within a *whether*-island and a *that* clause. Kluender and Kutas (1993a,b) reported LAN effects when the gap is within a grammatical embedded clause (see also King & Kutas 1995) as well as when it occurs inside a *wh*-island.

¹ While some readers may argue that the gap could/should be placed immediately after *who* in the matrix gap conditions, recall that we are using ‘gap-position’ to indicate the earliest point in the sentence at which the reader knows that the clause contains a gap.

Both of these predictions were confirmed in the present study. These LAN effects are compared in section 6.4.2.1.7 and discussed in section 6.4.2.2.4.

6.2.3 Sustained LAN (position 4 or later)

The LAN elicited after the filler has been encountered has sometimes been reported to constitute a sustained effect in English (Kluender & Kutas 1993a,b; King & Kutas 1995; Phillips et al. 2005 – but see McKinnon & Osterhout 1996; Kaan et al. 2000 for non-replications). This sustained effect has been claimed to reflect the ongoing cost of holding a filler in memory until it is associated with its gap. However, since the current materials (Table 6-1, also section 6.3.2) did not differ in the position of the filler (position 1), the start of this sustained effect would not be visible in the results. However, it was still possible that a difference might appear starting after position 3, where the MATRIX GAP conditions can complete a filler-gap dependency (_ *openly* at position 3), but the EMBEDDED GAP conditions must still wait for the gap site (*the sailor* at position 3). If sustained anterior negativity indexes the ongoing cost of holding a filler in working memory, then this should be visible at or after this point, since a filler needs to be held from this point on only in the EMBEDDED GAP conditions.

This effect was not found in the current experiment, at least partially due to other differences in the sentences, such as the lexical differences at the gap site (see section 6.2.1 above). However, a long-lasting effect was found for the post-gap LANs. These ‘lingering LANs’ are discussed in section 6.4.2.2.5.

6.2.4 Clause boundary N400 (position 5, *that/whether*)

Comparing a clause boundary headed by *who* to one headed by *that*, Kluender and Kutas (1993b) reported an increased N400 response to *who*. This N400 difference was only present in clauses embedded in yes/no questions, and not those embedded in *wh*-questions. They speculated that the increased processing load of the long-distance dependency in a *wh*-question “somehow overrides the lexical semantic effects seen in yes/no questions” (pg 601). Thus, in the current experiment, it was possible that there would be no N400 difference between the *that* and *whether* clause boundaries when the filler-gap dependency crosses over that clause boundary (EMBEDDED GAP condition). It did seem possible, however, that this difference would emerge when the gap is in the matrix clause. When the gap is in the matrix clause (position 3 in Table 6-1) the filler-gap dependency will have been already resolved by the time the clause boundary is encountered. Since this dependency is already resolved, the ‘override’ mentioned by Kluender and Kutas (1993b) should not be present.

This N400 difference was not found in either case, however. An earlier effect (250-350 msec) is reported in section 6.4.2.1.4. The lack of an N400 effect is discussed in section 6.4.2.2.6.

6.2.5 Pre-gap P600 (position 7, *befriended*)

Multiple studies have reported a P600 effect at the position prior to a gap (Kaan et al. 2000; Fiebach, Schlesewsky & Friederici 2002; Phillips et al. 2005 and Gouvea et al. 2010). This effect, which has been interpreted as an index of syntactic integration difficulty, should be present at the pre-gap position in the embedded clause (position 7, *befriended*, Table 6-1). It was not expected at the pre-gap position of the matrix clause since at that point the sentences are all identical (*Who had...*) and the parser has no way to predict whether a gap is immediately upcoming.

However, we did find late, broad positivities at both the matrix and embedded pre-gap positions. This raises questions over whether this is an index of integration difficulty, or simply recognition of the gap starting at the next position. This is discussed in more detail in section 6.4.2.2.2.

6.2.6 Embedded gap position lexical differences and embedded post-gap LAN (positions 8 and 9)

These predictions have already been discussed above, but are mentioned again here so that they are highlighted in the chronological order of predictions through the sentence.

First, recall from section 6.2.1 that for position 8 (*the sailor* / *_ openly*) we predict a ‘flipped’ pattern of the lexical differences observed at position 3 (*_ openly* / *the sailor*), since the same lexical items are used, but in different condition. Any

effects not evident at both sentence positions must be assumed to be more than a simple lexical difference between *the sailor* and *openly*. These differences, namely an increased N400 and LAN to *the sailor* compared to *openly*, are compared in section 6.4.2.1.6 and discussed in section 6.4.2.2.4.

Second, recall from section 6.2.2 that for position 9 (*before*) we predict a post-gap LAN, just as in the post-gap position 4 (*assumed/inquired*), though again, the conditions will be ‘flipped’ based on the experimental conditions. These LAN effects are compared in section 6.4.2.1.7 and discussed in section 6.4.2.2.4.

6.2.7 Sentence-final N400 (position 12, *hearing?*)

Sentence-final N400s have been reported at the end of ungrammatical sentences (e.g. Osterhout & Holcomb 1992; Hagoort, Brown & Groothusen 1993; McKinnon & Osterhout 1996) as well as at the end of syntactically complex but grammatical garden-path sentences (e.g. Osterhout 1990). This pattern was also reported for half the participants in Kluender and Kutas (1993b), with the most negative response elicited by the island violation. Thus, we expected to see a similar pattern in the current results: the island violation condition was the most negative at the sentence-final position 12.

The basic effects of the sentence-final position are presented in section 6.4.2.1.8, but this effect is discussed after the cognitive measures are considered first (section 6.4.3.2.3).

6.2.8 Processing cost of *whether*-island violation (multiple possible positions)

As in the reading time study (Chapter 5), we expected to see evidence of an on-line processing difficulty for the unacceptable island violation condition. That is, we expected an interaction of GAP position and clause STRUCTURE. Based on the self-paced reading data, we expected this to occur at the clause boundary. However, should ERPs prove more sensitive to the retrieval of the filler from memory than self-paced reading turned out to be, we might see an interaction surrounding the embedded gap site (i.e. an additional effect in the P600 or LAN responses) – especially if there is processing difficulty with a content-addressable memory process (i.e. due to similarity-based interference).

We did find an expected interaction of GAP and STRUCTURE at the embedded gap site, but it was an N400 rather than a P600 or LAN effect. Section 6.4.2.2.7 discusses why this effect is unlikely to be due to a difficulty with retrieval and is more likely to reflect a difference in predictability of the gap, stemming from differences with the clause boundary (i.e. a gap is less predictable within a *whether*-island). The lack of effect at the clause boundary is discussed in section 6.4.2.2.6.

6.2.9 Cognitive measures co-variation (multiple possible positions)

As with the self-paced reading study (Chapter 5), any of the above effects that are found will be examined to see if they co-vary with the cognitive measures participants completed (Chapter 3, section 3.3). In particular, it is expected that the LAN could co-vary with reading span, as the LAN has been previously associated with working memory processes (Kluender and Kutas 1993a,b; King & Kutas 1995; Chapter 2, section 2.2.5.2.1), although the P600 could also be implicated under the view that it reflects syntactic integration (Kaan et al. 2000). Considering a similarity-based interference view of island processing difficulty (Chapter 2, section 2.3.3.2), any effect found at the embedded gap could be expected to co-vary with the memory lure task. However, neither of these patterns was found. Section 6.4.3.2 discusses the results of the cognitive measure co-variation analyses.

6.3 Methods

6.3.1 Participants

32 undergraduate students from UC San Diego participated in this experiment (19 female, mean age: 20.8). All were right-handed native English speakers with no known history of neurological disorder and gave informed consent. All procedures were done in compliance with the University of California, San Diego Human

Research Protections Program. Participants received course credit for their participation of up to two hours and/or were paid at the rate of \$8.00/hr.

6.3.2 Materials

The design of the experimental sentences is detailed in Chapter 3 (section 3.2), but is briefly summarized here for convenience. Full materials can be found in Appendix 2. The experimental sentences manipulated two factors of *whether*-islands. The factor GAP (two levels: EMBEDDED, MATRIX), indicating which clause the gap was located in, was crossed with the factor STRUCTURE (two levels: ISLAND, NON-ISLAND), indicating the nature of the embedded clause boundary (*whether* or *that*, respectively). There were 40 items for each of these four conditions, as well as 80 fillers, for a total of 240 sentences in the experiment. These were arranged in a Latin square design, forming four lists. The stimuli were pseudo-randomized such that no individual level of a factor (ex. EMBEDDED) was presented more than twice in a row. Additionally, the 240 sentences were split into 10 blocks of 24 sentences each, counter-balanced by conditions. See Table 6-2 for sample sentences.

Table 6-2: Experiment 3 sample stimuli set. Manipulations of STRUCTURE are indicated in bold while manipulations of GAP are indicated by italics. No specific claims are intended by the placement of the gap, which is meant only to indicate the on-line point of disambiguation of the gap position.

		STRUCTURE	
		NON-ISLAND	ISLAND
GAP	MATRIX	Condition 1: Who had <i>_ openly</i> assumed [that the captain befriended <i>the sailor</i> before the final mutiny hearing?]	Condition 2: Who had <i>_ openly</i> inquired [whether the captain befriended <i>the sailor</i> before the final mutiny hearing?]
	EMBEDDED	Condition 3: Who had <i>the sailor</i> assumed [that the captain befriended <i>_ openly</i> before the final mutiny hearing?]	Condition 4: Who had <i>the sailor</i> inquired [whether the captain befriended <i>_ openly</i> before the final mutiny hearing?]

Filler sentences varied in whether they were eight or ten positions long, in order to provide length variation in the entire experiment. The fillers were also all bi-clausal questions beginning with *who had* and balanced in use of *that* or *whether* at the clause boundary. Ten matrix verbs were used - *advised, asked, informed, instructed, notified, questioned, quizzed, reminded, and told* - and balanced across *that* and *whether* subordinate clauses. All of the *that*-clause sentences had matrix gaps and were ungrammatical. All of the *whether*-clause sentences had embedded gaps and were grammatical. Thus, when combined with the experimental sentences, there were an equal number of matrix and embedded gaps (120 each) and an equal number of ungrammatical *that* and *whether*-clause sentences (40 each; thus 1/3 of the sentences were ungrammatical). Unlike the experimental items, filler sentences were not specifically designed to be plausible (i.e. *a cartoonist, programmer and fisherman* in

the same sentence) or to have the most common vocabulary (i.e. *spelunker* and *coxwain* were used). This was consciously done so that the experimental items, by comparison, would seem even more plausible than they already were.² A full list of fillers can be found in Appendix 2.

6.3.3 Procedure

Participants completed the cognitive measures task before EEG capping and recording. After the recording session, participants completed a short acceptability judgment survey. Participants took, on average, 2 hours and 15 minutes to complete all three portions of the experiment.

6.3.3.1 Cognitive measures

Prior to the acceptability rating task, the e-prime software program (Schneider, Eschman, and Zuccolotto 2002) was used to administer four cognitive individual differences measures to the participants in the following order: reading span, n-back, flanker and memory interference (see Chapter 3, section 3.3 for details).

² When debriefed, participants often mentioned that some of the sentences were very strange. They then would recall some of the situations in these filler items, but the experimental items were never used as examples.

6.3.3.2 Electrophysiological recording

Following the completion of the individual cognitive differences measures, participants completed the ERP experiment. EEG was recorded using 29 tin electrodes mounted in a mesh Electro-Cap and according to the international 10-20 configuration (Figure 6-1). Additional loose lead electrodes were placed at the outer canthi of each eye and below the left eye to record eye movements (including blinks). Electrical impedance was kept below 5k Ω . EEG was amplified with an SA Instrumentation bioelectric amplifier and digitized online at 250 Hz.

Words at each position appeared for a duration of 300 msec, followed by 200 msec of blank screen before the next words appeared, for a 500 msec SOA. Following 25% of trials, a true/false comprehension prompt appeared. Presentation of stimuli would commence after the participant responded to the prompt or 20,000 msec elapsed, whichever came first. Participants were advised that they could rest or blink during this time before continuing. Trials were separated by a 1800 msec 'Blink' reminder on screen and a 1500 msec black screen for a total of 3300 msec between sentences (when there was no comprehension check). After each block of 24 sentences the participant was given a short break. The EEG recording portion of the experiment lasted an average of 50 minutes.

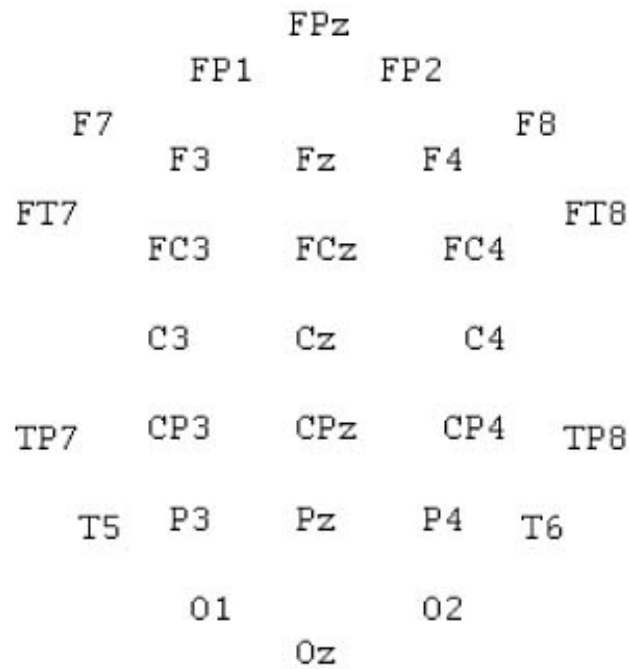


Figure 6-1: Electrode locations

6.3.3.3 Post-ERP acceptability judgments

After the EEG recording session, participants completed an acceptability judgment study with paper and pen. Participants were given 24 sentences from one of the other lists used in the ERP experiment. For each experimental condition, there were 4 items. 8 fillers were added to these 16 experimental sentences. Sentences were rated on a 7-point Likert scale (as in Experiment 1, Chapter 4).

Results from the acceptability survey were analyzed following the same procedure used in Chapter 4. Raw responses were transformed into z-scores and a linear mixed-effects model was constructed with PARTICIPANTS and ITEMS as random

factors. The linguistic factors GAP and STRUCTURE were included as fixed effects. Markov chain Monte Carlo sampling was used to estimate p-values in the languageR package for R (Baayen 2007, Baayen et al. 2008, R Development Core Team 2009, see also SWP).

6.3.4 EEG Analysis

EEG was referenced online to the left mastoid and re-referenced off-line to the average of the left and right mastoids. ERPs were timelocked to the onset of each critical position in each sentence (see Table 6-1). Artifacts due to eye movement and channel blocking were removed from the analysis below (13.3% of trials removed).

Mean amplitude was measured in standard latency windows for predicted components: 300-600 msec post-stimulus onset (LAN/N400) and 600-900 msec post-stimulus onset (P600/late positivity).

If visual inspection suggested that these standard windows were not suitably capturing a possible effect, windows were modified in an attempt to capture the potential difference, rounded to the nearest 50 msec. This procedure yielded a statistically reliable result only for the clause boundary (position 5, *that / whether*), which did not reveal the predicted N400 effect (see 6.2.2). The N400 window (300-600 msec) was modified to capture what appeared to be early positive (150-250 msec) and negative (250-350 msec) responses. These findings are reported (see 6.4.2.1.4)

and discussed (see 6.4.2.2.6) below for completeness, but no strong inferences are drawn from these data as non-standard windows are used.

ERP mean amplitudes were first submitted to an omnibus repeated measures ANOVA with the factors GAP (2 levels: MATRIX, EMBEDDED) x STRUCTURE (2 levels: ISLAND, NON-ISLAND) x ELECTRODE (29 levels, one for each electrode). If a significant interaction was found between ELECTRODE and any other factor, a distributional analysis was performed. The distributional analysis consisted of three repeated measures ANOVAs in order to capture data from all 29 electrodes while keeping the analysis symmetrical.³ Midline electrodes were submitted to a repeated measures ANOVA with 7 levels of ANTERIORITY (FPz, Fz, FCz, Cz, CPz, Pz, and Oz). Medial electrodes were submitted to a repeated measures ANOVA with 7 levels of ANTERIORITY and 2 levels of HEMISPHERE (LEFT, RIGHT; FP1/2, F3/4, FC3/4, C3/4, CP3/4, P3/4 and O1/2). Lateral sites were submitted to a repeated measures ANOVA with 4 levels of ANTERIORITY and 2 levels of HEMISPHERE (F7/8, FT7/8, TP7/8 and T5/6). All distributional analyses also included the linguistic variables GAP (2 levels: MATRIX, EMBEDDED) and STRUCTURE (2 levels: ISLAND, NON-ISLAND). For all ANOVAs, violations of sphericity (cf. Mauchly 1940) were corrected by the Huynh-Feldt correction (1976). Post-hoc comparisons were corrected by the Holm-Bonferroni

³ This was an issue due to missing cells in the electrode array.. For example, there were three pre-frontal electrodes (FP1/z./2) but five frontal electrodes (F7/3/z/4/8). The lack of right and left lateral pre-frontal channels created missing cells in the statistical analysis. ANOVAs were conducted in the ezANOVA package for R (Lawrence 2013, R Development Core Team 2009) which requires a symmetrical arrangement of cells. This three ANOVA distributional proceeded is not uncommon in the literature (e.g. Kaan et al. 2000, Boudreau, McCubbins & Coulson 2009).

correction (Holm 1979). Corrected p -values and original degrees of freedom are reported below.

In three of the sentence positions reported on below it appears that both a LAN effect and an N400 effect were elicited in the same 300-600 msec time window. In order to differentiate between these effects, two post-hoc distributional comparisons were used. This quadrant analysis excluded the midline (FPz, Fz, FCz, Cz, CPz, Pz, and Oz) and central (C3, Cz, C4) electrodes. The remaining 20 electrodes were submitted to a repeated measures ANOVA with two levels of ANTERIORITY (ANTERIOR, POSTERIOR) and two levels of HEMISPHERE (LEFT, RIGHT), in addition to the linguistic manipulations above. The interaction of these two factors resulted in four quadrants: left anterior (FP1, F7, F3, FT7, FC3), right anterior (FP2, F4, F8, FC4, FT8), left posterior (TP7, CP3, T5, P3, O1) and right posterior (CP4, TP8, P4, T6, O2).

The center analysis excluded the lateral (F7/8, FT7/8, TP7/8, T5/6), pre-frontal (FP1/z/2) and occipital (O1/z/2) electrodes in order to focus on the center of the scalp. The remaining 15 electrodes were submitted to a repeated measures ANOVA with 5 levels of ANTERIORITY and 3 levels of LATERALITY (F3/z/4, FC3/z/4, C3/z/4, CP3/z/4, P3/z/4).

6.4 Results and Discussion

In this section I first present and discuss the results of the post-ERP acceptability judgments (6.4.1) followed by the basic effects (6.4.2) and then the effects involving the cognitive measures (6.4.3).

6.4.1 Post-ERP acceptability judgments

6.4.1.1 Results

The results of the linear mixed-effects model revealed significant main effects of GAP ($p < 0.001$) and STRUCTURE ($p < 0.001$) and a marginal interaction of the two ($p = 0.068$). There were no significant interactions with cognitive measures. Results are shown in Figure 6-2 and mean z-scores are reported in Table 6-3).

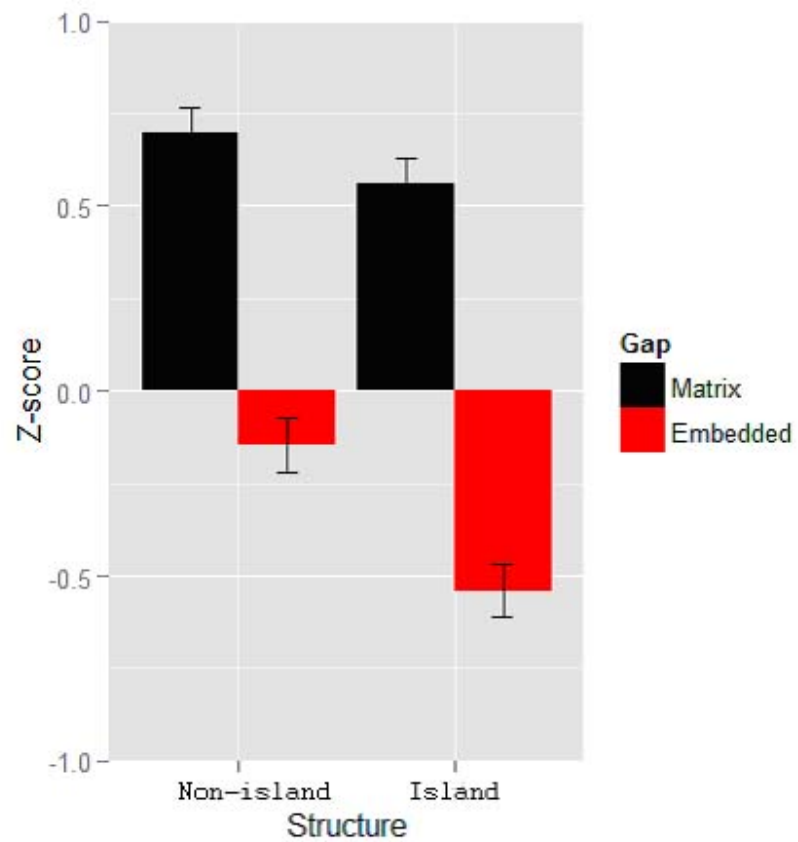


Figure 6-2: Post ERP acceptability scores

Table 6-3: Post ERP acceptability z-score transformed data. Means (standard deviation)

		STRUCTURE		
		NON-ISLAND	ISLAND	
GAP	MATRIX	0.697 (0.373)	0.561 (0.377)	0.629 (0.378)
	EMBEDDED	-0.146 (0.415)	-0.540 (0.41)	-0.343 (0.455)
		0.276 (0.578)	0.010 (0.678)	

6.4.1.2 Discussion

The results from the post-ERP acceptability judgments largely replicate the basic findings from the acceptability judgment in Experiment 1 (Chapter 4). EMBEDDED GAPS are rated as less acceptable than MATRIX GAPS, and ISLAND STRUCTURES are rated as less acceptable than NON-ISLAND STRUCTURES. There was a marginal interaction of GAP and EXTRACTION, with the EMBEDDED ISLAND being rated the least acceptable (see Figure 6-2). While this interaction was significant in Experiment 1, the marginal interaction here is not very surprising as there are a number of differences between the two experiments. Experiment 1 had eight items per condition while the post-ERP test only had four. There were many more (and more varied) fillers in Experiment 1. Additionally, the participants for the post-ERP acceptability judgments rated these sentences after having read 40 similar items for each condition. The crucial result then, is that participants have not satiated on the *whether*-island effect, even after reading many such sentences during the ERP experiment (cp. Sprouse 2009⁴). Thus, the results reported below should be interpreted as the traditional island violation patterns and not from a point of view that assumes that participants have lost the ability to distinguish the acceptability between these sentence types (i.e. we have no evidence that they have undergone syntactic satiation).

⁴ Other researchers have reported satiation for *whether*-islands (e.g. Snyder 2000; Hiramatsu 2000; Francom 2009; Crawford 2012), but these studies look at *whether*-islands in isolation, rather than relative to close (non-island) controls. Thus, it is plausible that the current participants would have shown a satiation pattern for *whether*-islands had they been asked to judge those sentences at the beginning of the experiment as well, but this does not change the overall pattern of how the *whether*-islands are rated with respect to the other sentences (possibly because those sentences have satiated an equal amount).

6.4.2 Basic effects

In this section I present and then discuss the ERP effects before consideration of the cognitive measures is included in the analysis.

6.4.2.1 Results

In the following sections I present the results of the ERP analysis for each sentence position before summarizing the findings in section 6.4.2.1.9. Table 6-1 is repeated below as Table 6-4, showing the sentence positions discussed below.

Table 6-4: Critical comparisons within the stimulus sentences, indicating both numbering and labels relative to the gap position in both the matrix and embedded clauses

		<i>pre-gap position</i>	<i>gap position</i>	<i>post-gap position</i>			
Matrix clause:							
Position:	1	2	3	4			
	Who	had	_ openly/ the sailor	assumed/ inquired			
Embedded clause:							
Position:	5	6	7	8	9	10/11	12
	that/ whether	the captain	befriended	the sailor/ _ openly	before	the final/ mutiny	hearing?

6.4.2.1.1 Position 2 (matrix pre-gap position: *had*)

Recall from section 6.2.5 that we had predicted P600 effects at the pre-gap position of the embedded clause. Additionally, in the 600-900 msec time window of the pre-gap position in the matrix clause (*had*, position 2), the omnibus ANOVA revealed a main effect of GAP ($F(1,31) = 15.42, p < 0.001$), with the MATRIX GAP condition being more positive (average 2.56 μV) than the EMBEDDED GAP condition (average 1.96 μV). There was no statistical interaction with electrode site. That is, there was a broad positivity across the scalp to *had* when it preceded the gap position in the matrix clause (see Figures 6-3 and 6-4).

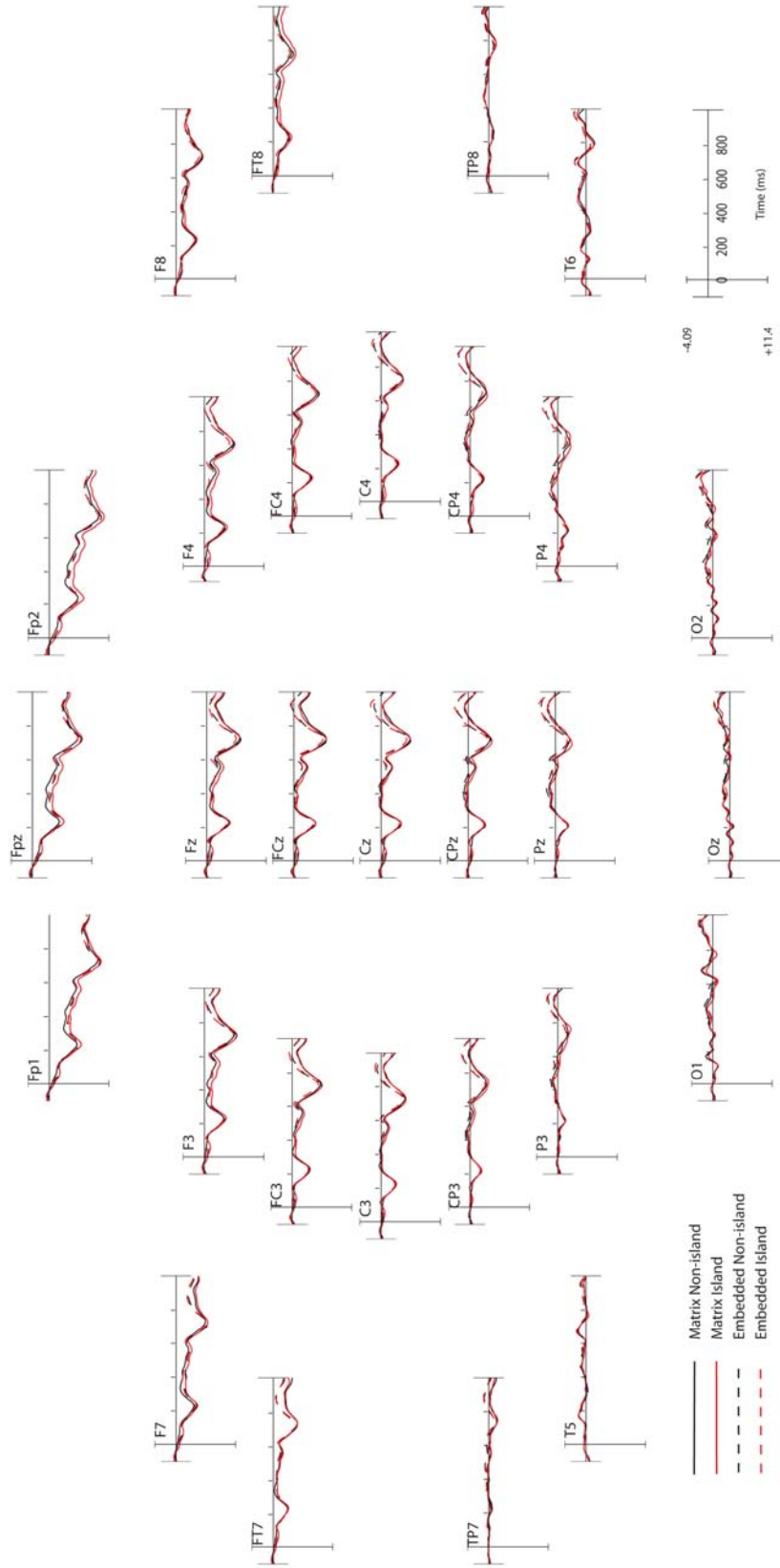


Figure 6-3: Position 2, pre-gap matrix clause (*had*) whole head ERPs.

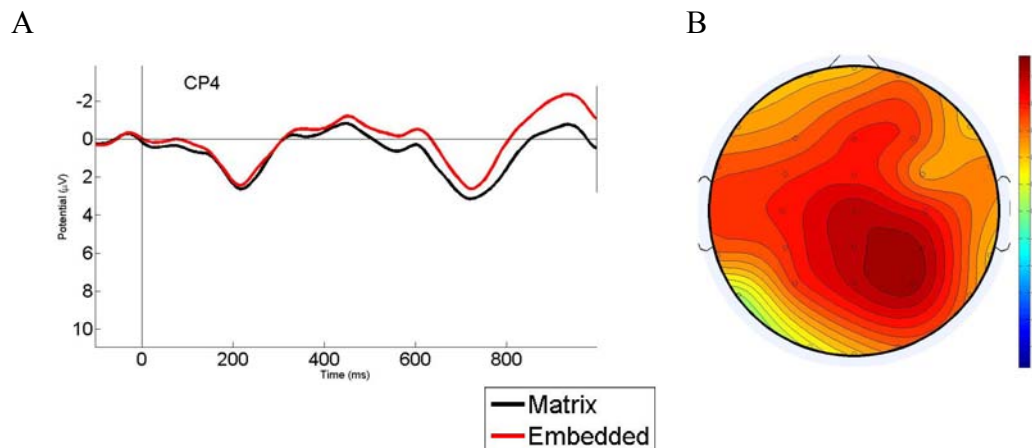


Figure 6-4: Position 2 (*had*) late positivity shown at CP4 (A) and in topographic isovoltage map showing MATRIX (preceding *_openly*) - EMBEDDED (preceding *the sailor*) from 600-900 msec (B).

6.4.2.1.2 Position 3 (matrix gap position: *_openly / the sailor*)

Recall from section 6.2.1 that even though care was taken to control the adverbs and nouns appearing around the gap positions for frequency, since we were comparing words of different grammatical categories (one of which includes a definite determiner), we expected to see evidence of lexical differences between them nonetheless. Crucially, whatever lexical differences are found here should also be found (in the opposite GAP conditions) at position 8, where they likewise occur. Any differences not found in both position 3 and position 8 will not be interpreted as lexical differences. This comparison is presented in section 6.4.2.1.7 below.

In the 300-600 msec time window of the matrix clause gap position (*_openly / the sailor*, position 3), the omnibus ANOVA revealed a main effect of GAP ($F(1,31) = 11.87, p = 0.002$), and a GAP x ELECTRODE interaction ($F(28,868) = 10.90, p = 0.003$).

The EMBEDDED GAP conditions (*the sailor*) were generally more negative than the MATRIX GAP conditions (*_ openly*). This is plotted in Figure 6-5. The distributional analysis (Table 6-5) suggested the possibility of both LAN and N400 effects in the 300-600 msec time window. The distributional analysis indicated that this negativity was strongly left lateralized over anterior regions of scalp, consistent with a Left Anterior Negativity (LAN), but also mildly right-lateralized over posterior regions, and therefore consistent with an N400.

Table 6-5: Position 3 (*_ openly / the sailor*) 300-600 msec window

Analysis:		<i>F</i>	<i>p</i>	
Omnibus:	GAP	$F(1,31) = 11.87$	$p = 0.002$	**
	GAP x ELECTRODE	$F(28,868) = 10.90$	$p = 0.003$	**
Midline:	GAP	$F(1,31) = 13.00$	$p = 0.001$	**
	GAP x ANTERIORITY	$F(6,186) = 4.28$	$p = 0.016$	*
Medial:	GAP	$F(1,31) = 9.58$	$p < 0.001$	***
	GAP x ANTERIORITY	$F(6,186) = 3.61$	$p = 0.043$	*
	GAP x ANTERIORITY x HEMISPHERE	$F(6,186) = 5.37$	$p < 0.001$	***
Lateral:	GAP	$F(1,31) = 8.20$	$p = 0.007$	**
	GAP x HEMISPHERE	$F(1,31) = 5.55$	$p = 0.025$	**
	GAP x ANTERIORITY x HEMISPHERE	$F(6,186) = 11.08$	$p < 0.001$	***

The topographic isovoltage map (Figure 6-6) further supported the possibility of both LAN and N400 effects in the 300-600 msec time window. The post-hoc quadrant analysis (Table 6-6) confirmed a GAP x ANTERIORITY x HEMISPHERE interaction ($F(1,31) = 20.59, p < 0.001$), lending further support to this conclusion. The difference between *the sailor* and *openly* was largest over the left anterior quadrant ($0.78 \mu\text{V}, t(627.13) = -3.56, p < 0.001$) and smallest over the right anterior quadrant ($0.26 \mu\text{V}, t(628.66) = -1.14, p = 0.256$), consistent with a LAN. The posterior quadrants were less differentiated, with only a slightly larger difference over

the right posterior quadrant ($0.68 \mu\text{V}$, $t(636.69) = -3.39$, $p = 0.001$) than over the left posterior quadrant ($0.54 \mu\text{V}$, $t(632.52) = -4.46$, $p < 0.001$). The post-hoc center analysis (Table 6-6) confirmed an effect of GAP independent of the lateral electrodes ($F(1,31) = 14.64$, $p < 0.001$), as well as an interaction between GAP, ANTERIORITY and LATERALITY ($F(8,248) = 14.55$, $p < 0.001$). This negativity was largest over central scalp sites (Figure 6-6 C). For both the LAN and N400 effects, *the sailor* elicited a more negative response than *openly*. As will be discussed in section 6.4.2.2.2, the lexical LAN is attributable to the presence of the determiner, and the lexical N400 is attributable to differences in word categories.

Table 6-6: Position 3 post-hoc analyses (_ *openly* /*the sailor*) 300-600 msec window

Analysis:		<i>F</i>	<i>p</i>	
Quadrant:	GAP	$F(1,31) = 9.14$	$p = 0.005$	**
	GAP x ANTERIORITY x HEMISPHERE	$F(1,31) = 20.06$	$p < 0.001$	***
Center:	GAP	$F(1,31) = 14.64$	$p < 0.001$	***
	GAP x ANTERIORITY x LATERALITY	$F(8,248) = 14.55$	$p < 0.001$	***

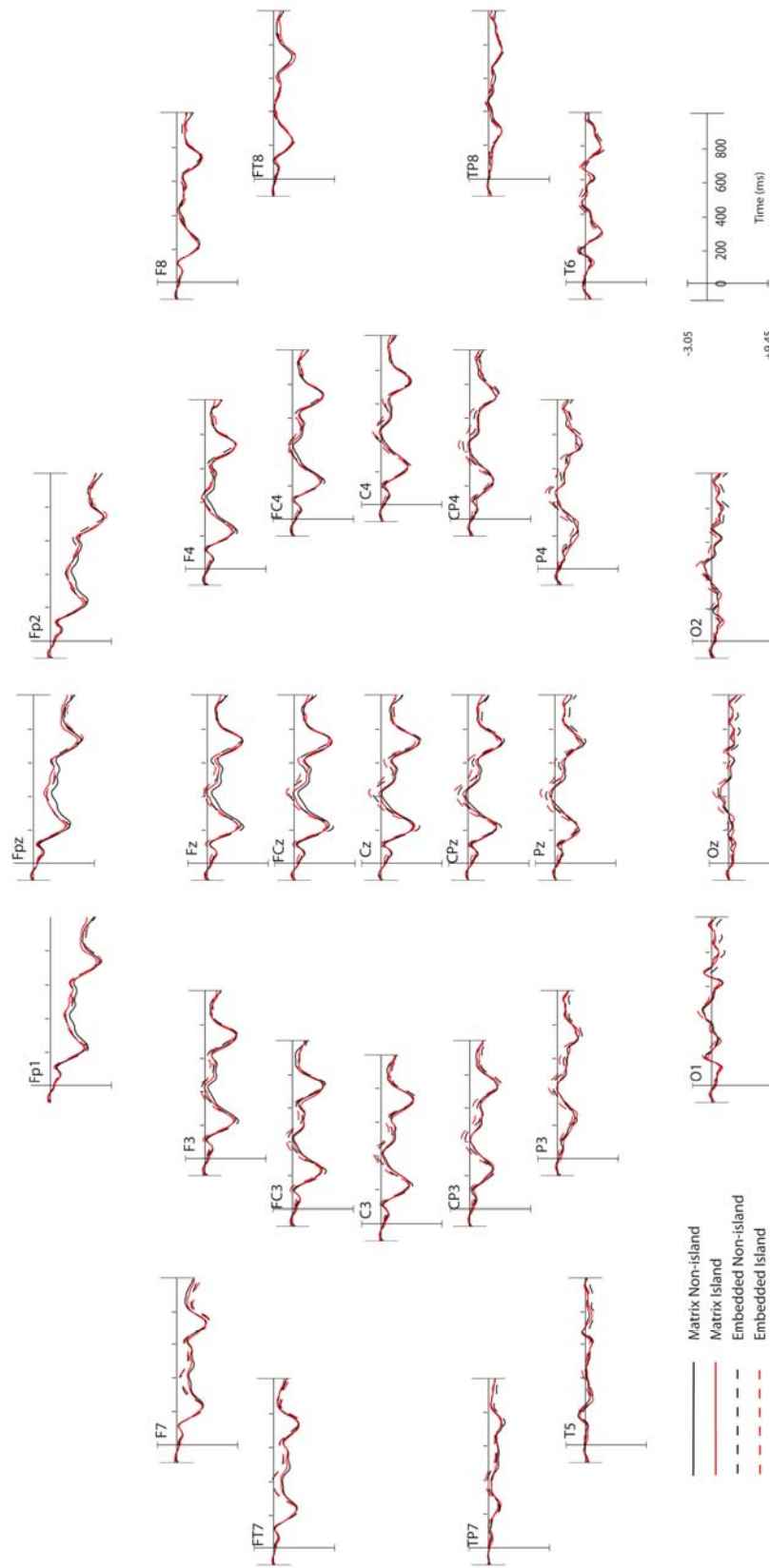


Figure 6-5: Position 3, gap position matrix clause (_ openly /the sailor) whole head ERPs.

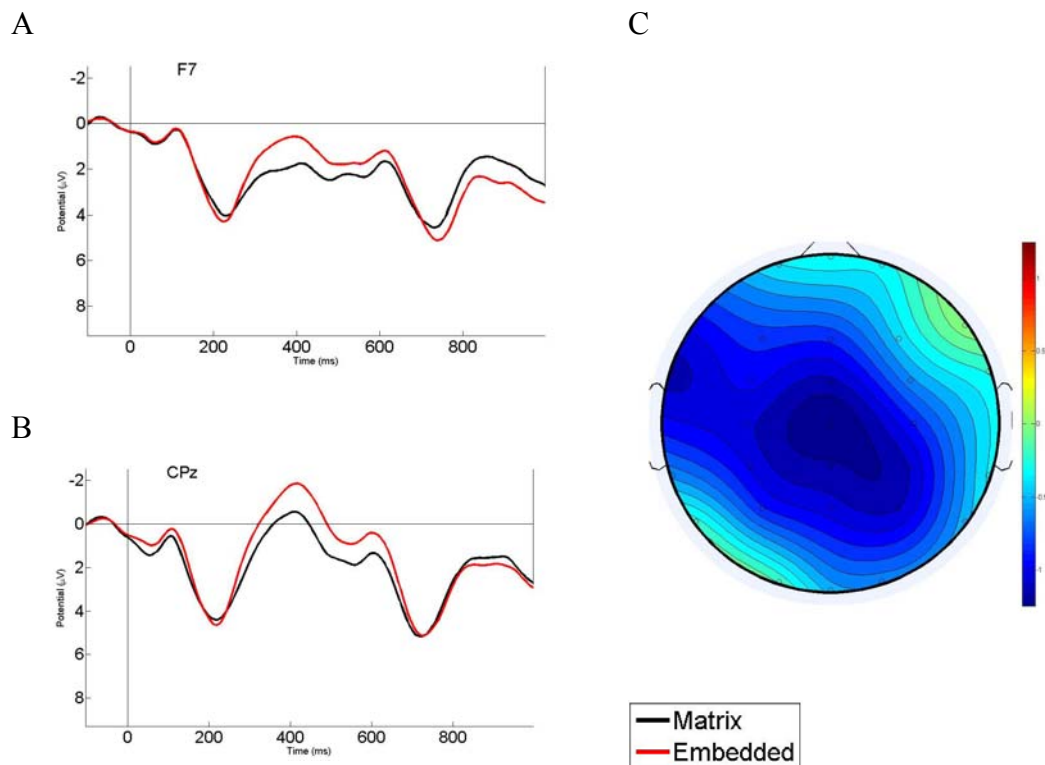


Figure 6-6: Position 3 (*_ openly /the sailor*) negativities shown at F7 (A) and CPz (B) with topographic isovoltage map showing EMBEDDED (*the sailor*) - MATRIX (*_ openly*) from 300-600 msec (C).

6.4.2.1.3 Position 4 (matrix post-gap position: *assumed / inquired*)

Recall now that we had predicted LAN effects at any post-gap positions when a filler and gap have successfully been associated (section 6.2.2). Accordingly, in the 300-600 msec time window of the matrix clause post-gap position (*assumed/inquired*, position 4), the omnibus ANOVA revealed a main effect of GAP ($F(1,31) = 26.20, p < 0.001$), and a GAP x ELECTRODE interaction ($F(28,868) = 7.76, p < 0.001$). The distributional analysis revealed that the response to the MATRIX GAP conditions was

more negative than the response to the EMBEDDED GAP conditions at anterior electrodes (all three distributional analyses in Table 6-7) and over the left hemisphere (medial and lateral analyses, Table 6-7).

Table 6-7: Position 4 (assumed / inquired) 300-600 msec window

Analysis:		<i>F</i>	<i>p</i>	
Omnibus:	GAP	$F(1,31) = 26.20$	$p < 0.001$	***
	GAP x ELECTRODE	$F(28,868) = 7.76$	$p < 0.001$	***
Midline:	GAP	$F(1,31) = 27.44$	$p < 0.001$	***
	GAP x ANTERIORITY	$F(6,186) = 6.27$	$p = 0.001$	**
Medial:	GAP	$F(1,31) = 22.6$	$p < 0.001$	***
	GAP x ANTERIOR	$F(6,186) = 4.94$	$p = 0.011$	*
	GAP x HEMISPHERE	$F(1,31) = 4.88$	$p = 0.035$	*
	GAP x ANTERIORITY x HEMISPHERE	$F(6,186) = 4.52$	$p = 0.002$	**
Lateral:	GAP	$F(1,31) = 26.34$	$p < 0.001$	***
	GAP x HEMISPHERE	$F(1,31) = 30.18$	$p < 0.001$	***
	GAP x ANTERIORITY x HEMISPHERE	$F(3,93) = 23.84$	$p < 0.001$	***

The topographic isovoltage map (Figure 6-8 C) again suggests the possibility of both LAN and N400 effects in the 300-600 time window. This position was thus submitted to the post-hoc distributional analyses (Table 6-8). The post-hoc quadrant analysis revealed a GAP x ANTERIORITY x HEMISPHERE interaction ($F(1,31) = 40.53$, $p < 0.001$), indicating that the effect was largest over the left anterior quadrant (1.39 μ V difference; Figures 6-7 & 6-8), consistent with a LAN effect. The post-hoc center analysis confirmed an effect of GAP independent of the lateral electrodes ($F(1,31) = 29.07$, $p < 0.001$), as well as an interaction between GAP, ANTERIORITY and LATERALITY ($F(8,248) = 4.28$, $p < 0.001$). This negativity was largest over central scalp sites (Figure 6-8 C). Note that while a LAN response was predicted for this position (section 6.2.2), an N400 was not. This N400 effect cannot be interpreted as a

lexical difference since it is elicited by both *assumed* and *inquired*. The N400 is elicited following the matrix gap (_ *openly*) and will be discussed in terms of how these verbs are less predictable following *openly* than following *the sailor* (section 6.4.2.2.7).

Table 6-8: Position 4 post-hoc (*assumed* / *inquired*) 300-600 msec window

Analysis:		<i>F</i>	<i>p</i>	
Quadrant:	GAP	$F(1,31) = 24.32$	$p < 0.001$	***
	GAP x ANTERIORITY x HEMISPHERE	$F(1,31) = 40.53$	$p < 0.001$	***
Center:	GAP	$F(1,31) = 29.07$	$p < 0.001$	***
	GAP x ANTERIORITY x LATERALITY	$F(8,248) = 4.28$	$p < 0.001$	***

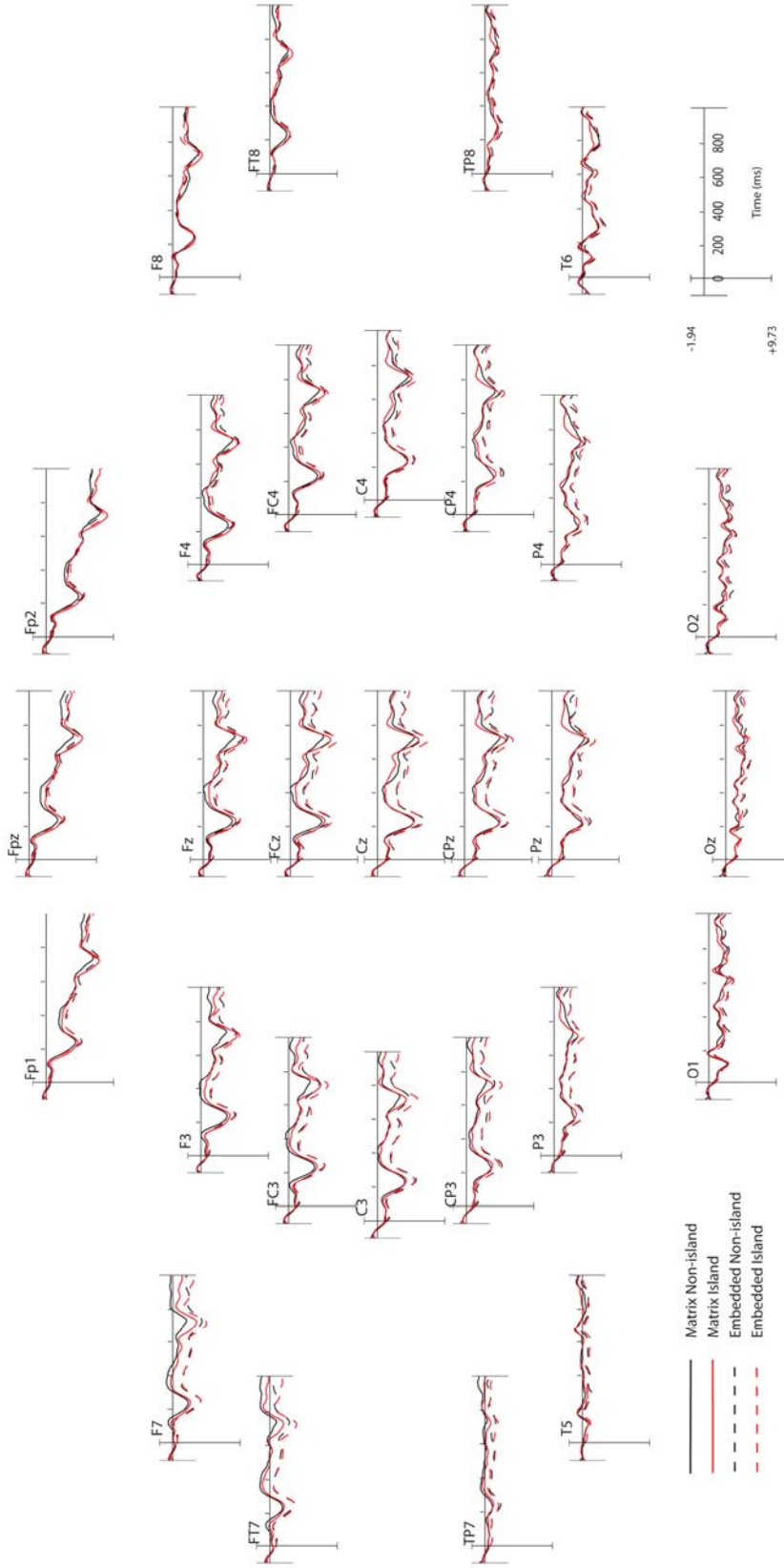


Figure 6-7: Position 4, post-gap matrix clause (*assumed / inquired*) whole head ERPs.

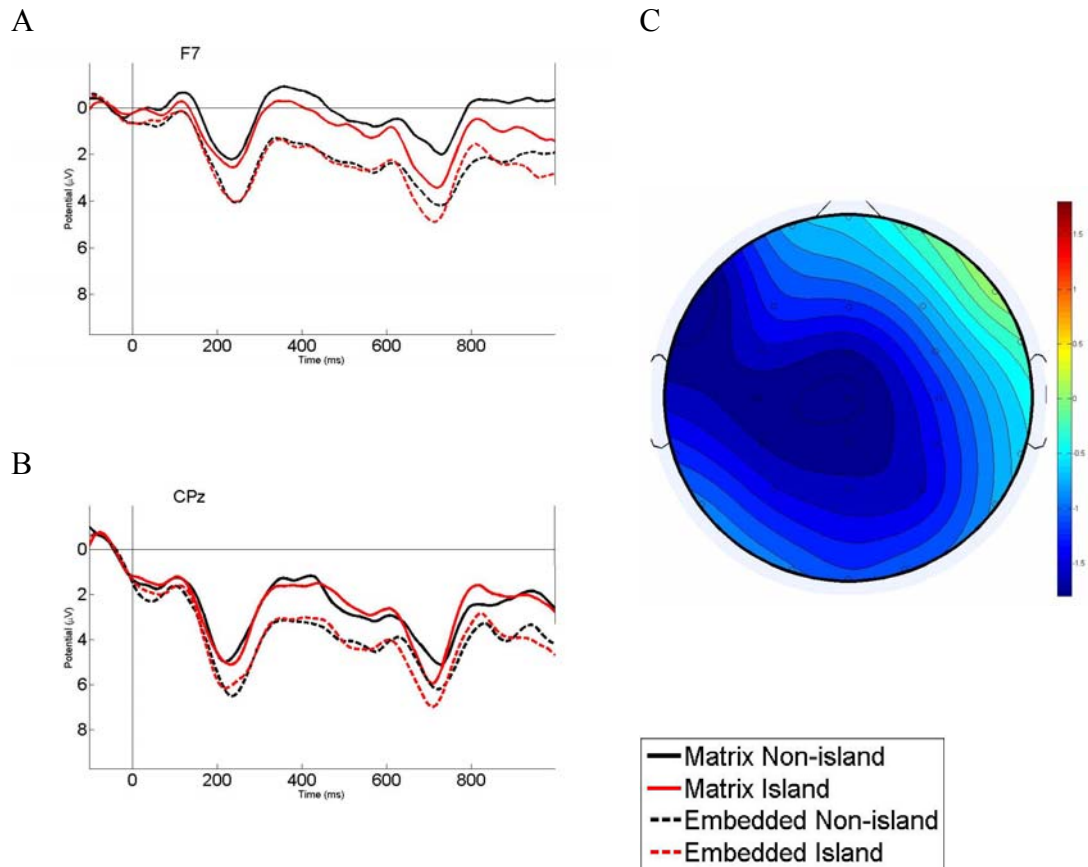


Figure 6-8: Position 4 (assumed/inquired) negativities shown at F7 (A) and CPz (B) and topographic isovoltage map showing MATRIX (after *_openly*) - EMBEDDED (after *the sailor*) from 300-600 msec (C).

6.4.2.1.4 Position 5 (clause boundary: *that/whether*)

Recall that we had predicted N400 effects at the clause boundary (section 6.2.4). However, the omnibus ANOVA revealed no significant effects in the 300-600 msec time window. Inspection of the waveforms (Figure 6-9, 6-10) suggested the possibility of earlier differences, such as a P200 response. Based on these inspections, three post-hoc windows were examined further. A 150-250 msec window was used to

capture a possible P200 response. Following this positivity, visual inspection indicated a short-duration negativity, possibly an N350 (Neville et al. 1992, Hauk & Pulvermüller 2004, and Ueno and Kluender 2009). A window from 250-350 msec was used for this since the effect appeared earlier than 350 msec. Finally, 400-600 msec was measured to examine the remainder of the standard N400 epoch (300-600 msec) to see if any effects could be observed when these early responses were excluded from analysis.

In the 150-250 msec window (surrounding a typical P200 latency), there was a main effect of STRUCTURE ($F(1,31) = 6.24, p = 0.018$) with the ISLAND (*whether*) condition more positive (2.68 μV) than the NON-ISLAND (*that*) condition (2.07 μV). This pattern reversed in the 250-350 msec window, with the ANOVA indicating a main effect of STRUCTURE ($F(1,31) = 7.23, p = 0.014$) with the ISLAND (*whether*) condition more negative (0.63 μV) than the NON-ISLAND (*that*) condition (1.12 μV). Finally, the pattern reversed again in the 400-600 msec window (the remainder of the N400 window used elsewhere), with the ANOVA indicating a main effect of STRUCTURE ($F(1,31) = 7.15, p = 0.011$) with the ISLAND (*whether*) condition more positive (1.06 μV) than the NON-ISLAND (*that*) condition (0.57 μV). None of these effects had a statistically significant interaction with ELECTRODE or GAP. The progression of differences is shown in the topographic isovoltage maps in Figure 6-10.

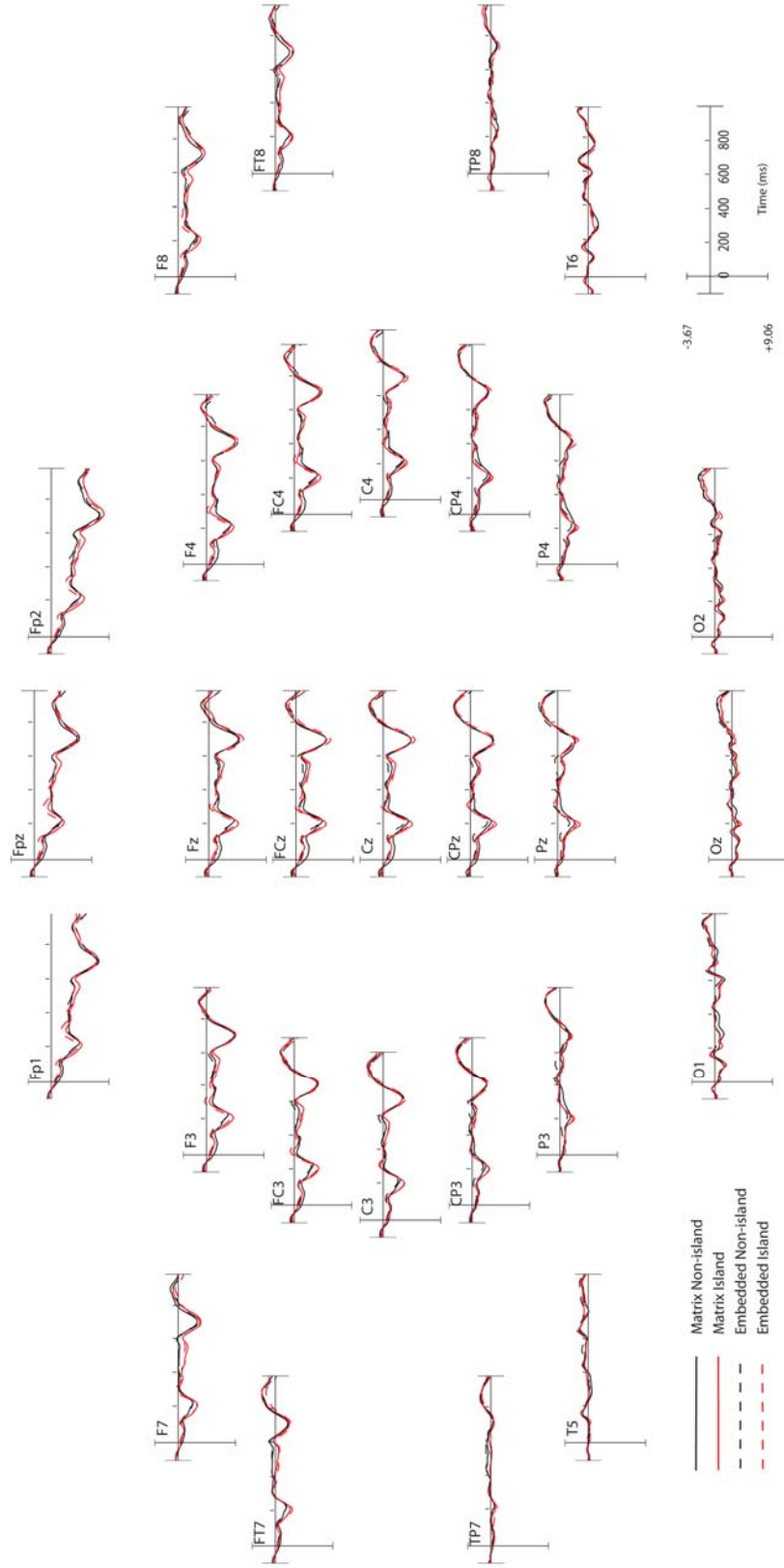


Figure 6-9: Position 5, clause boundary (*that / whether*) whole head ERPs.

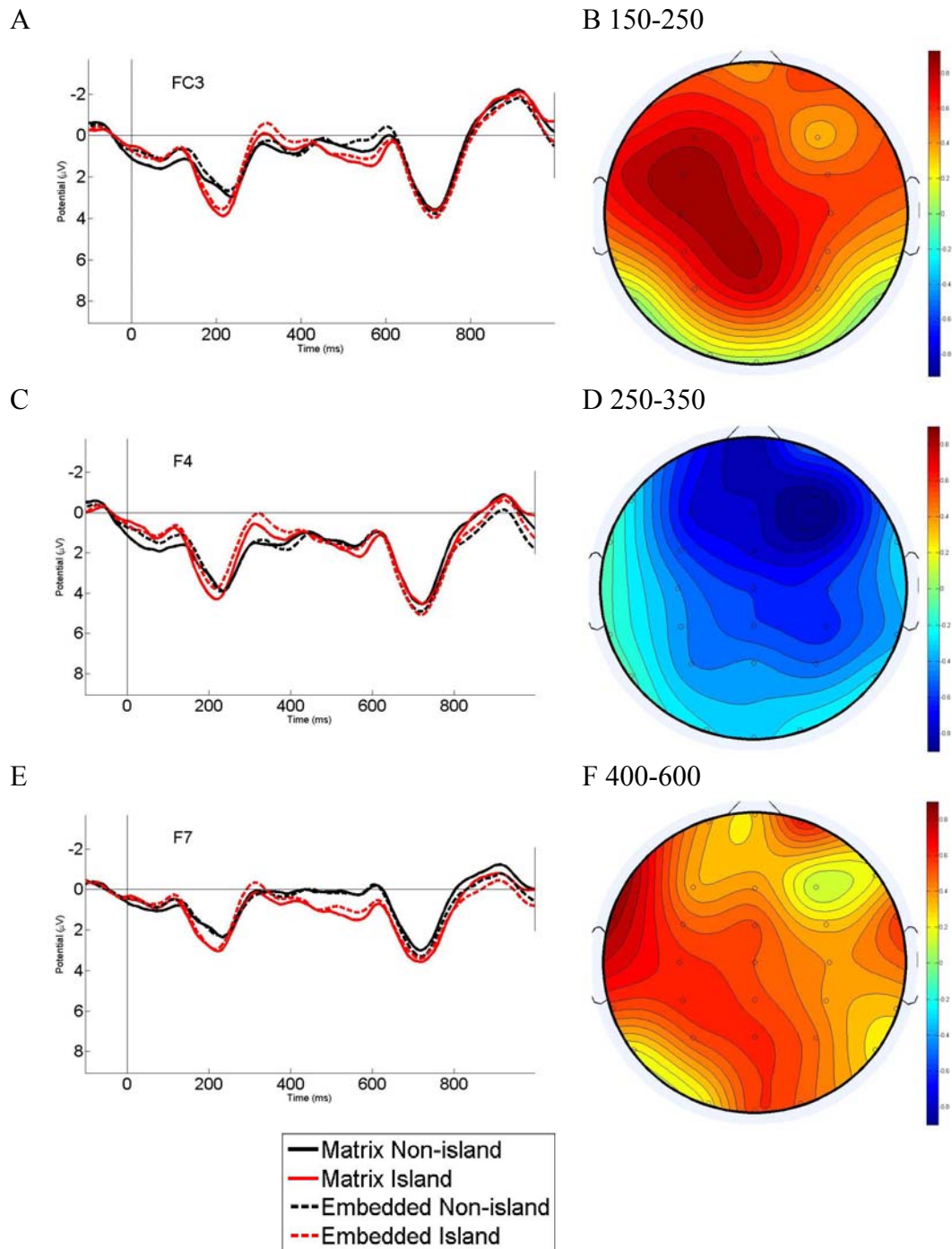


Figure 6-10: Position 5 (*that/whether*). Select electrodes shown with topographic isovoltage maps of ISLAND (*whether*) - NON-ISLAND (*that*) in time windows labeled above.

6.4.2.1.5 Position 7 (embedded pre-gap position: *befriended*)

Recall from section 6.2.5 that we had predicted P600 effects at the embedded clause pre-gap position. In the 600-900 msec latency window of the embedded clause pre-gap position (*befriended*, position 7), the omnibus ANOVA revealed a main effect of GAP ($F(1,31) = 5.19, p = 0.03$): the EMBEDDED GAP condition was more positive (2.91 μV) than the MATRIX GAP condition (2.35 μV). There was no statistical interaction with electrode site. Thus there was a broad positivity in response to *befriended* when it preceded the gap position (Figures 6-11, 6-12).

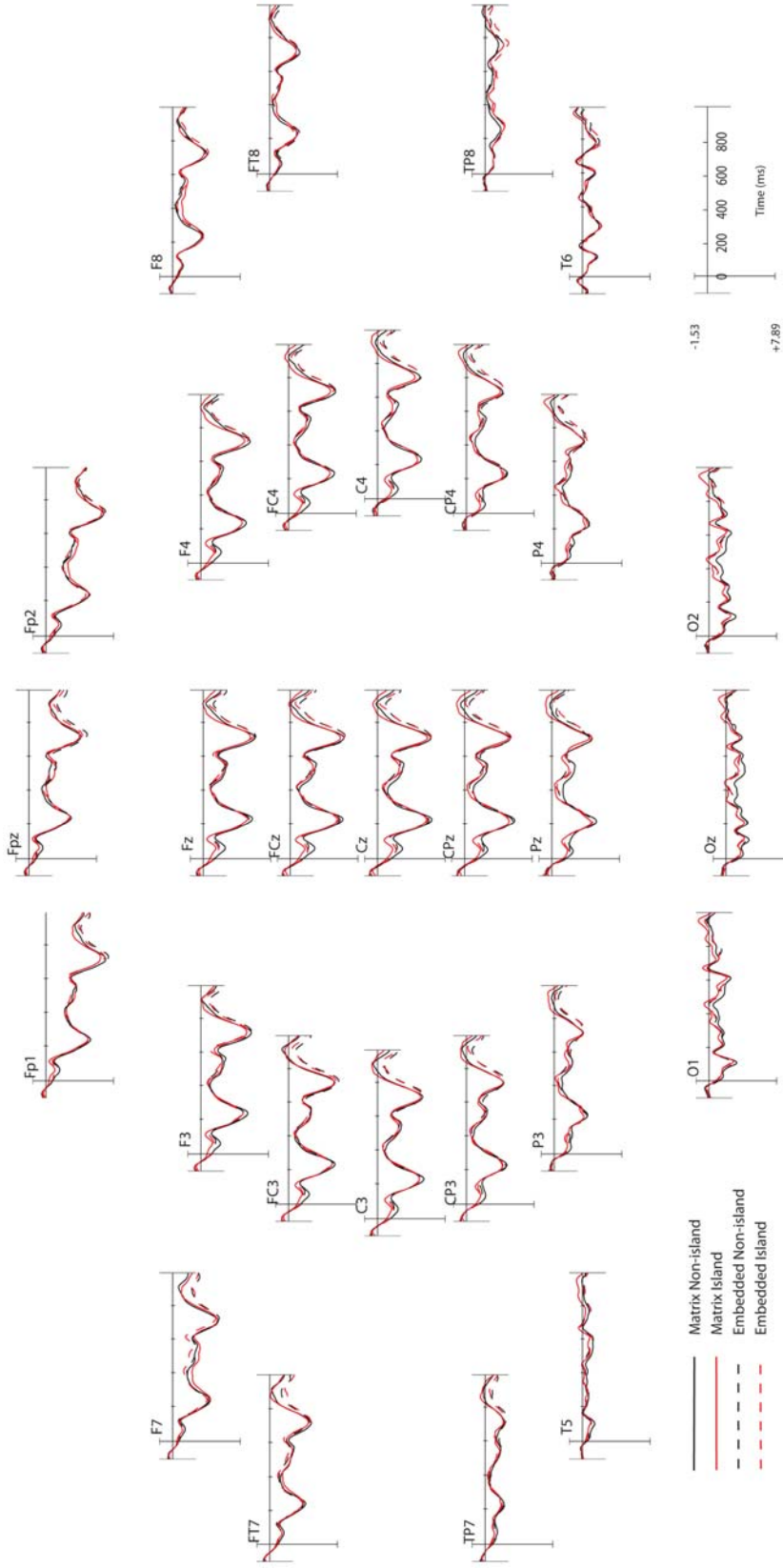


Figure 6-11: Position 7, pre-gap embedded clause (*befriended*) whole head ERPs.

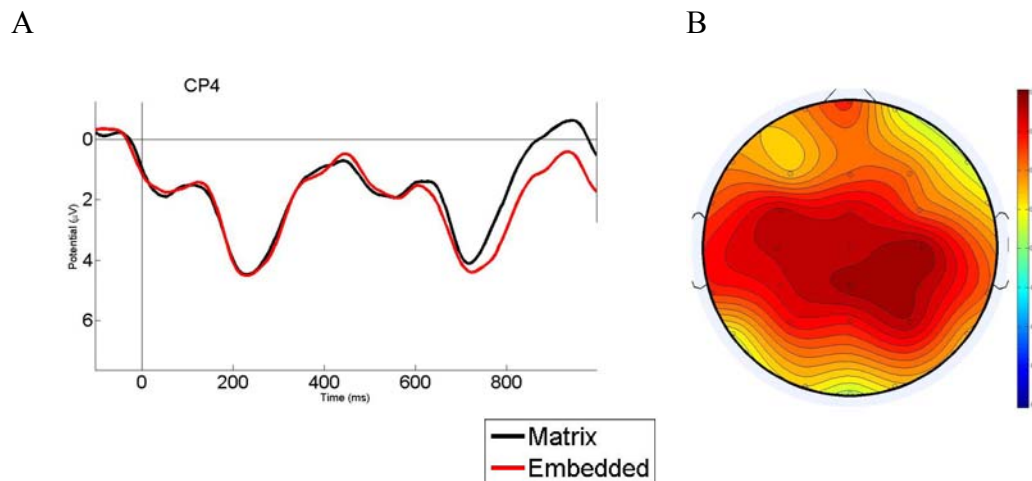
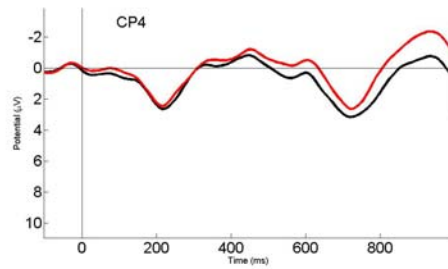


Figure 6-12: Position 7 (*befriended*) late positivity shown at CP4 (A) and topographic isovoltage map showing EMBEDDED (before *_openly*) – MATRIX (before *the sailor*) from 600-900 msec (B).

This broad positivity for position 7 (embedded pre-gap position) echoes the findings for position 2 (matrix pre-gap position). Figure 6-13 presents these findings side-by-side. In each case the broad positivity is elicited by the condition that is immediately followed by a gap (MATRIX for the matrix clause, position 2; EMBEDDED for the embedded clause, position 7). While this is the pattern that was predicted for the embedded gap position based on prior studies (Kaan et al. 2000, Phillips et al. 2005, Gouvea et al. 2010), the interpretation of these data is problematic for those studies. This is addressed in the discussion section 6.4.2.2.1.

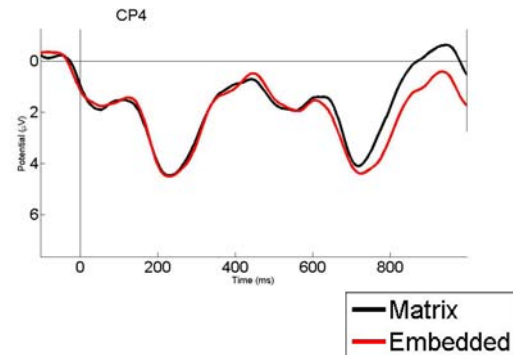
Position 2 (*had-* matrix clause)

A



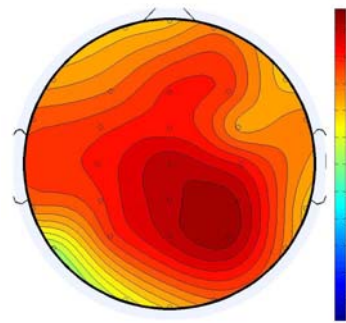
Position 7 (*befriended-* embedded clause)

B



C

MATRIX (preceding *_openly*) –
EMBEDDED (preceding *the sailor*)



D

EMBEDDED (preceding *_openly*) –
MATRIX (preceding *the sailor*)

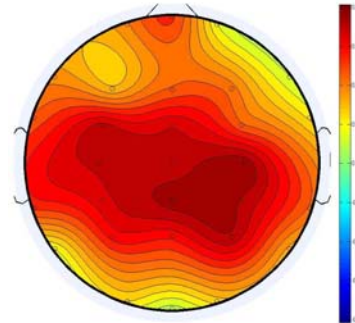


Figure 6-13: Comparison of positions 2 and 7. CP4 (A and B) and topographic isovoltage map showing [position immediately preceding the gap (*_openly*)] – [position immediately preceding *the sailor*] from 600-900 msec (C, D).

6.4.2.1.6 Position 8 (embedded gap position: *the sailor* / *_openly*)

Recall from section 6.2.1 that since we were comparing words of different grammatical categories (one of which included a definite determiner), we expected to find lexical differences between them. This same lexical comparison was made for position 3 (section 6.4.2.1.2), except that while for position 3 *the sailor* was present in the EMBEDDED GAP condition, at position 8 *the sailor* was present in the MATRIX GAP

condition. We thus expected that any differences that were directly due to lexical differences between *the sailor* and *openly* would be present at both position 3 and position 8 (thus the pattern should be ‘flipped’ when looking at the GAP manipulation). Any other differences between positions 3 and 8 therefore cannot be interpreted as lexical differences.

In the 300-600 msec time window of the embedded clause gap position (*the sailor/_openly*, position 8), the omnibus ANOVA again revealed a main effect of GAP ($F(1,31) = 14.60, p < 0.001$), and a GAP x ELECTRODE interaction ($F(28,868) = 3.75, p < 0.001$). The MATRIX GAP conditions (*the sailor*) were generally more negative than the EMBEDDED GAP conditions (*_openly*, Figure 6-14) (N.B. compare to position 3, section 6.4.2.1.2, where the lexical items *the sailor* and *_openly* were associated with the opposite conditions in the GAP manipulation). The distributional analysis (Table 6-9) again suggested the possibility of both LAN and N400 effects. The analysis indicated that the negativity was strongly left-lateralized over anterior scalp regions, consistent with a LAN, but also mildly-right lateralized over posterior regions, consistent with an N400. Like position 3, the topographic isovoltage map (Figure 6-15) again supports an analysis of both LAN and N400 effects in this time window.

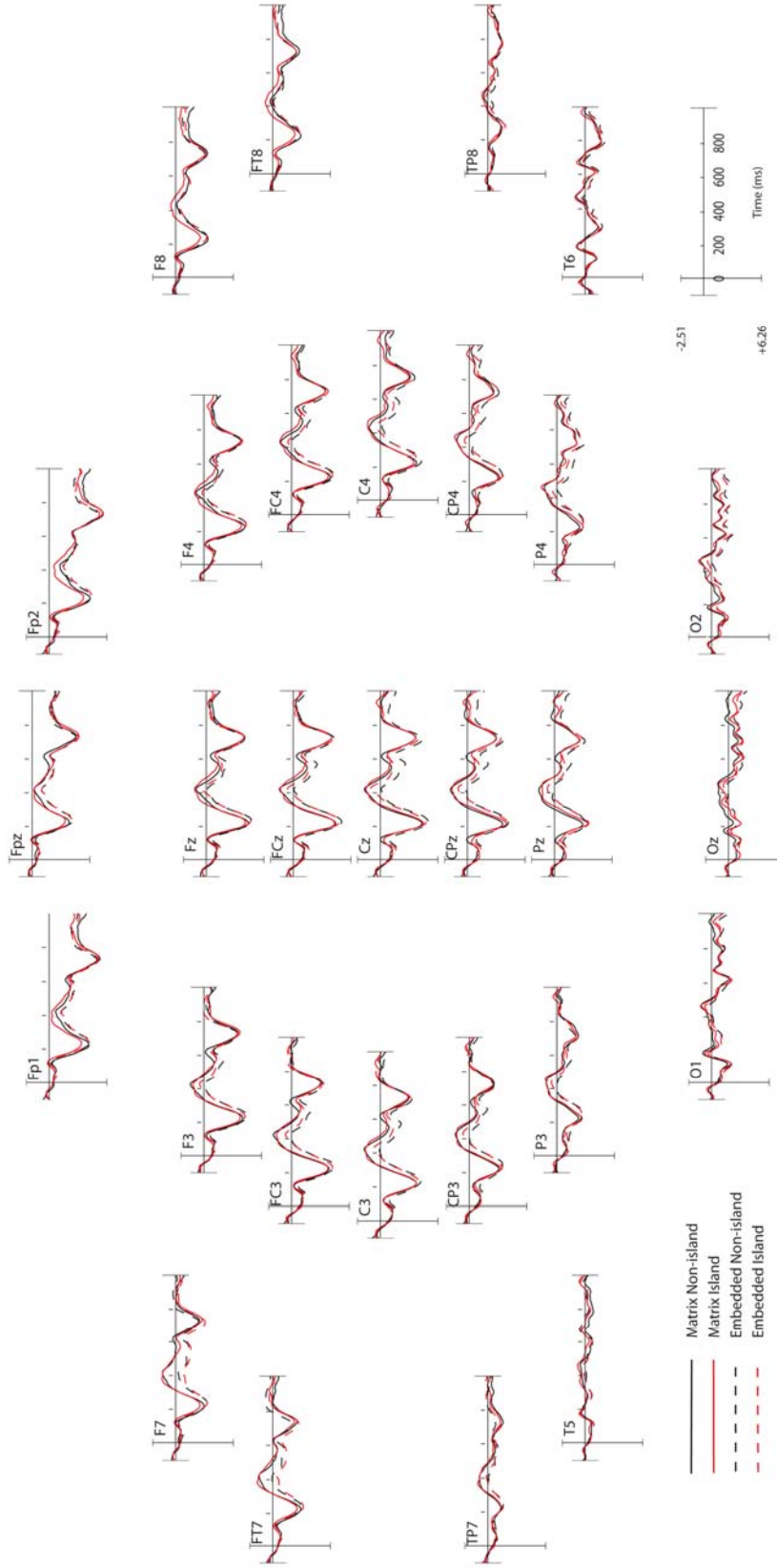


Figure 6-14: Position 8, gap position embedded clause (the sailor /_openly) whole head ERPs.

Table 6-9: Position 8 (*the sailor* / *_openly*) 300-600 msec window

Analysis:		<i>F</i>	<i>p</i>	
Omnibus:	GAP	$F(1,31) = 14.60$	$p < 0.001$	***
	GAP x ELECTRODE	$F(28,868) = 3.75$	$p < 0.001$	***
Midline:	GAP	$F(1,31) = 12.39$	$p = 0.009$	**
Medial:	GAP	$F(1,31) = 12.96$	$p = 0.001$	**
	GAP x ANTERIORITY x HEMISPHERE	$F(6,186) = 5.23$	$p < 0.001$	***
Lateral:	GAP	$F(1,31) = 18.62$	$p < 0.001$	***
	GAP x ANTERIORITY	$F(3,93) = 3.62$	$p = 0.036$	*
	GAP x HEMISPHERE	$F(3,93) = 11.83$	$p < 0.001$	***

The post-hoc quadrant analysis confirmed a GAP x ANTERIORITY x HEMISPHERE interaction ($F(1,31) = 20.59, p < 0.001$), lending further support to separate LAN and N400 responses. The difference between *the sailor* and *_openly* was largest over the left anterior quadrant (1.20 μV) and smallest over the right anterior quadrant (0.64 μV). *The sailor* elicited a greater negativity in the left anterior quadrant than the right anterior quadrant ($t(637.98) = -3.72, p < 0.001$), consistent with a LAN. The posterior regions were again less differentiated (compare position 3, section 6.4.2.1.2), with only a slightly larger difference over the right posterior quadrant (0.85 μV) than the left posterior quadrant (0.71 μV). The post-hoc center analysis confirmed an effect of GAP independent of the lateral electrodes ($F(1,31) = 12.95, p < 0.001$), as well as an interaction between GAP, ANTERIORITY and LATERALITY ($F(8,248) = 2.29, p = 0.048$). In both the LAN and N400 effects, *the sailor* elicited a more negative response than *_openly*.

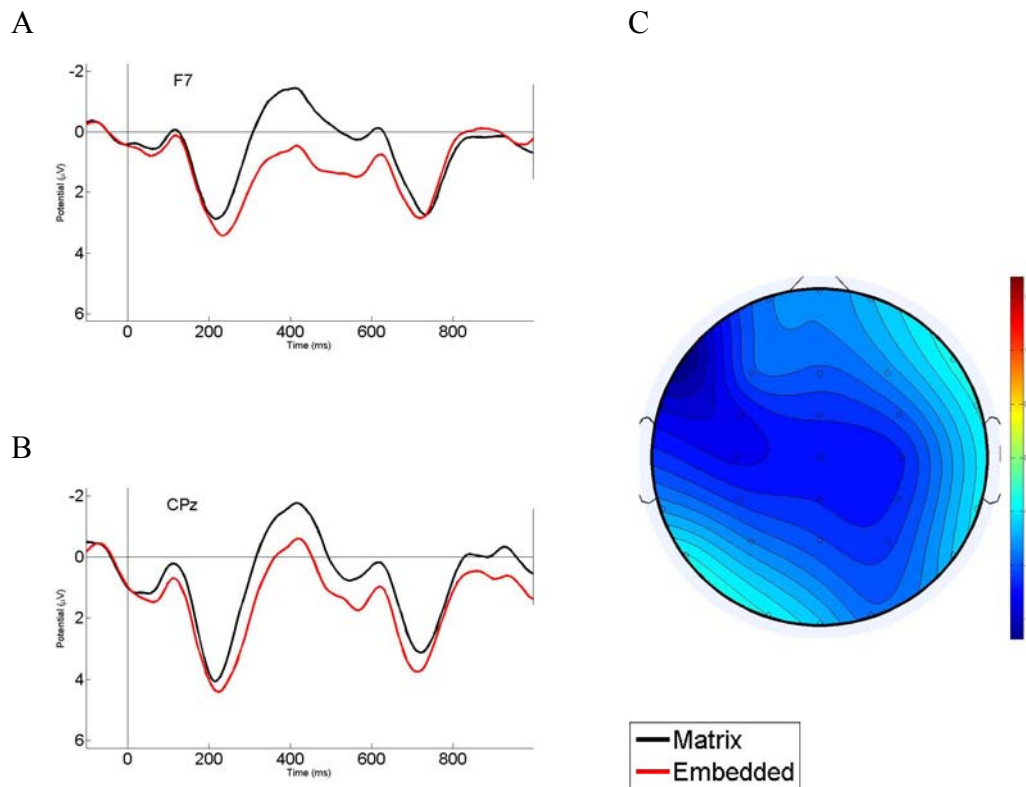


Figure 6-15: Position 8 (*the sailor* / *_ openly*) negativities shown at F7 (A) and CPz (B) and topographic isovoltage map showing showing MATRIX (*the sailor*) – EMBEDDED (*_ openly*) from 300-600 msec (C).

In addition to the effects of GAP (*the sailor* vs. *_ openly*), both the quadrant and center analyses revealed an interaction with STRUCTURE (Table 6-10). In the quadrant analysis, there was an interaction of GAP x STRUCTURE x ANTERIORITY x HEMISPHERE ($F(1,31) = 6.36, p = 0.017$). In the center analysis there was an interaction of GAP x STRUCTURE x ANTERIORITY x LATERALITY ($F(8,248) = 3.06, p = 0.005$). Examination of the means indicated that in addition to the effect of GAP that emerged in the earlier analyses (with *the sailor* eliciting a larger N400 effect than *_ openly*), when comparing *_ openly* to *_ openly* in the ISLAND and NON-ISLAND conditions, there was a greater negativity in the ISLAND condition, largest over Cz

(0.75 μ V) and slightly larger at left medial than at right medial electrodes (Figure 6-16).

Table 6-10: Position 8 post-hoc (*the sailor* / *_openly*) 300-600 msec window

Analysis:		<i>F</i>	<i>p</i>	
Quadrant:	GAP	$F(1,31) = 14.77$	$p = 0.001$	**
	GAP x ANTERIORITY x HEMISPHERE	$F(1,31) = 20.59$	$p < 0.001$	***
	GAP x STRUCTURE x ANTERIORITY x HEMISPHERE	$F(1,31) = 6.36$	$p = 0.017$	*
Center:	GAP	$F(1,31) = 12.95$	$p = 0.001$	**
	GAP x ANTERIORITY x LATERALITY	$F(8,248) = 2.29$	$p = 0.048$	*
	GAP x STRUCTURE x ANTERIORITY x LATERALITY	$F(8,248) = 3.06$	$p = 0.005$	**

In order to determine whether the interaction between GAP and STRUCTURE involved the LAN effect, the N400 effect, or both, the left anterior quadrant was analyzed separately, revealing only a main effect of GAP ($F(1,31) = 24.27$, $p < 0.001$) and no interaction with STRUCTURE.

Post-hoc pairwise comparisons (Table 6-11) revealed that for both the left anterior quadrant and the central analysis region the MATRIX conditions were more negative than the EMBEDDED conditions (*the sailor* more negative than *_openly*, all $p < 0.001$). Only in the center analysis was the EMBEDDED ISLAND more negative than the EMBEDDED NON-ISLAND ($p < 0.001$). There is no effect of STRUCTURE in the LAN region, only the N400 region.

Table 6-11: Position 8 post-hoc (*the sailor* / *_openly*) 300-600 msec window paired comparisons

Region:	Comparisons:	<i>F</i>	<i>p</i>	
Left Anterior Quadrant:	Matrix vs. Embedded Non-islands	<i>t</i> (304) = 5.81	<i>p</i> < 0.001	***
	Matrix vs. Embedded Islands	<i>t</i> (312.82) = 5.15	<i>p</i> < 0.001	***
	Non-island vs. Island Matrix	<i>t</i> (315.02) = -0.04	<i>p</i> = 0.964	
	Non-island vs. Island Embedded	<i>t</i> (307.59) = 0.05	<i>p</i> = 0.964	
Center:	Matrix vs. Embedded Non-islands	<i>t</i> (906.82) = 11.84	<i>p</i> < 0.001	***
	Matrix vs. Embedded Islands	<i>t</i> (825.12) = 7.53	<i>p</i> < 0.001	***
	Non-island vs. Island Matrix	<i>t</i> (908.01) = -0.23	<i>p</i> = 0.821	
	Non-island vs. Island Embedded	<i>t</i> (955) = -3.49	<i>p</i> < 0.001	***

Thus, when comparing *_openly* to *_openly* in the ISLAND and NON-ISLAND conditions, there was a greater negativity in response to the ISLAND condition in the central but not the left anterior analyses. This represents a second N400 effect independent of the lexical differences reported above (Figure 6-16).

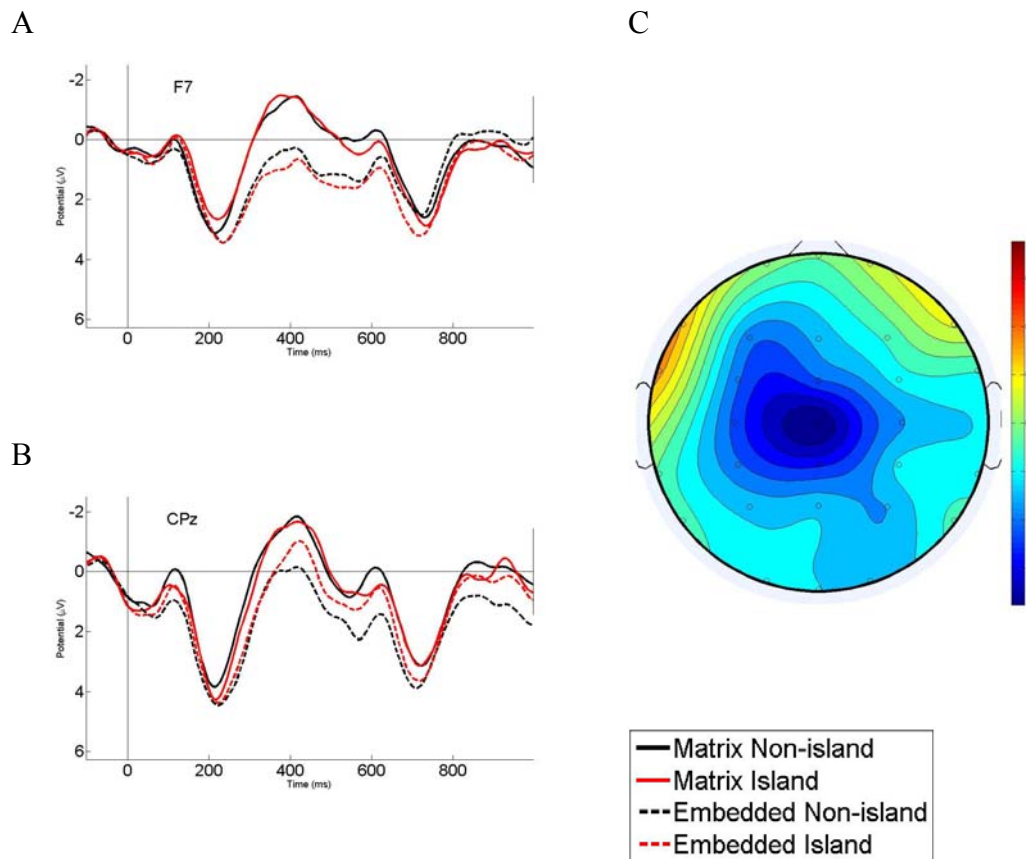


Figure 6-16: Position 8 (*the sailor* / *_ openly*) main effect of GAP shown at F7 (A), interaction of GAP x STRUCTURE shown at CPz (B) and topographic isovoltage map showing EMBEDDED ISLAND (*_ openly*) – EMBEDDED NON-ISLAND (*_ openly*) from 300-600 msec (C).

Note that this ‘additional N400’ (‘additional’ to, and independent from, the N400 caused by lexical word-class differences between *the sailor* and *openly*) is not an artefact of the baselining procedure. The previous word in all conditions is identical (*befriended*). Additionally, Figure 6-17 demonstrates that the interaction pattern persists even in a multi-word average of the embedded clause (and is still statistically significant $p = 0.017$). Starting from *the captain* (the first position in the embedded clause that is identical across conditions) we see that when position 8 (*the sailor* / *_ openly*) is encountered (starting at 1000 msec), the ‘additional N400’ effect is still

clearly visible with *_ openly* in the ISLAND condition (red dashed line) and more negative than *_ openly* in the NON-ISLAND condition (black dashed line) between 1300-1600 msec.

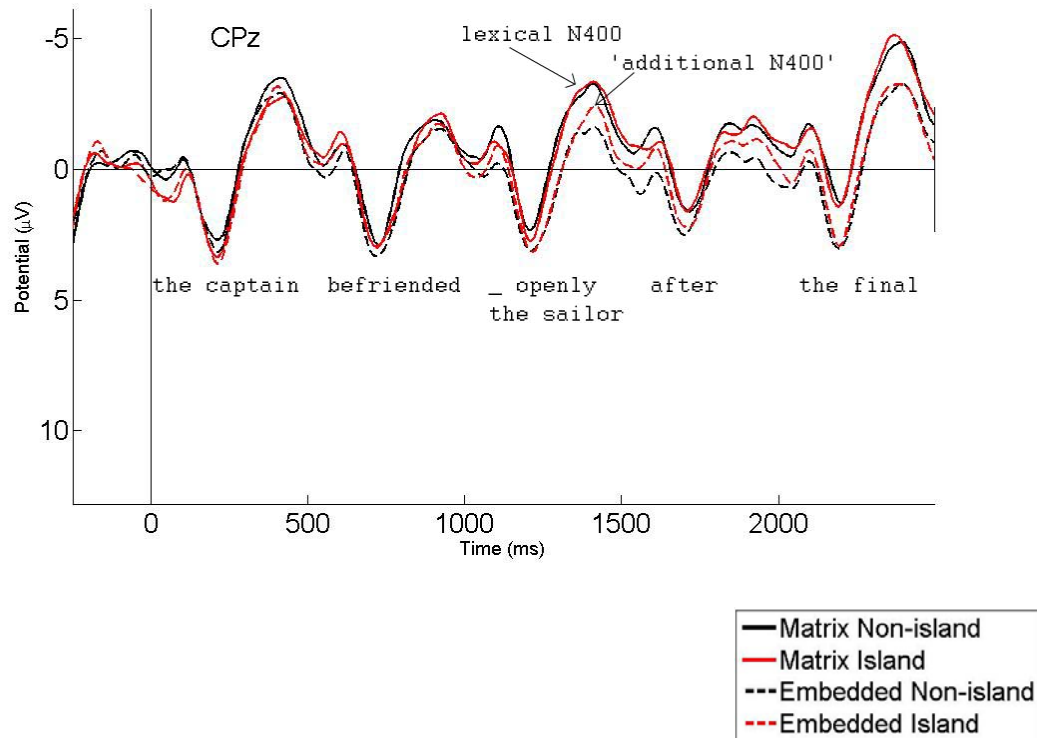
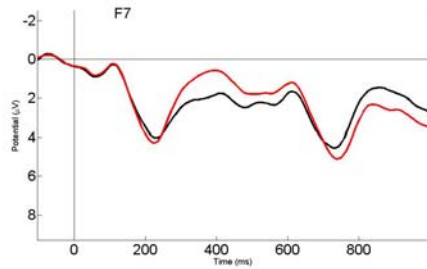


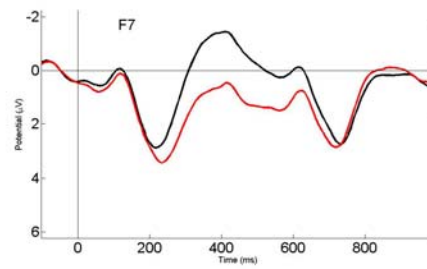
Figure 6-17: Five word average starting at position 6 (*the captain*): Point of interest is interaction at { *_ openly* / *the sailor* }

While the analysis of position 8 revealed an interaction of GAP and STRUCTURE not found at position 3, the lexical differences (i.e., main effects of GAP) between *the sailor* and *_ openly* can still be compared across these two positions. Figure 6-18 shows the main effects of GAP at two electrodes (F7 for the LAN and CPz for the N400). Figure 6-19 shows side-by-side topographic isovoltage maps.

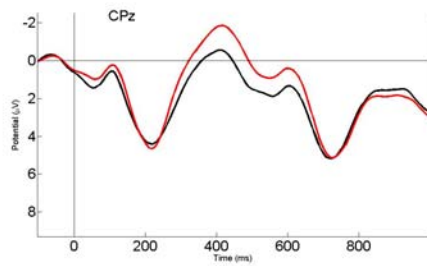
Position 3 (EMBEDDED = *the sailor*)
A



Position 8 (MATRIX = *the sailor*)
B



C



D

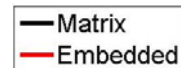
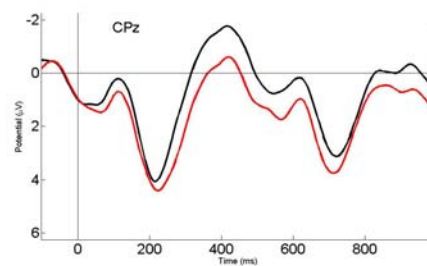
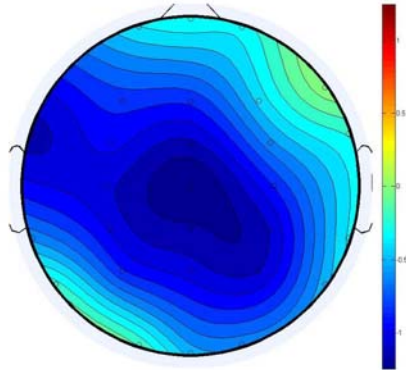


Figure 6-18: Comparison of main effect of GAP (lexical difference of *the sailor* vs. *_openly*) in positions 3 and 8: F7 (A and B) and CPz (C and D)

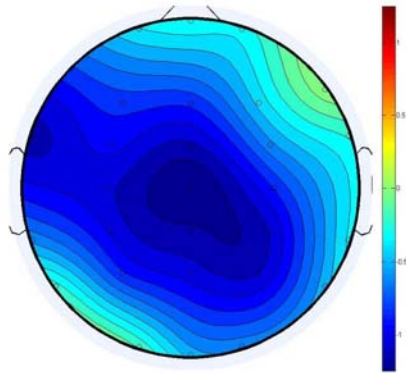
The main effects are more difficult to observe in the topographic isovoltage maps due to the interaction of GAP and STRUCTURE found at position 8. Figure 6-19 first compares the overall effects from position 2 (Figure 6-19 A) with the overall effects from position 8 (Figure 6-19 B). Note that the N400 response in Figure 6-19 B is not as robust as in Figure 6-19 A. When the interaction at position 8 is taken into consideration and compensated for by removing the island conditions (which caused the additional interaction), however, we see that the lexical differences in the NON-ISLAND conditions (Figure 6-19 C) more closely resemble the overall pattern observed at position 3 (Figure 6-19 A), where island effects do not play a role. It is because of

the additional N400 effect in the ISLAND conditions reported above for position 8 that the lexical difference appears washed out (Figure 6-19 D). Figure 6-16 B also shows how the interaction of GAP and STRUCTURE reduces the size of the lexical effect in the ISLAND conditions.

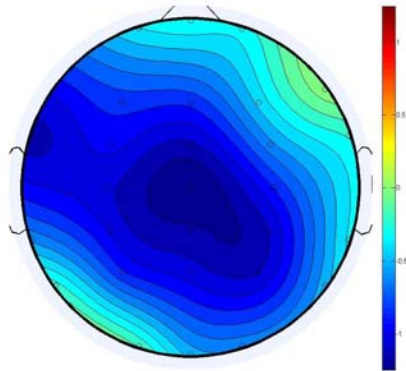
Position 3 *the sailor* - *_openly*
A



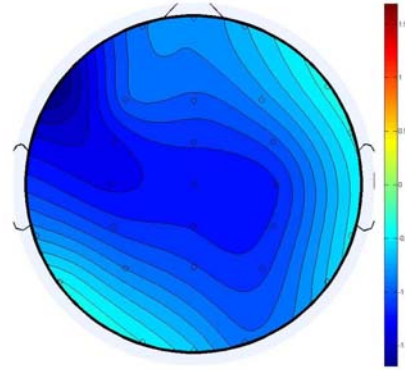
A'



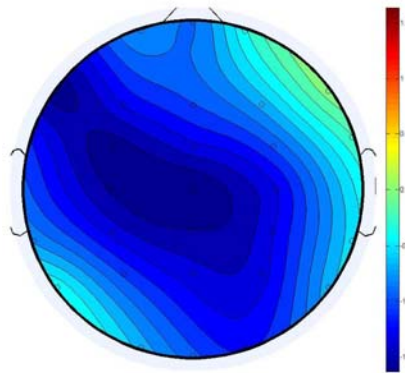
A''



Position 8 *_openly* - *the sailor*
B



C Non-island conditions only



D Island conditions only

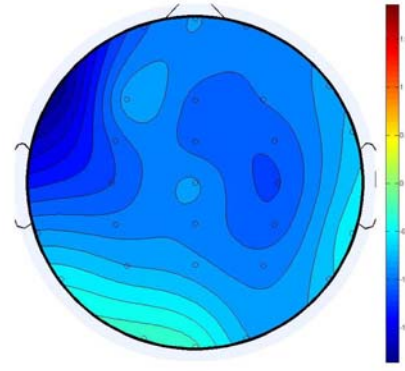


Figure 6-19: Comparison of positions 3 and 8. Topographic isovoltage map showing [the condition including the lexical item *the sailor*] – [the condition including the lexical item *_openly*] from 600-900 msec. A, A' and A'' are all the identical comparison from position 3 and are repeated for ease of comparison with position 8. Position 8 is shown in its entirety (B), in only the NON-ISLAND conditions (C) and in only the ISLAND conditions (D).

To briefly summarize the findings for position 8, the conditions with *the sailor* elicited a LAN and N400 response compared to conditions with *_ openly*, just like the pattern for word 3. These are thus taken to be lexical effects between these words. As previously mentioned for position 3, and as will be discussed in section 6.4.2.2.2, the lexical LAN is attributable to the presence of the determiner, and the lexical N400 is attributable to differences in word categories. However, in addition to the lexical effects, a larger N400 was elicited in the EMBEDDED ISLAND condition than the EMBEDDED NON-ISLAND condition. This can not be due to a lexical effect as the same lexical items were compared (*_ openly* in both cases). This additional N400 is thus interpreted as due to a syntactic manipulation rather than a lexical effect. More specific discussion for what process(es) this response is a reflection of is found in section 6.4.2.2.7.

6.4.2.1.9 Position 9 (embedded post-gap position: *before*)

Recall now that we had predicted LAN effects at any post-gap positions where a filler and gap have successfully been associated (section 6.2.2). Accordingly, in the 300-600 msec time window of the embedded clause post-gap position (*before*, position 9), the omnibus ANOVA again revealed a main effect of GAP ($F(1,31) = 10.28, p = 0.003$), and a GAP x ELECTRODE interaction ($F(28,868) = 7.94, p < 0.001$). The distributional analysis similarly revealed that the EMBEDDED GAP conditions were more negative than the MATRIX GAP conditions at anterior electrodes (all three

distributional analyses in Table 6-12) and over the left hemisphere (medial and lateral analyses, Table 6-12). In the post-hoc quadrant analysis, there was a GAP x ANTERIORITY interaction ($F(1,31) = 6.85, p = 0.014$), as well as a GAP x ANTERIORITY x HEMISPHERE interaction that just missed significance ($F(1,31) = 4.02, p = 0.053$), suggesting that the response to the EMBEDDED GAP conditions was maximal over the left anterior quadrant (1.52 μ V difference; Figures 6-20 and 6-21), and thus again consistent with a LAN effect.

Table 6-12: Position 9 (before) 300-600 msec window

Analysis:		<i>F</i>	<i>p</i>	
Omnibus:	GAP	$F(1,31) = 10.28$	$p = 0.003$	**
	GAP x ELECTRODE	$F(28,868) = 7.94$	$p < 0.001$	***
Midline:	GAP	$F(1,31) = 4.22$	$p = 0.048$	*
	GAP x ANTERIORITY	$F(6,186) = 5.60$	$p = 0.004$	**
Medial:	GAP	$F(1,31) = 8.17$	$p = 0.008$	**
	GAP x ANTERIORITY	$F(1,31) = 6.90$	$p = 0.013$	*
	GAP x HEMI	$F(6,186) = 7.00$	$p = 0.002$	**
	GAP x ANTERIORITY x HEMISPHERE	$F(6,186) = 4.59$	$p = 0.003$	**
Lateral:	GAP	$F(1,31) = 26.32$	$p < 0.001$	***
	GAP x ANT	$F(1,31) = 20.26$	$p < 0.001$	***
	GAP x HEMI	$F(3,93) = 7.25$	$p = 0.004$	**
	GAP x ANTERIORITY x HEMISPHERE	$F(3,93) = 6.87$	$p = 0.001$	**

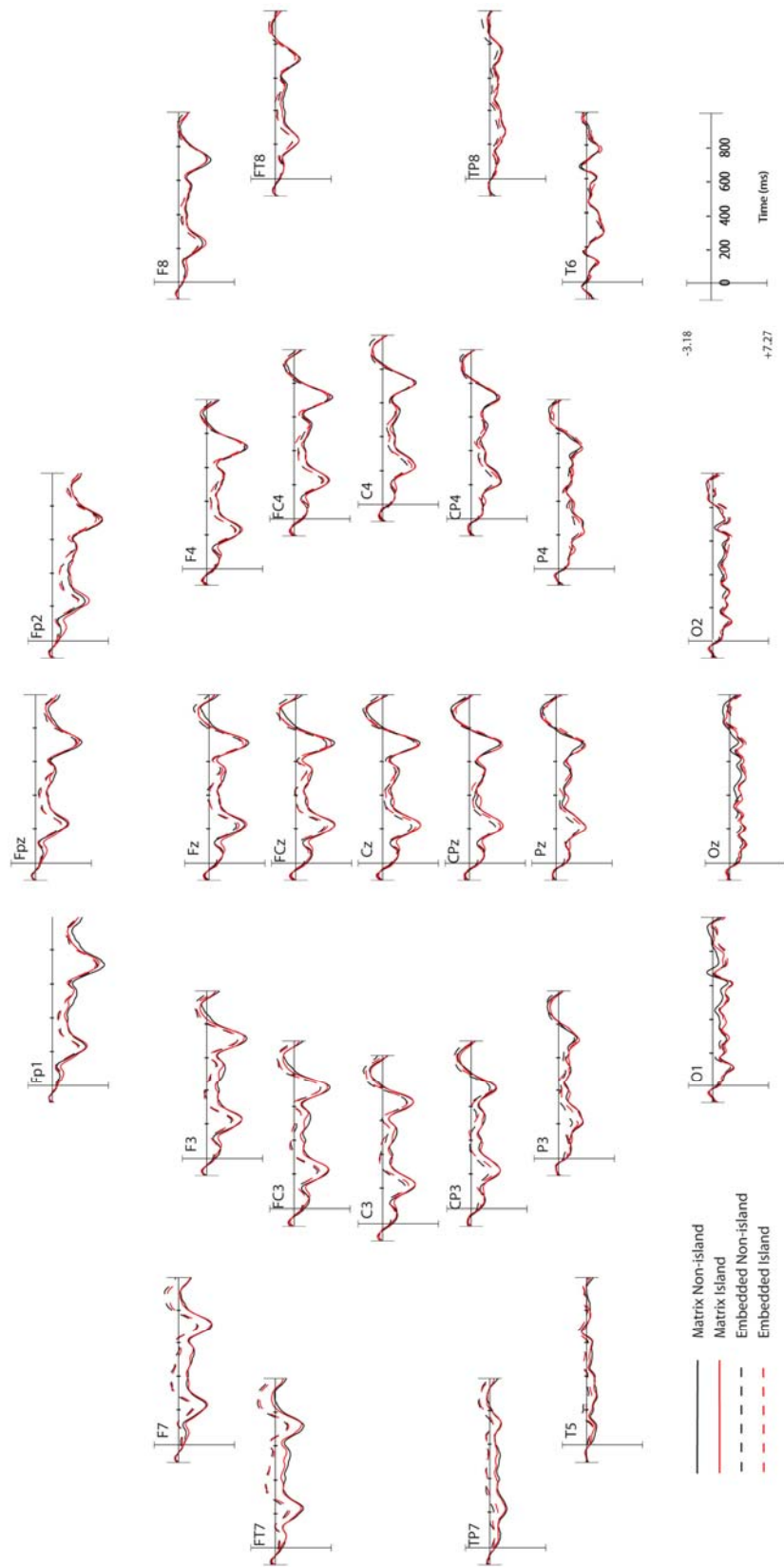


Figure 6-20: Position 9, post-gap position embedded clause (*before*) whole head ERPs.

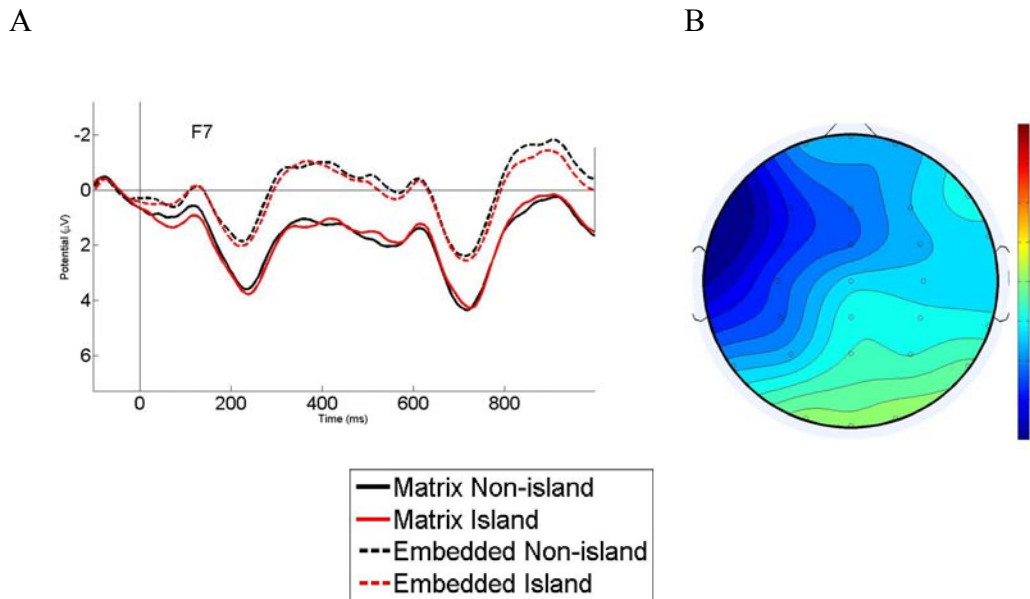


Figure 6-21: Position 9 (*before*) negativity shown at F7 (A) and topographic isovoltage map showing EMBEDDED (following *the sailor*) – MATRIX (following *_ openly*) from 300-600 msec (B).

As we did for the ‘additional N400’ at the previous position (section 6.4.2.1.6), we can examine a longer epoch to ensure that this post-gap LAN is not an artefact of the baselining procedure. In this case, the immediately preceding words do differ (*the sailor* for MATRIX GAPS and *_ openly* for EMBEDDED GAPS). Figure 6-22 presents the response at F7 starting at position 7 (*befriended*), lasting through the lexical difference LAN (solid lines, representing *the sailor*, are more negative from 800-1100 msec), but then reversing for the post-gap LAN (dashed lines following the gap are more negative from 1300 to 1600 msec) and continuing on through the following words (see section 6.4.2.1.9). We can see that the post-gap LAN is visible even without re-baselining at the prior words. We in fact observe a reversal of the patterns, with the solid lines

(MATRIX GAP) more negative for the lexical difference and the dashed lines (EMBEDDED GAP) more negative for the post-gap LAN difference.

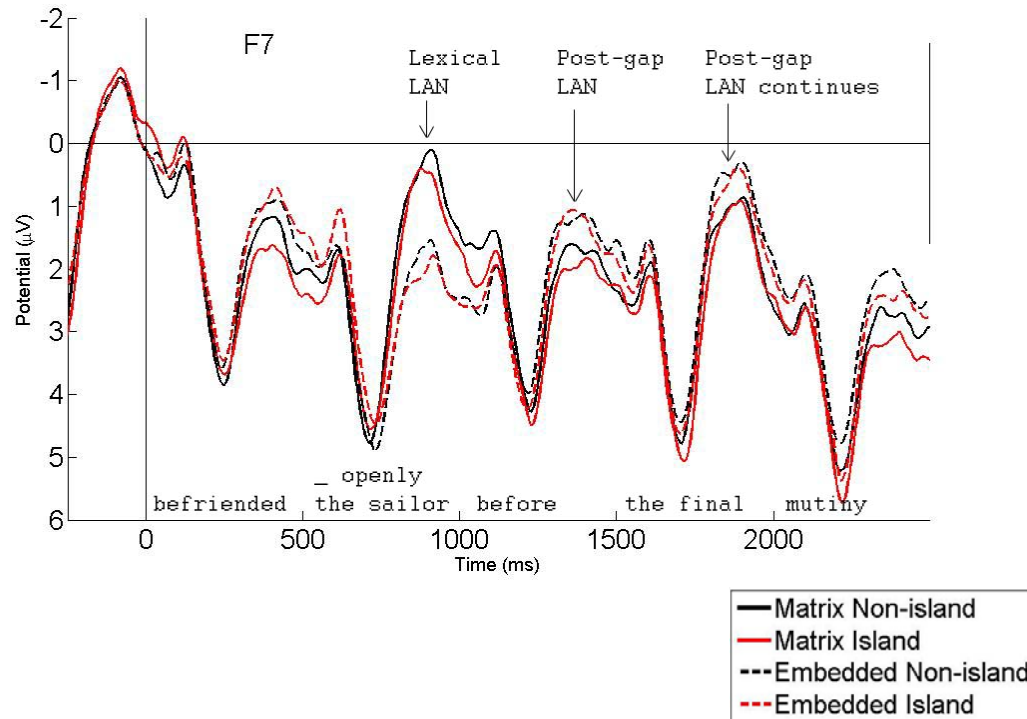


Figure 6-22: Five word average starting at position 7 (*befriended*): Point of interest is reversal of more negative conditions from lexical LAN { *_ openly / the sailor* } to post-gap LAN (*before*)

Recall that the post-gap matrix clause position (position 4) also elicited a LAN effect after the gap. While position 4 also elicited a N400 effect, no such response is evident in position 9. The interpretation of this N400 effect is discussed in more detail in section 6.4.2.2.4. For the current comparison, however, note that both post-gap positions (4 and 9) elicited a LAN effect (Figure 6-23). This LAN effect occurred following both matrix and embedded gaps, and inside both the island clause and non-island clause.

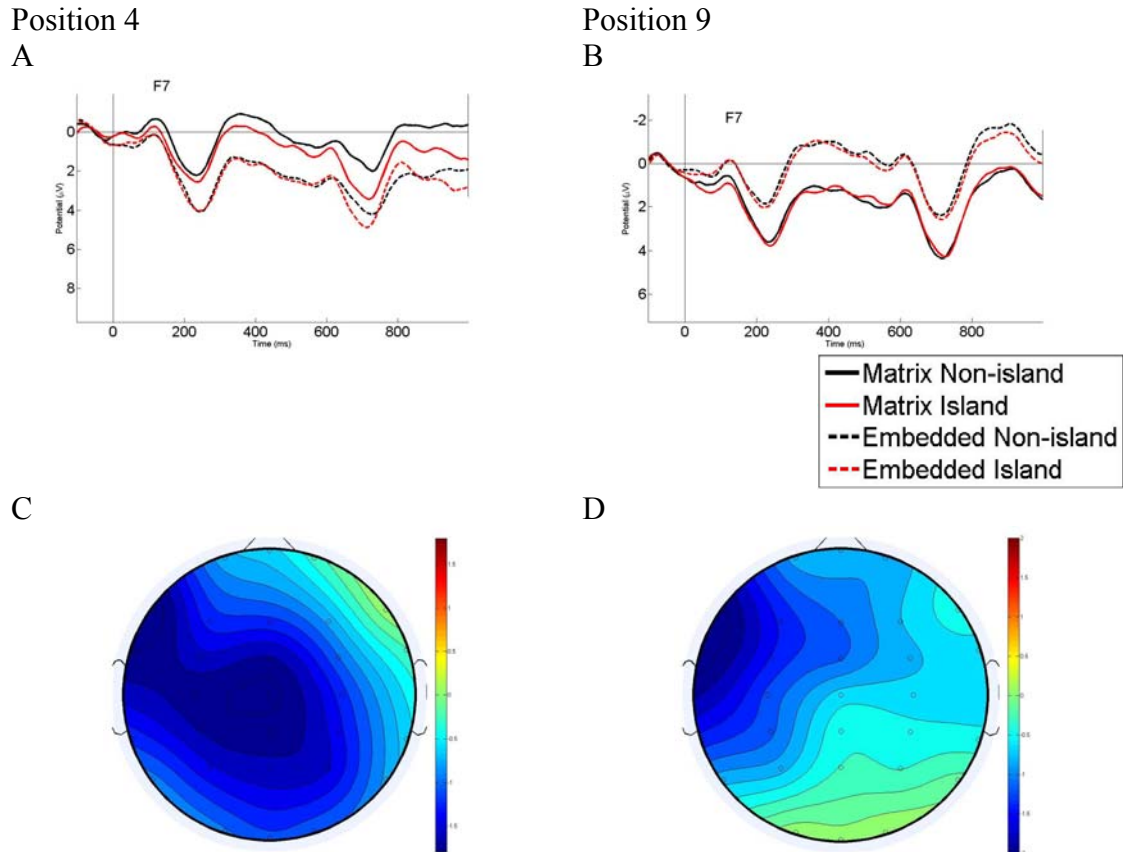


Figure 6-23: Comparison of LAN responses at positions 4 and 9. F7 (A and B) and topographic isovoltage map showing [the condition after the gap (*_openly*)] – [the condition after *the sailor*] from 300-600 msec (C, D).

6.4.2.1.8 Position 12 (sentence-final position: *hearing?*)

Recall from section 6.2.7 that N400s have been reported at the sentence-final position following both ungrammatical sentences (e.g. Osterhout & Holcomb 1992) and syntactically complex sentences (Osterhout 1990). Thus it was possible that the island violation condition would elicit an N400 effect at position 12. However, the 300-600 msec window omnibus ANOVA revealed only a main effect of GAP ($F(1,31) = 4.62, p = 0.039$) in which the final word was more negative after the long-distance

filler-gap dependency (EMBEDDED: 1.612 μV) than after the short-distance filler-gap dependency (MATRIX: 2.015 μV). See Figures 6-24 and 6-25. Discussion of these results will be delayed until after the analysis including the cognitive measures is presented (section 6.4.3.1.3 for results, section 6.4.3.2.3 for discussion).

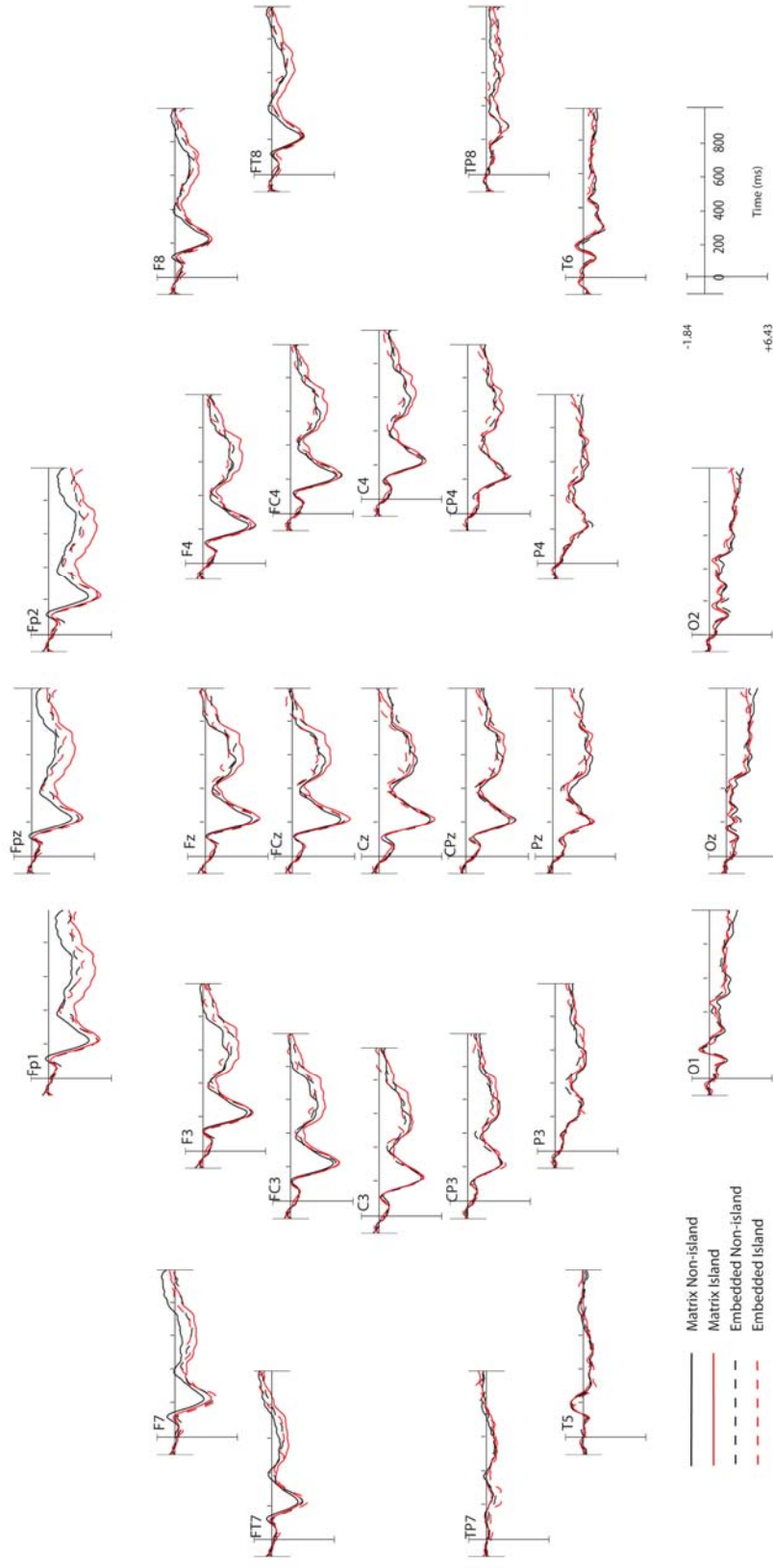


Figure 6-24: Position 12, sentence-final position (*hearing?*) whole head ERPs.

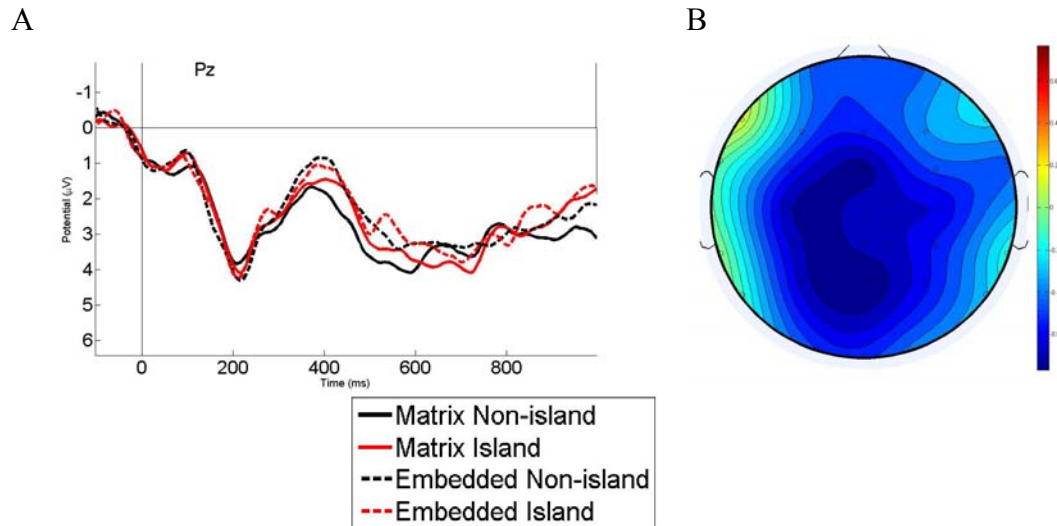


Figure 6-25: Position 12 (*hearing?*) broad negativity shown at Pz (A) and topographic isovoltage map showing showing EMBEDDED - MATRIX from 300-600 msec (B).

6.4.2.1.9 Slow wave: Sustained negativity

As the examination of sustained activity requires looking across multiple sentence positions, I repeat Table 6-1 as Table 6-13 for reference. Prior research has reported a sustained anterior negativity starting at the filler and continuing to the gap site (Kluender & Kutas 1993a,b; King & Kutas 1995; Phillips et al. 2005 but not McKinnon & Osterhout 1996; Kaan et al. 2000). The current materials (Table 6-1) do not differ in where the filler is located (*who*, position 1), so the start of this sustained effect would not be visible in the current results. However, it was thought that a difference might be found starting at or after position 3, where the MATRIX GAP conditions complete a filler-gap dependency, but the EMBEDDED GAP conditions must still wait for the gap site.

Table 6-13: Critical comparisons within the stimulus sentences, indicating both numbering and labels relative to the gap position in both the matrix and embedded clauses.

		<i>pre-gap position</i>	<i>gap position</i>	<i>post-gap position</i>			
Matrix clause:							
Position:	1	2	3	4			
	Who	had	_ openly/ the sailor	assumed/ inquired			
Embedded clause:							
Position:	5	6	7	8	9	10/11	12
	that/ whether	the captain	befriended	the sailor/ _ openly	before	the final/ mutiny	hearing?

Unfortunately, any possible sustained distinctions here are obscured by the other effects at and following this part of the sentence. The sustained negativity should be larger in response to the condition that still has a filler to associate with a gap (EMBEDDED GAP). The lexical differences at position 3 do result in a greater (left anterior) negativity in response to the EMBEDDED GAP conditions, but note that this can't be a sustained negativity (associated with working memory cost), as the same lexical items elicit this negativity at position 8, where no remaining sustained negativity is expected (see section 6.4.2.1.6 for comparison). There is a LAN effect at the following position (4, *assumed/inquired*, matrix clause post-gap position), but it is in response to the MATRIX GAP, not the EMBEDDED GAP (i.e. the opposite condition expected for the sustained negativity). That is, the post-gap LAN after the matrix clause gap prevents any attempt to isolate a sustained negativity for the incomplete filler-gap dependency. Additionally, the previously reported post-gap LAN effects

(sections 6.4.2.1.3, 6.4.2.1.7) are sustained across multiple word positions when not re-baselined (cp. Phillips 2006; Figure 6-26). Following the matrix gap (figure 6-26 A), there is a main effect of GAP, with MATRIX GAP more negative than EMBEDDED GAP from 300 msec through 2000 msec ($F(1,31) = 32.19, p < 0.001$). That is, the conditions that have just encountered the gap (Figure 6-26 A, MATRIX conditions; solid lines) are the most negative, and not the conditions that still have an unresolved filler-gap dependency during this portion of the sentence. Thus, under the view that the sustained negativity is elicited by the active holding of a word in memory, the wrong conditions are showing a sustained negative response.

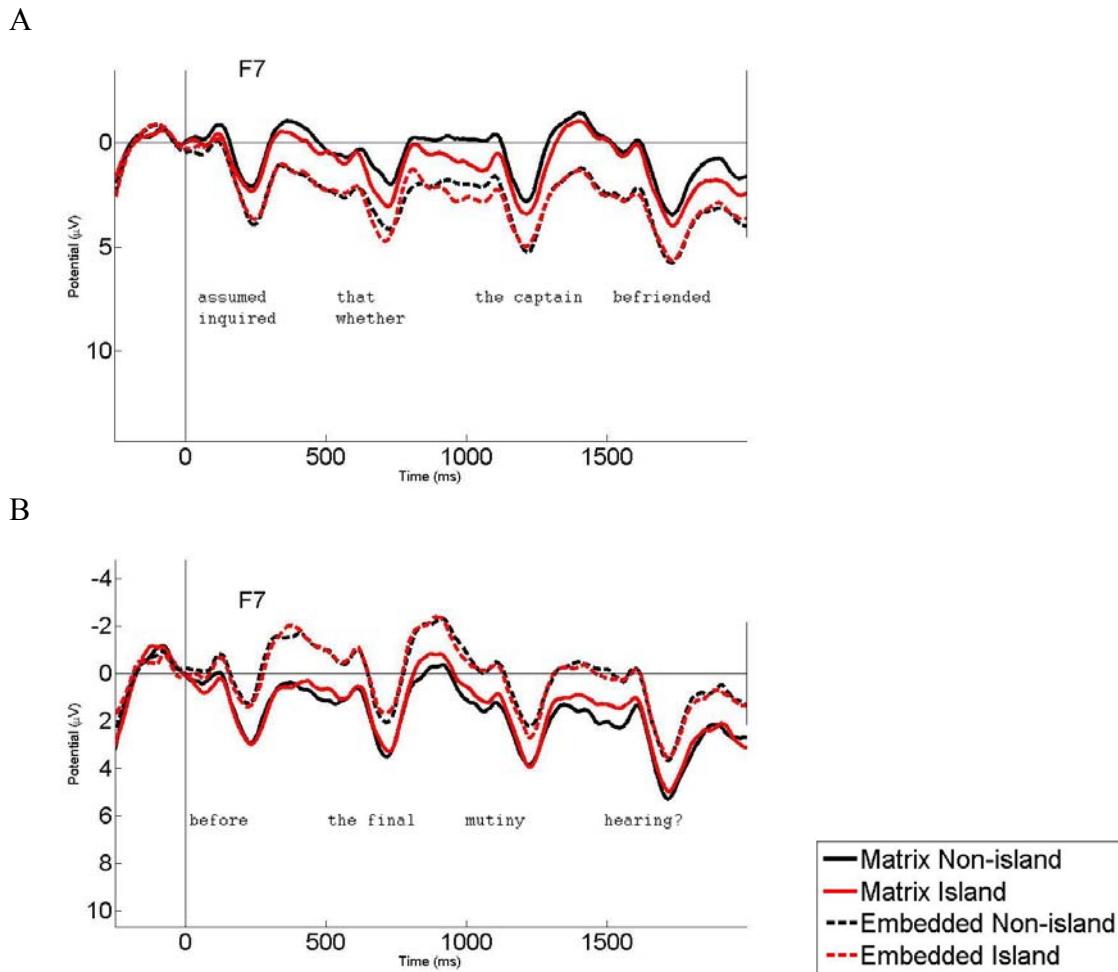


Figure 6-26: Four word averages starting at post-gap positions. Position 4 through 7 (A), position 9 through 12 (B)

As seen in Figure 6-26 B, the same pattern holds after the embedded gap, though conditions are ‘flipped.’ There is a main effect of GAP, with EMBEDDED GAP more negative than MATRIX GAP from 300 msec through 2000 msec ($F(1,31) = 18.35$, $p < 0.001$). The post-gap LAN is thus also sustained through multiple sentence positions. This latter effect especially cannot be interpreted as the cost of holding a filler in working memory, or the cost of having an unresolved filler-gap dependency.

All fillers and gaps have been encountered at this point, and yet we still observe a sustained negativity.

6.4.2.1.10 Summary

The results presented above can be summarized as five different findings (or lack thereof). Three effects occurred in both the matrix and embedded clauses, surrounding the gap. First, every pre-gap position, whether in the matrix or embedded clause, elicited a broad positivity. While consistent with previous results (Kaan et al. 2000, Phillips et al. 2005, Gouvea et al. 2010), the fact that this positivity occurred even before a gap at position 2, where every condition was identical up to this point (*Who had*), is problematic for previous interpretations of this positivity. This is discussed further in section 6.4.2.2.1. Second, the lexical differences between *the sailor* and *_ openly*, namely an N400 and a LAN, were visible at both the matrix and embedded gap positions and are discussed in section 6.4.2.2.2. Third, every post-gap position, again whether in the matrix or embedded clause, elicited a LAN effect. This occurred even in the EMBEDDED ISLAND conditions. This effect is discussed in section 6.4.2.2.4. Both of these LAN effects had a sustained duration, which is discussed in section 6.4.2.2.5.

Additionally, while an N400 response was predicted at the clause boundary based on Kluender and Kutas (1993b), this was not found in the 300-600 msec time window used for other N400 effects in this study. Instead, an earlier negativity (250-

350 msec) was found, surrounded by positivities. This is discussed in section 6.4.2.2.6.

Finally, independent of the lexical differences observable at both gap positions, there is an additional N400 effect at the embedded gap position. This N400 effect is larger in the EMBEDDED ISLAND condition compared to the EMBEDDED NON-ISLAND. The response to *_ openly* (the lexical item for both conditions) is more negative when inside a *whether*-island clause than when it is inside a non-island (*that*) clause. Recall that in addition to this N400 response, there was an N400 response to the matrix verb (position 4: *assumed / inquired*) when following a gap. Ideally, these two (non-lexical) N400 responses can be interpreted in a uniform manner. In section 6.4.2.2.7 I compare two possible interpretations of the N400 responses: semantic integration and predictability, ultimately arguing for the latter.

6.4.2.2 Discussion

In the following sections I discuss the results of the ERP results reported above. I proceed largely in chronological order throughout the sentence, but as some effects occur surrounding both the matrix and embedded gap position, I discuss these positions together when appropriate. In section 6.4.2.2.1 I discuss the pre-gap positivities elicited at positions 2 and 7. In section 6.4.2.2.2 I discuss the lexical differences elicited by *the sailor* and *_ openly* at the gap positions (3 and 8). In section 6.4.2.2.3 I briefly discuss the (non-lexical) N400 effects at position 4, but save the

majority of this discussion for section 6.4.2.2.7, where these N400 effects are discussed with the embedded gap ‘additional N400.’ In section 6.4.2.2.4 I discuss the post-gap LANs elicited at positions 4 and 9. In section 6.4.2.2.5 I discuss the sustained nature of the post-gap LANs. In section 6.4.2.2.6 I discuss the lack of an N400 effect as well as the early negativity elicited at the clause boundary (position 5). In Section 6.4.2.2.7 I discuss the only effect that resulted from an interaction of GAP and STRUCTURE, namely the (non-lexical) N400 effects at positions 4 and the ‘additional N400’ effect at position 8. Section 6.4.2.2.8 summarizes the discussion.

6.4.2.2.1 P600

A pre-gap P600 was first reported by Kaan et al. (2000), who interpreted this response as an index of syntactic integration difficulty. This was followed by studies by Phillips et al. (2005) and Gouvea et al. (2010), who also reported this late positivity at pre-gap positions. None of these three studies report on post-gap positions (i.e. there are no LAN effects measured). If the post-gap LAN indexes the retrieval of the filler from memory, as is commonly assumed and as we assume here, it is unclear how the integration process works such that an index of integration difficulty, as the P600 is commonly assumed to be, occurs before an index of retrieval. The results from the current experiment avoid this discussion, in part, by raising questions about how reliably the pre-gap P600 indexes syntactic integration difficulty.

In the current experiment, a late positivity was elicited before each of the four gap sites present in the material (just as a LAN is elicited after each of the four, section 6.4.2.2.4). At first glance, this appears to suggest support for the close relationship of the P600 with filler-gap processing, but the fact that the late positivity was elicited before the matrix gap position is a problem for the syntactic integration account. Consider the beginning of the experimental sentences in (6.1).

(6.1 a) MATRIX GAP: Who **had** openly assumed ...

(6.1 b) EMBEDDED GAP: Who **had** the sailor assumed ...

In the matrix clause, the pre-gap position is *had*. The key issue is that all four conditions are exactly identical up to this point: *Who had...* ? There is no way to predict which conditions will have a gap after *had* and which ones won't. The late positivity effect reported here must be a response to the following position. If we consider the conditions in two-position pairs, *had openly* (6.1 a) is more positive than *had the sailor* (6.1 b)⁵ in the 600-900 msec window (post-*had* onset). The 500 msec SOA used here means that the 600-900 msec window used to measure the pre-gap positivities⁶ corresponds to the 100-400 msec window of the following word. In the 300-600 msec window of position 3, *openly* is more positive than *the sailor* (a LAN and N400 are elicited at *the sailor*, see section 6.4.2.1.2). Prior claims that the P600 indexes the difficulty of syntactic integration (e.g. Kaan et al. 2000, see below) rely on

⁵ Recall that *the sailor* is presented simultaneously, as one 'word' position.

⁶ Differences are not significant using earlier windows.

the parser being able to predict where a gap would be. Syntactic integration is implausible for the current effect since the gap can't be predicted on the basis of *Who had... ?* Thus, in the current data, it is more plausible that the increased positivity of the response to *had openly* in (6.1 a) in the late 600-900 msec window is due to an early response to *openly*.

If this is true for the matrix gaps, the interpretation of the embedded gaps, which are followed by the same exact lexical items, immediately becomes suspect as well. A conservative analysis of the current data, then, would be that these late pre-gap positivities are artefacts of differences in the next position. These differences are still substantial and informative, but they are not predictive. The late positivities here are reflections of gap identification, but it is unlikely that they are indexing syntactic integration, particularly in the 100-400 msec time window of the gap disambiguation position.

Based on this conservative approach, we can say that the gaps are identified in all four cases (matrix and embedded gaps, and both embedded inside an island and non-island clause), but no further inferences will be drawn from these effects. But what about the previous studies that have reported pre-gap P600s? Are these also open to an interpretation where the effects are being driven by the next word? While it appears that the next word can have an influence on this pre-gap P600, this can't explain all of the prior findings.

Kaan et al. (2000, Experiment 1) compared three sentence types, shown in (6.2 a-c).

(6.2 a) Emily wondered who _____ the performer in the concert

had **imitated** _____ for the audience's amusement.

(6.2 b) Emily wondered whether _____ the performer in the concert

had **imitated** a pop star for the audience's amusement.

(6.2 c) Emily wondered which pop star the performer in the concert

had **imitated** _____ for the audience's amusement.

(modified from Kaan et al. 2000, 2a-c)

The pre-gap position is *imitated*. The following lexical items are *for* (actually a variety of words throughout the materials) or *a* (*some* in other lexicalizations). Kaan et al. consider the possibility that their pre-gap effect is being influenced by the following word but dismiss it for three reasons. First, they measure in two time windows, 500-700 and 700-900 msec. They claim that the 500-700 msec window, which would correspond to the 0-200 msec window of the following position, is too soon to show lexical effects. While this claim is questionable (N100 and P200 responses could index lexical effects), it may be more informative to examine which comparisons are significant in each of the two time windows. Specifically, in the early time window (500-700 msec) the d-linked *which pop star* sentences (6.2 c) are more positive than the other two conditions. However, the *who* sentences (6.2 a) are not more positive than the *whether* sentences (6.2 c) this early in the epoch. The *who* sentences (6.2 a) are only significantly different from the *whether* sentences (6.2 c) in the 700-900 msec window, "though the effect was statistically weak" (Kann et al. 2000, pg 171). The

most robust effect then is a positivity in the d-linked, *which pop star* sentences (6.2 c), while the bare, *who* sentences (6.2 a) show a later and weaker effect. So it may be that the key distinction here is that d-linked fillers elicit a pre-gap positivity early and robustly. If the pre-gap P600 can be the result of differences at the following word (as is the case in the current experiment), then this is more likely to be the case in the bare filler (6.2 a) conditions.

Second, Kaan et al. (2000) claim that the direction of the potential effect that could be caused by lexical differences of the following word are in the opposite direction of the attested effects at *imitated*. They argue that *a*, as a high frequency, closed-class word which is only alternating with one other lexical item (*some*) in the materials should have a reduced N400 response compared to the less predictable *for*. If we assume a more negative response to *for*, this cannot explain a more positive response to the prior word (*imitated*, 6.2 a, c). While this argument is fairly convincing, it would have been more informative to have shown the actual difference between *a* and *for* rather than relying solely on this thought experiment.⁷

The final argument put forth by Kaan et al. is that lexical differences do not explain the difference between (6.2 c) and (6.2 a), which are both followed by *for*. I do not dispute this, but note again that the comparison between (6.2 c) and (a) is comparing a d-linked filler (*which pop star*) with a bare filler (*who*). It may be that the more robust pre-gap effect is observed based on this d-linking manipulation rather than the presence/absence of the gap itself. Hofmeister (2007) presents reading time

⁷ Note that the current experiment's materials differ from Kaan et al.'s here in a significant way: while Kaan et al. present only the determiner following the potential gap site, the current experiment presents the determiner plus noun.

data that demonstrates processing facilitation for d-linked⁸ fillers at the gap site compared to bare fillers. If the P600 reflects syntactic integration, and the P600 response to d-linked fillers is more robust, then it becomes difficult to reconcile Kaan et al. (2000)'s claim that a P600 reflects greater integration difficulty with Hofmeister (2007)'s data showing facilitation for these fillers. On the other hand, if the P600 simply reflects gap identification, it could be that part of the d-linked fillers retrieval/integration facilitation is that they make it easier to identify a gap. This could be because the d-linked filler is heavier and thus carries a larger processing burden, making it a priority to discharge this cost (e.g. Gibson 2000). Or it could be because a more well defined filler generates a stronger/more certain prediction for a gap; there is less uncertainty about whether the filler was noise (e.g. Levy et al. 2009).

Phillips et al. (2005) also report a pre-gap positivity in the 700-900 msec window when comparing a gap filled by a d-linked *which accomplice* phrase to a lack of a gap. Again, it is unclear if the effect is being driven by the d-linked *which accomplice* or the basic comparison of a filler-gap dependency vs. a non-dependency control. Additionally, the lexical items in the Phillips et al. materials differ after the gap (*the* vs. *in*), leaving open the possibility that the pre-gap late positivity was influenced by lexical differences of the following word. Gouvea et al. (2010) sought to control these lexical differences. In order to do so, Gouvea et al. had to use sentences with a gap in an indirect object position (6.3) following the direct object, but still

⁸ Hofmeister avoids the term d-linking, referring to such phrases as 'more explicit *wh*-phrases.'

measured the positivity at the pre-gap verb (which was followed by the direct object NP).

(6.3 a) The patient met the doctor while the nurse with the white dress **showed** the chart during the meeting.

(6.3 b) The patient met the doctor to whom [the nurse with the white dress **showed** the chart __ during the meeting.]

(Modified from Gouvea et al. 2009, Table 1)

Gouvea et al. reported a marginal positivity in the 500-700 msec window in response to *showed* in (6.3 b) compared to (6.3 a).⁹ While they took care to control for lexical items surrounding the point of comparison, the discussion surrounding the lack of a more robust difference centered on the filler *to whom* “carrying information” compared to a less informative *who*. While this may be a factor, it seems at odds with earlier patterns in Kaan et al. (2000) and Phillips et al. (2005), in which the strongest pre-gap effects of positivity were in sentences in which the filler was d-linked, and thus also “carrying information.” It is plausible then, that the reason Gouvea et al. find only a marginal P600 is because of a lack of lexical differences in the following word.

While this discussion is by no means conclusive, it suggests that there are at least two factors that can contribute to a pre-gap P600: (i) lexical differences at the following position (as seen here) and (ii) a d-linked, rather than bare filler. As the current study used only bare *wh*-phrase fillers, it is unsurprising then that rather than

⁹ Gouvea et al. (2009) do not report any measurements at *chart*, the actual pre-gap word.

indexing an integration effect, the current results seem to reflect lexical differences at the next sentence position. The results from Gouvea et al. (2010) suggest that closely controlled following words may reduce the pre-gap positivity effect. The predicted direction of the lexical differences in Kaan et al. (2000) remains an outstanding issue (i.e. are the actual influences of *a* vs. *for* as predicted?). For the purposes of this dissertation, the more conservative interpretation of these P600/late positivities will be used, namely that they represent the identification of a gap.

6.4.2.2.2 Lexical differences

The lexical differences between words used at the gap positions are not the direct focus of inquiry for this experiment. The importance of identifying these differences is instead being able to identify effects above and beyond these lexical differences. Recall that the design of the materials allows us to compare these lexical differences at two word positions: both the matrix and embedded gap sites (section 6.4.2.1.6). This serves as an invaluable check that observed differences are attributable to lexical differences and not the sentential position that they are appearing in. We do see one such non-lexical variation occurring in addition to the lexical differences at the embedded gap site, but this is discussed below in section 6.4.2.2.4. For now, we focus only on the lexical differences independent of sentence position.

Two lexical differences were observed in the 300-600 msec window when comparing *the sailor* with *openly*: a LAN and an N400, both larger in response to *the*

sailor. These effects were observed in two positions: at position 3 (the gap disambiguation position) and position 8 (the embedded gap position). For ease of reference, and to distinguish these effects from others in the sentence, I will call these ‘lexical LAN’ and ‘lexical N400’ effects, respectively. I do not intend for there to be any particular interpretation of the processes underlying these effects by my use of the word ‘lexical.’ I simply mean to identify these as effects that arise when comparing the different lexical items *the sailor* to *openly*.

The differences between *the sailor* and *openly* should be obvious even to non-experts. *The sailor* is two words: the definite determiner and a noun. *Openly* is an adverb. It would not be surprising for the presence of the definite determiner and the difference in grammatical word category to elicit different brain responses, even when the words are controlled for length and frequency (section 3.2).

Open class words have long been known to elicit larger amplitude N400 effects than closed class words (e.g. Kutas & Van Petten 1988, Kutas, Van Petten & Besson 1988, Van Petten and Kutas 1991). Adverbs are not closed class words, but they are a more restricted class of words than nouns, so it is unsurprising to find a smaller N400 response to this more restricted class of words.

While the N400 difference is likely due to word category differences, the LAN is likely due to the presence of the definite determiner *the*. The determiner *the* (compared to *a*) is frequently used to refer back to a previously mentioned discourse referent. In the materials for the current experiment *the* appears frequently, but does not refer to any previous participant in the discourse. Ambiguity in referential

processing has been associated with a sustained frontal negative shift (called Nref by Van Berkum and colleagues), which is similar to a LAN in polarity, latency and scalp distribution (Van Berkum et al. 2007; Barkley, Kluender & Kutas 2011). Thus it is plausible that the left anterior negativity observed to *the sailor* reflects additional referential processing triggered by the definite determiner. Additionally, nouns following the word *the* (compared to *a*) have been shown to elicit a LAN effect (Anderson & Holcomb 2005). Crucially, however, it is not the case that the LAN effect observed to *the sailor* is a working memory related LAN (e.g. Kluender & Kutas 1993 a,b; King & Kutas 1995; see section 6.4.2.2.4 for discussion).

If a working memory related LAN were observed at either gap position (matrix: 3 or embedded: 8), it should be observed to the condition with a gap in the corresponding clause (the condition with *openly*). Kluender & Kutas (1993 a,b) report LAN responses following both fillers and gaps. As discussed previously, the fillers are identical across all conditions for the current experiment, so the post-filler effect should not be visible. It is possible that the post-gap LAN response would occur immediately after the matrix (*had*) or embedded (*befriended*) verbs, but this post gap LAN response should occur to *openly* (the condition containing the gap), not *the sailor*, as attested here. Further, the predicted post-gap LAN effect (with the LAN elicited in the *openly* conditions) occurs at the following position (4 and 9, section 6.4.2.2.4).

Another possibility is that the LAN observed at position 3 is a glimpse at the sustained post-filler LAN reported in the literature (e.g. King & Kutas 1995; Phillips

et al. 2005; see section 6.4.2.2.5 for discussion). While this may be plausible for position 3, it does not explain why the same effect is present at position 8 but not the word positions in between. That is, if the LAN measured at *the sailor* in the embedded clause reflected an ongoing sustained anterior negativity initiated at the filler, then this same effect should be visible in positions prior to this one as well. This is not the case. The best supported interpretation of the LAN elicited by *the sailor* in two different sentence positions is that, like the N400 effect elicited by the same word, it is due to lexical differences with *openly*.

In sum, it is not surprising that *the sailor* elicited substantially different brain responses than *openly*. But these differences were crucially not due to working memory considerations. While these differences are not the focus of the present inquiry, the design of the experiment (namely counter-balancing the appearance of these lexical items across the GAP manipulation) allows us to identify these effects and thus observe any differences that arise in addition to these effects, such as the N400 effect discussed in the section 6.4.2.2.7.

6.4.2.2.3 Matrix verb N400

The matrix clause post-gap N400 effect (section 6.4.2.1.3) was not predicted. The N400 response was larger to both matrix verbs introducing an embedded declarative (*assumed*) and those introducing an embedded interrogative (*inquired*) when these verbs immediately following *openly* (compared to following *the sailor*).

Since this is an effect of GAP position rather than STRUCTURE (which determined which verb was used), this is not an effect due to lexical differences. Instead this effect must be due to how these verbs relate to the previous words in the sentence. This issue is dealt with in more detail in section 6.4.2.2.7 when this effect is considered together with the ‘additional N400’ effect at the embedded gap position. Ultimately, I argue that both of these effects are due to differences in predictability. Verbs that subcategorize for embedded clauses are less predictable following *Who had openly...* (which can be completed with an intransitive verb) than they are following *Who had the sailor...* (which can not be straightforwardly completed with an intransitive verb).¹⁰

6.4.2.2.4 LAN

As discussed in section 6.2.2, when elicited by filler-gap dependencies, the LAN has been interpreted as an index of both “the storage of a filler in working memory and its subsequent retrieval” (Kluender and Kutas 1993a, pg 205). The current materials were not suited to examine the storage of fillers since all conditions have a *wh*-filler in the same position (namely in sentence-initial position), but they are well suited to examine the ‘subsequent retrieval’ of those fillers. Note that this differs from other ERP experiments on filler-gap dependencies in *wh*-questions, which often

¹⁰ Modifications of the predicate with prepositions could allow for valid continuations: *Who had the sailor danced for _ ?*

are not designed to make comparisons at the post-gap position (e.g. Kaan et al. 2000, Phillips et al., 2005).

Based on Kluender and Kutas (1993b) and subsequent research (see Chapter 2, section 2.2.5.2.1) we predicted a LAN effect following gap positions. The experimental conditions of the current experiment generated four gap positions: two MATRIX GAPS and two EMBEDDED GAPS. LAN effects were elicited after all four gaps, and did not interact with the factor STRUCTURE. To our knowledge, this represents the first ERP evidence for filler-gap association in a matrix clause subject gap.

Kluender and Kutas (1993b) did not observe LAN responses to subject filler-gap dependencies, because there should be no need for a working memory cost when the gap is detected and resolved immediately following the filler (at *tried* in 6.1 a). However, in the current materials, there is separation between where the filler is encountered, and where the location of the gap position is disambiguated (at *openly* in 6.1 b). The incremental nature of the parser means that it is not possible to know definitively where the gap will be when only *Who had* has been encountered (note that this is independent of theoretical syntactic considerations of where the gap should ultimately be represented). On the other hand, *Who had openly* (and *who tried*) both indicate a matrix clause subject gap. Thus, in the current materials, the subject filler is separated from the disambiguating gap position, if only by one word (*had*), and a post-gap LAN effect was elicited (as predicted by Kluender and Kutas 1993b, pg. 207-8). The differences between the materials are shown in (6.4).

(6.4 a) Can't you remember [who **tried** to scare him...

Modified from Kluender and Kutas (1993b, 35)

(6.4 a) Who had **openly** assumed that ...

Current experiment: MATRIX GAP (NON-ISLAND)

Furthermore, as predicted, the current experiment elicited post-gap LAN effects in the embedded clause, whether the gap was embedded in a NON-ISLAND (*that*) or an ISLAND (*whether*). Thus, we replicated Kluender and Kutas' (1993b) finding that post-gap LANs are not attributable to the unacceptability of a sentence. However, we did not replicate Kluender and Kutas' finding that the amplitude of the LAN varies with the type of embedded clause. We have no statistically significant interaction of GAP and STRUCTURE at this position (Figure 6-18 A). This is unexpected given that Kluender and Kutas report that (6.5 b) and (6.5 c) elicit a larger LAN at *by* than (6.5 a).

(6.5 a) What_a do you suppose [that they caught him at _a **by** accident]?

(6.5 b) ? What_a do you wonder [if they caught him at _a **by** accident]?

(6.5 c) * What_a do you wonder [who_b they caught _b at _a **by** accident]?

Modified from Kluender and Kutas (1993b)

(6.6 a) Who_a had the sailor assumed

[that the captain befriended _a **openly** ...

(6.6 b) * Who_a had the sailor inquired

[whether the captain befriended _a **openly** ...

Current experiment: EMBEDDED GAP conditions

The materials used for the current study (6.6) differ from those in (6.5) in a few ways. First, the current experiment exclusively uses animate fillers (*who*) instead of a mix of animate and inanimate fillers (*what*). Second, the embedded gaps in the current experiment are consistently direct objects of the embedded verb (*befriended* openly...) instead of sometimes being the object of a preposition (*caught him at* by...) in (6.5). This combination of differences generates different possible linguistic structures that could be being questioned, as opposed to questioning the identity of a person with *who*. Consider some possible answers to (6.5 a): *they caught him at the cigarette machine; they caught him at swindling the elderly*. It is possible that the availability of these different reading and structures (gap as the object of a preposition or direct object) result in Kluender and Kutas' participants being less certain of where the gap will be and/or how to interpret it. Third, there is only ever one filler-gap dependency in a given sentence in the current experiment. Island violations are due to an embedded *whether* clause, rather than an embedded *who* or *what* clause as in Kluender and Kutas (1993b), which introduces an additional filler-gap dependency. However, *if* (6.5 b), like *whether* (6.6 b), introduces an interrogative clause without an

additional filler-gap dependency. Kluender and Kutas still find a larger post-gap LAN after *if* than *that*. Thus the difference between the experiments' results cannot only be due to an additional filler-gap dependency in (6.5 b).

The fact that we see no reliable differences in the current experiment to the LAN elicited after EMBEDDED NON-ISLAND GAPS and EMBEDDED ISLAND GAPS indicates that the LAN is not being modulated by the fact the gap is inside of a *whether*-island. The same filler-gap association process is reflected in both the acceptable *that*-clause and unacceptable *whether*-clause conditions, and so we cannot attribute this difference in acceptability to the process underlying the LAN. Taking the canonical view that the post-gap LAN reflects the process of retrieving the filler from memory, then we have no evidence that participants are having any more difficulty retrieving the filler for a gap embedded in an unacceptable island clause than in an acceptable non-island (*that*) clause. This is incompatible with the similarity-interference account of islands, which predicts the difficulty in processing islands to be in the retrieval process (Chapter 2, section 2.3.3.2).

6.4.2.2.6 Sustained LAN

As discussed in section 6.4.2.1.9, the current design did not afford a clear view of the sustained (left) anterior negativity previously reported in the literature (e.g. King & Kutas 1995; Phillips et al. 2005). However, we do see a sustained response following both the matrix and embedded post gap LAN responses. As shown in Figure

6-26, this effect persists for several words past the gap site when subsequent sentence position epochs are not rebaselined (cp. Phillips et al. 2005). I will call this a ‘lingering LAN’ to distinguish it from the sustained LAN previously reported in the literature. In (6.7) the predicted location for the sustained post-filler LAN and the attested location for the lingering post-gap LAN are schematized for the EMBEDDED GAP conditions.

(6.7) Who had the sailor ... befriended _ openly before ... mutiny hearing?

|--possible sustained LAN--- |---lingering LAN-----

While the sustained LAN has been claimed to reflect the ongoing cost of holding a filler in memory, there is no similar cost that can be associated with the ‘lingering LAN,’ since it is elicited after both the filler and gap have been encountered (and presumably filled/associated with each other based on the above discussion); there is no longer a need to maintain the filler in working memory. If both of these continuing effects (the sustained LAN following the filler and the lingering LAN following the gap) are actually the same response, this calls into question the interpretation of the sustained LAN indexing the maintenance/storage cost of the filler.

A similar ‘lingering LAN’ effect was reported in Kluender and Kutas (1993a), who found a LAN two words after a grammatical, embedded gap, even with rebaselining. They note that “the LAN effect apparently did not subside immediately after the filler had been retrieved from working memory and assigned to its gap”

(Kluender and Kutas 1993a, pg 206). In the current experiment, the ‘lingering LAN’ is not found if later positions are re-baselined, but only when longer epochs are examined. The evidence here confirms that the LAN effect is persistent, even when a filler no longer needs to be held in working memory.

It is unclear at this point how this ‘lingering LAN’ might be related to the sustained LAN reported elsewhere. Since both the filler and gap elicit LAN responses, and both of those responses have been reported to have an ongoing effect, it raises the question of whether the sustained LAN is reflective of a maintenance/storage cost. The difficulty of maintaining a filler in memory presumably increases over time as additional words are encountered (Gibson 1998, 2000). Phillips et al. (2005) argue that the sustained LAN does not reflect this pattern of increasing difficulty throughout the sentence by demonstrating that it disappears if it is re-baselined at each word position. This is the same pattern as is seen with the ‘lingering LAN’ in the current experiment. If these two LAN responses both ‘linger,’ then it would be problematic to associate one with the cost of maintaining/storing a filler in working memory, but not the other. Note that this does not undermine the LAN’s association with working memory processes (i.e. encoding following the filler and retrieval following the gap), but only the association of the sustained nature of the LAN with a maintenance cost.

While the current result suggest that the apparent ongoing maintenance cost (the sustained LAN) may be the same ‘lingering LAN’ response reported here, this deserves to be tested explicitly in future research. As discussed previously, the current materials did not allow an examination of the sustained LAN, so a comparison is not

possible here. Future research could examine these two multiword effects to see if the sustained and lingering LANs have different properties. It may be that both types of sustained/lingering LAN effects elicited by filler-gap dependencies simply reflect the placing of the filler into memory (at the filler) and then the subsequent retrieval of that gap (at the gap site), but not the ongoing maintenance of that filler.

6.4.2.2.5 Clause boundary

An effect was predicted at the clause boundary based on two facts. First, Kluender and Kutas (1993b) reported an increased N400 to *who* compared to *that* at the clause boundary, but only in yes/no questions, not *wh*-questions. Even though the current materials are *wh*-questions, it is possible that we would detect such a difference here. Second, the interaction of GAP and STRUCTURE in the self-paced reading experiment presented in Chapter 5 occurred at the clause boundary. Thus we expected an ERP response that would pattern like that behavioral interaction (an interaction of GAP and STRUCTURE). Neither of these effects was found in the results of the current experiment.

With regard to the first issue, as already mentioned, Kluender and Kutas (1993b) reported finding an N400 at the clause boundary only for yes/no questions. It might have been the case that the matrix gap questions in the current study (6.8) would have been like those yes/no questions, as in both cases there is no outstanding filler-

gap dependency at the clause boundary (having never existed in the case of yes/no questions, and having been previously resolved in the case of matrix *wh*-questions).

(6.8 a) Who had the openly assumed [**that** the captain ...

(6.8 b) Who had the openly assumed [**whether** the captain ...

However, no N400 difference was found between *whether* in (6.8 b) and *that* in (6.8 a). Instead, an earlier negativity was elicited by *whether*, measured in the 250-350 msec window. It is likely that this early response is due to length differences between *that* and *whether*. Longer words have been reported to elicit a more negative response than shorter words following the P200 (Neville et al. 1992, Hauk and Pulvermüller 2004, and Ueno and Kluender 2009).

There are three reasons why it is unlikely that this 250-350 msec negativity is actually an N400 with a somewhat earlier latency than expected. First, consider why this N400 could plausibly be occurring earlier than the standard 300-600 msec window. One possibility is that the clause boundary (i.e. the complementizers *that* and *whether*) could be predicted by the preceding matrix verb. However, the two-fifths of the experimental materials include verbs that can and do precede both *that* and *whether* (i.e. *said that* and *said whether* were both present in this experiment, see Chapter 3, section 3.2), making the complementizer less predictable overall. Secondly, this negativity did not show any particular scalp topography, while all other N400 effects elicited in the experiment showed significant interactions with ELECTRODE and

exhibited central maxima (occasionally also extending over parietal regions of scalp or right lateralized, but always at least central). While this early negativity showed no significant interaction with ELECTRODE in the omnibus ANOVA, the topographic isovoltage map (Figure 6-10 D) gave the visual impression that the effect was largest over the right anterior regions, unlike any other N400 effect in the data. Finally, all other N400 effects in the current experiment were observed in the canonical 300-600 msec latency window, and as discussed above, there is not a clear reason why it should be earlier just in this case.

The second reason why an effect at the clause boundary was predicted was the fact that there was an interaction of the factors GAP and STRUCTURE in the self-paced reading experiment at the clause boundary (Chapter 5, section 5.4.2.1.1). The largest slowdown in reading times occurred at the clause boundary in the EMBEDDED ISLAND condition, suggesting that processing difficulty for *whether*-island violations was greatest in the clause boundary region. Furthermore, this interaction varied with reading span scores (section 5.4.3.1.1). Low span readers made a three-way distinction between the EMBEDDED ISLAND condition (island violation, read slowest), MATRIX ISLAND condition (island structure, no violation) and both NON-ISLAND conditions (read the fastest). High span readers, on the other hand, did not distinguish between the MATRIX ISLAND condition and the NON-ISLAND conditions; they slowed down only for the EMBEDDED ISLAND at the clause boundary. If the negativity elicited by *whether* from 250-350 msec is the ERP reflection of this reading time result, we expect the ERP response to pattern with the reading times with respect to (i) the interaction and

(ii) the individual differences.¹¹ Neither of these patterns are found for this 250-350 msec negativity, making it unlikely that this is the brain response correlate of the reading time effects.

Why then do we see an interaction of GAP and STRUCTURE in the reading times and not in the event-related-potentials? Recall that while the design of the experimental sentences in all three experiments reported in this dissertation was the same (Chapter 3, section 3.2), because of the need for a higher signal-to-noise ratio in the ERP experiment, many more experimental sentences were presented to the participants (40 per condition rather than 8 per condition). Also, in order to keep the total number of sentences as low as possible so as to not exhaust the participants, different fillers were used for the ERP experiment than the self-paced reading experiment. The fillers for the ERP experiment all included sentences with *that* or *whether* clauses, while balancing for a number of other factors (see section 6.3.2). This resulted in the participants in the ERP study reading 240/240 sentences containing either a *that* or *whether* clause, while the participants in the self-paced reading experiment read only 32/200 such sentences. Thus, it seems reasonable that distinctions that participants made between *that* and *whether* might have lessened over the course of repeated exposure to them. Furthermore, the self-paced reading participants' experimental lists were organized such that they only saw two sentences with *that* and two sentences with *whether* (one example of each experimental condition) per twenty-five sentences. Thus, while the ERP participants were

¹¹ A median split analysis of the ERP data and discussion of the analysis is presented in section 6.4.3. but it is worth noting at this point that there was no co-variation with reading span at the clause boundary in the ERP experiment.

bombarded with *that* and *whether* sentences, the self-paced reading participants' exposure to them was much less overall, and spread out amongst other sentences. It is known that repetition of lexical items over the course of an experiment leads to decreased N400 amplitudes (Van Petten et al. 1991), and that closed-class words elicit much smaller N400s than open-class words, so it is perhaps unsurprising that a potential N400 difference in a comparison of closed-class words was not found when using the current materials.

This methodological explanation for the lack of a GAP x STRUCTURE interaction at the clause boundary for the ERP experiment makes a straightforward prediction: If the self-paced reading experiment were repeated using the materials and fillers from the ERP experiment, no GAP x STRUCTURE interaction at the clause boundary would be observed either.

While we did not find direct evidence for the importance of the clause boundary in the online processing of *whether*-island violations in the ERP experiment, this does not automatically lead to the conclusion that the clause boundary is unimportant. The next section (6.4.2.2.7) discusses the 'additional' N400 effect at the embedded gap position, which crucially only appears after the *whether* clause boundary and not after the *that* clause boundary. Thus, while design issues may have obscured effects at the clause boundary itself, their influence is nevertheless observed at the embedded gap site.

6.4.2.2.7 N400 effects

In addition to the lexical differences discussed in section 6.4.2.2.2, there was an ‘additional N400’ at the embedded gap site elicited by the ISLAND condition compared to the NON-ISLAND condition (section 6.4.2.1.6). I refer to this as an ‘additional N400’ in order to distinguish it from the lexical effects found at the same sentence position. Recall that this ‘additional N400’ cannot be a lexical effect as the key comparison is between identical lexical items *openly* and *openly*. The difference in the two conditions is in whether the gap is found embedded in an island (*whether*-clause) or not (*that*-clause).

This effect represents the crucial interaction of GAP and STRUCTURE that we had predicted would distinguish the processing of the *whether*-island violation from related control sentences. All of the previously described effects have been main effects, mostly of GAP POSITION, and thus reflecting the difference between whether a gap was present at that point of the sentence, but *not* how that gap was processed differently when it was within a *whether*-island.

That the interaction of GAP and STRUCTURE took the form of an N400 effect was unexpected, however. As discussed in section 6.2.7, we expected that this interaction might appear in the pre-gap P600 or post-gap LAN, effects that have been claimed to reflect syntactic integration (e.g. Kaan et al. 2000) and filler-gap association (e.g. Kluender & Kutas 1993a,b), respectively. In language, the N400 is more commonly associated with semantic, rather than syntactic, phenomena, though it

has strong associations beyond language (see Kutas & Federmeier 2011 for a review). What process is the N400 effect indexing here, then?

Two hypotheses will be considered here, with the evidence to favor the second one. First, the N400 could be indexing the process of semantic integration of the filler with the gap (e.g. Brown & Hagoort 1993, 1999; Chwilla, Brown, & Hagoort 1995; Hagoort et al. 2009). In this ‘integration hypothesis’ the larger N400 response in the ISLAND condition would be due to the integration being more difficult in this condition than in the NON-ISLAND *that*-clause. Second, the N400 could be in response to a difference in predictability of the gap in each clause (e.g. Kutas & Hillyard 1984; DeLong et al. 2005; Lau, Holcomb & Kuperberg 2013). In the ‘predictability hypothesis’ the larger N400 response in the ISLAND condition would be due to the gap being less predictable in this condition than in the NON-ISLAND *that*-clause. In order to decide between these two hypotheses, we will consider the ‘additional N400’ effect at the embedded gap position together with the unpredicted N400 responses elicited after the matrix gap position.

There was also a non-lexical N400 effect elicited in the matrix clause of the current experiment. A larger N400 is elicited in the MATRIX GAP condition at position 4 for both *assumed* and *inquired*. These same lexical items (*assumed/inquired*) occur at position 4 in the EMBEDDED GAP condition as well, so this cannot be a lexical difference. Table 6-14 schematizes the relevant non-lexical N400 effects for the discussion below.

Table 6-14: Location of (non-lexical) N400 effects shaded in gray. Critical indicators of condition are underlined. Note: the same experimental conditions are not represented by the matrix clause (MATRIX GAP) and embedded clause (EMBEDDED ISLAND).

<i>pre-gap position</i>	<i>gap position</i>	<i>post-gap position</i>
-----------------------------	-------------------------	------------------------------

Matrix clause: N400s after the gap in both MATRIX GAP conditions

Position:	1	2	3	4
	Who	had	_ <u>openly</u>	assumed/ inquired

Embedded clause: N400 at the gap only in the EMBEDDED ISLAND condition

Position:	5	6	7	8	9	10/11	12
	<u>whether</u>	the captain	befriended	_ <u>openly</u>	before	the final/ mutiny	hearing?

Recall that at position 4, after the matrix gap position, a post-gap LAN was elicited (just as it was after the embedded gap, see section 6.4.2.1.7). However, at this same position (position 4), there was also an increased N400 in the post-gap position for both NON-ISLAND and ISLAND conditions (section 6.4.2.1.3). Additionally, in the embedded clause, an increased N400 is elicited at the gap position *only* in the ISLAND condition. The ‘additional N400’ in the embedded clause differs from those found in the matrix clause in two ways: it occurs at the gap position (and not the post-gap position as in the matrix clause) and it only occurs in the ISLAND condition. The similarity between these matrix and embedded N400 responses is that they occur after *both* (i) the gap position and (ii) the verb that assigns the filler/gap its thematic role are encountered. The discussion below aims for a unified account for these three effects (two in the matrix clause, one in the embedded clause) as well as an explanation for

why there is no N400 effect visible at the EMBEDDED NON-ISLAND gap (which is why there is an interaction of GAP and STRUCTURE at this position).

Under the integration hypothesis, we note that these N400 effects occur after both the gap and the verb it is (thematically/structurally) associated with have been encountered. If both of these linguistic elements are present, it could be argued that the N400 represents the semantic integration of the filler with its gap. There are two issues with such an interpretation, however.

The first issue with the integration hypothesis is based on the relative timing of this N400 with respect to the LAN. It is not necessarily problematic that the N400 is elicited at the verb in the matrix clause and at the gap position in the embedded clause since the order of the verb and gap position differ in these clauses (Table 6-14). What is problematic is that the LAN response is consistently found at the post-gap position (position 4 or 9). If the N400 reflects semantic integration of the filler and gap, it is unclear that the canonical view of the LAN indexing the working memory process of either retrieving or discharging the filler from memory can be sustained.¹² More troubling is that the LAN response is simultaneous with the N400 response in the matrix clause (post-gap position 4) while it is one position later than the N400 in the embedded clause (position 9). If the ‘integration’ N400 must await both the verb and gap to appear, why is the LAN response consistently relative only to the gap? That is, the commonality described above for the N400 effects under consideration is two elements (the gap position and verb) need to be encountered to elicit the N400, and the

¹² ‘Retrieving’ under a content-addressable similarity-based interference view of the process, and ‘discharging’ under an active storage/ constrained capacity view of the process.

N400 is elicited to the latter of those two elements. The LAN, however is always elicited at the post-gap position, irrespective of how whether that position is the verb itself or occurs after the verb.

The second issue with the integration hypothesis is that while there are four gaps across the experimental materials, there are only three N400s following the verb-gap or gap-verb pairs. Conspicuously missing is an ‘integration’ N400 in the embedded *that*-clause (EMBEDDED NON-ISLAND condition, 6.9 a). (6.9) presents the two EMBEDDED GAP conditions for reference.

(6.9 a) Who had the sailor assumed [that the captain befriended _ openly

(6.9 b) * Who had the sailor inquired [whether the captain befriended _ openly

If the N400 is indexing an integration cost, why isn’t there an integration cost for (6.9 a) at *openly*? It is plausible that the integration cost is larger inside an island than a non-island clause, so the effect in (6.9 b) would be larger than in (6.9 a). It would seem strange though that the long-distance filler-gap dependency in (6.9 a) shows no N400 integration cost while the short-distance matrix gaps (see Table 6-14) do. One possibility is that the ‘integration N400’ is not visible in the EMBEDDED NON-ISLAND (6.9 a) because it is obscured by other effects. Even though this ‘additional N400’ response was not predicted (and the materials were not designed in a way to maximize our ability to measure such a response), it should, in theory, be possible to obtain suggestive evidence for an ‘integration N400’ within the *that*-clause (6.9 a).

Recall that the ‘additional N400’ at position 8 was elicited over and above the N400 effect elicited by lexical differences between *the sailor* and *openly* at the same position. These lexical N400 differences are clearly identifiable because they were also elicited at position 3. It was argued that effects that occur in both positions should be interpreted as lexical effects, while any effect that occurred in only one position should be interpreted as being influenced by the surrounding sentence (i.e. a syntactic influence). We can use similar logic for testing the presence of these ‘integration N400s.’

If the ‘additional/non-lexical’ N400s are indexing the integration cost of two elements (both a gap and associated verb), then the N400 should be larger to a word (*_ openly*) when it represents the second element to be integrated. That is, in the embedded clause, an integration cost should be observable to *_ openly* (which completes the set of elements to be integrated: *befriended _ openly*) compared to *_ openly* in the matrix clause (which is still missing the matrix verb to be associated with). However, making this comparison, we observe the ‘additional’ N400 at *openly* when embedded in a *whether*-island (Figure 6-27 A) but not when embedded in a NON-ISLAND *that*-clause (Figure 6-27 B). Thus we still see no evidence for an ‘integration N400’ in the EMBEDDED NON-ISLAND condition (6.9 a).

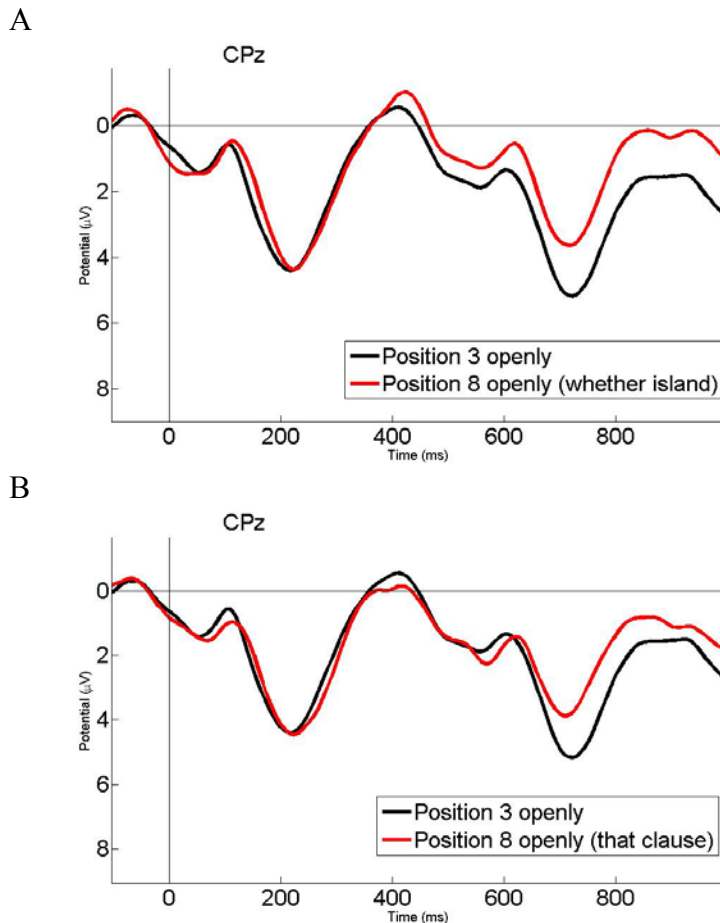


Figure 6-27: Matrix clause *_openly* (black trace) compared to embedded clause *_openly* (red trace) in a *whether-island* (A) and non-island *that-clause* (B)

While we see no evidence for the ‘additional N400’ embedded in a *that-clause* (figure 6-27 B), this could be due to differences in sentence position. Since the amplitude of the N400 response to open-class words decreases throughout the course of a sentence (Van Petten & Kutas 1990, 1991), this decline could be masking the ‘additional integration N400’ in the embedded *that-clause*. Again, the current experimental design was not constructed to test for this possibility, but we can examine how much the amplitude of the N400 elicited by *the sailor* decreases over the same exact positions in the sentence. As shown in Figure 6-28, the amplitude of the

N400 response to *the sailor* did not change between positions 3 and 8. This makes it less likely that the lack of a visible ‘additional’ N400 in Figure 6-27 B is due to sentence position effects.

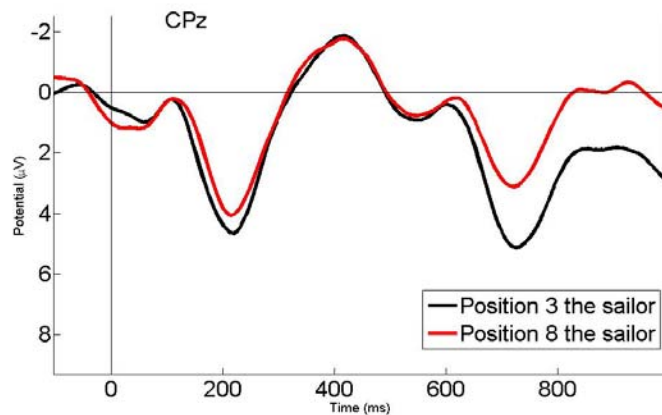


Figure 6-28: Comparison of N400 amplitudes of *the sailor* at positions 3 and 8

The lack of an ‘integration N400’ in response to the NON-ISLAND EMBEDDED GAP (the embedded *that* clause, 6.9 a) as well as the theoretical issues raised above combine to undermine the integration hypothesis for these N400 responses. We turn now to the predictability hypothesis, which does not predict an N400 effect in the NON-ISLAND EMBEDDED GAP.

Under the predictability hypothesis, the ‘additional N400’ effect that we see at the embedded gap would be due to this gap being less predictable in the ISLAND condition compared to the NON-ISLAND condition. This is because, upon encountering the island clause boundary (*whether*), the parser does not expect to encounter a filler within this clause, as this would be ungrammatical (e.g. Stowe 1986; Phillips 2006). I repeat (6.9) below for reference.

(6.9 a) Who had the sailor assumed [that the captain befriended _ openly

(6.9 b) * Who had the sailor inquired [whether the captain befriended _ openly

In (6.9 a), the parser first encounters the filler, *who*. The parser then predicts that it will soon encounter a gap that this filler can be associated with. Encountering *that* as a clause boundary does not alter this prediction. *Openly* is a cue to the parser that there is a gap present, and the parser proceeds with filler-gap association (indexed by the LAN, section 6.4.2.2.4).

In (6.9 b), the parser again encounters the filler and predicts a gap. However, upon encountering the *whether* clause boundary (an island) the parser modifies the prediction for a gap (Stowe 1986; Phillips 2006). A gap is less likely to occur within an island. Now, when *openly* (the cue that a gap is present) is encountered, this less predicted word elicits an N400.

Unlike the ‘integration n400’ account, which predicts that all four gaps should elicit an integration response, the predictability hypothesis requires that there be a difference between the EMBEDDED ISLAND condition (6.9 b) and the EMBEDDED NON-ISLAND condition (6.9 a), so there is no need to explain the lack of an ‘additional N400’ effect within the *that*-clause- there should not be one under the predictability hypothesis.

The challenge for the predictability hypothesis is explaining the N400s elicited at position 4, after the matrix gap. If these N400s do not reflect semantic integration

can they be explained by the predictability hypothesis? Since N400 amplitude correlates negatively with cloze probability (Kutas & Hillyard 1984), such an explanation would posit that the matrix verbs (*assumed / inquired*) are less predictable following *Who had openly* than they are following *Who had the sailor*. That is verbs like *assumed* should have a lower cloze probability in (6.10 a) than in (6.10 b).

(6.10 a) Predicted lower cloze: Who had openly assumed ... ?

(6.10 b) Predicted higher cloze: Who had the sailor assumed ... ?

To test this intuition, a pilot cloze completion task was conducted. Unfortunately, when presented with sentences like *who had openly...* and *who had the sailor...*, participants did not continue the sentences with sentential complements like the ones used in this study. Thus, the pilot study reveals a cloze probability of 0% for both conditions. While unfortunate for our present purposes, this is understandable, as the current materials are quite complex. What could be determined from the task, however, was that transitive verbs were used as a continuation following *who had the sailor...* much more frequently than after *who had openly...* (96% vs. 74%). While this may not appear to directly relate to the current sentences, we can see that participants overwhelmingly produced verbs that subcategorized for an additional NP argument following *who had the sailor...?* Following *who had openly...?*, the verb type is less predictable as it can also be an intransitive (e.g. *Who had openly cried?*).

In the current experimental materials, matrix verbs are used that subcategorize for a sentential complement, not only an additional NP. However, these sentential complement verbs are more similar to transitive verbs than they are to intransitives in that both sentential complements transitive verbs generate a prediction that there will be a noun phrase later in the sentence. This is relevant for sentences like (6.10 b), which are also expecting a noun phrase gap position. Thus these two predictions are in sync. Sentences starting like *who had the sailor... ?* generate a prediction for a verb that can either directly or indirectly (though an embedded sentential complement, for example) host a gap position. No such prediction is needed for sentences like *who had openly... ?*; without a strong prediction for a certain type of verb, when sentential complement verbs like *assume* or *inquire* are encountered, these unpredicted verbs elicit a larger N400 (6.10 a) compared to when those same verbs are more predicted by a different sentential context (6.10 b).

In summary, the data available from the current experiment favor a view of the ‘additional N400’ at the embedded gap position (6.9 b) as a response to the lower predictability of encountering a gap inside an island. There is also a difference in predictability when the matrix verbs follow an adverb (6.10 a: less predictable since no strong predictions are needed for the verb) compared to when these same verbs follow a noun phrase (6.10 b: more predictable since a verb is required that will allow for an NP gap). This view is favored over an integration (cost) account of the ‘additional N400’ for two reasons. First, it is not clear how the timing of an N400 integration cost would or should interact with the LAN’s retrieval/integration

response, and second, it is not clear why there should be no apparent effect of integration of the filler with the gap embedded inside the *that*-clause (6.9 a). The predictability view of the N400 avoids these issues.

Section 6.4.3.1.2 revisits this issue briefly when the ‘additional N400’ elicited at the EMBEDDED GAP position in *whether*-islands is shown to co-vary with reading span: the HIGH SPAN group has a more robust ‘additional N400’ effect while the lexically driven N400s at position 4 do not co-vary with reading span. This further supports the predictability view of the ‘additional N400.’

6.4.2.2.8 Summary

The two key findings from this section are (i) the robustness of the LAN effects elicited after each gap position, including the matrix gaps and the embedded gap in the *whether*-island violation condition (section 6.4.2.2.2) and (ii) the ‘additional N400’ elicited at the embedded gap position in the *whether*-island. This later effect is interpreted as a resulting from gaps being less predictable inside islands compared to non-island clauses (section 6.4.2.2.7).

Additionally, the expected interaction of GAP and STRUCTURE did not materialize at the clause boundary (section 6.4.2.2.6), nor was an observable sustained anterior negativity elicited by the current materials as expected (6.4.2.2.5). The classic interpretation of the positivity elicited by pre-gap positions as an index of syntactic integration (Kaan et al. 2000) is undermined by the fact that even the matrix pre-gap

position elicited a late positivity, i.e. when all conditions were still identical (*Who had...?*; see section 6.4.2.2.1). In the next section, we address the co-variation of these effects with the cognitive measures.

6.4.3 Median splits

In this section, I present and then discuss the findings of the ERP experiment including the scores from the cognitive measures.

6.4.3.1 Results

The following sections present the results for the median split analyses. In the following analyses, the data were examined for interactions between at least one of the median split groups and at least one of the linguistic manipulations (GAP and/or STRUCTURE). Effects that did not include a significant effect of median split group are not reported below (see section 6.4.2 for results without median splits).

No significant results (beyond the basic effects reported in section 6.4.2) were found for the LAN (positions 4, 9) or the sustained negativity originating with these LANs. Unlike in the self-paced reading experiment (Chapter 5, section 5.4.3.1.1) we found no effect of reading span at the clause boundary (but see section 6.4.2.2.5 for discussion of the lack of a clause boundary effect in the ERP experiment). Three co-variation effects were found between the linguistic manipulations and cognitive measures. The broad positivity/P600 elicited prior to the embedded gap showed an

interaction of N-BACK group with STRUCTURE (section 6.4.3.1.1). The ‘additional’ non-lexical N400 effect at the embedded gap site was larger in the HIGH SPAN group and smaller in the LOW SPAN group (section 6.4.3.1.2). Finally, the sentence-final negativity also showed different patterns in the HIGH and LOW SPAN groups.

6.4.3.1.1 Position 7 (*befriended*): Interaction of STRUCTURE x N-BACK group

Recall that the embedded clause pre-gap position elicited a broad positivity from 600-900 msec (section 6.4.2.1.5). The median split ANOVA analyses revealed a significant interaction of N-BACK group and STRUCTURE ($F(1,30) = 4.67, p = 0.039$). The HIGH N-BACK group did not differentiate between the ISLAND condition (2.35 μV) and the NON-ISLAND condition (2.23 μV). The LOW N-BACK group, however, showed a more positive response to the NON-ISLAND condition (3.35 μV) than to the ISLAND condition (2.69 μV ; $t(1621.97) = -5.2, p < 0.001$). The LOW N-BACK group’s response to the NON-ISLAND condition was also significantly greater than the HIGH N-BACK group’s response to the same condition ($t(1824.73) = 9.19, p < 0.001$). There were no significant interactions with ELECTRODE. These values are plotted in Figure 6-29.

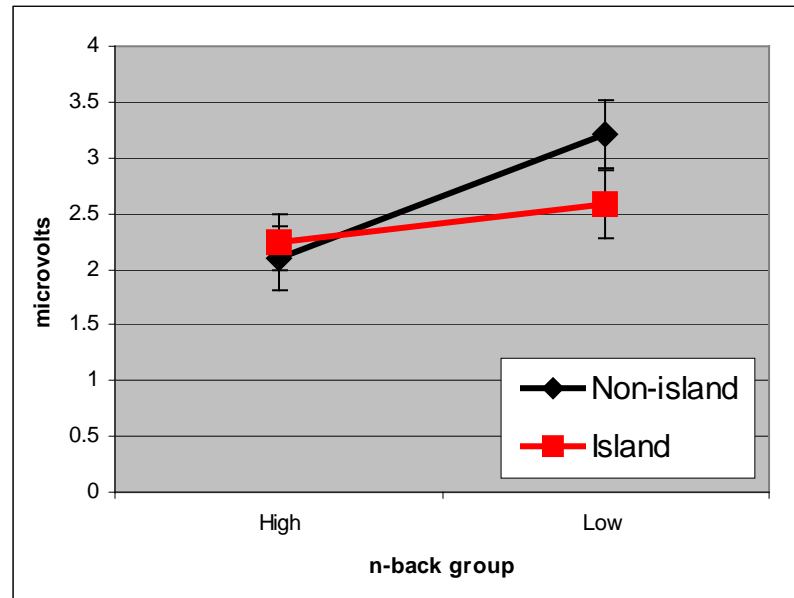


Figure 6-29: Position 7 (*befriended*) STRUCTURE x N-BACK group mean scalp voltage (600-900 msec). Error bars denote standard error

6.4.3.1.2 Position 8 (*the sailor* / *_openly*)

Recall that the embedded clause gap position elicited two types of responses. First, there were the lexical differences between *the sailor* and *openly*, which resulted in more negative LAN and N400 responses *to the sailor*. Second, there was an ‘additional’ N400 effect when comparing *openly* to *openly*. The additional N400 was elicited in the EMBEDDED ISLAND condition compared to the EMBEDDED NON-ISLAND condition (for all effects see section 6.4.2.1.6).

In the 300-600 msec time window the omnibus median split ANOVA analysis revealed a significant interaction of SPAN group and ELECTRODE ($F(28,840) = 3.59, p = 0.012$), a significant interaction of GAP and ELECTRODE ($F(28,840) = 3.84, p <$

0.001), and a marginal interaction of SPAN group, GAP and ELECTRODE ($F(28,840) = 1.73, p = 0.099$). The distributional analysis revealed a significant interaction of SPAN and ANTERIORITY over the midline and an interaction of SPAN and GAP over medial sites (Table 6-15). There were no significant effects over lateral sites.

Table 6-15: Position 8: *the sailor* / *_openly* (300-600)

Analysis:		<i>F</i>	<i>p</i>	
Omnibus:	SPAN x ELECTRODE	$F(28,840) = 3.59$	$p = 0.012$	*
	GAP x ELECTRODE	$F(28,840) = 3.84$	$p < 0.001$	***
	SPAN x GAP x ELECTRODE	$F(28,840) = 1.73$	$p = 0.099$.
Midline:	SPAN x ANTERIORITY	$F(6,180) = 4.59$	$p = 0.012$	*
Medial:	SPAN x GAP	$F(1,30) = 4.47$	$p = 0.043$	*
Lateral:	--	--		

Since both LAN and N400 effects were previously reported for this site (section 6.4.2.1.6), the quadrant and center post-hoc analyses were used to determine if the SPAN interaction was occurring predominantly with the LAN effect, the N400 effect or both (Table 6-16). The quadrant analysis revealed no statistically significant differences. As can be seen in Figure 6-30 B, the effect in question was concentrated over central regions of the scalp for the HIGH SPAN group. As the central electrodes were not used in the quadrant analysis, it is unsurprising that no effects were found to be significant in it. However, the center analysis revealed an interaction of SPAN x STRUCTURE x ANTERIORITY x LATERALITY ($F(8,240) = 3.27, p = 0.008$) and a marginal interaction of SPAN x GAP x STRUCTURE x ANTERIORITY x LATERALITY ($F(8,240) = 1.93, p = 0.065$). The lack of lateral effects in the distributional analysis and the failure to find significant results in the quadrant analysis indicate that SPAN was not interacting with the LAN effect.

Table 6-16: Position 8 post-hoc (*the sailor* / *_openly*) 300-600 msec window

Analysis:		<i>F</i>	<i>p</i>	
Quadrant:	--	--		
	SPAN x STRUCTURE x ANTERIORITY x LATERALITY	<i>F</i> (8,240) = 3.27	<i>p</i> = 0.008	**
Center:	SPAN x GAP x STRUCTURE x ANTERIORITY x LATERALITY	<i>F</i> (8,240) = 1.93	<i>p</i> = 0.065	.

The interaction with SPAN is due to the HIGH SPAN group showing a robust ‘additional N400’ effect over central regions of scalp, where the response to the EMBEDDED ISLAND is more negative than that to the EMBEDDED NON-ISLAND¹³ (Figure 6-30 A, B), while the LOW SPAN group shows a much smaller (Figure 6-30 C) and less centralized (Figure 6-30 D) effect. In the HIGH SPAN group, the EMBEDDED ISLAND condition is significantly more negative than the EMBEDDED NON-ISLAND ($t(597.61) = -3.64$ $p < 0.001$) averaged over all electrodes used for the center analysis. This same comparison was not significant in the LOW SPAN group ($p = 0.4$).

¹³ Recall that this is independent of the lexical effect (solid vs. dashed lines in Figure 6-30 A, C; see section 6.4.2.1.6)

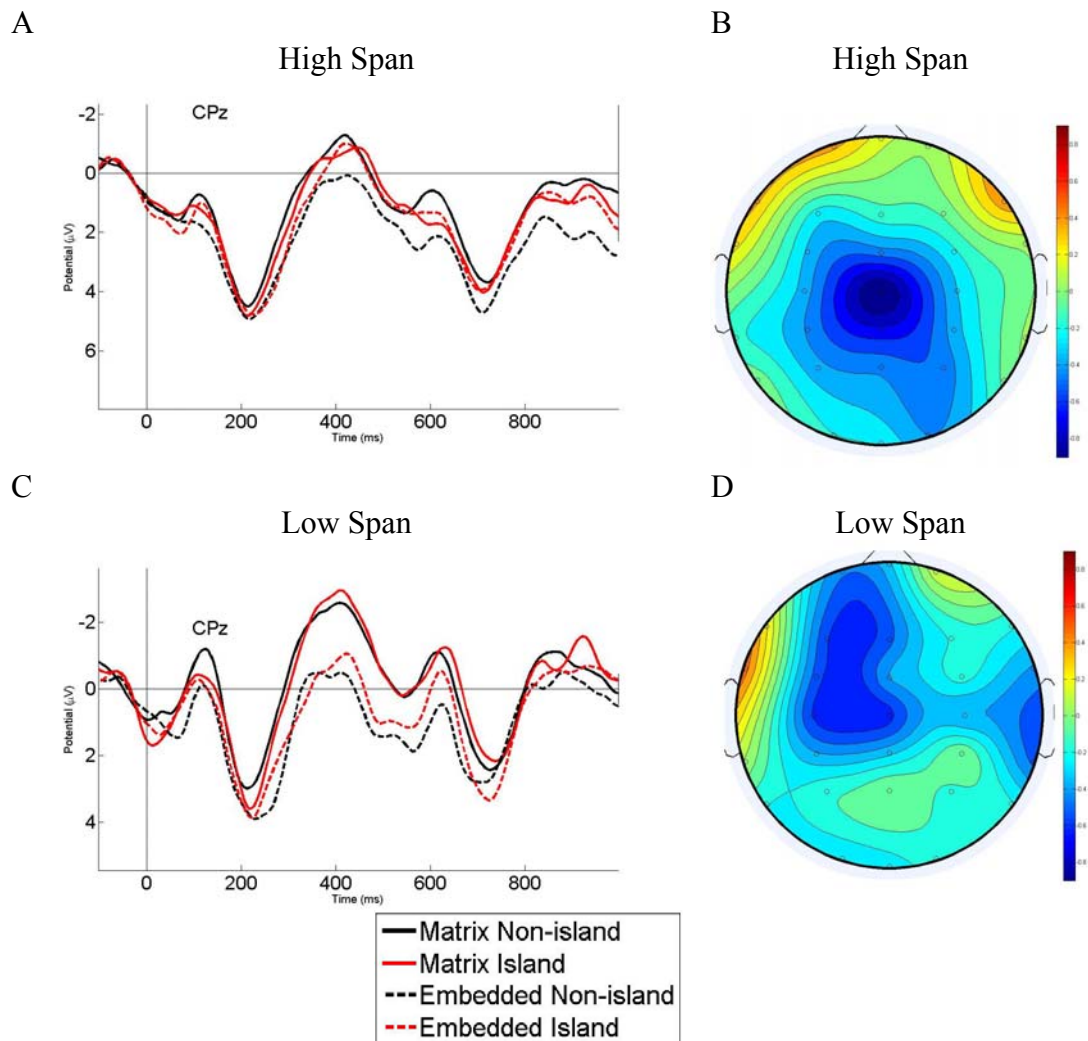


Figure 6-30: Position 8: *the sailor / openly* GAP x STRUCTURE interaction at CPz in high (A) and low (C) span groups and topographic isovoltage map showing EMBEDDED ISLAND (*_openly*) – EMBEDDED NON-ISLAND (*_openly*) from 300-600 msec in high (B) and low (D) span groups. (Compare Figure 6-16)

6.4.3.1.2 Position 12 (*hearing?*) sentence final negativity

Recall that in section 6.4.2.1.8, an overall negativity was reported for the EMBEDDED GAP conditions (long-distance filler-gap dependencies) compared to MATRIX GAP conditions (short distance filler-gap dependencies) at the sentence final

position (*hearing?*). It was predicted that the *whether*-island violation condition would elicit a sentence-final N400 (6.2.4), but this was not the case. When the same region (300-600 msec) was submitted to the median split ANOVAs, a significant interaction of SPAN x ELECTRODE was found ($F(28,840) = 2.87, p = 0.026$), as well as a marginal interaction of SPAN x GAP x ELECTRODE ($F(28,840) = 2.21, p = 0.099$). Although this later effect was only marginal, we decided to explore further, applying the distributional analysis, because (i) we originally predicted an effect at this location and (ii) since this was the sentence-final position, it contained more noise than other positions, as participants were no longer waiting for another word (the end of the sentence was marked with a question mark).

Table 6-17: Position 12 (*hearing?*) 300-600 msec window

Analysis:		<i>F</i>	<i>p</i>	
Omnibus:	SPAN x ELECTRODE	$F(28,840) = 2.87$	$p = 0.026$	*
	SPAN x GAP x ELECTRODE	$F(28,840) = 2.21$	$p = 0.099$.
Midline:	SPAN x STRUCTURE x ANTERIORITY	$F(6,180) = 2.11$	$p = 0.053$.
Medial:	--	--		
Lateral:	--	--		

The distributional analysis revealed an interaction of SPAN x STRUCTURE x ANTERIORITY along the midline that just missed significance ($F(6,180) = 2.11, p = 0.053$). The medial and lateral analyses revealed no significant findings. The central nature of the effect was further supported by a marginal SPAN x STRUCTURE x ANTERIORITY x LATERALITY interaction in the center analysis (Table 6-18).

Table 6-18: Position 12 post-hoc (hearing?) 300-600 msec window

Analysis:		<i>F</i>	<i>p</i>
Quadrant:	--	--	
Center:	SPAN x ANTERIORITY	$F(4,120) = 3.45$	$p = 0.053$
	SPAN x STRUCTURE x ANTERIORITY x LATERALITY	$F(8,240) = 1.87$	$p = 0.099$

There was a more negative response between 300 and 600 msec in the HIGH SPAN group over midline electrodes to the EMBEDDED GAP conditions (1.35 mV) compared to the MATRIX GAP conditions (2.02 mV; $t(351.6) = 2.88$, $p = 0.004$). The morphology of the waveform in Figure 6-31A suggests an N400 response. The LOW SPAN group showed a similar trend, though it was more anterior (Figure 6-32B) and was not statistically significant (EMBEDDED: 2.8 mV, MATRIX 3.23 mV; $t(252.65) = 1.17$, $p = 0.24$).

Additionally, the HIGH SPAN group showed a nonsignificant trend ($p = 0.22$) over the same midline electrodes, where the ISLAND conditions were more negative than the NON-ISLAND conditions. The LOW SPAN group did not show this trend (compare Figure 6-32C, D).

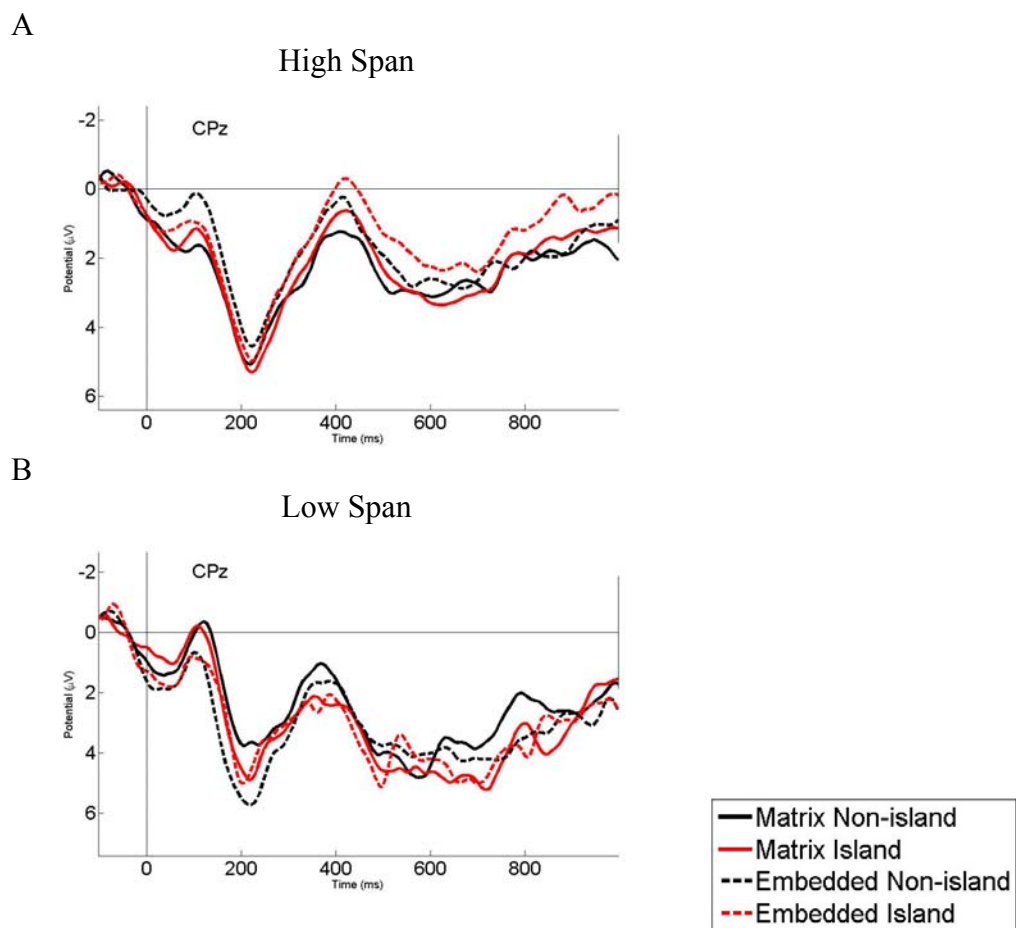


Figure 6-31: Position 12 (*hearing?*) potential N400 responses at Pz for high (A) and low (B) span groups

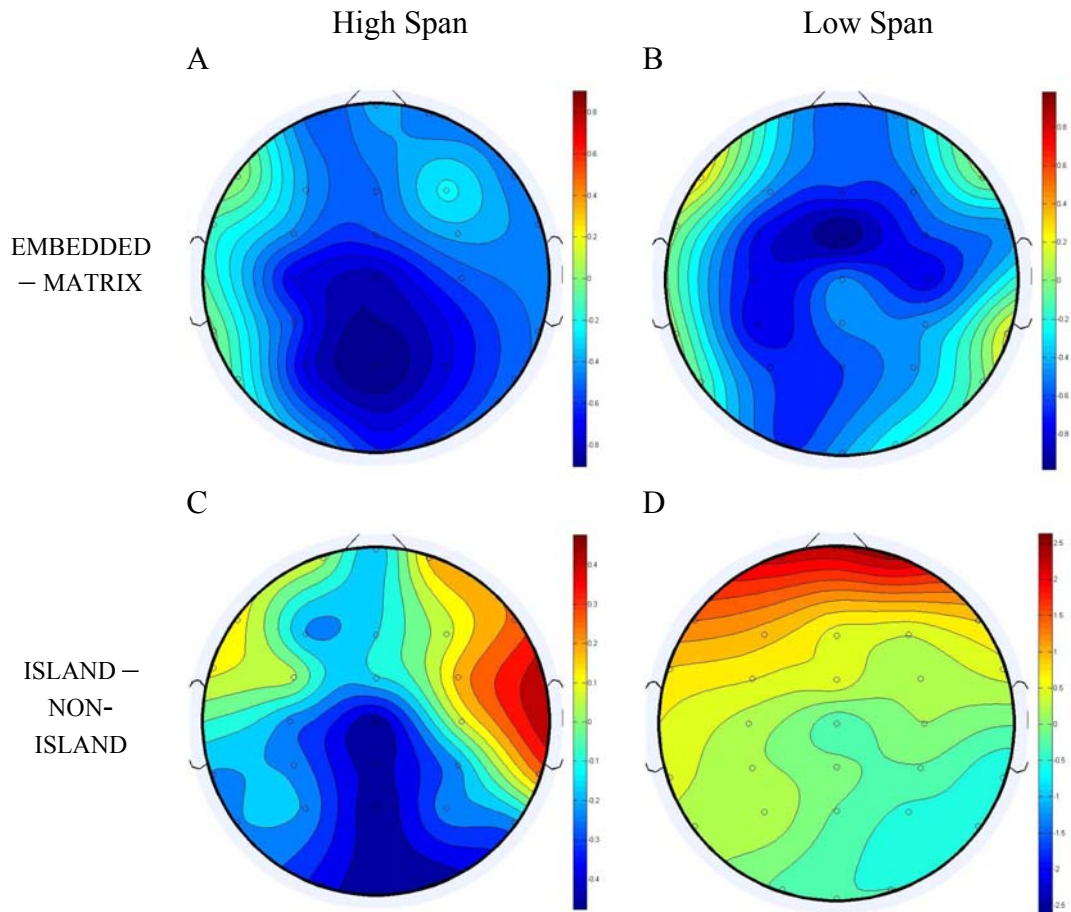


Figure 6-32: Position 12 (*hearing?*) topographic isovoltage map showing 300-600 msec. High (A,C) and low (B,D) span groups showing EMBEDDED – MATRIX (A,B) and ISLAND – NON-ISLAND (C,D).

6.4.3.2 Discussion

Before discussing the findings below, it may be useful to first note what we did not find. We did not find any co-variation of the LAN or the ongoing ‘lingering LAN’ with the cognitive measures. This may seem strange considering the close association the LAN has with working memory processes (CH 2, section 2.2.5.2.1). However, also note that in the current experiment, the amplitude of the LAN does not vary according

to the STRUCTURE manipulation (ISLAND / NON-ISLAND). The LAN appears steady and robust, varying neither with cognitive measures nor with linguistic manipulations other than the presence of a gap. This presents a view of the LAN as involved with associating a filler and gap, which presumably must include some kind of working memory component (though it is not clear from only this pattern if this component is the release of the filler from storage or the cue-based retrieval of the filler from recent memory). In any case, the LAN is not being modulated. It appears more like an on/off light switch rather than a dimmer. The intensity of the LAN does not vary with either the amount of ‘light’ needed or the ‘capacity’ of the circuit. Instead the LAN is ‘flipped on’ when the parser needs to ‘see’ the filler at the gap position. To continue the metaphor, the ‘lingering LAN’ from section 6.4.2.2.6 indicates that the light is left on, even after the need for it has ended.

6.4.3.2.1 Position 7 (*befriended*)

Analysis of position 7 (the embedded pre-gap position) revealed an interaction of STRUCTURE and N-BACK group, with LOW scorers showing a larger positivity in the 600-900 msec window in the NON-ISLAND conditions compared to ISLAND conditions. The HIGH scorers showed no such difference. Note that this interaction of STRUCTURE and N-BACK group occurred only in the embedded clause pre-gap positivity (position 7, *befriended*), and not in the matrix clause pre-gap positivity (position 2, *had*). As noted in section 6.4.2.2.1, it is problematic for the syntactic integration difficulty view

of the P600 (Kaan et al. 2000) that the pre-gap position in the matrix clause elicited late positivity before the conditions could be differentiated. Instead, we take a more conservative view of the pre-gap P600, namely that it indexes gap identification (section 6.4.2.2.1).

The modulation of the embedded pre-gap positivity does not provide evidence that would warrant a change in our position. Note especially that the co-variation of N-BACK and STRUCTURE reported here does not interact with GAP POSITION. That is, unlike the pre-gap positivies discussed in section 6.4.2.2.1, this positivity occurs both before the gap position *_ openly* (6.11 c) and its control *the sailor* (6.11 d) in ISLAND conditions.

(6.11 a) Who had the sailor assumed

[that the captain befriended *_ openly* ... ?

(6.11 b) Who had *_ openly* assumed

[that the captain befriended the sailor ... ?

(6.11 c) Who had the sailor inquired

[whether the captain befriended *_ openly* ... ?

(6.11 d) Who had *_ openly* inquired

[whether the captain befriended the sailor ... ?

This effect cannot be interpreted as either a filler-gap integration cost (Kaan et al. 2000) or gap identification (section 6.4.2.2.1) since it is elicited when no gap

follows it (6.11 c). It is possible that the LOW N-BACK group finds the verb *befriended* to be less preferred within an island structure (independent of gap location), but it is unclear why this would be. P600 effects have been reported to garden path sentences (e.g. Osterhout & Holcomb 1992; Osterhout et al. 1994), but there does not appear to be a reason to reanalyze (6.11 c) and (6.11 d) but not the corresponding (6.11 a) and (6.11 b).

As this effect is not straightforwardly interpretable, and considering the many analyses done for the median split ANOVAs, none of which in this or the self-paced reading experiment (Chapter 5) produced an interaction with n-back scores, we will not pursue it further here. It is possible that this effect is due to noise and/or a spurious result. Presentation and discussion of this data is retained here for completeness.

6.4.3.2.2 Position 8 (*the sailor* / *_openly*)

Recall that at position 8, the basic analysis (i.e. not including cognitive measures, section 6.4.2.1.6) revealed both lexical differences between *the sailor* and *openly* as well as an ‘additional N400’ that we argued reflects the low predictability of a gap inside a *whether*-island (section 6.4.2.2.7). The analysis including cognitive measures finds co-variation only with that ‘additional N400,’ i.e. the predictability effect.

This ‘additional N400’ effect was more robust in the HIGH SPAN group than the LOW SPAN group. Under the predictability interpretation of this ‘additional N400’

effect, this would indicate that HIGH SPAN readers are more sensitive to the *whether*-island clause boundary - specifically in recognizing that they are entering an island domain - than the LOW SPAN group. After having encountered the *whether*-island clause boundary, the expectations of HIGH SPAN readers for encountering a gap within the island appear to be lower than those of LOW SPAN readers; in other words, the gap is less predictable for HIGH SPAN readers. When the gap appears within this island domain nonetheless, a larger N400 response is elicited in the HIGH SPAN group precisely because they appear to be more sensitive to the fact that gaps are unlikely to be found within islands. The LOW SPAN group, on the other hand, appear not to be as sensitive to this fact (as a group) and so do not substantially alter their predictions about whether a gap should or should not appear inside an island domain. The LOW SPAN group is less surprised by the presence of the island gap because they are apparently not as sensitive to the information provided by the clause boundary, namely that they were entering an island domain and that gaps are unlikely to be found there. Note that there are actually three separate possible differences in processing: (i) the LOW SPAN group is less aware of the island boundary (and thus does not alter its predictions for the presence of a gap), (ii) the LOW SPAN group is less aware/has less experience with the fact that gaps are unlikely to occur in an island domain (that is, they are aware of the island boundary, but they do not appreciate what that implies for the filler-gap dependency), and (iii) they may be aware of both the island boundary and the unacceptability/low likelihood of finding a gap in an island, but they are less

able to adjust to and act on this information in real time. Interestingly, there are data from other sections of this dissertation that may bear on this distinction.

Recall that in the self-paced reading experiment (Chapter 5), while there was co-variation of reading times at the clause boundary with SPAN group, the variation occurred only with how the HIGH and LOW SPAN groups read the *whether* boundary in the MATRIX GAP condition (i.e. when there was no outstanding filler-gap dependency, and therefore what followed would not constitute an island violation). For the island violation condition (the same condition for which we are considering the ‘additional N400’ effect in the ERP experiment), there was no difference discernable between the HIGH and LOW SPAN readers (section 5.4.3.1.1). It would appear then that both HIGH and LOW SPAN readers are aware of the island boundary in real time (contra option (i) in the previous paragraph).¹⁴

Recall also that participants completed an acceptability judgment task following the ERP experiment. Participants still rated the *whether*-island violations as the most unacceptable, even after prolonged exposure to them in the experiment. There were no differences in these acceptability scores between the HIGH and LOW SPAN groups. This indicates that both groups are able to recognize that gaps within *whether*-islands are unacceptable in an off-line measure (contra option (ii) above).

If the LOW SPAN group matches the HIGH SPAN group both in how they (i) process the clause boundaries of *whether*-island violation sentences and (ii) rate the overall acceptability of *whether*-island violation sentences, then it seems that the most

¹⁴ This reading time pattern was not reflected in the ERP responses at this point, however. See section 6.4.2.2.5 for discussion.

likely point of divergence between the HIGH and LOW SPAN groups with respect to the ‘additional N400’ response is in (iii) how they are able to use the information that the clause boundary provides *in real time* to update and alter their expectations.

The LOW SPAN group demonstrates that in the off-line measure, they recognize that gaps in *whether*-islands are unacceptable. However, this pattern does not seem to inform the on-line processing of the sentence as much as it does for the HIGH SPAN group. That is, the LOW SPAN group does not appear to alter its predictions for where the gap will be located in the face of other processing demands (e.g. the computational costs of processing incoming material as the sentence continues). Once these processing demands are completed at sentence end, however, the LOW SPAN group can use the information provided by the clause boundary in the same way as the HIGH SPAN group, and rate the acceptability of the sentences similarly.

Thus we see a pattern in the co-variation data that is consistent with the predictability account of the ‘additional N400,’ as well as the reading time and offline acceptability data. Let us briefly compare how the same data would appear in light of the integration hypothesis discussed in section 6.4.2.2.4.

Under the integration account, the ‘additional N400’ effect in the *whether*-island domain reflects more difficulty integrating the filler with the gap. Why would the HIGH SPAN group be having more difficulty with this? It would be much more intuitive if the LOW SPAN group showed a larger effect, because they would need to commit more resources/effort for this integration process. In order to reconcile this, we could claim that it is only the HIGH SPAN group that is actually integrating the filler

with the gap inside the island. The LOW SPAN group either does not recognize that this is a possibility or does not put in the effort to attempt this integration (since it is beyond their ability). This type of account could potentially align well with a capacity-constrained processing account of island phenomena (Kluender 1991, 1998; Kluender and Kutas 1993a,b; Hofmeister 2008; Sag et al. 2007) in that it is only those with more cognitive resources who can successfully process the island violations. While this line of explanation may be promising in isolation, it is not consistent with other data we have available.

First, if the HIGH SPAN, and not the LOW SPAN, group is able to integrate the filler with the island-internal gap, then why doesn't the HIGH SPAN group rate these sentences as more acceptable? Thus, even if there is a difference in processing, it does not appear tenable that this processing difference is responsible for the low acceptability ratings given to island violation sentences.

Second, it is unclear how to reconcile this integration N400 view, according to which only the HIGH SPAN group is integrating the filler and gap inside the island, with the fact that a LAN effect is elicited at the post-gap position in both groups, within both a *whether*-island and a *that*-clause. If only the HIGH SPAN group is integrating filler and gap, then presumably the LAN effect should be seen only in the HIGH SPAN group. This is a special case of the 'relative timing' difficulty already discussed in section 6.4.2.2.4, namely that if the 'additional N400' is taken to be reflective of integration of the filler and the gap, then it is unlikely that the LAN can be assumed to

have its canonical interpretation of reflecting a working memory process associated with fillers and gaps.

That the LAN is consistently elicited in both *whether*- and *that*-clauses in both reading span groups is not a problem for the predictability hypothesis, since it makes no claims about whether the filler and gap will be associated (canonically indexed by the LAN), only how predictable the gap is. Thus, both the LOW SPAN group (which does not alter its predictions for a gap) and the HIGH SPAN group (which presumably has the resources to associate a filler even with an unpredicted gap when necessary) both show the LAN response following the gap. Again, the predictability hypothesis appears more consistent with the data than the integration hypothesis.

6.4.3.2.3 Position 12 (*hearing?*)

Recall that we predicted (section 6.2.7) a larger N400 to the sentence-final word in the EMBEDDED ISLAND condition (the island violation), as it has previously been reported that ungrammatical sentences with mid-sentence morphosyntactic violations elicit a sentence-final N400 (e.g. Osterhout & Holcomb 1992). The basic analysis revealed only a broad negativity to the EMBEDDED GAP condition with no statistically significant distributional findings (section 6.4.2.1.8).

The median split analysis additionally revealed an interaction with SPAN group. In addition to the negativity elicited by the EMBEDDED GAP conditions, the ISLAND conditions elicited a centro-posterior negativity compared to the NON-ISLAND

conditions only in the HIGH SPAN group. While the expected interaction of GAP x STRUCTURE (x SPAN) for the sentence-final N400 was still not statistically significant, the EMBEDDED ISLAND (island violation) condition elicited the most negative response for the HIGH GROUP at least.

While many of the effects for this position are only marginally significant, together they suggest that the HIGH SPAN, but not the LOW SPAN, group is exhibiting a pattern like the predicted sentence-final N400 to the island violation. It appears that the HIGH GROUP is showing an end of sentence effect based on immediately recognizing the island violation (EMBEDDED ISLAND condition) as unacceptable. However, note that both the HIGH and LOW SPAN groups rate the acceptability for these sentences in the same way, including rating the island violation as the least acceptable (Chapter 4; Chapter 6, section 6.4.1). So while the LOW SPAN group may not immediately exhibit a sentence-final N400 response to the least acceptable sentence, they still eventually rate the island violation as least acceptable. Thus, like the ‘additional N400’ effect (section 6.4.3.1.2), while LOW SPAN readers do not seem to be processing as quickly/efficiently as the HIGH SPAN readers, they still arrive at the same acceptability judgments.

6.5 Summary

The first key finding from this chapter is the discovery of an N400 effect elicited by the embedded gap position only in the island violation condition (section

6.4.2.2.7). We argued that this ‘additional N400’ effect was a predictability effect rather than an integration cost effect (sections 6.4.2.2.7, 6.4.3.1.2). This prediction N400 was most robust in the HIGH SPAN group (section 6.4.3.1.2). This was interpreted as the HIGH SPAN group being able to alter their predictions of whether a gap should occur in an island domain as soon as they encountered the *whether*-clause boundary. The LOW SPAN group (based on the self-paced reading data) also showed sensitivity to the *whether*-clause boundary, and (based on off-line acceptability ratings) found gaps within *whether*-islands to be unacceptable, but were not as able to act on this information under the time pressures of the incoming sentence materials to adjust their predictions as to whether a gap should occur or not. This modulation of predictability did not appear to influence the second key finding of the chapter, namely the consistent elicitation of LAN effects at post-gap positions (section 6.4.2.2.4). LAN effects were elicited following both matrix and embedded clause gaps, even within an island violation. This represents (i) the first ERP evidence we are aware of for filler-gap association in a matrix clause subject gap, and (ii) evidence that filler-gap association can occur within an island domain just as easily as in a non-island control sentence.

In addition to these two findings, we reported two unexpected results that deserve additional investigation. First, a P600 effect was elicited by positions preceding both the matrix and embedded clause gaps, even when the conditions had not diverged in the matrix clause, raising issues about the interpretation of pre-gap P600 effects (section 6.4.2.2.1). Second, the post-gap LAN effects were followed by a

sustained negativity in both matrix and embedded clauses. This raises questions about the interpretation of sustained negativities previously reported following fillers (section 6.4.2.2.5).

6.6 Conclusion

Perhaps the most striking result of the ERP experiment was the fact that the island violation condition did not elicit a differential response in either of the two ERP components traditionally associated with syntactic processes (i.e. the P600 or LAN), but instead elicited an N400 at the embedded gap position. Discussed in more detail throughout section 6.4.2.2, this will not be rehashed here except to highlight two points.

First, there is no large, syntactic ‘ungrammaticality’ response to the *whether*-island violations, either at the clause boundary or elsewhere. Yet the participants still rate these sentences as the most unacceptable. One would expect, based either on a grammatical account for islands (Chapter 2, section 2.3.1) or Kluender’s capacity-constrained processing account of islands where the processing load is largest at the clause boundary (Kluender 1991, 1998; Kluender & Kutas 1993a,b; Chapter 2, section 2.3.3.1) that there would be a large brain response, especially in a component previously associated with syntactic or working memory processes (i.e. a P600 or LAN). Instead, the unacceptability of *whether*-islands, at least in terms of what causes processing difficulties, appear to be an accumulation of more subtle effects, such as in the view of Ross (1987). These subtle effects are not enough to elicit substantially

different brain responses from those to parallel control sentences (long-distance filler-gap dependencies into non-island clauses, for example). Instead we simply see a difference in predictability: gaps are less predictable in *whether*-islands. Even so, when that less predictable gap is encountered, the parser still associates the filler with it, as evidenced by the post-gap LAN response in all four conditions. *Whether*-islands are unexpected, but not unprocessable.

Second, the lack of a large processing difficulty for the *whether*-island violations means that we do not have a clear test of the capacity-constrained and similarity-interference accounts of islands (Chapter 2, section 2.3.3). The self-paced reading experiment (Chapter 5) provided initial evidence for the capacity-constrained view of working memory as applied to island violations, but a similar result was not found here (see section 6.4.2.2.6 for methodological explanation). Neither was there found to be greater difficulty in associating the filler with the gap inside a *whether*-island than in an acceptable *that*-clause (i.e. modulation of the LAN effect), as one might expect if the difficulty in processing an island violation lay in the retrieval of the filler. Instead we see a difference in the pattern of predictive processing, which is independent from and compatible with both views of working memory, as well as with both views of the processing difficulty in *whether*-islands. Drawing on data from across all three experiments, I present this as the ‘gap predictability account of islands’ in the final chapter of the dissertation, in section 7.5.

Chapter 7: Discussion and Conclusion

This dissertation tested the capacity-constrained and similarity-interference accounts of islands phenomena using three different methods (acceptability judgments, self-paced reading and ERPs) to examine *whether*-islands and carefully constructed control sentences. The experiments presented in the previous chapters have resulted in a rich and complex dataset. Interpreting the combined results of these experiments have implications broader than those available from any of the experiments individually.

In order to discuss these implications, I first briefly review the previous chapters, highlighting the innovations and major findings from each. These innovations and findings in turn inform the discussion that follows on the possibility of re-interpreting some classical ERP components (section 7.2), updating working memory models in sentence processing in general (section 7.3) as well as processing accounts of islands in particular (section 7.4). The ‘gap predictability account of processing islands’ conceptualizes the findings of all three experiments (section 7.5). Questions are raised about the relationship between the parser and grammar (section 7.6) and about how to best leverage the current findings for future research (section 7.7). Section 7.8 concludes the chapter.

7.1 Dissertation summary

This dissertation is focused on examining *whether*-islands as a phenomenon to leverage an individual differences approach in order to better understand working memory in sentence processing. It is important to note that, contrary to impressions of *whether*-islands as ‘weak’ or less strict than other types of islands, acceptability experiments have clearly and consistently found that *whether*-island violations are judged the least acceptable compared to relevant controls (Sprouse, Wagers and Phillips 2012; this dissertation Chapter 4, Chapter 6). Participants’ experimental acceptability judgments are as robust as judgments to other island types (e.g. Sprouse, Wagers and Phillips 2012). Thus, in the summary that follows, while the specific examples from this dissertation are *whether*-islands, it is predicted that the pattern of results will hold for other island types and possibly even for other types of long-distance dependencies.

7.1.1 Introduction of the similarity-interference account of islands

In Chapter 2, I presented the basic data and various accounts of island phenomena. The key focus was on comparing the capacity-constrained account of islands (e.g. Kluender 1991, 1998; Kluender and Kutas 1993a,b; Hofmeister 2008; Sag et al. 2007), based on the Just and Carpenter (1992) model of working memory, with my newly introduced similarity-interference account of islands, based on prior

research on the importance of similarity-based interference in sentence processing (e.g. Gordon, Hendrick and Johnson 2001; Gordon, Hendrick and Levine 2002; Lewis and Vasishth 2005; Gordon et al. 2006; Lewis, Vasishth and Van Dyke 2006; Van Dyke and McElree 2006). These approaches to the processing of islands make different predictions for where the locus of greatest processing difficulty in an island violation should occur (capacity-constrained: the clause boundary; similarity-interference: the embedded gap position), and what cognitive measures are expected to co-vary with the linguistic data (capacity-constrained: reading span; similarity-interference: memory lure).

7.1.2 Development of a balanced factorial design

In Chapter 3, I presented the design of the stimulus sentences used in the three experiments reported in the dissertation. The factorial design allows for the examination of effects due to the distance between the filler and gap (GAP POSITION), whether the clause boundary is a potential island (STRUCTURE), and the interaction of the two (most importantly in the island violation condition). Additionally, the use of adverbial ‘placeholders’ at the gap position allowed for controlled comparisons across conditions (Table 7-1).

Table 7-1: Sample stimulus set. Manipulation of STRUCTURE indicated in bold. Manipulation of GAP indicated by italics. No specific claims are intended by the placement of the gap, which is meant only to indicate the on-line point of disambiguation of the gap position.

		STRUCTURE	
		NON-ISLAND	ISLAND
GAP	MATRIX	Condition 1: Who had <i>_ openly</i> assumed [that the captain befriended <i>the sailor</i> before the final mutiny hearing?]	Condition 2: Who had <i>_ openly</i> inquired [whether the captain befriended <i>the sailor</i> before the final mutiny hearing?]
	EMBEDDED	Condition 3: Who had <i>the sailor</i> assumed [that the captain befriended <i>_ openly</i> before the final mutiny hearing?]	Condition 4: Who had <i>the sailor</i> inquired [whether the captain befriended <i>_ openly</i> before the final mutiny hearing?]

This novel design allowed for both the pre-gap and post-gap positions to be matched, which permitted the ERP experiment to measure both positions without a lexical confound. Many prior experiments controlled only for one of these positions (e.g. Kluender & Kutas 1993a,b; King & Kutas 1995; Kaan et al. 2000; Phillips et al. 2005), preventing both a pre-gap P600 effect and post-gap LAN from being examined in the same experiment.

7.1.3 Advances in acceptability/processing frameworks: The Cognitive Co-variation Intuition and the Processing Benefit Schedule

In Chapter 4, I presented the results of an acceptability judgment study that tested for the possible co-variation of participants' responses with various cognitive measures. I first reviewed the findings of Sprouse, Wagers and Phillips (2012), which

also tested the acceptability of sentences similar to those in Table 7-1 for co-variation with cognitive scores. Hofmeister, Staum-Casasanto and Sag (2012a,b) raised a number of issues with the Sprouse, Wagers and Phillips (2012) study and their interpretation of its results. The experiment and analysis in Chapter 4, while undertaken independently from Sprouse, Wagers and Phillips (2012), represents improvements on the Sprouse, Wagers and Phillips (2012) study in terms of their choice of cognitive measures, interpretation of null results, interpretation of R^2 values, and reliance on DD scores. The findings and analysis of Chapter 4 largely replicate those of Sprouse, Wagers and Phillips (2012), in that no convincing evidence was found that supports the capacity-constrained processing account of islands. However, this should not in itself be taken as evidence against such an account. In this chapter I also described three conceptual devices that are important to consider when interpreting results based on individual cognitive differences: the Cognitive Co-variation Intuition (CCI, Michel 2013), rating task differences, and the Processing Benefits Schedule (PBS, this dissertation, Chapter 4).

The CCI, repeated in (7.1), represents a more developed and clearer specification of the basic intuition that led to the current line of research. The initial (basic) intuition is that if islands are best explained with a working-memory-based processing account (e.g. Kluender 1991, 1998; Kluender and Kutas 1993a,b; Hofmeister 2007; Sag et. al. 2007) then we should see high working memory capacity readers rate island violations as more acceptable than low working memory capacity readers do. This same intuition was shared by Sprouse, Wagers and Phillips (2012),

but I demonstrated that both prior data, and data from this dissertation required a more nuanced version of this intuition. This is presented in (7.1) as the Cognitive Co-variation Intuition (CCI), which represents an advancement in our understanding of the relationship between acceptability, processing and cognitive differences.

(7.1) Cognitive Co-variation Intuition (CCI) applied to island phenomena

- a) If the unacceptability of a sentence (here specifically an island violation) is due to processing difficulties
- b) And these processing difficulties arise from constraints on measurable cognitive resources (such as WM)
- c) Then those individuals with a measurably greater cognitive score are expected to process the sentences in question differently than lower scorers
- d) And this will result in these high-scoring individuals rating these difficult to process sentences as more acceptable, assuming there are no rating task differences between scorers

One reason why acceptability scores might not co-vary with working memory scores is the presence of rating task differences (7.1 d; Chapter 4, section 4.2.2.2). While some amount of acceptability rating may be automatic when we process a sentence, the task of assigning a score on a scale to represent that level of acceptability is a separate process. Differences can occur between groups both in how the sentences are processed and in making such explicit judgments. Thus, it need not be the case that relative difficulties in processing a sentence will transfer transparently into relative

differences in rating a sentence's acceptability. In fact there are clear cases where this does not appear to hold (d-linking: Michel 2010; center-embedding: Sprouse 2009; Hofmeister, Staum-Casasanto and Sag 2012a). The results from Chapter 4 present yet another such example, with high scorers on the form lure task rating island violation sentences lower, not higher, than the low scorers.

The possibility of rating task differences is only one issue in the larger concern of how we should expect cognitive scores and processing benefits to manifest. I presented the Processing Benefit Schedule (PBS) as a framework for considering the possible logical options. The PBS is described in detail in Chapter 4, section 4.2.2.1, and repeated in Table 7-2.

Table 7-2: Processing Benefit Schedule (PBS): Expectations of processing benefits for individuals with greater cognitive resources / higher cognitive scores (i.e. working memory, attention)

Higher cognitive resources benefit...	Does apply to difficult to process sentences (a 'push the limits' view)	Does not apply to difficult to process sentences
Does apply to easy to process sentences	(A) Global benefits: All sentences become easier to process	(B) Simple (only) benefit: Difficult sentences are at ceiling for everyone: no benefit available. Room available for benefit only in simple sentences.
Does not apply to easy to process sentences	(C) Complex (only) benefit: Difficult sentences require more resources that, if present, allow faster resolution of difficulties. Simple sentences do not need nor can they benefit from these extra resources.	(D) No benefits: Cognitive co-variation is irrelevant to processing

The acceptability results do not pattern in a way predicted by the ‘push the limits’ view of processing benefits, as assumed by Sprouse, Wagers and Phillips (2012). As was seen in the self-paced reading experiment of Chapter 5, the *whether*-islands and control sentences under examination follow a ‘simple (only) benefit’ when a processing benefit is observable.

7.1.4 Individual differences in the self-paced reading of islands

Chapter 5 presented the self-paced reading results of the sentences in Table 7-1. All readers slowed down at the clause boundary (*whether* read more slowly than *that*) in the island violation condition, indicating immediate sensitivity to the island boundary when there was an unresolved filler-gap dependency. However, only the low reading span participants showed a slowdown at the clause boundary when there was no unresolved dependency. That is, the low span readers showed an independent processing cost for the more complex clause boundary (*whether* read more slowly than *that*), while the high span readers had no apparent difficulties with this (relatively) easy condition. This represents a Processing Schedule Benefit (PBS) ‘simple (only) benefit’ for the high span readers. The high span readers showed the same processing cost as the low span readers when this *whether* clause boundary was encountered with an unresolved filler-gap dependency. That is, all readers had processing difficulty in the most difficult *whether*-island condition – there was no benefit here for the high

span readers – while only the low span readers had difficulty processing a *whether* clause boundary when no unresolved filler-gap dependency was present.

In addition to this reading span effect, it was found that participants with high form lure scores slowed down at the clause boundary compared to low scorers. This is again contrary to a ‘push the limits’ view in the Processing Benefit Schedule (PBS, Table 7-2). High form lure scorers also rated the difference between the MATRIX GAP and EMBEDDED GAP as larger than low scorers did in the acceptability judgment experiment (Chapter 4). Taken together, it appears that participants with high form lure scores are slowing down to note a relevant difference (i.e. when there is an unresolved filler-gap dependency crossing a clause boundary), and they (subsequently) rate this difference more distinctly. Comparing the results from both the reading time and acceptability studies, the slowdown and rating differences taken together appear to be similar to a speed-accuracy tradeoff: the high form lure scorers process this critical juncture in the sentence more slowly but are more precise in assessing its consequences for overall acceptability.

7.1.5 ERP responses to gap prediction and filler-gap association

The two key results revealed by the ERP experiment presented in Chapter 6 were the post-gap LAN and the embedded gap N400 responses. A post-gap LAN was elicited by all four post-gap positions, both matrix and embedded, including inside a *whether*-island. This consistent response indicated that (i) the process of filler-gap

association previously reported for fairly long-distance dependencies also occurs when a matrix clause subject gap cannot be immediately resolved, and (ii) the same filler-gap association process occurs in both an unacceptable *whether*-island and its acceptable control. That is, the presence of an island boundary did not prevent filler-gap association.

Additionally, an N400 response was found to be larger at the embedded gap position in an island compared to a non-island (this effect was not due to lexical differences, since the same lexical item, *openly*, occurred in both conditions). This response was interpreted as the gap (and/or the word that confirms the presence of the gap, namely *openly* in the example materials) being less predictable inside an island. This is because once the island clause boundary was encountered, the expectation for an upcoming gap was reduced. This modulation of predictions was stronger in high span readers than in low span readers. However, modulating the prediction for the gap did not influence the process of associating the filler and gap (there was no subsequent span co-variation with the LAN response) and it did not influence how participants rated similar sentences after the experiment (which also did not vary with span).

7.1.6 Predictability in processing islands

Taken together, these results indicate the importance of reading span with respect to how predictions are made online, but do not indicate a difference in the actual associating of the filler and gap. There was no evidence of a failed parse or

reanalysis in any condition or in any group of participants in the ERP experiment, yet the participants still rated island violations as the least acceptable sentences. Even though the brain responses to the island and non-island embedded gaps indicated the same basic responses (modulated only for predictability of the gap), the acceptability responses still differed. There is no apparent evidence of a large processing cost in the ERP results that would account for this difference in acceptability.

We did observe a processing cost for island violations at the clause boundary in the self-paced reading experiment, which is expected under the capacity-constrained account of islands. While part of this clause boundary cost co-varied with reading span, the off-line judgments did not. Reading span also co-varied with the N400 response at the embedded gap, interpreted as a predictability effect. While it is tempting to associate these two reading span effects, they do not co-vary in the same way. Both high and low span readers read the island violation condition equally slowly at the *whether* clause boundary. It was only in the condition with no embedded gap that the high span group demonstrated a processing benefit reading the clause boundary *whether*. On the other hand, the gap predictability effect was in response to the island violation condition (unlike the reading time effect just described). While it cannot be claimed that one (clause boundary reading time) leads to the other (gap predictability), they share a common basis in that they can be explained using the notion of predictability. I present the ‘gap predictability account of islands’ in section 7.5.

7.2 New questions arising from the ERP findings

Before advancing the gap predictability account of islands, it is useful to review how the results of the current experiments call into question some of the claims of the capacity-constrained and similarity-interference views of working memory, and how those views are applied to island phenomena. To do this, I start with two effects from the ERP experiment that do not align with prior claims.

In addition to the main findings of the ERP experiment reviewed above, there were two additional findings that will be worth pursuing further. First, the pre-gap P600, as discussed in detail in section 6.4.2.2.1, was found not only before the embedded gap position, but also before the matrix gap position, when it was impossible to predict which conditions would include a gap because they had not yet diverged at that point (*Who had...?*). Considering that the P600 is measured at a latency when the following word (i.e. the gap position: __ *openly*) has already been presented to the participants, I interpreted this P600 effect, at least in response to the current materials, as reflecting gap identification rather than integration difficulty (Kaan et al 2000). If the N400 effect at the embedded gap position reviewed above (section 7.1.5; see Chapter 6, section 6.4.2.2.7 for detailed discussion) does indeed reflect a difference in predictability, it would seem counterintuitive to observe an index of syntactic integration before this predictability effect. As discussed in section 6.4.2.2.1, there are differences between the current materials and prior studies that

report pre-gap P600 effects, such that an integration interpretation may still be plausible in these cases. This is a line of research that deserves further investigation.

Secondly, while there was no observable effect of sustained anterior negativity following the filler in the current ERP experiment, there was a sustained left anterior negativity following the post-gap LAN (section 6.4.2.2.5). Post-filler sustained LANs have been previously associated with the cost of holding a filler in working memory (e.g., King & Kutas 1995), but it has been shown that (in English) this sustained effect does not increase over time, as one might expect for an index of storage (e.g. King & Kutas 1995, Phillips et al. 2005, but see Fiebach et al. 2002 for a counter-example in German). Based on the current results, it is possible that having a sustained effect is simply a characteristic of many (working-memory-related) LANs. The sustained post-gap LAN reported in this dissertation cannot be interpreted as the cost of holding a filler in working memory, since it occurs after the filler and gap are associated. Thus, the finding of a sustained LAN that is clearly not due to the cost of actively holding an item in working memory undermines the interpretation of the post-filler sustained anterior negativity as an index of holding an item actively in working memory. This leads to additional evidence that the storage cost component of the capacity-constrained account of islands, and the capacity-constrained view of working memory in general, need to be updated.

7.3 Updating working memory models

The evidence for an active cost of holding a filler in working memory has been shaken by (i) the disappearance of the sustained anterior negativity effect under re-baselining (King & Kutas 1995, Phillips et al. 2005) and (ii) the appearance of a similar effect, the ‘lingering LAN,’ when no active storage cost should be expected (sections 6.4.2.2.5, 7.2). Thus the importance of the alleged storage cost of the filler should be re-evaluated. The lack of an active storage cost (and corresponding brain response) is consistent with a similarity-based interference view of working memory in sentence processing (e.g. Lewis & Vasishth 2005; Lewis, Vasishth & Van Dyke 2006), which does not have an active storage component. However, while similarity-based interference may be involved with long-distance dependencies in general, we found no evidence that island violations themselves can be accounted for by such a view of working memory (see section 7.4). If the concept of working memory is to be included in our understanding of how islands are processed, then verbal working memory in sentence processing needs to be reconceptualized with regard to (i) the cost of storing a filler (which appears to be negligible) and (ii) the relative ease in retrieving a filler from memory in the face of an interfering island boundary (for which there was no evidence that such a retrieval was more difficult than non-island controls in the three experiments of this dissertation). Based on the results from this dissertation, this reconception of working memory should include the ability to make and alter on-line predictions.

While the traditional capacity-constrained view of working memory has been undercut, the fact is that the reading span task still produced co-variation with the online measures in this dissertation, as well as in previous research. If the capacity account is not completely correct, then what do scores of reading span (traditionally used as a capacity measure) reflect? Crucially, the ‘capacity’ of the Capacity Constrained Comprehension Theory (Just and Carpenter 1992) is for both storage and processing costs. If we do not rely on the storage component, then it would appear that the co-variation of reading span and sentence processing is due to the processing component. Of the cognitive measures used in this dissertation, the reading span task most resembles ‘normal’ sentence processing, but with an additional memory task. Other tasks include a memory or attention component without the need to process a sentence. One might expect, then, that participants who are able to process a sentence with the added complexity of performing a simultaneous memory task are better able to process complexity in sentences. We see evidence for this view of reading span in the co-variation with the N400 effect at the embedded gap position. I argued (Chapter 6, section 6.4.3.2.2) that while both low and high span readers were sensitive to the island boundary and rated the island violations as the least acceptable, only the high span readers demonstrated the ability to act on the island boundary information in real time. That is, while both groups were able to recognize, and ultimately rate the importance of, the island boundary, only the high span group demonstrated that they had enough *processing capacity* to incorporate this information in their online parse of

the sentence. Thus, we see a need to distinguish the idea of *storage capacity* from *processing capacity*, with only the latter directly implicated in the current results.

As an alternative (though not an exclusive one), consider that Daneman and Carpenter (1980)'s original presentation of the reading span task demonstrated that it was co-variant with reading comprehension (also Daneman & Merikele 1996). Only later was this measure associated with a capacity-constrained theory of working memory (Just & Carpenter 1992). Recent work by Van Dyke et al. (2014) claims that reading span does not represent a connection to working memory ability per se, but is instead implicated only spuriously through its correlation with other cognitive measures, such as IQ. If this turns out to be the case, then one would expect that the pattern of these results could be replicated using co-variation with IQ score. If so, then we would want to know if this is an influence of general intelligence (i.e. 'g,' Spearman 1904), as is often assumed with IQ scores, or a specific ('s') verbal intelligence. Presumably, either of these options would include an idea of measurable *processing capacity*, either in terms of processing speed, efficiency or simultaneous processing. Further research is needed to disentangle these different types of processing capacity. For now, a conservative view appears to be that reading span tasks should be considered a co-variant of the ability (capacity) to process and comprehend sentences, but not necessarily due to differences in storage capacity limits.

Removing the focus on *storage capacity* from our interpretation of reading span and instead focusing on *processing capacity* requires revisiting how we

understand the active-filler hypothesis (Frazier & Clifton 1989). The active-filler hypothesis states that fillers should be assigned to gaps as soon as possible. This has, among other options, previously been thought of in terms of (storage-focused) capacity-constrained working memory (Just & Carpenter 1992; Gibson 1998). Under this view, the parser would benefit from discharging the filler from storage, as that would free more cognitive resources for other tasks. If we do not need to depend on the notion of a storage cost, however, while the active-filler strategy itself is still necessary, the memory-based motivation for this strategy is no longer tenable (see Aoshima, Phillips & Weinberg 2002 for an account of the active filler strategy that does not rely on storage costs).

Just as we saw the importance of reading span, but not necessarily for the traditional capacity-constrained view of working memory, we also saw that while memory interference scores co-varied with the linguistic data, both in acceptability judgments (Chapter 4, section 4.4.2) and reading times (Chapter 5, section 5.4.3.2.2), they did not do so in a way that indicated that a retrieval process was involved. These results revealed that high scorers on the memory interference task were more sensitive to whether there was an incomplete filler-gap dependency when a clause boundary was encountered, rating such sentences lower and reading them more slowly. Note that since this reading time effect was measured at the clause boundary, it doesn't make sense to identify this effect as a result of cue-based retrieval. The embedded cue (gap position) has not yet been encountered, and presumably no retrieval has occurred. It may simply be that individuals who score well on memory interference tasks are

more attuned to (linguistic) detail and therefore are also more attuned to recognizing and/or rating differences in sentences.

The importance of similarity-based interference has been supported by multiple researchers (e.g. Gordon, Hendrik and Johnson 2001; Gordon, Hendrick and Levine 2002; Lewis and Vasishth 2005; Gordon et al. 2006; Lewis, Vasishth and Van Dyke 2006; Van Dyke and McElree 2006). While the similarity-interference account of islands that I introduced in Chapter 2 (section 2.3.3.2) appeared to be a good candidate to add support to this view of working memory, the results of three experiments did not produce evidence in favor of this account (see section 7.4 below). This is relevant to other researchers working in a similarity-interference framework, as island effects, previously described as intervention effects of similar elements or features (i.e. the A-over-A condition, Chomsky 1964; Relativized Minimality, Rizzi 1990) do not appear amenable to a similarity-interference explanation. This is surprising considering that Lewis, Vasishth and Van Dyke (2006) assume that structural positions are relevant features for interference (see discussion in Chapter 2, section 2.3.3.2), just as grammatical accounts have used structural definitions in accounting for islands. Lewis, Vasishth and Van Dyke's (2006) example of a relevant structural position (subject) is fairly basic compared to those present in many syntactic architectures, but it is possible that 'subject' is a relevant feature for interference while the position of 'Spec-CP' is not. It would be useful, moving forward, to determine if island types that do not make use of the position 'Spec-CP' (for example subject islands) are more amenable to a similarity-based interference account.

While this dissertation did not find that a similarity-based interference view of working memory could explain the pattern of *wh*-island acceptability, it does provide additional evidence for its importance (in the form of how susceptible readers are to similarity-based interference) to the processing of clause boundaries. Recall that, compared to low memory lure scorers, readers with high memory lure scores slowed down more at the clause boundary and made a greater distinction between the GAP conditions (Chapter 5, section 5.4.3.2.2). Additionally, high scorers rated the difference between the GAP conditions as larger than low scorers did (Chapter 4, section 4.4.2.3.2). Taken together, these results show a greater sensitivity to whether a filler-gap dependency is unresolved when a new clause is encountered. This greater sensitivity takes the form of a more extreme processing difference as well as more extreme acceptability judgments. That is, this represents a straightforward example of greater processing difficulty (in this case due to heightened sensitivity) corresponding to greater acceptability judgment differences. It is important to reiterate, however, that this processing effect is independent of the STRUCTURE (ISLAND vs. NON-ISLAND) manipulation. While there is evidence that susceptibility to similarity-based interference is implicated in the processing of filler-gap dependencies in general (including the co-variation between memory lure and the distance manipulation reported in Chapter 5, section 5.4.3.1.2), there is no such evidence for the recognition of island boundaries. It would not be surprising if future research finds that this susceptibility to similarity-based interference will be implicated in other, more broad parsing processes (other long-distance dependencies, for example anaphors and

agreement), but the lack of sensitivity to the island/non-island nature of the clause boundary indicates that differences in susceptibility to similarity-based interference do not simply reduce to greater overall care or sensitivity in processing.

As a final note, it is important to point out that our understanding of how to characterize working memory does not need to be categorically based on either the capacity-constrained or the similarity-interference model. In the results presented in this dissertation, we have seen evidence for aspects of both views. While neither fully account for the *whether*-island data, neither can be fully discarded, either. We see different aspects of the sentences co-varying with different individual differences, indicating that certain processes rely more on one type/view of working memory than another. The next section specifically discusses how the results of this dissertation specifically inform the processing accounts of islands.

7.4 Processing accounts of islands

The stated intent of this dissertation was to test the capacity-constrained and similarity-interference accounts of island phenomena. This has been accomplished, and neither account is fully supported. Here I review the results of the three experiments of this dissertation as they relate to the similarity-interference and capacity-constrained accounts and briefly touch on how these results relate to other accounts of islands. The following section will present my gap predictability account of islands.

The similarity-interference account of islands that I introduced in this dissertation predicted that the locus of difficulty would be at or near the embedded gap position (where the retrieval process should be initiated), and that the difficulty should co-vary with the memory lure task. While there was co-variation of both acceptability judgments and self-paced reading with the memory lure task scores, this was only for the general distance effect of how far the filler was from the gap, and was not particular to the island violation. Additionally, the reading time difficulty appeared at the clause boundary, not the embedded gap site. Finally, there was no ERP evidence of the predicted retrieval difficulty at the gap position. There was an N400 effect at the gap for the island violation condition, but this was interpreted as a predictability effect and not an integration effect as (i) it precedes the LAN, taken to be reflective of filler-gap association (Kluender and Kutas 1993a,b; see Chapter 6, section 6.4.2.2.7 for additional arguments), and (ii) the effect is significant only in high span readers (see Chapter 6, section 6.4.3.2.2 for discussion). Even if this N400 effect were to be interpreted as related to the retrieval process, the post-ERP off-line judgments do not differ between the high and low span groups. It is important to note that these effects do not argue against the importance of similarity-based interference processes in general, but it is unlikely that they can account for why island violations are deemed unacceptable.

The results of the three experiments in this dissertation also do not directly support a capacity-constrained account of *whether*-islands, which predicted that the locus of difficulty would be at the island clause boundary, where the active memory

cost for the filler and additional processing cost of the island boundary combine, and that the difficulty should co-vary with the reading span task. While the self-paced reading experiment in Chapter 5 indicated that the locus of processing difficulty was at the clause boundary (though this was not replicated in the ERP experiment), as predicted by the capacity-constrained account, we have seen other evidence (discussed above in section 7.3) that it is unlikely that an active storage cost is involved (as assumed by the capacity-constrained account). The reading span data likewise do not support the capacity-constrained account, although in each case there are plausible reasons why this evidence was not present. No co-variation with reading span was found with acceptability judgments (Chapter 4, section 4.4.2), but this can be explained with reference to the Cognitive Co-variation Intuition (CCI; example 7.1, above). The co-variation pattern in the self-paced reading experiment (Chapter 5, section 5.4.3.2.1) showed a PBS ‘simple (only) benefit’ (Table 7-2), indicating the possibility of a ceiling effect at the clause boundary. Finally, in the ERP experiment, co-variation with reading span was found only at the embedded gap position, interpreted as a predictability effect (Chapter 6, section 6.4.3.2.2); the expected clause boundary effect was not present, arguably due to methodological issues (see section 6.4.2.2.5). While none of these results act as support for the capacity-constrained account of islands as previously presented (e.g. Kluender 1991, 1998), they do not necessarily undermine the idea that the unacceptability of *whether*-islands (and other islands) is due to processing difficulties. Instead, these results give us reason to reevaluate the nature of these processing difficulties. Specifically, we have seen

evidence that active storage is not a defining factor is island processing, but that island recognition and subsequent processes are. A major contribution of the current dissertation is found in the examination of what occurs after the island boundary is encountered.

In terms of a general view of processing islands, Phillips claims that there is a widespread view that “island constraints are immediately effective in parsing” (2006, pg 800). This is a claim that participants do not posit gaps within islands. The results from the ERP experiment indicate that this is partially true. Some participants (high span readers) do not appear to predict/posit gaps within islands (or at least predict them with lower probability). The results from the self-paced reading experiment indicate that all participants recognize the island clause boundary. Results from the acceptability judgment experiments (Experiment 1 and the shorter version conducted after Experiment 3) indicate that all participants also rate the island violations as the least acceptable. However, low span readers do not appear to be able to use the information from encountering the island boundary to successfully reduce their expectation for a gap while burdened by the demands of continuing to process the sentence. Even with these differences in prediction strength for an upcoming gap, the data reported in the ERP experiment indicate that both groups successfully identify the gap (P600: Chapter 6, section 6.4.2.2.1) and associate the filler and gap (LAN: Chapter 6, section 6.4.2.2.4). It is unlikely that the search for the gap position was entirely suspended or abandoned if filler-gap association occurs so readily in both groups and in both conditions. One might argue that readers are simply acquiescing to

the stream of RSVP words being presented to them. Consider however, (i) that visual ERP experiments have repeatedly reported successful differentiation between grammatical and ungrammatical conditions, even when participants are ‘required’ to process the ungrammatical sentences due to RSVP presentation and (ii) that there were also no self-paced reading time differences at and following the embedded gap between the ISLAND and NON-ISLAND conditions. Thus, it does not appear that encountering an island boundary results in island constraints shutting down or even substantially pausing the parser’s active search for a gap. Instead, the island boundary serves as a signal to high span readers to modulate their prediction for a gap. This is presented in the next section as the gap predictability account of islands.

7.5 Gap predictability account of processing islands

While neither the capacity-constrained nor the similarity-interference account of islands was fully supported by the current set of results, co-variation was found between the on-line processing of *whether*-islands and reading span. In order to distinguish it from the two accounts above, I will refer to the following as the ‘gap predictability account of processing islands.’ Note that this is an account of processing islands, and not a processing account of islands. We have seen no direct evidence that the online differences described below result in differences in acceptability judgments.

Whether-islands are not processed substantially differently from non-island controls except for (i) a slowdown in reading times at the island clause boundary for

both high and low span readers and (ii) the fact that high span readers modify their predictions of whether a gap is upcoming based on having encountered the island boundary. The gap predictability account of processing islands is schematized in (7.2).

(7.2) The gap predictability account of processing islands

- a) If there is an unresolved filler-gap dependency in a sentence, upon encountering an island boundary, the parser revises its predictions that a gap will be forthcoming.
- b) High span readers are better able to revise/modulate this prediction.
- c) If evidence for a gap is encountered within an island, it is straightforwardly identified and associated with a filler.
- d) Neither of these processes (b or c) directly influences the acceptability ratings assigned to an island violation.

I have outlined the gap predictability account of processing islands in a very general form, such that it is not specific to *whether*-islands. This represents the strongest form of the account. The account represents a claim about which processes are related and which are independent. Encountering evidence for an island structure triggers a revision of gap predictability, but neither (i) encountering the island structure nor (ii) lowered predictions for a gap prevents filler-gap association. Additional research is needed to determine if there is variability across island types with respect to the gap predictability account. Thus far, the behavioral data using filled-gaps and plausibility manipulations have supported the idea that gaps are not

predicted in ‘strong’ subject islands (Stowe 1986; Traxler & Pickering 1996). The current research allows us to support the idea that predictability matters in ‘weak’ *whether*-islands as well (though the predictability effect in *whether*-islands is modulated by reading span; it is still to be determined if this is also true for subject islands). At a first pass, then, it seems that it is not unreasonable to think that (7.2 a) would hold across multiple island types. It remains to be seen whether (7.2 b-d) do so as well. With the schema of the gap predictability account of processing islands outlined in (7.2), we can review how the current findings for *whether*-islands fit into this schema.

Upon encountering a *wh*-filler, the parser predicts that a gap will occur later in the sentence (Active Filler Strategy, Frazier & Clifton 1989). If a gap is identified (P600, section 6.4.2.2.1) in the matrix clause, the filler and gap can be associated (LAN, 6.4.2.2.4). On the other hand, if the filler-gap dependency extends into an embedded clause, the parser (of both high and low span participants) is immediately sensitive to whether the clause boundary represents an island structure (slower RTs, section 5.4.2.2.1). Low span readers show some difficulty at this island structure even if there is not an incomplete filler-gap dependency in the parse (slower RTs, section 5.4.3.2.1). The high span group thus shows a Processing Benefits Schedule (PBS) ‘easy (only) benefit’ (sections 4.2.2.1, 5.4.3.2.1) at the clause boundary. Having encountered the island, the high span readers adjust their prediction for where/when the gap will occur (7.2 a). Specifically, the high span group predicts that the gap is less likely to occur within the island domain (7.2 b). The low span group does not modify

its original prediction. The gap is identified, just as it would be in a non-island domain (P600, section 6.4.2.2.1). This is unpredicted by the high span group (N400, 6.4.3.2.7), which had modified its predictions for a gap occurring within the island based on the clause boundary. The low group, not having modified any predictions, shows no such effect. Both groups are able to associate the identified gap with the filler, again, just as in a non-island domain (7.2 c; LAN, 6.4.2.2.4). And both groups, despite differences in online predictions, rate the island violations as more unacceptable compared to control sentences (7.2 d; sections 4.4, 6.4.1.2).

One final effect reported in this dissertation worth discussing at this point comes from the self-paced reading experiment. While not necessary for the gap predictability account of processing islands, the reading times reported at the clause boundary (*that/whether*) are compatible with a general predictability-based account. Recall that all readers slowed down at the island-clause boundary (*whether*) when there was an unresolved filler-gap dependency in the sentence. Previously this was argued to reflect the combined costs of processing that clause boundary while holding a filler in memory (in support of the capacity-constrained account). If there is no active cost for holding a filler in memory (section 7.3), why is there a slowdown in reading time here? Rather than thinking of this effect as a memory cost, we can consider the possibility that it is a predictability effect. The filler generates a prediction for a gap (i.e. the active-filler hypothesis, Frazier & Clifton 1989). Part of that gap prediction includes what types of clauses can best contain a gap (in this case a declarative *that*-clause is preferred over an interrogative *whether*-clause). When the

interrogative *whether*-island clause boundary is encountered, the parser's incorrect prediction results in a processing cost. Both high and low span readers show this cost, though high span readers are better able to act on this information (7.2 b; Chapter 6, section 6.4.3.2.2). Note that the difference between the high and low span readers can still be thought of in terms of capacity, but it is crucially not *storage capacity*, but *processing capacity*.

The gap predictability account differs crucially from the capacity-constrained and similarity-interference accounts in that the differences in processing an island violation are not assumed to be the direct cause of lower acceptability ratings for that sentence. We saw evidence that the predictability effect (N400 difference between high and low span readers) is independent of filler-gap association (post-gap LAN). Additionally, the post-gap LAN response did not differ between island and non-island controls, even though offline acceptability ratings did differ in these conditions. Taken together, this suggests that the process of filler-gap association may be fairly automatic. There is no evidence of a failed parse or reanalysis. These island violations are unacceptable, but they do not appear to be unparseable.

7.6 Parser-grammar relationship

The gap predictability account of processing islands represents a claim about (i) what processes are occurring across different sentence positions in different populations, and (ii) which of those processes are independent from or influenced by

the others. However, this account makes no claim that these processes are the root cause of the unacceptability of island violations. Neither the capacity-constrained nor the similarity-interference processing accounts of islands were fully supported by the current data. Still, island violations were rated as the least acceptable compared to control sentences. Does this mean that a grammatical account of islands is more likely to be correct? The answer to this question depends, in part, on which data the grammar is potentially called upon to account for.

Islands are traditionally identified by native speaker judgments of grammaticality/ acceptability. If a grammatical account of islands relies strictly on informal judgments, then the current dissertation cannot shed light on this approach, since such informal judgments were not the current focus of inquiry. However, if a grammatical account of islands seeks to account for the more fine-grained and gradient data obtained by experimental acceptability judgments, then the acceptability results reported in Experiments 1 and 3 are relevant. As discussed previously (see Chapter 4, section 4.2.1), since the acceptability data were compared directly with online processing data in this dissertation, they bring additional nuance to the acceptability judgments under discussion in the grammar vs. processing debate of Sprouse, Wagers and Phillips (2012) and Hofmiester, Staum-Casasanto and Sag (2012a,b). Namely, we must now think of the acceptability results not in isolation, but with respect to the pattern of processing throughout the sentence. Before doing so, it is important to note that how one thinks about the relationship of island-violations to related controls informs our understanding of the data.

In this dissertation, and in Sprouse, Wagers and Phillips (2012), island violations are compared to controls along two dimensions: whether there is an island structure present, and whether the filler is outside and the gap inside that structure. There is no claim by grammatical theory that a gap that is further from its filler should be judged as less acceptable. Nor is there a grammatical claim that a *whether* clause boundary should be inherently less acceptable than a *that* clause boundary. Both of these effects are explained outside of the grammar, namely in processing terms. Thus, even if it is decided that island violations are unacceptable due to a grammatical violation, this is not the complete story: there are processing factors at play in addition to this possible grammatical violation. We should not expect that the grammar would account for all of the acceptability judgment data, just as we must consider that processing factors on their own will not account for all the judgment data.

While the processing factors of filler-gap dependency length and the nature of the intervening material are important to keep in mind, it is the remainder of the acceptability effect (i.e. the superadditive effect) that grammatical accounts focus on. As this dissertation instead focused on contrasting two processing accounts of islands, no direct evidence for or against a grammatical account of islands was uncovered. However, there is nevertheless a pattern in the processing data that is problematic for a grammatical account.

One of the most remarkable findings from the ERP experiment is that there is no evidence of a failed parse, or of reanalysis, either at the clause boundary (upon recognition of the island boundary) or at the gap embedded in an island (upon

realization of the ungrammatical structure). It would be reasonable to expect a P600 response, which has been elicited by both syntactic violation paradigms (e.g. subcategorization violations: Neville et al. 1991; Osterhout & Holcomb 1992; Hagoort et al. 1993; Osterhout et al. 1994) and reanalysis/garden path paradigms (e.g. Osterhout & Holcomb 1992; Osterhout et al. 1994). Even if one were inclined to interpret the N400 effect reported in Experiment 3 as a recognition of ungrammaticality (contra the arguments presented in Chapter 6, section 6.4.2.2.7), consider that this (i) is not the expected response to a syntactic violation, (ii) was a small but significant effect (one might reasonably expect a larger response to ungrammaticality), and (iii) was only significant in the high span group. We are in no position to claim that only high span readers have access to grammatical knowledge (especially since the acceptability judgments of high and low span readers do not differ). Even without a robust ERP response to perceived ungrammaticality, participants still rated *whether*-island violation sentences as the least acceptable.

This highlights an intriguing pattern of results across the dissertation's experiments. Readers show an immediate effect of encountering the *whether* clause boundary in the island-violation condition, but show no subsequent difficulty in associating the filler and gap. In section 7.4, this was discussed in terms of filler-gap association being a fairly automatic process. Even though the ERP evidence indicates that this process is completed in an island as well as in non-island controls, participants still rate the island violations as less acceptable. We can consider whether there is a grammatical explanation for this fact.

The question of whether a structure is representable but still ungrammatical has recently been raised again in the literature (Phillips 2013a,b; Lewis & Phillips 2014). A grammar that generates structures, and then applies grammatical constraints to them as a filter, such as Government-Binding Theory (Chomsky 1981) or Optimality Theory (Prince & Smolensky 2004), allows for structures to be represented in the grammar before they are ruled out as ungrammatical. This is in contrast to a grammar that applies grammatical constraints to the derivation of structures, such as in the Minimalist Program (Chomsky 1993), in which case the structure in question would never be represented in the grammar. The question is whether (*whether*) island violations represent a structure that is representable but still ungrammatical.

The allure of treating island violations as representable but ungrammatical is that it is a potential way of reconciling both (i) the lack of a large ERP response to perceived ungrammaticality and (ii) the lack of difference in filler-gap association between island violations and grammatical controls with the fact that the island violation is rated as less acceptable than control sentences. Under this analysis, the acceptability rating results are explained by the grammatical filter, and the lack of online differentiation is explained because the structures are treated as representable by the grammar, and presumably by the parser as well.

In order to examine this possibility, we must assume that there is a close relationship between the grammar and parser, such that ‘representableness’ in the grammar corresponds to a certain permissiveness in the parser (see Lewis & Phillips 2014 for discussion of the grammar and parser as closely related systems or separate

systems). This correspondence amounts to an understanding of how the timing of grammatical derivation and the timing of online processing interact. In the current situation, we have to assume that the grammatical filter(s) for islands apply late, after the parser has already associated the filler and gap, but before an acceptability judgment is given. This is problematic, however, as we have evidence from the self-paced reading experiment that the parser has responded to the island violation as early as the clause boundary. That is, we have evidence that the island constraints/filters are already active at the clause boundary, so it is unclear why they are not still active at the gap position. It seems improbable that island constraints/filters would be active at the clause boundary and offline, but inactive (or turned off) at the embedded gap position between them. That is, the sequencing of effects in the current set of experiments would require the constraints/filters be active and inactive at different points of the sentence for unclear reasons. Thus, while potentially alluring, claiming that islands are representable does not solve the mismatch between online and offline data.

There may be two ways to salvage a ‘representable’ analysis, though neither is a simple solution, as both would require a detailed understanding of how the grammar and parser should relate to each other. Such an understanding is not currently available. One option is to claim that while a grammatical filter is responsible for the ‘island effect,’ the reading time effect measured at the clause boundary is purely a processing effect, and *not* due to the filter. Just as is claimed in the constrained-capacity account of islands, having an unresolved filler-gap dependency at a clause

boundary causes a strain on the parser, as does having a more complex interrogative *whether* clause. This creates a processing bottleneck in the island violation condition, resulting in slower reading times. However, this bottleneck is *not* responsible for the (superadditive) low acceptability ratings assigned to island violations. (The bottleneck can still be responsible for the additive effects of dependency length and clause type, however.) There are no other processing bottlenecks in the sentence: the embedded clauses of both the island and control sentences are representable and are processed smoothly. It is only upon the application of the grammatical filters (after the embedded gap is read, but before acceptability judgments are given), that the island violation is judged as unacceptable. In essence, this approach bypasses the ‘sequencing of effects’ complication described above by claiming that the first effect (at the clause boundary) is not due to the grammar.

But how does one determine which processing effects are ‘purely’ processing effects and which are the processing consequences of the application of grammatical filters? How does one determine the extent to which processing bottlenecks (vs. grammatical filters) influence acceptability? Clearly, an understanding of the relationship between what the grammar represents (and when) and what the parser can process, as well as how this relates to acceptability, needs to be detailed for this approach to be feasible.

Second, it could be that grammatical filters apply at various times in a sentence and only apply to the part of the sentence it has encountered thus far. For example, it would be plausible for them to apply at clause boundaries. In such a scenario it would

be possible for the filter relevant for islands to apply upon encountering the clause boundary (*whether/that*), causing the parser to slow down in order to process an ungrammatical sentence (reading time effect). However, the filter applied at the clause boundary does not apply to the embedded gap position, which has not yet been encountered. In this case the parser does not treat the embedded clause material as ungrammatical (but it is representable) until the filter is applied again at the next clause boundary (the end of the sentence). Here the ‘sequencing of effects’ complication is addressed by having the grammatical filter apply at different times, influencing only a limited portion of the parse at a time.

As above, this salvage of a ‘representable’ account of islands is unconvincing without a much more developed understanding of the grammar-parser relationship. It needs to be explained why the application of the first filter does not influence the parser for the remainder of the sentence. If the application of the first filter results in the reading time effects reported at the clause boundary, why are there no effects present when the second filter applies (i.e. end of sentence wrap up effects)? How quickly do ‘filter effects’ fade once they are applied? That is, how is the grammar-parser relationship arranged such that it is opaque to effects that it was susceptible to only a moment before? And why would a sentence that was ungrammatical at the clause boundary no longer be treated as ungrammatical later in the sentence?

These types of questions are only appropriate, of course, if one assumes a grammar that constrains structures that have already been built, but not if one assumes a grammar that constrains/prevents the generation of such structures. This serves to

highlight a need to be explicit in which grammatical theory is being examined, as each will have differences in their assumptions and architectures. There is value, then, in considering multiple grammatical theories (just as this dissertation sought to test multiple processing accounts of islands) to see how well their assumptions and architectures account for the data. Further, an examination of how these various grammatical theories interact with various processing theories is required. This is a complex undertaking beyond the scope of the current discussion, but it may be that in order to make substantial headway in debates of grammar vs. processing for various phenomena, we must first lay a groundwork defining which grammar and which parser are under consideration as well as how those two theories interface. Syntacticians have put much effort into various syntax-X interfaces. I would argue that the next interface that needs to be mapped out is the grammar-parser interface.

7.7 Future research

This dissertation used an individual differences approach to test the capacity-constrained and similarity-interference accounts of island phenomena. Strong, direct evidence was not found for either processing account, and no evidence was found for on-line processing difficulties being the cause of the unacceptability of island violations. Instead we saw a certain amount of independence between the on-line processing data and the off-line acceptability judgments. For example, we observed that while high and low span readers may process the sentences differently on-line,

they did not rate the acceptability of the sentences differently off-line. The Processing Benefits Schedule (PBS) was provided as a framework with which to understand these relationships. The Cognitive Co-variation Intuition (CCI) was outlined in order to provide a clear understanding of the relationships between on-line processing and off-line judgments. Both of these frameworks can be applied to different datasets and used to sharpen future research questions.

Instead of either storage capacity limits or retrieval interference explaining individual differences in the on-line processing of island phenomena, we found that the key difference was in how the parser modified its prediction of where a gap could be expected. This finding adds to the growing evidence in support of the importance of predictive processing (Altmann and Kamide 1999; 2009; Kamide, Altmann & Haywood 2003; Federmeier 2007; Pickering and Garrod 2007) and is compatible with probability-based approaches to parsing (e.g. MacDonald & Christensen 2002; Hale 2006; Levy 2008). Additional research is needed to determine if the individual differences pattern found here (that high span readers are better at modulating predictions) also holds for other predictive processes in other sentence types. If predictability is thought of as a top-down process, then it appears that high reading span participants are better able to make use of that top-down process. This suggests that 'better' reading and comprehension ability is linked to efficiency in top-down predictions, even if this efficiency resulted in a temporarily incorrect prediction in the current materials (i.e. predicting that a gap would not be forthcoming in the island when, in the current experiments, it in fact was).

The ERP experiment reaffirmed the importance of the LAN in filler-gap dependencies. The LAN elicited by the short-distance matrix subject gap was a novel finding providing evidence that the same processes are involved in relatively simple and short filler-gap dependencies as are involved in more complex dependencies into embedded clauses. At the same time, the results here raised new questions about the LAN and P600. Does the LAN elicited by fillers and gaps naturally have a sustained duration, making the sustained negativity a part of the LAN response rather than an index of storage cost? Does the P600 index an integration cost prior to the gap site, recognition of the gap site or some combination thereof? The resolution of these issues will need careful additional experimentation.

The gap predictability account of processing islands opens up new avenues of research. In addition to testing the account with respect to other islands, it will be interesting, moving forward, to research how gap predictability interacts with ameliorations of island phenomena such as d-linking and (non-)finiteness. By using the electrophysiological response (the gap position N400, which appears to reflect gap predictability) reported in this dissertation, we can explicitly test whether and to what extent the nature of the filler (d-linking) and/or intervening material (finiteness/non-finiteness) modulate gap predictability. Prior research comparing finite and non-finite sentences that were potentially subject island violations (i.e. involving parasitic gaps, Phillips 2006) suggests that finiteness does influence on-line parsing of filler-gap dependencies (see also Michel & Goodall 2013). While we saw no evidence that gap predictability itself modulated the acceptability of islands and related control

sentences, recall that individual differences co-variation frequently occurred on the easier to process sentences. Thus, we may be able to observe a link between on-line predictability and off-line acceptability when the ameliorating effects of d-linking and (non-)finiteness are used to manipulate the relative ease with which the parser can make the predictions for a gap.

7.8 Concluding remarks

This dissertation has demonstrated the value of examining the incremental processing of islands using an individual differences approach. Where previous research has focused on how the parser reacts to an island boundary, the experiments presented here have allowed us to examine the behavior of the parser before, at, and after this point, including off-line acceptability judgments. We have seen that while the parser immediately recognizes an island boundary, this does not prevent the association of the filler and gap inside that island.

Additionally, the results from the experiments in this dissertation have demonstrated the presence of different effects of individual differences at different processes at different parts of the sentence. Susceptibility to similarity-based interference interacted with dependency length at the clause boundary, but reading span interacted with predictions for a gap inside an island. If we consider the sentence as unfolding in time, we saw that the high and low span readers had similar initial processes, diverged at the clause boundary and then converged again at the embedded

gap position. By the end of the sentence, the high and low span readers give the same acceptability ratings for sentences that they had processed (at least momentarily) differently. Here we have strong evidence that to best understand how a sentence is processed, we cannot limit ourselves to examining only one position in the sentence nor can we limit ourselves to one view of working memory in sentence processing.

The findings presented here represent contributions not only to our understanding of *whether*-islands specifically, but also make predictions about island phenomena in general. The findings also touch on filler-gap dependencies more broadly with regard to how readers respond to them, both behaviorally and in terms of brain responses. The relationship between classic cognitive measures and the views of working memory that they are associated with has been questioned, but at the same time, I have demonstrated how an individual differences approach can assist in interpreting results within and across experiments. Overall, the individual difference approach taken in this dissertation has deepened our understanding of a complex phenomenon, and it is my hope that the findings reported here and the thoughts developed within represent another step in our continual search for better understandings of the inner workings of human language.

Appendix 1: Materials for acceptability judgment and self-paced reading experiments (Experiments 1 and 2, respectively)

Item: Sentences:

- 1 Who / had / calmly / deduced / that / the soldier / assisted / the medic / as / the / enemy / forces / approached?
Who / had / calmly / deduced / whether / the soldier / assisted / the medic / as / the / enemy / forces / approached?
Who / had / the medic / deduced / that / the soldier / assisted / calmly / as / the / enemy / forces / approached?
Who / had / the medic / deduced / whether / the soldier / assisted / calmly / as / the / enemy / forces / approached?
- 2 Who / had / honestly / assumed / that / the sailor / befriended / the captain / although / the / betrayal / would / come?
Who / had / honestly / inquired / whether / the sailor / befriended / the captain / although / the / betrayal / would / come?
Who / had / the captain / assumed / that / the sailor / befriended / honestly / although / the / betrayal / would / come?
Who / had / the captain / inquired / whether / the sailor / befriended / honestly / although / the / betrayal / would / come?
- 3 Who / had / loudly / said / that / the custodian / berated / the plumber / although / the / solution / was / working?
Who / had / loudly / said / whether / the custodian / berated / the plumber / although / the / solution / was / working?
Who / had / the plumber / said / that / the custodian / berated / loudly / although / the / solution / was / working?
Who / had / the plumber / said / whether / the custodian / berated / loudly / although / the / solution / was / working?
- 4 Who / had / on Thursday / assumed / that / the salesman / bit / the manager / after / the / group / got / drunk?
Who / had / on Thursday / inquired / whether / the salesman / bit / the manager / after / the / group / got / drunk?
Who / had / the manager / assumed / that / the salesman / bit / on Thursday / after / the / group / got / drunk?
Who / had / the manager / inquired / whether / the salesman / bit / on Thursday / after / the / group / got / drunk?
- 5 Who / had / mockingly / contended / that / the baker / called / the farmer / although / the / phone / was / disconnected?
Who / had / mockingly / wondered / whether / the baker / called / the farmer / although / the / phone / was / disconnected?
Who / had / the farmer / contended / that / the baker / called / mockingly / although / the / phone / was / disconnected?
Who / had / the farmer / wondered / whether / the baker / called / mockingly / although / the / phone / was / disconnected?

- 6 Who / had / nicely / assumed / that / the beautician / copied / the trainee / because / the / result / was / entrancing?
 Who / had / nicely / inquired / whether / the beautician / copied / the trainee / because / the / result / was / entrancing?
 Who / had / the trainee / assumed / that / the beautician / copied / nicely / because / the / result / was / entrancing?
 Who / had / the trainee / inquired / whether / the beautician / copied / nicely / because / the / result / was / entrancing?
- 7 Who / had / on Tuesday / assumed / that / the decorator / annoyed / the carpenter / when / the / deadline / was / missed?
 Who / had / on Tuesday / inquired / whether / the decorator / annoyed / the carpenter / when / the / deadline / was / missed?
 Who / had / the carpenter / assumed / that / the decorator / annoyed / on Tuesday / when / the / deadline / was / missed?
 Who / had / the carpenter / inquired / whether / the decorator / annoyed / on Tuesday / when / the / deadline / was / missed?
- 8 Who / had / abruptly / contended / that / the dentist / divorced / the accountant / after / the / draining / lawsuit / ended?
 Who / had / abruptly / wondered / whether / the dentist / divorced / the accountant / after / the / draining / lawsuit / ended?
 Who / had / the accountant / contended / that / the dentist / divorced / abruptly / after / the / draining / lawsuit / ended?
 Who / had / the accountant / wondered / whether / the dentist / divorced / abruptly / after / the / draining / lawsuit / ended?
- 9 Who / had / on Wednesday / declared / that / the artist / endangered / the critic / while / the / sculpture / swayed / unsteadily?
 Who / had / on Wednesday / speculated / whether / the artist / endangered / the critic / while / the / sculpture / swayed / unsteadily?
 Who / had / the critic / declared / that / the artist / endangered / on Wednesday / while / the / sculpture / swayed / unsteadily?
 Who / had / the critic / speculated / whether / the artist / endangered / on Wednesday / while / the / sculpture / swayed / unsteadily?
- 10 Who / had / noisily / contended / that / the gardener / fired / the florist / when / the / roses / all / wilted?
 Who / had / noisily / wondered / whether / the gardener / fired / the florist / when / the / roses / all / wilted?
 Who / had / the florist / contended / that / the gardener / fired / noisily / when / the / roses / all / wilted?
 Who / had / the florist / wondered / whether / the gardener / fired / noisily / when / the / roses / all / wilted?
- 11 Who / had / publicly / said / that / the bride / followed / the groom / because / the / press / was / watching?
 Who / had / publicly / said / whether / the bride / followed / the groom / because / the / press / was / watching?
 Who / had / the groom / said / that / the bride / followed / publicly / because / the / press / was / watching?

Who / had / the groom / said / whether / the bride / followed / publicly / because / the / press / was / watching?

- 12 Who / had / viciously / declared / that / the waiter / hated / the busboy / since / the / two / never / spoke?

Who / had / viciously / speculated / whether / the waiter / hated / the busboy / since / the / two / never / spoke?

Who / had / the busboy / declared / that / the waiter / hated / viciously / since / the / two / never / spoke?

Who / had / the busboy / speculated / whether / the waiter / hated / viciously / since / the / two / never / spoke?

- 13 Who / had / wisely / deduced / that / the janitor / helped / the mechanic / while / the / sprinklers / were / leaking?

Who / had / wisely / deduced / whether / the janitor / helped / the mechanic / while / the / sprinklers / were / leaking?

Who / had / the mechanic / deduced / that / the janitor / helped / wisely / while / the / sprinklers / were / leaking?

Who / had / the mechanic / deduced / whether / the janitor / helped / wisely / while / the / sprinklers / were / leaking?

- 14 Who / had / sadly / declared / that / the painter / hugged / the model / because / the / funding / was / eliminated?

Who / had / sadly / speculated / whether / the painter / hugged / the model / because / the / funding / was / eliminated?

Who / had / the model / declared / that / the painter / hugged / sadly / because / the / funding / was / eliminated?

Who / had / the model / speculated / whether / the painter / hugged / sadly / because / the / funding / was / eliminated?

- 15 Who / had / seriously / declared / that / the runner / hurt / the cyclist / as / the / scene / was / inspected?

Who / had / seriously / speculated / whether / the runner / hurt / the cyclist / as / the / scene / was / inspected?

Who / had / the cyclist / declared / that / the runner / hurt / seriously / as / the / scene / was / inspected?

Who / had / the cyclist / speculated / whether / the runner / hurt / seriously / as / the / scene / was / inspected?

- 16 Who / had / on Sunday / assumed / that / the musician / kicked / the conductor / before / the / last / concert / began?

Who / had / on Sunday / inquired / whether / the musician / kicked / the conductor / before / the / last / concert / began?

Who / had / the conductor / assumed / that / the musician / kicked / on Sunday / before / the / last / concert / began?

Who / had / the conductor / inquired / whether / the musician / kicked / on Sunday / before / the / last / concert / began?

- 17 Who / had / on Monday / contended / that / the reporter / kidnapped / the photographer / while / the / recorder / was / running?

Who / had / on Monday / wondered / whether / the reporter / kidnapped / the photographer / while / the / recorder / was / running?

Who / had / the photographer / contended / that / the reporter / kidnapped / on Monday / while / the / recorder / was / running?

Who / had / the photographer / wondered / whether / the reporter / kidnapped / on Monday / while / the / recorder / was / running?

- 18 Who / had / gently / contended / that / the stewardess / kissed / the pilot / before / the / private / plane / departed?

Who / had / gently / wondered / whether / the stewardess / kissed / the pilot / before / the / private / plane / departed?

Who / had / the pilot / contended / that / the stewardess / kissed / gently / before / the / private / plane / departed?

Who / had / the pilot / wondered / whether / the stewardess / kissed / gently / before / the / private / plane / departed?

- 19 Who / had / rashly / said / that / the maid / married / the butler / when / the / prior / spouse / died?

Who / had / rashly / said / whether / the maid / married / the butler / when / the / prior / spouse / died?

Who / had / the butler / said / that / the maid / married / rashly / when / the / prior / spouse / died?

Who / had / the butler / said / whether / the maid / married / rashly / when / the / prior / spouse / died?

- 20 Who / had / recently / declared / that / the butcher / met / the cook / although / the / alibi / was / checked?

Who / had / recently / speculated / whether / the butcher / met / the cook / although / the / alibi / was / checked?

Who / had / the cook / declared / that / the butcher / met / recently / although / the / alibi / was / checked?

Who / had / the cook / speculated / whether / the butcher / met / recently / although / the / alibi / was / checked?

- 21 Who / had / timidly / contended / that / the clerk / observed / the mailman / since / the / paychecks / were / lost?

Who / had / timidly / wondered / whether / the clerk / observed / the mailman / since / the / paychecks / were / lost?

Who / had / the mailman / contended / that / the clerk / observed / timidly / since / the / paychecks / were / lost?

Who / had / the mailman / wondered / whether / the clerk / observed / timidly / since / the / paychecks / were / lost?

- 22 Who / had / eagerly / contended / that / the lawyer / punched / the defendant / as / the / aggressive / questions / continued?

Who / had / eagerly / wondered / whether / the lawyer / punched / the defendant / as / the / aggressive / questions / continued?

Who / had / the defendant / contended / that / the lawyer / punched / eagerly / as / the / aggressive / questions / continued?

Who / had / the defendant / wondered / whether / the lawyer / punched / eagerly / as / the / aggressive / questions / continued?

- 23 Who / had / on Saturday / assumed / that / the banker / pushed / the teller / while / the / fire / alarm / rang?

- Who / had / on Saturday / inquired / whether / the banker / pushed / the teller / while / the / fire / alarm / rang?
 Who / had / the teller / assumed / that / the banker / pushed / on Saturday / while / the / fire / alarm / rang?
 Who / had / the teller / inquired / whether / the banker / pushed / on Saturday / while / the / fire / alarm / rang?
- 24 Who / had / on Friday / declared / that / the fisherman / rescued / the swimmer / after / the / dangerous / storm / ended?
 Who / had / on Friday / speculated / whether / the fisherman / rescued / the swimmer / after / the / dangerous / storm / ended?
 Who / had / the swimmer / declared / that / the fisherman / rescued / on Friday / after / the / dangerous / storm / ended?
 Who / had / the swimmer / speculated / whether / the fisherman / rescued / on Friday / after / the / dangerous / storm / ended?
- 25 Who / had / quickly / declared / that / the writer / saved / the editor / before / the / plagiarism / panel / met?
 Who / had / quickly / speculated / whether / the writer / saved / the editor / before / the / plagiarism / panel / met?
 Who / had / the editor / declared / that / the writer / saved / quickly / before / the / plagiarism / panel / met?
 Who / had / the editor / speculated / whether / the writer / saved / quickly / before / the / plagiarism / panel / met?
- 26 Who / had / moodily / declared / that / the policeman / scolded / the fireman / when / the / unfavorable / report / surfaced?
 Who / had / moodily / speculated / whether / the policeman / scolded / the fireman / when / the / unfavorable / report / surfaced?
 Who / had / the fireman / declared / that / the policeman / scolded / moodily / when / the / unfavorable / report / surfaced?
 Who / had / the fireman / speculated / whether / the policeman / scolded / moodily / when / the / unfavorable / report / surfaced?
- 27 Who / had / secretly / deduced / that / the dancer / stalked / the violinist / before / the / intense / police / questioning?
 Who / had / secretly / deduced / whether / the dancer / stalked / the violinist / before / the / intense / police / questioning?
 Who / had / the violinist / deduced / that / the dancer / stalked / secretly / before / the / intense / police / questioning?
 Who / had / the violinist / deduced / whether / the dancer / stalked / secretly / before / the / intense / police / questioning?
- 28 Who / had / rapidly / assumed / that / the guitarist / texted / the drummer / since / the / show / was / starting?
 Who / had / rapidly / inquired / whether / the guitarist / texted / the drummer / since / the / show / was / starting?
 Who / had / the drummer / assumed / that / the guitarist / texted / rapidly / since / the / show / was / starting?
 Who / had / the drummer / inquired / whether / the guitarist / texted / rapidly / since / the / show / was / starting?

- 29 Who / had / bluntly / assumed / that / the wrestler / tripped / the referee / as / the / crowd / wildly / cheered?
 Who / had / bluntly / inquired / whether / the wrestler / tripped / the referee / as / the / crowd / wildly / cheered?
 Who / had / the referee / assumed / that / the wrestler / tripped / bluntly / as / the / crowd / wildly / cheered?
 Who / had / the referee / inquired / whether / the wrestler / tripped / bluntly / as / the / crowd / wildly / cheered?
- 30 Who / had / quietly / deduced / that / the thief / undermined / the informant / after / the / local / authorities / arrived?
 Who / had / quietly / deduced / whether / the thief / undermined / the informant / after / the / local / authorities / arrived?
 Who / had / the informant / deduced / that / the thief / undermined / quietly / after / the / local / authorities / arrived?
 Who / had / the informant / deduced / whether / the thief / undermined / quietly / after / the / local / authorities / arrived?
- 31 Who / had / dimly / contended / that / the nurse / welcomed / the patient / because / the / companionship / was / needed?
 Who / had / dimly / wondered / whether / the nurse / welcomed / the patient / because / the / companionship / was / needed?
 Who / had / the patient / contended / that / the nurse / welcomed / dimly / because / the / companionship / was / needed?
 Who / had / the patient / wondered / whether / the nurse / welcomed / dimly / because / the / companionship / was / needed?
- 32 Who / had / warmly / said / that / the teacher / wrote / the parent / since / the / grades / had / improved?
 Who / had / warmly / said / whether / the teacher / wrote / the parent / since / the / grades / had / improved?
 Who / had / the parent / said / that / the teacher / wrote / warmly / since / the / grades / had / improved?
 Who / had / the parent / said / whether / the teacher / wrote / warmly / since / the / grades / had / improved?

Appendix 2: Materials for ERP experiment (Experiment 3)

Item: Sentences:

- 1 Who / had / unreservedly / assumed / that / the skipper / befriended / the mariner / before / the final / mutiny / hearing?
Who / had / unreservedly / inquired / whether / the skipper / befriended / the mariner / before / the final / mutiny / hearing?
Who / had / the mariner / assumed / that / the skipper / befriended / unreservedly / before / the final / mutiny / hearing?
Who / had / the mariner / inquired / whether / the skipper / befriended / unreservedly / before / the final / mutiny / hearing?
- 2 Who / had / secretly / assumed / that / the stylist / copied / the trainee / because / the cut / was / similar?
Who / had / secretly / inquired / whether / the stylist / copied / the trainee / because / the cut / was / similar?
Who / had / the trainee / assumed / that / the stylist / copied / secretly / because / the cut / was / similar?
Who / had / the trainee / inquired / whether / the stylist / copied / secretly / because / the cut / was / similar?
- 3 Who / had / publicly / said / that / the groomsman / embarrassed / the groom / because / the tuxedo / didn't / fit?
Who / had / publicly / said / whether / the groomsman / embarrassed / the groom / because / the tuxedo / didn't / fit?
Who / had / the groom / said / that / the groomsman / embarrassed / publicly / because / the tuxedo / didn't / fit?
Who / had / the groom / said / whether / the groomsman / embarrassed / publicly / because / the tuxedo / didn't / fit?
- 4 Who / had / immediately / declared / that / the driver / hit / the cyclist / when / the state / police / arrived?
Who / had / immediately / speculated / whether / the driver / hit / the cyclist / when / the state / police / arrived?
Who / had / the cyclist / declared / that / the driver / hit / immediately / when / the state / police / arrived?
Who / had / the cyclist / speculated / whether / the driver / hit / immediately / when / the state / police / arrived?
- 5 Who / had / prematurely / said / that / the heiress / married / the bodyguard / when / the divorce / wasn't / settled?
Who / had / prematurely / said / whether / the heiress / married / the bodyguard / when / the divorce / wasn't / settled?
Who / had / the bodyguard / said / that / the heiress / married / prematurely / when / the divorce / wasn't / settled?
Who / had / the bodyguard / said / whether / the heiress / married / prematurely / when / the divorce / wasn't / settled?
- 6 Who / had / impetuously / assumed / that / the regulator / blamed / the treasury / when / the stock / market / crashed?

- Who / had / impetuously / inquired / whether / the regulator / blamed / the treasury / when / the stock / market / crashed?
 Who / had / the treasury / assumed / that / the regulator / blamed / impetuously / when / the stock / market / crashed?
 Who / had / the treasury / inquired / whether / the regulator / blamed / impetuously / when / the stock / market / crashed?
- 7 Who / had / slyly / deduced / that / the dancer / stalked / the choreographer / before / the police / got / involved?
 Who / had / slyly / deduced / whether / the dancer / stalked / the choreographer / before / the police / got / involved?
 Who / had / the choreographer / deduced / that / the dancer / stalked / slyly / before / the police / got / involved?
 Who / had / the choreographer / deduced / whether / the dancer / stalked / slyly / before / the police / got / involved?
- 8 Who / had / readily / contended / that / the invalid / welcomed / the nurse / because / the companionship / was / needed?
 Who / had / readily / wondered / whether / the invalid / welcomed / the nurse / because / the companionship / was / needed?
 Who / had / the nurse / contended / that / the invalid / welcomed / readily / because / the companionship / was / needed?
 Who / had / the nurse / wondered / whether / the invalid / welcomed / readily / because / the companionship / was / needed?
- 9 Who / had / competently / deduced / that / the private / assisted / the medic / as / the enemy / forces / approached?
 Who / had / competently / deduced / whether / the private / assisted / the medic / as / the enemy / forces / approached?
 Who / had / the medic / deduced / that / the private / assisted / competently / as / the enemy / forces / approached?
 Who / had / the medic / deduced / whether / the private / assisted / competently / as / the enemy / forces / approached?
- 10 Who / had / mistakenly / contended / that / the baker / called / the confectioner / although / the phone / was / disconnected?
 Who / had / mistakenly / wondered / whether / the baker / called / the confectioner / although / the phone / was / disconnected?
 Who / had / the confectioner / contended / that / the baker / called / mistakenly / although / the phone / was / disconnected?
 Who / had / the confectioner / wondered / whether / the baker / called / mistakenly / although / the phone / was / disconnected?
- 11 Who / had / abruptly / contended / that / the landscaper / fired / the gardener / when / the roses / began / wilting?
 Who / had / abruptly / wondered / whether / the landscaper / fired / the gardener / when / the roses / began / wilting?
 Who / had / the gardener / contended / that / the landscaper / fired / abruptly / when / the roses / began / wilting?
 Who / had / the gardener / wondered / whether / the landscaper / fired / abruptly / when / the roses / began / wilting?

- 12 Who / had / unfairly / declared / that / the agency / defunded / the painter / because / the work / was / controversial?
 Who / had / sadly / speculated / whether / the agency / defunded / the painter / because / the work / was / controversial?
 Who / had / the painter / declared / that / the agency / defunded / unfairly / because / the work / was / controversial?
 Who / had / the painter / speculated / whether / the agency / defunded / unfairly / because / the work / was / controversial?
- 13 Who / had / meekly / contended / that / the stewardess / kissed / the aviator / before / the private / plane / departed?
 Who / had / meekly / wondered / whether / the stewardess / kissed / the aviator / before / the private / plane / departed?
 Who / had / the aviator / contended / that / the stewardess / kissed / meekly / before / the private / plane / departed?
 Who / had / the aviator / wondered / whether / the stewardess / kissed / meekly / before / the private / plane / departed?
- 14 Who / had / eagerly / contended / that / the defendant / punched / the solicitor / as / the aggressive / questions / continued?
 Who / had / eagerly / wondered / whether / the defendant / punched / the solicitor / as / the aggressive / questions / continued?
 Who / had / the solicitor / contended / that / the defendant / punched / eagerly / as / the aggressive / questions / continued?
 Who / had / the solicitor / wondered / whether / the defendant / punched / eagerly / as / the aggressive / questions / continued?
- 15 Who / had / moodily / declared / that / the chief / scolded / the fireman / when / the unfavorable / report / surfaced?
 Who / had / moodily / speculated / whether / the chief / scolded / the fireman / when / the unfavorable / report / surfaced?
 Who / had / the fireman / declared / that / the chief / scolded / moodily / when / the unfavorable / report / surfaced?
 Who / had / the fireman / speculated / whether / the chief / scolded / moodily / when / the unfavorable / report / surfaced?
- 16 Who / had / swiftly / deduced / that / the thief / blamed / the informant / after / the local / authorities / arrived?
 Who / had / swiftly / deduced / whether / the thief / blamed / the informant / after / the local / authorities / arrived?
 Who / had / the informant / deduced / that / the thief / blamed / swiftly / after / the local / authorities / arrived?
 Who / had / the informant / deduced / whether / the thief / blamed / swiftly / after / the local / authorities / arrived?
- 17 Who / had / unjustly / assumed / that / the carpenter / accused / the roofer / when / the deadline / was / missed?
 Who / had / unjustly / inquired / whether / the carpenter / accused / the roofer / when / the deadline / was / missed?
 Who / had / the roofer / assumed / that / the carpenter / accused / unjustly / when / the deadline / was / missed?

- Who / had / the roofer / inquired / whether / the carpenter / accused / unjustly / when / the deadline / was / missed?
- 18 Who / had / imprudently / assumed / that / the manager / demoted / the salesman / after / the wild / office / party?
 Who / had / imprudently / inquired / whether / the manager / demoted / the salesman / after / the wild / office / party?
 Who / had / the salesman / assumed / that / the manager / demoted / imprudently / after / the wild / office / party?
 Who / had / the salesman / inquired / whether / the manager / demoted / imprudently / after / the wild / office / party?
- 19 Who / had / outrageously / declared / that / the artist / threatened / the critic / while / the sculpture / was / panned?
 Who / had / outrageously / speculated / whether / the artist / threatened / the critic / while / the sculpture / was / panned?
 Who / had / the critic / declared / that / the artist / threatened / outrageously / while / the sculpture / was / panned?
 Who / had / the critic / speculated / whether / the artist / threatened / outrageously / while / the sculpture / was / panned?
- 20 Who / had / accurately / deduced / that / the janitor / advised / the mechanic / when / the sprinklers / started / leaking?
 Who / had / accurately / deduced / whether / the janitor / advised / the mechanic / when / the sprinklers / started / leaking?
 Who / had / the mechanic / deduced / that / the janitor / advised / accurately / when / the sprinklers / started / leaking?
 Who / had / the mechanic / deduced / whether / the janitor / advised / accurately / when / the sprinklers / started / leaking?
- 21 Who / had / thoughtlessly / contended / that / the reporter / agitated / the photographer / while / the recorder / was / running?
 Who / had / thoughtlessly / wondered / whether / the reporter / agitated / the photographer / while / the recorder / was / running?
 Who / had / the photographer / contended / that / the reporter / agitated / thoughtlessly / while / the recorder / was / running?
 Who / had / the photographer / wondered / whether / the reporter / agitated / thoughtlessly / while / the recorder / was / running?
- 22 Who / had / timidly / contended / that / the clerk / obeyed / the supervisor / when / the paychecks / went / missing?
 Who / had / timidly / wondered / whether / the clerk / obeyed / the supervisor / when / the paychecks / went / missing?
 Who / had / the supervisor / contended / that / the clerk / obeyed / timidly / when / the paychecks / went / missing?
 Who / had / the supervisor / wondered / whether / the clerk / obeyed / timidly / when / the paychecks / went / missing?
- 23 Who / had / irresponsibly / declared / that / the editor / saved / the ghostwriter / before / the plagiarism / was / uncovered?
 Who / had / irresponsibly / speculated / whether / the editor / saved / the ghostwriter / before / the plagiarism / was / uncovered?

- Who / had / the ghostwriter / declared / that / the editor / saved / irresponsibly / before / the plagiarism / was / uncovered?
 Who / had / the ghostwriter / speculated / whether / the editor / saved / irresponsibly / before / the plagiarism / was / uncovered?
- 24 Who / had / purposefully / assumed / that / the wrestler / struck / the referee / as / the crowd / went / wild?
 Who / had / purposefully / inquired / whether / the wrestler / struck / the referee / as / the crowd / went / wild?
 Who / had / the referee / assumed / that / the wrestler / struck / purposefully / as / the crowd / went / wild?
 Who / had / the referee / inquired / whether / the wrestler / struck / purposefully / as / the crowd / went / wild?
- 25 Who / had / specifically / said / that / the superintendent / berated / the comptroller / although / the project / was / completed?
 Who / had / specifically / said / whether / the superintendent / berated / the comptroller / although / the project / was / completed?
 Who / had / the comptroller / said / that / the superintendent / berated / specifically / although / the project / was / completed?
 Who / had / the comptroller / said / whether / the superintendent / berated / specifically / although / the project / was / completed?
- 26 Who / had / summarily / contended / that / the dentist / dismissed / the hygienist / after / the X-ray / machine / malfunctioned?
 Who / had / summarily / wondered / whether / the dentist / dismissed / the hygienist / after / the X-ray / machine / malfunctioned?
 Who / had / the hygienist / contended / that / the dentist / dismissed / summarily / after / the X-ray / machine / malfunctioned?
 Who / had / the hygienist / wondered / whether / the dentist / dismissed / summarily / after / the X-ray / machine / malfunctioned?
- 27 Who / had / viciously / declared / that / the waiter / mistreated / the busboy / because / the tips / were / misplaced?
 Who / had / viciously / speculated / whether / the waiter / mistreated / the busboy / because / the tips / were / misplaced?
 Who / had / the busboy / declared / that / the waiter / mistreated / viciously / because / the tips / were / misplaced?
 Who / had / the busboy / speculated / whether / the waiter / mistreated / viciously / because / the tips / were / misplaced?
- 28 Who / had / blatantly / assumed / that / the conductor / humiliated / the soloist / after / the final / dress / rehearsal?
 Who / had / blatantly / inquired / whether / the conductor / humiliated / the soloist / after / the final / dress / rehearsal?
 Who / had / the soloist / assumed / that / the conductor / humiliated / blatantly / after / the final / dress / rehearsal?
 Who / had / the soloist / inquired / whether / the conductor / humiliated / blatantly / after / the final / dress / rehearsal?
- 29 Who / had / recently / declared / that / the president / met / the chaplain / although / the press / was / unaware?

- Who / had / recently / speculated / whether / the president / met / the chaplain / although / the press / was / unaware?
 Who / had / the chaplain / declared / that / the president / met / recently / although / the press / was / unaware?
 Who / had / the chaplain / speculated / whether / the president / met / recently / although / the press / was / unaware?
- 30 Who / had / intentionally / declared / that / the crew / ignored / the navigator / before / the dangerous / storm / hit?
 Who / had / intentionally / speculated / whether / the crew / ignored / the navigator / before / the dangerous / storm / hit?
 Who / had / the navigator / declared / that / the crew / ignored / intentionally / before / the dangerous / storm / hit?
 Who / had / the navigator / speculated / whether / the crew / ignored / intentionally / before / the dangerous / storm / hit?
- 31 Who / had / feverishly / assumed / that / the publicist / texted / the press / after / the scandal / had / broken?
 Who / had / feverishly / inquired / whether / the publicist / texted / the press / after / the scandal / had / broken?
 Who / had / the press / assumed / that / the publicist / texted / feverishly / after / the scandal / had / broken?
 Who / had / the press / inquired / whether / the publicist / texted / feverishly / after / the scandal / had / broken?
- 32 Who / had / promptly / said / that / the teacher / notified / the magistracy / after / the violent / student / incident?
 Who / had / promptly / said / whether / the teacher / notified / the magistracy / after / the violent / student / incident?
 Who / had / the magistracy / said / that / the teacher / notified / promptly / after / the violent / student / incident?
 Who / had / the magistracy / said / whether / the teacher / notified / promptly / after / the violent / student / incident?
- 33 Who / had / hastily / assumed / that / the accountant / contacted / the treasurer / after / the errors / were / identified?
 Who / had / hastily / inquired / whether / the accountant / contacted / the treasurer / after / the errors / were / identified?
 Who / had / the treasurer / assumed / that / the accountant / contacted / hastily / after / the errors / were / identified?
 Who / had / the treasurer / inquired / whether / the accountant / contacted / hastily / after / the errors / were / identified?
- 34 Who / had / blithely / assumed / that / the school / defended / the jocks / although / the evidence / was / daunting?
 Who / had / blithely / inquired / whether / the school / defended / the jocks / although / the evidence / was / daunting?
 Who / had / the jocks / assumed / that / the school / defended / blithely / although / the evidence / was / daunting?
 Who / had / the jocks / inquired / whether / the school / defended / blithely / although / the evidence / was / daunting?

- 35 Who / had / nervously / assumed / that / the suspect / e-mailed / the eyewitness / before / the restraining / order / expired?
 Who / had / nervously / inquired / whether / the suspect / e-mailed / the eyewitness / before / the restraining / order / expired?
 Who / had / the eyewitness / assumed / that / the suspect / e-mailed / nervously / before / the restraining / order / expired?
 Who / had / the eyewitness / inquired / whether / the suspect / e-mailed / nervously / before / the restraining / order / expired?
- 36 Who / had / resentfully / assumed / that / the bartender / insulted / the patrons / since / the woman / was / crying?
 Who / had / resentfully / inquired / whether / the bartender / insulted / the patrons / since / the woman / was / crying?
 Who / had / the patrons / assumed / that / the bartender / insulted / resentfully / since / the woman / was / crying?
 Who / had / the patrons / inquired / whether / the bartender / insulted / resentfully / since / the woman / was / crying?
- 37 Who / had / quietly / assumed / that / the neighbors / discharged / the babysitter / because / the affair / was / discovered?
 Who / had / quietly / inquired / whether / the neighbors / discharged / the babysitter / because / the affair / was / discovered?
 Who / had / the babysitter / assumed / that / the neighbors / discharged / quietly / because / the affair / was / discovered?
 Who / had / the babysitter / inquired / whether / the neighbors / discharged / quietly / because / the affair / was / discovered?
- 38 Who / had / doggedly / assumed / that / the guard / chased / the racketeer / when / the shopkeeper / needed / assistance?
 Who / had / doggedly / inquired / whether / the guard / chased / the racketeer / when / the shopkeeper / needed / assistance?
 Who / had / the racketeer / assumed / that / the guard / chased / doggedly / when / the shopkeeper / needed / assistance?
 Who / had / the racketeer / inquired / whether / the guard / chased / doggedly / when / the shopkeeper / needed / assistance?
- 39 Who / had / repeatedly / assumed / that / the swindler / cheated / the investor / because / the funds / kept / disappearing?
 Who / had / repeatedly / inquired / whether / the swindler / cheated / the investor / because / the funds / kept / disappearing?
 Who / had / the investor / assumed / that / the swindler / cheated / repeatedly / because / the funds / kept / disappearing?
 Who / had / the investor / inquired / whether / the swindler / cheated / repeatedly / because / the funds / kept / disappearing?
- 40 Who / had / sympathetically / assumed / that / the bridesmaid / consoled / the bride / when / the ceremony / was / delayed?
 Who / had / sympathetically / inquired / whether / the bridesmaid / consoled / the bride / when / the ceremony / was / delayed?
 Who / had / the bride / assumed / that / the bridesmaid / consoled / sympathetically / when / the ceremony / was / delayed?

- Who / had / the bride / inquired / whether / the bridesmaid / consoled / sympathetically / when / the ceremony / was / delayed?
- 41 Who / had / loudly / said / that / the coach / reprimanded / the quarterback / after / the ball / was / intercepted?
 Who / had / loudly / said / whether / the coach / reprimanded / the quarterback / after / the ball / was / intercepted?
 Who / had / the quarterback / said / that / the coach / reprimanded / loudly / after / the ball / was / intercepted?
 Who / had / the quarterback / said / whether / the coach / reprimanded / loudly / after / the ball / was / intercepted?
- 42 Who / had / officially / said / that / the highschooler / outran / the undergrad / before / the error / was / discovered?
 Who / had / officially / said / whether / the highschooler / outran / the undergrad / before / the error / was / discovered?
 Who / had / the undergrad / said / that / the highschooler / outran / officially / before / the error / was / discovered?
 Who / had / the undergrad / said / whether / the highschooler / outran / officially / before / the error / was / discovered?
- 43 Who / had / gently / said / that / the mortician / painted / the godparent / before / the funeral / guests / arrived?
 Who / had / gently / said / whether / the mortician / painted / the godparent / before / the funeral / guests / arrived?
 Who / had / the godparent / said / that / the mortician / painted / gently / before / the funeral / guests / arrived?
 Who / had / the godparent / said / whether / the mortician / painted / gently / before / the funeral / guests / arrived?
- 44 Who / had / knowingly / said / that / the telemarketer / deceived / the pensioner / since / the product / never / arrived?
 Who / had / knowingly / said / whether / the telemarketer / deceived / the pensioner / since / the product / never / arrived?
 Who / had / the pensioner / said / that / the telemarketer / deceived / knowingly / since / the product / never / arrived?
 Who / had / the pensioner / said / whether / the telemarketer / deceived / knowingly / since / the product / never / arrived?
- 45 Who / had / reassuringly / declared / that / the counselor / encouraged / the classmates / after / the traumatic / school / shooting?
 Who / had / reassuringly / speculated / whether / the counselor / encouraged / the classmates / after / the traumatic / school / shooting?
 Who / had / the classmates / declared / that / the counselor / encouraged / reassuringly / after / the traumatic / school / shooting?
 Who / had / the classmates / speculated / whether / the counselor / encouraged / reassuringly / after / the traumatic / school / shooting?
- 46 Who / had / gracefully / declared / that / the politician / impressed / the media / during / the large / press / conference?
 Who / had / gracefully / speculated / whether / the politician / impressed / the media / during / the large / press / conference?

- Who / had / the tabloids / declared / that / the politician / impressed / gracefully / during / the large / press / conference?
 Who / had / the tabloids / speculated / whether / the politician / impressed / gracefully / during / the large / press / conference?
- 47 Who / had / urgently / declared / that / the rookie / inspired / the pitcher / because / the game / was / close?
 Who / had / urgently / speculated / whether / the rookie / inspired / the pitcher / because / the game / was / close?
 Who / had / the pitcher / declared / that / the rookie / inspired / urgently / because / the game / was / close?
 Who / had / the pitcher / speculated / whether / the rookie / inspired / urgently / because / the game / was / close?
- 48 Who / had / unfortunately / declared / that / the priest / hid / the snitch / while / the police / search / continued?
 Who / had / unfortunately / speculated / whether / the priest / hid / the snitch / while / the police / search / continued?
 Who / had / the snitch / declared / that / the priest / hid / unfortunately / while / the police / search / continued?
 Who / had / the snitch / speculated / whether / the priest / hid / unfortunately / while / the police / search / continued?
- 49 Who / had / briefly / declared / that / the gymnast / awed / the commentator / after / the impressive / vault / routine?
 Who / had / briefly / speculated / whether / the gymnast / awed / the commentator / after / the impressive / vault / routine?
 Who / had / the commentator / declared / that / the gymnast / awed / briefly / after / the impressive / vault / routine?
 Who / had / the commentator / speculated / whether / the gymnast / awed / briefly / after / the impressive / vault / routine?
- 50 Who / had / quickly / declared / that / the bishop / concealed / the fugitive / before / the angry / mob / arrived?
 Who / had / quickly / speculated / whether / the bishop / concealed / the fugitive / before / the angry / mob / arrived?
 Who / had / the fugitive / declared / that / the bishop / concealed / quickly / before / the angry / mob / arrived?
 Who / had / the fugitive / speculated / whether / the bishop / concealed / quickly / before / the angry / mob / arrived?
- 51 Who / had / appropriately / declared / that / the prosecutor / subpoenaed / the bureaucrat / as / the case / was / opening?
 Who / had / appropriately / speculated / whether / the prosecutor / subpoenaed / the bureaucrat / as / the case / was / opening?
 Who / had / the bureaucrat / declared / that / the prosecutor / subpoenaed / appropriately / as / the case / was / opening?
 Who / had / the bureaucrat / speculated / whether / the prosecutor / subpoenaed / appropriately / as / the case / was / opening?
- 52 Who / had / foolishly / declared / that / the governor / appointed / the donor / since / the position / wasn't / open?

- Who / had / foolishly / speculated / whether / the governor / appointed / the donor / since / the position / wasn't / open?
 Who / had / the donor / declared / that / the governor / appointed / foolishly / since / the position / wasn't / open?
 Who / had / the donor / speculated / whether / the governor / appointed / foolishly / since / the position / wasn't / open?
- 53 Who / had / coolly / deduced / that / the forger / framed / the typesetter / when / the documents / began / fading?
 Who / had / coolly / deduced / whether / the forger / framed / the typesetter / when / the documents / began / fading?
 Who / had / the typesetter / deduced / that / the forger / framed / coolly / when / the documents / began / fading?
 Who / had / the typesetter / deduced / whether / the forger / framed / coolly / when / the documents / began / fading?
- 54 Who / had / methodically / deduced / that / the revolutionaries / assaulted / the regime / when / the gunfire / was / heard?
 Who / had / methodically / deduced / whether / the revolutionaries / assaulted / the regime / when / the gunfire / was / heard?
 Who / had / the regime / deduced / that / the revolutionaries / assaulted / methodically / when / the gunfire / was / heard?
 Who / had / the regime / deduced / whether / the revolutionaries / assaulted / methodically / when / the gunfire / was / heard?
- 55 Who / had / spitefully / deduced / that / the psychiatrist / deceived / the insurer / because / the scam / was / personal?
 Who / had / spitefully / deduced / whether / the psychiatrist / deceived / the insurer / because / the scam / was / personal?
 Who / had / the insurer / deduced / that / the psychiatrist / deceived / spitefully / because / the scam / was / personal?
 Who / had / the insurer / deduced / whether / the psychiatrist / deceived / spitefully / because / the scam / was / personal?
- 56 Who / had / helpfully / deduced / that / the guest / accompanied / the hostess / when / the suspicious / package / arrived?
 Who / had / helpfully / deduced / whether / the guest / accompanied / the hostess / when / the suspicious / package / arrived?
 Who / had / the hostess / deduced / that / the guest / accompanied / helpfully / when / the suspicious / package / arrived?
 Who / had / the hostess / deduced / whether / the guest / accompanied / helpfully / when / the suspicious / package / arrived?
- 57 Who / had / excitedly / contended / that / the champion / surpassed / the challenger / before / the score / was / official?
 Who / had / excitedly / wondered / whether / the champion / surpassed / the challenger / before / the score / was / official?
 Who / had / the challenger / contended / that / the champion / surpassed / excitedly / before / the score / was / official?
 Who / had / the challenger / wondered / whether / the champion / surpassed / excitedly / before / the score / was / official?

- 58 Who / had / justly / contended / that / the landlord / misled / the constable / before / the criminal / gang / returned?
 Who / had / justly / wondered / whether / the landlord / misled / the constable / before / the criminal / gang / returned?
 Who / had / the constable / contended / that / the landlord / misled / justly / before / the criminal / gang / returned?
 Who / had / the constable / wondered / whether / the landlord / misled / justly / before / the criminal / gang / returned?
- 59 Who / had / warily / contended / that / the skycap / scammed / the globetrotter / although / the airline / denied / it?
 Who / had / warily / wondered / whether / the skycap / scammed / the globetrotter / although / the airline / denied / it?
 Who / had / the globetrotter / contended / that / the skycap / scammed / warily / although / the airline / denied / it?
 Who / had / the globetrotter / wondered / whether / the skycap / scammed / warily / although / the airline / denied / it?
- 60 Who / had / naturally / contended / that / the conman / duped / the tourist / since / the game / seemed / unwinnable?
 Who / had / naturally / wondered / whether / the conman / duped / the tourist / since / the game / seemed / unwinnable?
 Who / had / the tourist / contended / that / the conman / duped / naturally / since / the game / seemed / unwinnable?
 Who / had / the tourist / wondered / whether / the conman / duped / naturally / since / the game / seemed / unwinnable?
- 61 Who / had / harshly / contended / that / the supplier / gouged / the purchaser / since / the bill / was / unreasonable?
 Who / had / harshly / wondered / whether / the supplier / gouged / the purchaser / since / the bill / was / unreasonable?
 Who / had / the purchaser / contended / that / the supplier / gouged / harshly / since / the bill / was / unreasonable?
 Who / had / the purchaser / wondered / whether / the supplier / gouged / harshly / since / the bill / was / unreasonable?
- 62 Who / had / lovingly / contended / that / the caretaker / hugged / the orphan / when / the bad / news / arrived?
 Who / had / lovingly / wondered / whether / the caretaker / hugged / the orphan / when / the bad / news / arrived?
 Who / had / the orphan / contended / that / the caretaker / hugged / lovingly / when / the bad / news / arrived?
 Who / had / the orphan / wondered / whether / the caretaker / hugged / lovingly / when / the bad / news / arrived?
- 63 Who / had / surreptitiously / contended / that / the author / aided / the reviewer / because / the essay / was / uncharacteristic?
 Who / had / surreptitiously / wondered / whether / the author / aided / the reviewer / because / the essay / was / uncharacteristic?
 Who / had / the reviewer / contended / that / the author / aided / surreptitiously / because / the essay / was / uncharacteristic?

- Who / had / the reviewer / wondered / whether / the author / aided / surreptitiously / because / the essay / was / uncharacteristic?
- 64 Who / had / defiantly / contended / that / the accomplice / stalled / the bookkeeper / while / the rival / copied / documents?
 Who / had / defiantly / wondered / whether / the accomplice / stalled / the bookkeeper / while / the rival / copied / documents?
 Who / had / the bookkeeper / contended / that / the accomplice / stalled / defiantly / while / the rival / copied / documents?
 Who / had / the bookkeeper / wondered / whether / the accomplice / stalled / defiantly / while / the rival / copied / documents?
- 65 Who / had / covertly / assumed / that / the journalist / interviewed / the delegate / after / the suspicious / story / leaked?
 Who / had / covertly / inquired / whether / the journalist / interviewed / the delegate / after / the suspicious / story / leaked?
 Who / had / the delegate / assumed / that / the journalist / interviewed / covertly / after / the suspicious / story / leaked?
 Who / had / the delegate / inquired / whether / the journalist / interviewed / covertly / after / the suspicious / story / leaked?
- 66 Who / had / boldly / assumed / that / the apprentice / confronted / the foreman / after / the first / paycheck / bounced?
 Who / had / boldly / inquired / whether / the apprentice / confronted / the foreman / after / the first / paycheck / bounced?
 Who / had / the foreman / assumed / that / the apprentice / confronted / boldly / after / the first / paycheck / bounced?
 Who / had / the foreman / inquired / whether / the apprentice / confronted / boldly / after / the first / paycheck / bounced?
- 67 Who / had / accidentally / assumed / that / the anchor / informed / the meteorologist / since / the information / was / wrong?
 Who / had / accidentally / inquired / whether / the anchor / informed / the meteorologist / since / the information / was / wrong?
 Who / had / the meteorologist / assumed / that / the anchor / informed / accidentally / since / the information / was / wrong?
 Who / had / the meteorologist / inquired / whether / the anchor / informed / accidentally / since / the information / was / wrong?
- 68 Who / had / arrogantly / assumed / that / the doorman / snubbed / the chauffeur / since / the lobby / was / dark?
 Who / had / arrogantly / inquired / whether / the doorman / snubbed / the chauffeur / since / the lobby / was / dark?
 Who / had / the chauffeur / assumed / that / the doorman / snubbed / arrogantly / since / the lobby / was / dark?
 Who / had / the chauffeur / inquired / whether / the doorman / snubbed / arrogantly / since / the lobby / was / dark?
- 69 Who / had / charmingly / assumed / that / the tutor / motivated / the schoolmates / as / the grades / began / improving?
 Who / had / charmingly / inquired / whether / the tutor / motivated / the schoolmates / as / the grades / began / improving?

- Who / had / the schoolmates / assumed / that / the tutor / motivated / charmingly / as / the grades / began / improving?
 Who / had / the schoolmates / inquired / whether / the tutor / motivated / charmingly / as / the grades / began / improving?
- 70 Who / had / awkwardly / assumed / that / the gambler / misinformed / the bookie / when / the race / was / cancelled?
 Who / had / awkwardly / inquired / whether / the gambler / misinformed / the bookie / when / the race / was / cancelled?
 Who / had / the bookie / assumed / that / the gambler / misinformed / awkwardly / when / the race / was / cancelled?
 Who / had / the bookie / inquired / whether / the gambler / misinformed / awkwardly / when / the race / was / cancelled?
- 71 Who / had / innocently / assumed / that / the substitute / provoked / the sophomore / because / the word / was / ambiguous?
 Who / had / innocently / inquired / whether / the substitute / provoked / the sophomore / because / the word / was / ambiguous?
 Who / had / the sophomore / assumed / that / the substitute / provoked / innocently / because / the word / was / ambiguous?
 Who / had / the sophomore / inquired / whether / the substitute / provoked / innocently / because / the word / was / ambiguous?
- 72 Who / had / cheerfully / assumed / that / the organist / visited / the parishioner / while / the minister / was / unavailable?
 Who / had / cheerfully / inquired / whether / the organist / visited / the parishioner / while / the minister / was / unavailable?
 Who / had / the parishioner / assumed / that / the organist / visited / cheerfully / while / the minister / was / unavailable?
 Who / had / the parishioner / inquired / whether / the organist / visited / cheerfully / while / the minister / was / unavailable?
- 73 Who / had / warmly / said / that / the plumber / assisted / the electrician / after / the storm / blew / through?
 Who / had / warmly / said / whether / the plumber / assisted / the electrician / after / the storm / blew / through?
 Who / had / the electrician / said / that / the plumber / assisted / warmly / after / the storm / blew / through?
 Who / had / the electrician / said / whether / the plumber / assisted / warmly / after / the storm / blew / through?
- 74 Who / had / deliberately / said / that / the executive / betrayed / the financier / before / the stock / price / tumbled?
 Who / had / deliberately / said / whether / the executive / betrayed / the financier / before / the stock / price / tumbled?
 Who / had / the financier / said / that / the executive / betrayed / deliberately / before / the stock / price / tumbled?
 Who / had / the financier / said / whether / the executive / betrayed / deliberately / before / the stock / price / tumbled?
- 75 Who / had / clearly / said / that / the administrator / greeted / the receptionist / after / the talk / about / rudeness?

- Who / had / clearly / said / whether / the administrator / greeted / the receptionist / after / the talk / about / rudeness?
 Who / had / the receptionist / said / that / the administrator / greeted / clearly / after / the talk / about / rudeness?
 Who / had / the receptionist / said / whether / the administrator / greeted / clearly / after / the talk / about / rudeness?
- 76 Who / had / easily / said / that / the grappler / overwhelmed / the striker / since / the match / ended / early?
 Who / had / easily / said / whether / the grappler / overwhelmed / the striker / since / the match / ended / early?
 Who / had / the striker / said / that / the grappler / overwhelmed / easily / since / the match / ended / early?
 Who / had / the striker / said / whether / the grappler / overwhelmed / easily / since / the match / ended / early?
- 77 Who / had / miserably / declared / that / the nominee / grilled / the incumbent / as / the press / lost / interest?
 Who / had / miserably / speculated / whether / the nominee / grilled / the incumbent / as / the press / lost / interest?
 Who / had / the incumbent / declared / that / the nominee / grilled / miserably / as / the press / lost / interest?
 Who / had / the incumbent / speculated / whether / the nominee / grilled / miserably / as / the press / lost / interest?
- 78 Who / had / kiddingly / declared / that / the grandfather / spooked / the teenager / when / the lights / went / out?
 Who / had / kiddingly / speculated / whether / the grandfather / spooked / the teenager / when / the lights / went / out?
 Who / had / the teenager / declared / that / the grandfather / spooked / kiddingly / when / the lights / went / out?
 Who / had / the teenager / speculated / whether / the grandfather / spooked / kiddingly / when / the lights / went / out?
- 79 Who / had / overconfidently / declared / that / the mobster / taunted / the patrolman / because / the evidence / looked / weak?
 Who / had / overconfidently / speculated / whether / the mobster / taunted / the patrolman / because / the evidence / looked / weak?
 Who / had / the patrolman / declared / that / the mobster / taunted / overconfidently / because / the evidence / looked / weak?
 Who / had / the patrolman / speculated / whether / the mobster / taunted / overconfidently / because / the evidence / looked / weak?
- 80 Who / had / unexpectedly / declared / that / the principal / expelled / the valedictorian / after / the lockers / were / searched?
 Who / had / unexpectedly / speculated / whether / the principal / expelled / the valedictorian / after / the lockers / were / searched?
 Who / had / the valedictorian / declared / that / the principal / expelled / unexpectedly / after / the lockers / were / searched?
 Who / had / the valedictorian / speculated / whether / the principal / expelled / unexpectedly / after / the lockers / were / searched?
- 81 Who / had / angrily / declared / that / the biographer / rebuked / the proofreader / after / the

typos / were / found?

Who / had / angrily / speculated / whether / the biographer / rebuked / the proofreader / after / the typos / were / found?

Who / had / the proofreader / declared / that / the biographer / rebuked / angrily / after / the typos / were / found?

Who / had / the proofreader / speculated / whether / the biographer / rebuked / angrily / after / the typos / were / found?

82 Who / had / gleefully / declared / that / the columnist / slandered / the congressperson / before / the story / was / printed?

Who / had / gleefully / speculated / whether / the columnist / slandered / the congressperson / before / the story / was / printed?

Who / had / the congressperson / declared / that / the columnist / slandered / gleefully / before / the story / was / printed?

Who / had / the congressperson / speculated / whether / the columnist / slandered / gleefully / before / the story / was / printed?

83 Who / had / furiously / declared / that / the major / disciplined / the lieutenant / although / the accusation / was / false?

Who / had / furiously / speculated / whether / the major / disciplined / the lieutenant / although / the accusation / was / false?

Who / had / the lieutenant / declared / that / the major / disciplined / furiously / although / the accusation / was / false?

Who / had / the lieutenant / speculated / whether / the major / disciplined / furiously / although / the accusation / was / false?

84 Who / had / frantically / declared / that / the attorney / investigated / the eyewitness / since / the trial / was / near?

Who / had / frantically / speculated / whether / the attorney / investigated / the eyewitness / since / the trial / was / near?

Who / had / the eyewitness / declared / that / the attorney / investigated / frantically / since / the trial / was / near?

Who / had / the eyewitness / speculated / whether / the attorney / investigated / frantically / since / the trial / was / near?

85 Who / had / clumsily / deduced / that / the congressman / questioned / the colonel / as / the hearing / ended / early?

Who / had / clumsily / deduced / whether / the congressman / questioned / the colonel / as / the hearing / ended / early?

Who / had / the colonel / deduced / that / the congressman / questioned / clumsily / as / the hearing / ended / early?

Who / had / the colonel / deduced / whether / the congressman / questioned / clumsily / as / the hearing / ended / early?

86 Who / had / carefully / deduced / that / the investigator / photographed / the mayor / when / the documents / were / exchanged?

Who / had / carefully / deduced / whether / the investigator / photographed / the mayor / when / the documents / were / exchanged?

Who / had / the mayor / deduced / that / the investigator / photographed / carefully / when / the documents / were / exchanged?

Who / had / the mayor / deduced / whether / the investigator / photographed / carefully / when / the documents / were / exchanged?

- 87 Who / had / suspiciously / deduced / that / the director / undermined / the thespian / because / the play / was / unprofitable?
 Who / had / suspiciously / deduced / whether / the director / undermined / the thespian / because / the play / was / unprofitable?
 Who / had / the thespian / deduced / that / the director / undermined / suspiciously / because / the play / was / unprofitable?
 Who / had / the thespian / deduced / whether / the director / undermined / suspiciously / because / the play / was / unprofitable?
- 88 Who / had / surprisingly / deduced / that / the mogul / murdered / the butler / while / the FBI / knew / nothing?
 Who / had / surprisingly / deduced / whether / the mogul / murdered / the butler / while / the FBI / knew / nothing?
 Who / had / the butler / deduced / that / the mogul / murdered / surprisingly / while / the FBI / knew / nothing?
 Who / had / the butler / deduced / whether / the mogul / murdered / surprisingly / while / the FBI / knew / nothing?
- 89 Who / had / hesitantly / contended / that / the producer / acknowledged / the actress / after / the embarrassing / paparazzi / fiasco?
 Who / had / hesitantly / wondered / whether / the producer / acknowledged / the actress / after / the embarrassing / paparazzi / fiasco?
 Who / had / the actress / contended / that / the producer / acknowledged / hesitantly / after / the embarrassing / paparazzi / fiasco?
 Who / had / the actress / wondered / whether / the producer / acknowledged / hesitantly / after / the embarrassing / paparazzi / fiasco?
- 90 Who / had / cautiously / contended / that / the lobbyist / bribed / the congresswoman / during / the close / election / campaign?
 Who / had / cautiously / wondered / whether / the lobbyist / bribed / the congresswoman / during / the close / election / campaign?
 Who / had / the congresswoman / contended / that / the lobbyist / bribed / cautiously / during / the close / election / campaign?
 Who / had / the congresswoman / wondered / whether / the lobbyist / bribed / cautiously / during / the close / election / campaign?
- 91 Who / had / mysteriously / contended / that / the spy / recruited / the ambassador / although / the government / denied / it?
 Who / had / mysteriously / wondered / whether / the spy / recruited / the ambassador / although / the government / denied / it?
 Who / had / the ambassador / contended / that / the spy / recruited / mysteriously / although / the government / denied / it?
 Who / had / the ambassador / wondered / whether / the spy / recruited / mysteriously / although / the government / denied / it?
- 92 Who / had / cruelly / contended / that / the cheerleaders / harassed / the geeks / after / the competition / was / lost?
 Who / had / cruelly / wondered / whether / the cheerleaders / harassed / the geeks / after / the competition / was / lost?
 Who / had / the geeks / contended / that / the cheerleaders / harassed / cruelly / after / the competition / was / lost?

- Who / had / the geeks / wondered / whether / the cheerleaders / harassed / cruelly / after / the competition / was / lost?
- 93 Who / had / rudely / contended / that / the youth / annoyed / the shopkeeper / as / the other / kids / laughed?
 Who / had / rudely / wondered / whether / the youth / annoyed / the shopkeeper / as / the other / kids / laughed?
 Who / had / the shopkeeper / contended / that / the youth / annoyed / rudely / as / the other / kids / laughed?
 Who / had / the shopkeeper / wondered / whether / the youth / annoyed / rudely / as / the other / kids / laughed?
- 94 Who / had / seriously / contended / that / the assistant / resented / the costumer / when / the press / coverage / spiked?
 Who / had / seriously / wondered / whether / the assistant / resented / the costumer / when / the press / coverage / spiked?
 Who / had / the costumer / contended / that / the assistant / resented / seriously / when / the press / coverage / spiked?
 Who / had / the costumer / wondered / whether / the assistant / resented / seriously / when / the press / coverage / spiked?
- 95 Who / had / obnoxiously / contended / that / the caddy / teased / the golfer / because / the shot / was / terrible?
 Who / had / obnoxiously / wondered / whether / the caddy / teased / the golfer / because / the shot / was / terrible?
 Who / had / the golfer / contended / that / the caddy / teased / obnoxiously / because / the shot / was / terrible?
 Who / had / the golfer / wondered / whether / the caddy / teased / obnoxiously / because / the shot / was / terrible?
- 96 Who / had / wisely / contended / that / the utility / placated / the activists / before / the new / plant / opened?
 Who / had / wisely / wondered / whether / the utility / placated / the activists / before / the new / plant / opened?
 Who / had / the activists / contended / that / the utility / placated / wisely / before / the new / plant / opened?
 Who / had / the activists / wondered / whether / the utility / placated / wisely / before / the new / plant / opened?
- 97 Who / had / diligently / assumed / that / the sergeant / interrogated / the deserter / after / the mission / had / failed?
 Who / had / diligently / inquired / whether / the sergeant / interrogated / the deserter / after / the mission / had / failed?
 Who / had / the deserter / assumed / that / the sergeant / interrogated / diligently / after / the mission / had / failed?
 Who / had / the deserter / inquired / whether / the sergeant / interrogated / diligently / after / the mission / had / failed?
- 98 Who / had / frankly / assumed / that / the capitalist / approached / the shareholder / before / the contract / was / signed?
 Who / had / frankly / inquired / whether / the capitalist / approached / the shareholder / before / the contract / was / signed?

Who / had / the shareholder / assumed / that / the capitalist / approached / frankly / before / the contract / was / signed?

Who / had / the shareholder / inquired / whether / the capitalist / approached / frankly / before / the contract / was / signed?

- 99 Who / had / kindheartedly / assumed / that / the nun / fed / the outcast / although / the convent / was / shuttered?

Who / had / kindheartedly / inquired / whether / the nun / fed / the outcast / although / the convent / was / shuttered?

Who / had / the outcast / assumed / that / the nun / fed / kindheartedly / although / the convent / was / shuttered?

Who / had / the outcast / inquired / whether / the nun / fed / kindheartedly / although / the convent / was / shuttered?

- 100 Who / had / thoughtfully / assumed / that / the volunteer / sheltered / the drifter / because / the temperature / was / plummeting?

Who / had / thoughtfully / inquired / whether / the volunteer / sheltered / the drifter / because / the temperature / was / plummeting?

Who / had / the drifter / assumed / that / the volunteer / sheltered / thoughtfully / because / the temperature / was / plummeting?

Who / had / the drifter / inquired / whether / the volunteer / sheltered / thoughtfully / because / the temperature / was / plummeting?

- 101 Who / had / wearily / assumed / that / the vagrant / begged / the landlady / as / the winter / storm / approached?

Who / had / wearily / inquired / whether / the vagrant / begged / the landlady / as / the winter / storm / approached?

Who / had / the landlady / assumed / that / the vagrant / begged / wearily / as / the winter / storm / approached?

Who / had / the landlady / inquired / whether / the vagrant / begged / wearily / as / the winter / storm / approached?

- 102 Who / had / fortuitously / assumed / that / the dispatcher / telephoned / the cripple / when / the evacuation / was / ordered?

Who / had / fortuitously / inquired / whether / the dispatcher / telephoned / the cripple / when / the evacuation / was / ordered?

Who / had / the cripple / assumed / that / the dispatcher / telephoned / fortuitously / when / the evacuation / was / ordered?

Who / had / the cripple / inquired / whether / the dispatcher / telephoned / fortuitously / when / the evacuation / was / ordered?

- 103 Who / had / patiently / assumed / that / the locksmith / helped / the pupil / when / the dormitory / was / locked?

Who / had / patiently / inquired / whether / the locksmith / helped / the pupil / when / the dormitory / was / locked?

Who / had / the pupil / assumed / that / the locksmith / helped / patiently / when / the dormitory / was / locked?

Who / had / the pupil / inquired / whether / the locksmith / helped / patiently / when / the dormitory / was / locked?

- 104 Who / had / worriedly / assumed / that / the orderly / watched / the spinster / when / the fever / wouldn't / break?

Who / had / worriedly / inquired / whether / the orderly / watched / the spinster / when / the fever / wouldn't / break?
 Who / had / the spinster / assumed / that / the orderly / watched / worriedly / when / the fever / wouldn't / break?
 Who / had / the spinster / inquired / whether / the orderly / watched / worriedly / when / the fever / wouldn't / break?

- 105 Who / had / persistently / said / that / the stockholder / pestered / the broker / after / the stock / market / crashed?
 Who / had / persistently / said / whether / the stockholder / pestered / the broker / after / the stock / market / crashed?
 Who / had / the broker / said / that / the stockholder / pestered / persistently / after / the stock / market / crashed?
 Who / had / the broker / said / whether / the stockholder / pestered / persistently / after / the stock / market / crashed?
- 106 Who / had / carelessly / said / that / the bailiff / released / the inmate / before / the order / was / given?
 Who / had / carelessly / said / whether / the bailiff / released / the inmate / before / the order / was / given?
 Who / had / the inmate / said / that / the bailiff / released / carelessly / before / the order / was / given?
 Who / had / the inmate / said / whether / the bailiff / released / carelessly / before / the order / was / given?
- 107 Who / had / unintentionally / said / that / the teller / bothered / the lender / during / the tedious / branch / meeting?
 Who / had / unintentionally / said / whether / the teller / bothered / the lender / during / the tedious / branch / meeting?
 Who / had / the lender / said / that / the teller / bothered / unintentionally / during / the tedious / branch / meeting?
 Who / had / the lender / said / whether / the teller / bothered / unintentionally / during / the tedious / branch / meeting?
- 108 Who / had / correctly / said / that / the pundit / identified / the whistleblower / although / the information / was / confidential?
 Who / had / correctly / said / whether / the pundit / identified / the whistleblower / although / the information / was / confidential?
 Who / had / the whistleblower / said / that / the pundit / identified / correctly / although / the information / was / confidential?
 Who / had / the whistleblower / said / whether / the pundit / identified / correctly / although / the information / was / confidential?
- 109 Who / had / irritably / declared / that / the panhandler / nagged / the pedestrian / as / the lunch / hour / began?
 Who / had / irritably / speculated / whether / the panhandler / nagged / the pedestrian / as / the lunch / hour / began?
 Who / had / the pedestrian / declared / that / the panhandler / nagged / irritably / as / the lunch / hour / began?
 Who / had / the pedestrian / speculated / whether / the panhandler / nagged / irritably / as / the lunch / hour / began?

- 110 Who / had / irately / declared / that / the beggar / intimidated / the commuter / when / the bus / was / delayed?
 Who / had / irately / speculated / whether / the beggar / intimidated / the commuter / when / the bus / was / delayed?
 Who / had / the commuter / declared / that / the beggar / intimidated / irately / when / the bus / was / delayed?
 Who / had / the commuter / speculated / whether / the beggar / intimidated / irately / when / the bus / was / delayed?
- 111 Who / had / fearfully / declared / that / the housewife / incriminated / the deliveryman / because / the valuables / were / missing?
 Who / had / fearfully / speculated / whether / the housewife / incriminated / the deliveryman / because / the valuables / were / missing?
 Who / had / the deliveryman / declared / that / the housewife / incriminated / fearfully / because / the valuables / were / missing?
 Who / had / the deliveryman / speculated / whether / the housewife / incriminated / fearfully / because / the valuables / were / missing?
- 112 Who / had / cleverly / declared / that / the hustler / scammed / the passerby / after / the patrolwoman / had / departed?
 Who / had / cleverly / speculated / whether / the hustler / scammed / the passerby / after / the patrolwoman / had / departed?
 Who / had / the passerby / declared / that / the hustler / scammed / cleverly / after / the patrolwoman / had / departed?
 Who / had / the passerby / speculated / whether / the hustler / scammed / cleverly / after / the patrolwoman / had / departed?
- 113 Who / had / proudly / declared / that / the novice / defeated / the gymnast / after / the judges' / final / scores?
 Who / had / proudly / speculated / whether / the novice / defeated / the gymnast / after / the judges' / final / scores?
 Who / had / the gymnast / declared / that / the novice / defeated / proudly / after / the judges' / final / scores?
 Who / had / the gymnast / speculated / whether / the novice / defeated / proudly / after / the judges' / final / scores?
- 114 Who / had / condescendingly / declared / that / the illusionist / tricked / the emcee / before / the curtains / were / opened?
 Who / had / condescendingly / speculated / whether / the illusionist / tricked / the emcee / before / the curtains / were / opened?
 Who / had / the emcee / declared / that / the illusionist / tricked / condescendingly / before / the curtains / were / opened?
 Who / had / the emcee / speculated / whether / the illusionist / tricked / condescendingly / before / the curtains / were / opened?
- 115 Who / had / casually / declared / that / the captain / saluted / the longshoreman / although / the protocol / was / unnecessary?
 Who / had / casually / speculated / whether / the captain / saluted / the longshoreman / although / the protocol / was / unnecessary?
 Who / had / the longshoreman / declared / that / the captain / saluted / casually / although / the protocol / was / unnecessary?

Who / had / the longshoreman / speculated / whether / the captain / saluted / casually /
although / the protocol / was / unnecessary?

116 Who / had / fearlessly / declared / that / the courier / battled / the thug / since / the package /
was / vital?

Who / had / fearlessly / speculated / whether / the courier / battled / the thug / since / the
package / was / vital?

Who / had / the thug / declared / that / the courier / battled / fearlessly / since / the package /
was / vital?

Who / had / the thug / speculated / whether / the courier / battled / fearlessly / since / the
package / was / vital?

117 Who / had / haphazardly / deduced / that / the custodian / hindered / the trespasser / as / the
floors / were / slippery?

Who / had / haphazardly / deduced / whether / the custodian / hindered / the trespasser / as /
the floors / were / slippery?

Who / had / the trespasser / deduced / that / the custodian / hindered / haphazardly / as / the
floors / were / slippery?

Who / had / the trespasser / deduced / whether / the custodian / hindered / haphazardly / as /
the floors / were / slippery?

118 Who / had / graciously / deduced / that / the duke / rewarded / the servant / after / the term / of
/ employment?

Who / had / graciously / deduced / whether / the duke / rewarded / the servant / after / the
term / of / employment?

Who / had / the servant / deduced / that / the duke / rewarded / graciously / after / the term / of
/ employment?

Who / had / the servant / deduced / whether / the duke / rewarded / graciously / after / the
term / of / employment?

119 Who / had / cunningly / deduced / that / the visitor / hoodwinked / the sentry / because / the
key / was / stolen?

Who / had / cunningly / deduced / whether / the visitor / hoodwinked / the sentry / because /
the key / was / stolen?

Who / had / the sentry / deduced / that / the visitor / hoodwinked / cunningly / because / the
key / was / stolen?

Who / had / the sentry / deduced / whether / the visitor / hoodwinked / cunningly / because /
the key / was / stolen?

120 Who / had / discretely / deduced / that / the diplomat / deceived / the chancellor / while / the
cabinet / was / meeting?

Who / had / discretely / deduced / whether / the diplomat / deceived / the chancellor / while /
the cabinet / was / meeting?

Who / had / the chancellor / deduced / that / the diplomat / deceived / discretely / while / the
cabinet / was / meeting?

Who / had / the chancellor / deduced / whether / the diplomat / deceived / discretely / while /
the cabinet / was / meeting?

121 Who / had / properly / contended / that / the usher / seated / the contributor / after / the
orchestra / began / playing?

Who / had / properly / wondered / whether / the usher / seated / the contributor / after / the
orchestra / began / playing?

Who / had / the contributor / contended / that / the usher / seated / properly / after / the orchestra / began / playing?
 Who / had / the contributor / wondered / whether / the usher / seated / properly / after / the orchestra / began / playing?

- 122 Who / had / callously / contended / that / the ranger / abandoned / the camper / before / the wildfire / broke / out?
 Who / had / callously / wondered / whether / the ranger / abandoned / the camper / before / the wildfire / broke / out?
 Who / had / the camper / contended / that / the ranger / abandoned / callously / before / the wildfire / broke / out?
 Who / had / the camper / wondered / whether / the ranger / abandoned / callously / before / the wildfire / broke / out?
- 123 Who / had / excitedly / contended / that / the millionaire / sponsored / the researcher / although / the experiment / was / risky?
 Who / had / excitedly / wondered / whether / the millionaire / sponsored / the researcher / although / the experiment / was / risky?
 Who / had / the researcher / contended / that / the millionaire / sponsored / excitedly / although / the experiment / was / risky?
 Who / had / the researcher / wondered / whether / the millionaire / sponsored / excitedly / although / the experiment / was / risky?
- 124 Who / had / serenely / contended / that / the mediator / embraced / the emissary / since / the dispute / was / resolved?
 Who / had / serenely / wondered / whether / the mediator / embraced / the emissary / since / the dispute / was / resolved?
 Who / had / the emissary / contended / that / the mediator / embraced / serenely / since / the dispute / was / resolved?
 Who / had / the emissary / wondered / whether / the mediator / embraced / serenely / since / the dispute / was / resolved?
- 125 Who / had / reluctantly / contended / that / the tycoon / endorsed / the candidate / as / the poll / numbers / climbed?
 Who / had / reluctantly / wondered / whether / the tycoon / endorsed / the candidate / as / the poll / numbers / climbed?
 Who / had / the candidate / contended / that / the tycoon / endorsed / reluctantly / as / the poll / numbers / climbed?
 Who / had / the candidate / wondered / whether / the tycoon / endorsed / reluctantly / as / the poll / numbers / climbed?
- 126 Who / had / preemptively / contended / that / the officer / suspended / the appointee / when / the threatening / letter / arrived?
 Who / had / preemptively / wondered / whether / the officer / suspended / the appointee / when / the threatening / letter / arrived?
 Who / had / the appointee / contended / that / the officer / suspended / preemptively / when / the threatening / letter / arrived?
 Who / had / the appointee / wondered / whether / the officer / suspended / preemptively / when / the threatening / letter / arrived?
- 127 Who / had / impatiently / contended / that / the weatherman / chided / the anchorman / because / the segment / went / long?

Who / had / impatiently / wondered / whether / the weatherman / chided / the anchorman / because / the segment / went / long?

Who / had / the anchorman / contended / that / the weatherman / chided / impatiently / because / the segment / went / long?

Who / had / the anchorman / wondered / whether / the weatherman / chided / impatiently / because / the segment / went / long?

- 128 Who / had / stupidly / contended / that / the stuntman / disregarded / the cameraman / while / the lights / teetered / above?

Who / had / stupidly / wondered / whether / the stuntman / disregarded / the cameraman / while / the lights / teetered / above?

Who / had / the cameraman / contended / that / the stuntman / disregarded / stupidly / while / the lights / teetered / above?

Who / had / the cameraman / wondered / whether / the stuntman / disregarded / stupidly / while / the lights / teetered / above?

- 129 Who / had / willingly / assumed / that / the interviewer / recommended / the applicant / after / the long / interview / process?

Who / had / willingly / inquired / whether / the interviewer / recommended / the applicant / after / the long / interview / process?

Who / had / the applicant / assumed / that / the interviewer / recommended / willingly / after / the long / interview / process?

Who / had / the applicant / inquired / whether / the interviewer / recommended / willingly / after / the long / interview / process?

- 130 Who / had / rashly / assumed / that / the attendant / derided / the traveler / before / the facts / became / clear?

Who / had / rashly / inquired / whether / the attendant / derided / the traveler / before / the facts / became / clear?

Who / had / the traveler / assumed / that / the attendant / derided / rashly / before / the facts / became / clear?

Who / had / the traveler / inquired / whether / the attendant / derided / rashly / before / the facts / became / clear?

- 131 Who / had / nonchalantly / assumed / that / the kicker / lifted / the linebacker / during / the short / injury / timeout?

Who / had / nonchalantly / inquired / whether / the kicker / lifted / the linebacker / during / the short / injury / timeout?

Who / had / the linebacker / assumed / that / the kicker / lifted / nonchalantly / during / the short / injury / timeout?

Who / had / the linebacker / inquired / whether / the kicker / lifted / nonchalantly / during / the short / injury / timeout?

- 132 Who / had / instantly / assumed / that / the jock / selected / the nerd / because / the team / needed / diversity?

Who / had / instantly / inquired / whether / the jock / selected / the nerd / because / the team / needed / diversity?

Who / had / the nerd / assumed / that / the jock / selected / instantly / because / the team / needed / diversity?

Who / had / the nerd / inquired / whether / the jock / selected / instantly / because / the team / needed / diversity?

- 133 Who / had / naively / assumed / that / the attache / addressed / the dignitary / during / the high / level / gathering?
 Who / had / naively / inquired / whether / the attache / addressed / the dignitary / during / the high / level / gathering?
 Who / had / the dignitary / assumed / that / the attache / addressed / naively / during / the high / level / gathering?
 Who / had / the dignitary / inquired / whether / the attache / addressed / naively / during / the high / level / gathering?
- 134 Who / had / openly / assumed / that / the warden / mocked / the prisoner / when / the verdict / was / announced?
 Who / had / openly / inquired / whether / the warden / mocked / the prisoner / when / the verdict / was / announced?
 Who / had / the prisoner / assumed / that / the warden / mocked / openly / when / the verdict / was / announced?
 Who / had / the prisoner / inquired / whether / the warden / mocked / openly / when / the verdict / was / announced?
- 135 Who / had / amusingly / assumed / that / the clown / fooled / the stagehand / because / the audience / was / laughing?
 Who / had / amusingly / inquired / whether / the clown / fooled / the stagehand / because / the audience / was / laughing?
 Who / had / the stagehand / assumed / that / the clown / fooled / amusingly / because / the audience / was / laughing?
 Who / had / the stagehand / inquired / whether / the clown / fooled / amusingly / because / the audience / was / laughing?
- 136 Who / had / reasonably / assumed / that / the bouncer / doubted / the clubbers / as / the IDs / were / examined?
 Who / had / reasonably / inquired / whether / the bouncer / doubted / the clubbers / as / the IDs / were / examined?
 Who / had / the clubbers / assumed / that / the bouncer / doubted / reasonably / as / the IDs / were / examined?
 Who / had / the clubbers / inquired / whether / the bouncer / doubted / reasonably / as / the IDs / were / examined?
- 137 Who / had / entertainingly / said / that / the correspondent / kidded / the supermodel / after / the broadcast / had / ended?
 Who / had / entertainingly / said / whether / the correspondent / kidded / the supermodel / after / the broadcast / had / ended?
 Who / had / the supermodel / said / that / the correspondent / kidded / entertainingly / after / the broadcast / had / ended?
 Who / had / the supermodel / said / whether / the correspondent / kidded / entertainingly / after / the broadcast / had / ended?
- 138 Who / had / kindly / said / that / the negotiator / nursed / the detainee / before / the family / had / arrived?
 Who / had / kindly / said / whether / the negotiator / nursed / the detainee / before / the family / had / arrived?
 Who / had / the detainee / said / that / the negotiator / nursed / kindly / before / the family / had / arrived?

Who / had / the detainee / said / whether / the negotiator / nursed / kindly / before / the family / had / arrived?

- 139 Who / had / calculatingly / said / that / the legislator / discredited / the spokesman / although / the testimony / was / accurate?

Who / had / calculatingly / said / whether / the legislator / discredited / the spokesman / although / the testimony / was / accurate?

Who / had / the spokesman / said / that / the legislator / discredited / calculatingly / although / the testimony / was / accurate?

Who / had / the spokesman / said / whether / the legislator / discredited / calculatingly / although / the testimony / was / accurate?

- 140 Who / had / inexplicably / said / that / the girlfriend / scorned / the bachelor / during / the recent / anniversary / dinner?

Who / had / inexplicably / said / whether / the girlfriend / scorned / the bachelor / during / the recent / anniversary / dinner?

Who / had / the bachelor / said / that / the girlfriend / scorned / inexplicably / during / the recent / anniversary / dinner?

Who / had / the bachelor / said / whether / the girlfriend / scorned / inexplicably / during / the recent / anniversary / dinner?

- 141 Who / had / uncharitably / declared / that / the blogger / disparaged / the popstar / as / the uncomfortable / interview / continued?

Who / had / uncharitably / speculated / whether / the blogger / disparaged / the popstar / as / the uncomfortable / interview / continued?

Who / had / the popstar / declared / that / the blogger / disparaged / uncharitably / as / the uncomfortable / interview / continued?

Who / had / the popstar / speculated / whether / the blogger / disparaged / uncharitably / as / the uncomfortable / interview / continued?

- 142 Who / had / compassionately / declared / that / the physician / guided / the refugee / when / the stressful / exodus / began?

Who / had / compassionately / speculated / whether / the physician / guided / the refugee / when / the stressful / exodus / began?

Who / had / the refugee / declared / that / the physician / guided / compassionately / when / the stressful / exodus / began?

Who / had / the refugee / speculated / whether / the physician / guided / compassionately / when / the stressful / exodus / began?

- 143 Who / had / gloomily / declared / that / the businessman / ousted / the worker / because / the board / wouldn't / relent?

Who / had / gloomily / speculated / whether / the businessman / ousted / the worker / because / the board / wouldn't / relent?

Who / had / the worker / declared / that / the businessman / ousted / gloomily / because / the board / wouldn't / relent?

Who / had / the worker / speculated / whether / the businessman / ousted / gloomily / because / the board / wouldn't / relent?

- 144 Who / had / secretly / declared / that / the juror / consulted / the convict / when / the trial / was / delayed?

Who / had / secretly / speculated / whether / the juror / consulted / the convict / when / the trial / was / delayed?

Who / had / the convict / declared / that / the juror / consulted / secretly / when / the trial / was / delayed?

Who / had / the convict / speculated / whether / the juror / consulted / secretly / when / the trial / was / delayed?

- 145 Who / had / curiously / declared / that / the publisher / rejected / the novelist / after / the promising / book / proposal?
 Who / had / curiously / speculated / whether / the publisher / rejected / the novelist / after / the promising / book / proposal?
 Who / had / the novelist / declared / that / the publisher / rejected / curiously / after / the promising / book / proposal?
 Who / had / the novelist / speculated / whether / the publisher / rejected / curiously / after / the promising / book / proposal?
- 146 Who / had / regrettably / declared / that / the counsel / slighted / the parliamentarian / during / the very / important / presentation?
 Who / had / regrettably / speculated / whether / the counsel / slighted / the parliamentarian / during / the very / important / presentation?
 Who / had / the parliamentarian / declared / that / the counsel / slighted / regrettably / during / the very / important / presentation?
 Who / had / the parliamentarian / speculated / whether / the counsel / slighted / regrettably / during / the very / important / presentation?
- 147 Who / had / respectfully / declared / that / the seniors / petitioned / the commissioner / during / the town / hall / meeting?
 Who / had / respectfully / speculated / whether / the seniors / petitioned / the commissioner / during / the town / hall / meeting?
 Who / had / the commissioner / declared / that / the seniors / petitioned / respectfully / during / the town / hall / meeting?
 Who / had / the commissioner / speculated / whether / the seniors / petitioned / respectfully / during / the town / hall / meeting?
- 148 Who / had / grudgingly / declared / that / the architect / requested / the contractor / since / the project / was / behind?
 Who / had / grudgingly / speculated / whether / the architect / requested / the contractor / since / the project / was / behind?
 Who / had / the contractor / declared / that / the architect / requested / grudgingly / since / the project / was / behind?
 Who / had / the contractor / speculated / whether / the architect / requested / grudgingly / since / the project / was / behind?
- 149 Who / had / expertly / deduced / that / the stockbroker / defrauded / the businesswoman / as / the stock / price / fell?
 Who / had / expertly / deduced / whether / the stockbroker / defrauded / the businesswoman / as / the stock / price / fell?
 Who / had / the businesswoman / deduced / that / the stockbroker / defrauded / expertly / as / the stock / price / fell?
 Who / had / the businesswoman / deduced / whether / the stockbroker / defrauded / expertly / as / the stock / price / fell?
- 150 Who / had / shrewdly / deduced / that / the gangster / blackmailed / the inspector / during / the short / probation / hearing?

Who / had / shrewdly / deduced / whether / the gangster / blackmailed / the inspector / during / the short / probation / hearing?

Who / had / the inspector / deduced / that / the gangster / blackmailed / shrewdly / during / the short / probation / hearing?

Who / had / the inspector / deduced / whether / the gangster / blackmailed / shrewdly / during / the short / probation / hearing?

- 151 Who / had / ingeniously / deduced / that / the wholesaler / manipulated / the antiquer / because / the vase / was / valuable?
 Who / had / ingeniously / deduced / whether / the wholesaler / manipulated / the antiquer / because / the vase / was / valuable?
 Who / had / the antiquer / deduced / that / the wholesaler / manipulated / ingeniously / because / the vase / was / valuable?
 Who / had / the antiquer / deduced / whether / the wholesaler / manipulated / ingeniously / because / the vase / was / valuable?
- 152 Who / had / suddenly / deduced / that / the traitor / distracted / the detective / while / the thieves / got / away?
 Who / had / suddenly / deduced / whether / the traitor / distracted / the detective / while / the thieves / got / away?
 Who / had / the detective / deduced / that / the traitor / distracted / suddenly / while / the thieves / got / away?
 Who / had / the detective / deduced / whether / the traitor / distracted / suddenly / while / the thieves / got / away?
- 153 Who / had / continuously / contended / that / the retiree / monitored / the ruffians / after / the recent / neighborhood / robberies?
 Who / had / continuously / wondered / whether / the retiree / monitored / the ruffians / after / the recent / neighborhood / robberies?
 Who / had / the ruffians / contended / that / the retiree / monitored / continuously / after / the recent / neighborhood / robberies?
 Who / had / the ruffians / wondered / whether / the retiree / monitored / continuously / after / the recent / neighborhood / robberies?
- 154 Who / had / constantly / contended / that / the analyst / avoided / the boss / before / the official / daily / briefing?
 Who / had / constantly / wondered / whether / the analyst / avoided / the boss / before / the official / daily / briefing?
 Who / had / the boss / contended / that / the analyst / avoided / constantly / before / the official / daily / briefing?
 Who / had / the boss / wondered / whether / the analyst / avoided / constantly / before / the official / daily / briefing?
- 155 Who / had / frequently / contended / that / the aide / punished / the youngster / although / the school / denied / it?
 Who / had / frequently / wondered / whether / the aide / punished / the youngster / although / the school / denied / it?
 Who / had / the youngster / contended / that / the aide / punished / frequently / although / the school / denied / it?
 Who / had / the youngster / wondered / whether / the aide / punished / frequently / although / the school / denied / it?

- 156 Who / had / smoothly / contended / that / the retailer / undercut / the inventor / since / the product / wasn't / selling?
 Who / had / smoothly / wondered / whether / the retailer / undercut / the inventor / since / the product / wasn't / selling?
 Who / had / the inventor / contended / that / the retailer / undercut / smoothly / since / the product / wasn't / selling?
 Who / had / the inventor / wondered / whether / the retailer / undercut / smoothly / since / the product / wasn't / selling?
- 157 Who / had / skeptically / contended / that / the homeowner / queried / the housekeeper / when / the wallet / was / missing?
 Who / had / skeptically / wondered / whether / the homeowner / queried / the housekeeper / when / the wallet / was / missing?
 Who / had / the housekeeper / contended / that / the homeowner / queried / skeptically / when / the wallet / was / missing?
 Who / had / the housekeeper / wondered / whether / the homeowner / queried / skeptically / when / the wallet / was / missing?
- 158 Who / had / unhappily / contended / that / the widow / slapped / the executor / when / the secret / was / revealed?
 Who / had / unhappily / wondered / whether / the widow / slapped / the executor / when / the secret / was / revealed?
 Who / had / the executor / contended / that / the widow / slapped / unhappily / when / the secret / was / revealed?
 Who / had / the executor / wondered / whether / the widow / slapped / unhappily / when / the secret / was / revealed?
- 159 Who / had / anxiously / contended / that / the litigator / scrutinized / the arbitrator / because / the decision / was / monumental?
 Who / had / anxiously / wondered / whether / the litigator / scrutinized / the arbitrator / because / the decision / was / monumental?
 Who / had / the arbitrator / contended / that / the litigator / scrutinized / anxiously / because / the decision / was / monumental?
 Who / had / the arbitrator / wondered / whether / the litigator / scrutinized / anxiously / because / the decision / was / monumental?
- 160 Who / had / suggestively / contended / that / the killer / eyeballed / the deputy / when / the recess / was / called?
 Who / had / suggestively / wondered / whether / the killer / eyeballed / the deputy / when / the recess / was / called?
 Who / had / the deputy / contended / that / the killer / eyeballed / suggestively / when / the recess / was / called?
 Who / had / the deputy / wondered / whether / the killer / eyeballed / suggestively / when / the recess / was / called?

Filler sentences:

Filler: Sentences:

- 1 Who / had / the jockey / advised / whether / the horse / would / certainly / win / the derby?
- 2 Who / had / the bricklayer / advised / whether / the rain / would / surely / ruin / the walkway?
- 3 Who / had / the pollster / advised / whether / the measure / would / fail?

- 4 Who / had / the witch / advised / whether / the curse / could / work?
- 5 Who / had / the dean / asked / whether / the revolutionaries / had / violently / destroyed / the tea?
- 6 Who / had / the novice / asked / whether / the forester / framed / the barmaid / vindictively / again?
- 7 Who / had / the ninja / asked / whether / the shogun / invited / the foreigner?
- 8 Who / had / the paparazzi / asked / whether / the movie star / slugged / the hairdresser?
- 9 Who / had / the sportscaster / informed / whether / the team / scored / barely / enough / to win?
- 10 Who / had / the messenger / informed / whether / the bigwigs / approved / the startup / rather / supportively?
- 11 Who / had / the birdwatcher / informed / whether / the sparrows / flew / south?
- 12 Who / had / the sommelier / informed / whether / the wine / was / aromatic?
- 13 Who / had / the professor / instructed / whether / the French / supported / the upstarts / quite / happily?
- 14 Who / had / the scholar / instructed / whether / the government / should / enthusiastically / subsidize / the rich?
- 15 Who / had / the speaker / instructed / whether / the chairman / was / stubborn?
- 16 Who / had / the farmer / instructed / whether / the animals / were / hungry?
- 17 Who / had / the hunter / notified / whether / the scout / shot / the deer / callously / before?
- 18 Who / had / the driver / notified / whether / the mechanic / fixed / the expensive / car / slowly?
- 19 Who / had / the mascot / notified / whether / the band / was / lost?
- 20 Who / had / the promoter / notified / whether / the prima donna / sang / beautifully?
- 21 Who / had / the emancipator / questioned / whether / the confederates / would / gratefully / accept / the terms?
- 22 Who / had / the boy / questioned / whether / the examiner / joyfully / failed / the intelligent / student?
- 23 Who / had / the curator / questioned / whether / the botanist / wrestled / the glassblower?
- 24 Who / had / the zookeeper / questioned / whether / the translator / bamboozled / the seamstress?
- 25 Who / had / the disc jockey / quizzed / whether / the musician / broke / the sales / record / delightfully?
- 26 Who / had / the comedian / quizzed / whether / the saxophonist / precisely / reached / the difficult / note?
- 27 Who / had / the archeologist / quizzed / whether / the pharaoh / invaded / the barbarians?
- 28 Who / had / the ventriloquist / quizzed / whether / the breeder / raised / the poodle?
- 29 Who / had / the snorkeler / reminded / whether / the arborist / fled / the lifeguard / very / clumsily?
- 30 Who / had / the general / reminded / whether / the airmen / loathed / the press / snooping / intrusively?
- 31 Who / had / the scrapbooker / reminded / whether / the surfer / wooed / the kayaker?
- 32 Who / had / the explorer / reminded / whether / the sailor / appreciated / the cook?
- 33 Who / had / the jester / told / whether / the count / granted / the knight / homesteading / rights?
- 34 Who / had / the instructor / told / whether / the shepherd / grabbed / the mountaineer / intensely / today?
- 35 Who / had / the barista / told / whether / the fumigator / tackled / the candlemaker?
- 36 Who / had / the rapper / told / whether / the keyboardist / overshadowed / the breakdancers?

- 37 Who / had / the philanthropist / warned / whether / the fencer / grazed / the chambermaid / on / the ear?
- 38 Who / had / the cartoonist / warned / whether / the programmer / struck / the fisherman / quite / solidly?
- 39 Who / had / the strongman / warned / whether / the brewer / shoved / the cobbler?
- 40 Who / had / the epidemiologist / warned / whether / the juggler / introduced / the interns?
- 41 Who / had / beautifully / advised / that / the sculptor / charming / the glazier / for / the reception?
- 42 Who / had / politely / advised / that / the jogger / tripped / the technician / saw / the struggle?
- 43 Who / had / judgmentally / advised / that / the astronaut / slept / the moon?
- 44 Who / had / wearily / advised / that / the whaler / revenged / the Eskimo?
- 45 Who / had / doubtfully / asked / that / the infiltrator / snuck / the hostage / past / security?
- 46 Who / had / optimistically / asked / that / the concierge / bet / the rancher / on / the match?
- 47 Who / had / nearly / asked / that / the cowboy / lassoed / the dentist?
- 48 Who / had / queasily / asked / that / the veterinarian / examined / the milkman?
- 49 Who / had / gratefully / informed / that / the carver / hated / the maid / although / the party?
- 50 Who / had / fortunately / informed / that / the coxswain / motivated / the rowers / since / only?
- 51 Who / had / partially / informed / that / the spelunker / charge / the runningback?
- 52 Who / had / annually / informed / that / the librarian / shush / the hooligans?
- 53 Who / had / interestingly / instructed / that / the leader / practiced / the goon / before / the thunderstorm?
- 54 Who / had / gladly / instructed / that / the vegan / serving / the homeless / as / well?
- 55 Who / had / mechanically / instructed / that / the robot / mimicking / the operator?
- 56 Who / had / brightly / instructed / that / the rascal / terrorizing / the octogenarian?
- 57 Who / had / frankly / notified / that / the coordinator / disappointed / the guys / before / the playoffs?
- 58 Who / had / merrily / notified / that / the goalie / outplayed / the offense / after / the rest?
- 59 Who / had / bleakly / notified / that / the twins / evaded / the security?
- 60 Who / had / hungrily / notified / that / the vegetarian / slighted / the chef?
- 61 Who / had / reproachfully / questioned / that / the underling / wrong / the ruler / after / the ceremony?
- 62 Who / had / dreamingly / questioned / that / the justice / ruled / the accused / although / the alibi?
- 63 Who / had / coaxingly / questioned / that / the prospector / locating / the miners?
- 64 Who / had / famously / questioned / that / the lover / miss / the betrothed?
- 65 Who / had / jubilantly / quizzed / that / the civilians / outsmarted / the reserves / before / starving?
- 66 Who / had / evenly / quizzed / that / the monarch / acquiesced / the peasants / after / the uprising?
- 67 Who / had / keenly / quizzed / that / the emissary / exasperated / the reception?
- 68 Who / had / sedately / quizzed / that / the pilot / attacked / the base?
- 69 Who / had / unabashedly / reminded / that / the chemist / echoed / the physicist / after / the lecture?
- 70 Who / had / longingly / reminded / that / the debugger / refused / the coder / before / the launch?
- 71 Who / had / bitterly / reminded / that / the environmentalist / kidnapped / the logger?
- 72 Who / had / adventurously / reminded / that / the guide / mistrailed / the hikers?

- 73 Who / had / lazily / told / that / the mathematician / collaborate / the engineer / before / the conference?
- 74 Who / had / kookily / told / that / the hippies / adore / the wealthy / before / the breakdown?
- 75 Who / had / yawningly / told / that / the tailor / knee / the magnate?
- 76 Who / had / vivaciously / told / that / the ladies / struggled / the management?
- 77 Who / had / noisily / warned / that / the sniper / shooting / the lookout / after / the dawn?
- 78 Who / had / zealously / warned / that / the ruffian / assault / the vicar / after / the celebration?
- 79 Who / had / generally / warned / that / the vaulter / throw / the skater?
- 80 Who / had / fiercely / warned / that / the cinematographer / film / the downtrodden?

References

- Allen, M., Badecker, W., & Osterhout, L. (2003). Morphological analysis in sentence processing: An ERP study. *Language and Cognitive Processes, 18*(4), 405–430. doi:10.1080/01690960244000054
- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition, 73*(3), 247–264. doi:10.1016/S0010-0277(99)00059-1
- Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition, 111*(1), 55–71. doi:10.1016/j.cognition.2008.12.005
- Ambridge, B., & Goldberg, A. E. (2008). The island status of clausal complements: Evidence in favor of an information structure explanation. *Cognitive Linguistics, 19*(3), 357–389. doi:10.1515/COGL.2008.014
- Anderson, J. E., & Holcomb, P. J. (2005). An electrophysiological investigation of the effects of coreference on word repetition and synonymy. *Brain and Language, 94*(2), 200–216. doi:10.1016/j.bandl.2005.01.001
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An Integrated Theory of the Mind. *Psychological Review, 111*(4), 1036–1060. doi:10.1037/0033-295X.111.4.1036
- Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R* (1 edition.). Cambridge, UK ; New York: Cambridge University Press.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. doi:10.1016/j.jml.2007.12.005
- Baddeley, A. D. (1983). Working Memory. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 302*(1110), 311–324. doi:10.1098/rstb.1983.0057
- Baddeley, A. D. (2002). Is Working Memory Still Working? *European Psychologist, 7*(2), 85–97. doi:10.1027//1016-9040.7.2.85

- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic Press.
- Bader, M., & Häussler, J. (2010). Toward a model of grammaticality judgments. *Journal of Linguistics*, *46*(02), 273–330. doi:10.1017/S0022226709990260
- Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459. doi:10.3758/BF03193014
- Barber, H., Vergara, M., & Carreiras, M. (2004). Syllable-frequency effects in visual word recognition: evidence from ERPs. *NeuroReport*, *15*(3), 545–548.
- Barkley, C. (2011). LAN P600 Chart.
- Barkley, C., Kluender, R., & Kutas, M. (2011). *Event-related potential (ERP) indices of referential dependency formation*. Poster presented at the 24th Annual Conference on Human Sentence Processing, Stanford University.
- Bentin, S. (1987). Event-related potentials, semantic processes, and expectancy factors in word recognition. *Brain and Language*, *31*(2), 308–327.
- Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology*, *60*(4), 343–355. doi:10.1016/0013-4694(85)90008-2
- Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echallier, J. F., & Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. *Journal of Cognitive Neuroscience*, *11*(3), 235–260.
- Bever, T. G., & McElree, B. (1988). Empty Categories Access Their Antecedents during Comprehension. *Linguistic Inquiry*, *19*(1), 35–43.
- Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb Argument Structure in Parsing and Interpretation: Evidence from wh-Questions. *Journal of Memory and Language*, *34*(6), 774–806. doi:10.1006/jmla.1995.1034
- Bornkessel, I. D., Fiebach, C. J., & Friederici, A. D. (2004). On the cost of syntactic ambiguity in human language comprehension: an individual differences approach. *Cognitive Brain Research*, *21*(1), 11–21. doi:10.1016/j.cogbrainres.2004.05.007

- Borovsky, A., Elman, J., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology, 112*(4), 417–436. doi:10.1016/j.jecp.2012.01.005
- Boudreau, C., McCubbins, M. D., & Coulson, S. (2009). Knowing when to trust others: An ERP study of decision making after receiving information from unknown people. *Social Cognitive and Affective Neuroscience, 4*(1), 23–34. doi:10.1093/scan/nsn034
- Bourdages, J. (1992). Parsing complex NPs in French. In H. Goodluck & M. Rochemont (Eds.), *Island Constraints* (Vol. 15, pp. 61–88). Netherlands: Kluwer.
- Brown, C., & Hagoort, P. (1993). The Processing Nature of the N400: Evidence from Masked Priming. *Journal of Cognitive Neuroscience, 5*(1), 34–44. doi:10.1162/jocn.1993.5.1.34
- Brown, C., & Hagoort, P. (1999). The cognitive neuroscience of language: Challenges and future directions. In C. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 3–14). Oxford: Oxford University Press.
- Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *The Behavioral and Brain Sciences, 22*(1), 77–94; discussion 95–126.
- Chao, L. L., Nielsen-Bohlman, L., & Knight, R. T. (1995). Auditory event-related potentials dissociate early and late memory processes. *Electroencephalography and Clinical Neurophysiology, 96*(2), 157–168.
- Chapman, R. M., & Bragdon, H. R. (1964). Evoked responses to numerical and nonnumerical visual stimuli while problem solving. *Nature, 203*, 1155–1157.
- Chomsky, N. (1964). *Current Issues in Linguistic Theory*. Mouton.
- Chomsky, N. (1973). Conditions on transformations. In S. Anderson & P. Kiparsky (Eds.), *A festschrift for Morris Halle* (pp. 232–286). New: Holt, Rinehart, and Winston.
- Chomsky, N. (1977). On Wh-Movement. In P. Culicover, T. Wasow, & A. Akmajian (Eds.), *Formal Syntax*. New York: Academic Press.
- Chomsky, N. (1981). Lectures on Government and Binding. Dordrecht, Foris, 1981. *Studies in Generative Grammar, 9*.

- Chomsky, N. (1986). *Barriers*. MIT Press.
- Chomsky, N. (1993). A minimalist program for linguistic theory. In K. Hale & S. J. Keyser (Eds.), *The view from Building 20* (pp. 1–52). Cambridge, Massachusetts: MIT Press.
- Chomsky, N. (2001). Derivation by phase. In K. Kenstowicz (Ed.), *Ken Hale: A life in language* (pp. 1–52). Cambridge, MA: MIT.
- Chomsky, N. (2004). Beyond explanatory adequacy. In A. Belletti (Ed.), *Structures and beyond* (pp. 104–131). Oxford: Oxford University Press.
- Chomsky, N., & Miller, G. A. (1963). Introduction to the formal analysis of natural languages. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2, pp. 269–321). New York: Wiley.
- Chung, S. (1994). Wh-Agreement and “Referentiality” in Chamorro. *Linguistic Inquiry*, 25(1), 1–44.
- Chwilla, D. J., Brown, C. M., & Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology*, 32(3), 274–285. doi:10.1111/j.1469-8986.1995.tb02956.x
- Cinque, G. (1990). *Type of A-Dependencies*. Cambridge, MA: MIT Press.
- Cole, P., & Hermon, G. (1994). Is There LF Wh-Movement? *Linguistic Inquiry*, 25(2), 239–262.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user’s guide. *Psychonomic Bulletin & Review*, 12(5), 769–786. doi:10.3758/BF03196772
- Coulson, S., King, J., & Kutas, M. (1998a). ERPs and Domain Specificity: Beating a Straw Horse. *Language and Cognitive Processes*, 13(6), 653–672. doi:10.1080/016909698386410
- Coulson, S., King, J. W., & Kutas, M. (1998b). Expect the Unexpected: Event-related Brain Response to Morphosyntactic Violations. *Language and Cognitive Processes*, 13(1), 21–58. doi:10.1080/016909698386582
- Cowan, N. (1995). *Attention and Memory: An Integrated Framework* (1 edition.). New York : Oxford: Oxford University Press.

- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87–114.
- Cowan, N. (2004). *Working Memory Capacity*. Psychology Press.
- Cowan, N. (2008). Chapter 20 What are the differences between long-term, short-term, and working memory? In J.-C. L., Vincent F. Castellucci and Sylvie Belleville Wayne S. Sossin (Ed.), *Progress in Brain Research* (Vol. Volume 169, pp. 323–338). Elsevier. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0079612307000209>
- Cowart, W. (1997). *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. Thousand Oaks, Calif: SAGE Publications, Inc.
- Crain, S., & Fodor, J. D. (1985). How can grammars help parsers? In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural Language Parsing: Psychological, Computational and Theoretical Perspectives*. Cambridge: Cambridge University Press.
- Crawford, J. (2012). Using Syntactic Satiation to Investigate Subject Islands. In J. Choi, E. A. Hogue, J. Punske, D. Tat, J. Schertz, & A. Truman (Eds.), *Proceedings of the 29th West Coast Conference on Formal Linguistics* (pp. 38–45). Somerville, MA: Cascadilla Proceedings Project.
- Daltrozzo, J., & Schön, D. (2009). Is conceptual processing in music automatic? An electrophysiological approach. *Brain Research*, 1270, 88–94. doi:10.1016/j.brainres.2009.03.019
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450–466.
- Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, 3(4), 422–433.
- Deane, P. (2009). Limits to attention: A cognitive theory of island phenomena. *Cognitive Linguistics (includes Cognitive Linguistic Bibliography)*, 2(1), 1–64. doi:10.1515/cogl.1991.2.1.1
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, 58(1), 17–22. doi:10.1037/h0046671

- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117–1121.
- Donchin, E., Ritter, W., & McCallum, W. D. (1978). Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting, & S. H. Koslow (Eds.), *Event-Related Brain Potentials in Man* (pp. 349–411). New York: Academic Press.
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception and Psychophysics*, 16(1), 143–149.
- Eriksen, C. W., & Schultz, D. W. (1979). Information processing in visual search: A continuous flow conception and experimental results. *Perception & Psychophysics*, 25(4), 249–263. doi:10.3758/BF03198804
- Erteschik-Shir, N., & Lappin, S. (2009). Dominance and the functional explanation of island constraints. *Theoretical Linguistics*, 6(1-3), 41–86. doi:10.1515/thli.1979.6.1-3.41
- Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44(4), 491–505. doi:10.1111/j.1469-8986.2007.00531.x
- Felser, C., Clahsen, H., & Münte, T. F. (2003). Storage and integration in the processing of filler-gap dependencies: an ERP study of topicalization and wh-movement in German. *Brain and Language*, 87(3), 345–354.
- Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2001). Syntactic Working Memory and the Establishment of Filler-Gap Dependencies: Insights from ERPs and fMRI. *Journal of Psycholinguistic Research*, 30(3), 321–338. doi:10.1023/A:1010447102554
- Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: the processing of German WH-questions. *Journal of Memory and Language*, 47(2), 250–272. doi:10.1016/S0749-596X(02)00004-9
- Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry*, 9, 427–473.
- Fodor, J. D. (1983). Phrase structure parsing and the island constraints. *Linguistics and Philosophy*, 6(2), 163–223. doi:10.1007/BF00635643

- Francom, J. (2009). *Experimental Syntax: Exploring the effect of repeated exposure to anomalous syntactic structure- evidence from rating and reading tasks*. Doctoral Dissertation, University of Arizona.
- Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, 4(2), 93–126.
doi:10.1080/01690968908406359
- Freedman, S. E., & Forster, K. I. (1985). The psychological status of overgenerated sentences. *Cognition*, 19(2), 101–131.
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6(2), 78–84.
- Fukuda, S., Goodall, G., Michel, D., & Beecher, H. (2012). Is Magnitude Estimation worth the trouble? In J. Choi, E. A. Hogue, J. Punske, D. Tat, J. Schertz, & A. Truman (Eds.), *Proceedings of the 29th West Coast Conference on Formal Linguistics* (pp. 328–336). Somerville, MA: Cascadilla Proceedings Project.
- Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Brain Research. Cognitive Brain Research*, 16(2), 123–144.
- Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for “common sense”: an electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, 8(2), 89–106.
doi:10.1162/jocn.1996.8.2.89
- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68(1), 1–76.
- Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O’Neil (Eds.), *Papers from the first mind articulation project symposium* (pp. 95–126). Cambridge, MA: MIT Press.
- Goldberg, A. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford ; New York: Oxford University Press.
- Goodall, G. (2011). Syntactic Satiation and the Inversion Effect in English and Spanish Wh-Questions. *Syntax*, 14(1), 29–47. doi:10.1111/j.1467-9612.2010.00148.x

- Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Gordon Hendrick Johnson 2001 Mem Int dur Lang Proc JEPLMC.pdf. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(6), 1411–1423.
- Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(6), 1304.
- Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-Load Interference in Syntactic Processing. *Psychological Science*, 13(5), 425–430. doi:10.1111/1467-9280.00475
- Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes*, 25(2), 149–188. doi:10.1080/01690960902965951
- Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology*, 34(6), 660–676.
- Hagoort, P., Baggio, G., & Willems, R. (2009). Semantic unification. In M. Gazzaniga (Ed.), *The Cognitive Neurosciences* (4th ed., pp. 819–836). Boston, MA: MIT Press.
- Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (sps) as an erp measure of syntactic processing. *Language and Cognitive Processes*, 8(4), 439–483. doi:10.1080/01690969308407585
- Hagoort, P., & Brown, C. M. (1999). Gender electrified: ERP evidence on the syntactic nature of gender processing. *Journal of Psycholinguistic Research*, 28(6), 715–728.
- Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science*, 30(4), 609–642.
- Hampton Wray, A., & Weber-Fox, C. (2013). Specific aspects of cognitive and language proficiency account for variability in neural indices of semantic and syntactic processing in children. *Developmental Cognitive Neuroscience*, 5, 149–171. doi:10.1016/j.dcn.2013.03.002
- Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 115(5), 1090–1103. doi:10.1016/j.clinph.2003.12.020

- Hiramatsu, K. (2000). *Assessing Linguistic Competence: Evidence from children's and Adults' Acceptability Judgments*. Doctoral Dissertation, University of Connecticut.
- Hoeks, J. C. J., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: the interaction of lexical and sentence level information during reading. *Brain Research. Cognitive Brain Research*, *19*(1), 59–73.
doi:10.1016/j.cogbrainres.2003.10.022
- Hofmeister, P. (2007). *Representational Complexity and Memory Retrieval in Language Comprehension*. Doctoral Dissertation, Stanford University.
- Hofmeister, P., Casasanto, L. S., & Sag, I. A. (2012a). How do individual cognitive differences relate to acceptability judgments?: A reply to Sprouse, Wagers, and Phillips. *Language*, *88*(2), 390–400.
- Hofmeister, P., Casasanto, L. S., & Sag, I. A. (2012b). Misapplying working-memory tests: A reductio ad absurdum. *Language*, *88*(2), 408–409.
- Hofmeister, P., Casasanto, L. S., & Sag, I. A. (2014). Processing effects in linguistic judgment data: (super-)additivity and reading span scores. *Language and Cognition*, *6*(01), 111–145. doi:10.1017/langcog.2013.7
- Hofmeister, P., & Sag, I. A. (2010). Cognitive Constraints and Island Effects. *Language*, *86*(2), 366–415.
- Holcomb, P. J. (1988). Automatic and attentional processing: an event-related brain potential analysis of semantic priming. *Brain and Language*, *35*(1), 66–85.
- Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology*, *30*(1), 47–61.
doi:10.1111/j.1469-8986.1993.tb03204.x
- Holcomb, P. J., Grainger, J., & O'Rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience*, *14*(6), 938–950.
doi:10.1162/089892902760191153
- Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: an electrophysiological investigation. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *25*(3), 721–742.

- Holcomb, P. J., & Neville, H. J. (1990). Auditory and Visual Semantic Priming in Lexical Decision: A Comparison Using Event-related Brain Potentials. *Language and Cognitive Processes*, 5(4), 281–312. doi:10.1080/01690969008407065
- Holcomb, P. J., & Neville, H. J. (1991). Natural speech processing: An analysis using event-related brain potentials. *Psychobiology*, 19(4), 286–300. doi:10.3758/BF03332082
- Honcoop, M. (1998). *Dynamic excursions on weak islands*. Holland Academic Graphics.
- Huang, C. T. J. (1982). *Logical relations in Chinese and the theory of grammar*. Doctoral Dissertation, MIT.
- Jackendoff, R. S., & Culicover, P. (1971). A reconsideration of dative movements. *Foundations of Language*, 7, 397–412.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122–149.
- Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, 15(2), 159–201. doi:10.1080/016909600386084
- Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of Syntactic and Semantic Information in Predictive Processing: Cross-Linguistic Evidence from German and English. *Journal of Psycholinguistic Research*, 32(1), 37–55. doi:10.1023/A:1021933015362
- Keffala, B., & Goodall, G. (2013). *On processing difficulty and the acceptability of resumptive pronouns*. Talk presented at the Linguistic Evidence - Berlin Special, Humboldt-Universität zu Berlin.
- Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language*, 89(1), 253–260. doi:10.1016/S0093-934X(03)00335-3
- Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, 52(2), 205–225. doi:10.1016/j.jml.2004.10.002

- Kim, B. (2010). L2 Wh-Movement: That-trace Effect and Subject/Object Preferences. In D. Bailey & V. Teliga (Eds.), *Proceedings of the 39th Western Conference on Linguistics* (pp. 118–130). Fresno, CA: Department of Linguistics, California State University.
- King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, 30(5), 580–602. doi:10.1016/0749-596X(91)90027-H
- King, J. W., & Kutas, M. (1995). Who Did What and When? Using Word- and Clause-Level ERPs to Monitor Working Memory Usage in Reading. *Journal of Cognitive Neuroscience*, 7(3), 376–395. doi:10.1162/jocn.1995.7.3.376
- Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, 55(4), 352.
- Kluender, R. (1991). *Cognitive Constraints on Variables in Syntax*. Doctoral Dissertation, University of California, San Diego.
- Kluender, R. (1998). On the distinction between strong and weak islands: a processing perspective. In P. Culicover & L. McNally (Eds.), *Syntax and Semantics 29: The Limits of Syntax*. San Diego, CA: Academic Press.
- Kluender, R., & Kutas, M. (1993a). Bridging the Gap: Evidence from ERPs on the Processing of Unbounded Dependencies. *Journal of Cognitive Neuroscience*, 5(2), 196–214. doi:10.1162/jocn.1993.5.2.196
- Kluender, R., & Kutas, M. (1993b). Subjacency as a processing phenomenon. *Language and Cognitive Processes*, 8(4), 573–633. doi:10.1080/01690969308407588
- Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nature Neuroscience*, 7(3), 302–307. doi:10.1038/nn1197
- Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, 38(3), 557–577.
- Kopp, B., Mattler, U., & Rist, F. (1994). Selective attention and response competition in schizophrenic patients. *Psychiatry Research*, 53(2), 129–139.
- Kounios, J., & Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 20(4), 804–823.

- Kuno, S. (1976). Subject theme and speaker's empathy – a reexamination of relativization phenomena. In C. Li (Ed.), *Subject and Topic* (pp. 419–444). New York: Academic Press.
- Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: evidence from event-related potentials. *Brain and Language*, *100*(3), 223–237. doi:10.1016/j.bandl.2005.12.006
- Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Brain Research. Cognitive Brain Research*, *17*(1), 117–129.
- Kutas, M., & Federmeier, K. D. (2009). N400. *Scholarpedia*, *4*, 7790.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. doi:10.1146/annurev.psych.093008.131123
- Kutas, M., & Hillyard, S. A. (1980a). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology*, *11*(2), 99–116.
- Kutas, M., & Hillyard, S. A. (1980b). Reading between the lines: event-related brain potentials during natural sentence processing. *Brain and Language*, *11*(2), 354–373.
- Kutas, M., & Hillyard, S. A. (1980c). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science (New York, N.Y.)*, *207*(4427), 203–205.
- Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & Cognition*, *11*(5), 539–550.
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163.
- Kutas, M., & Hillyard, S. A. (1989). An electrophysiological probe of incidental semantic association. *Journal of Cognitive Neuroscience*, *1*(1), 38–49. doi:10.1162/jocn.1989.1.1.38

- Kutas, M., Neville, H. J., & Holcomb, P. J. (1987). A preliminary comparison of the N400 response to semantic anomalies during reading, listening and signing. *Electroencephalography and Clinical Neurophysiology. Supplement*, 39, 325–330.
- Kwon, N. (2008). *Processing of Syntactic and Anaphoric Gap-filler Dependencies in Korean: Evidence from Self-paced Reading Time, ERP and Eye-tracking Experiments*. Doctoral Dissertation, University of California, San Diego.
- Kwon, N., Kluender, R., Kutas, M., & Polinsky, M. (2013). Subject/object processing asymmetries in Korean relative clauses: Evidence from ERP data. *Language*, 89(3), 537–585. doi:10.1353/lan.2013.0044
- Lasnik, H., & Saito, M. (1984). On the nature of proper government. *Linguistic Inquiry*, 15, 235–289.
- Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience*, 25(3), 484–502.
- Lau, E., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain and Language*, 98(1), 74–88. doi:10.1016/j.bandl.2006.02.003
- Lawrence, M. A. (2013). *Package “ez.”* Retrieved from <http://cran.r-project.org/web/packages/ez/ez.pdf>
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177. doi:10.1016/j.cognition.2007.05.006
- Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences of the United States of America*, 106(50), 21086–21090. doi:10.1073/pnas.0907664106
- Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3), 375–419.
- Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, 10(10), 447–454. doi:10.1016/j.tics.2006.08.007
- Lewis, S., & Phillips, C. (2014) Aligning grammatical theories and processing models. To appear: *Journal of Psycholinguistic Research*.

- Macdonald, M. C. (1989). Priming effects from gaps to antecedents. *Language and Cognitive Processes*, 4(1), 35–56. doi:10.1080/01690968908406356
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, 109(1), 35–54. doi:10.1037/0033-295X.109.1.35
- Matzke, M., Mai, H., Nager, W., Rüsseler, J., & Münte, T. (2002). The costs of freedom: an ERP -- study of non-canonical sentences. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 113(6), 844–852.
- McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research*, 29(2), 111–123.
- McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, 48(1), 67–91.
- McKinnon, R., & Osterhout, L. (1996). Constraints on Movement Phenomena in Sentence Processing: Evidence from Event-related Brain Potentials. *Language and Cognitive Processes*, 11(5), 495–524. doi:10.1080/016909696387132
- Malhorta, S. (2009). *On wh-quantifier interactions*. Manuscript, University of Maryland.
- Michel, D. (In Prep). *Additivity in acceptability judgments*. Manuscript.
- Michel, D. (2010). *Relating working memory and d-linked wh-islands: An experimental syntax study*. Talk presented at the 84th Annual Meeting of the Linguistic Society of America, Baltimore, MD.
- Michel, D., & Goodall, G. (2013). Finiteness and the nature of island constraints. In N. Goto, K. Otaki, A. Sato, & K. Takita (Eds.), *Proceedings of GLOW in Asia IX 2012: The main session* (pp. 187–197). Mie University, Japan.
- Müller, H. M., King, J. W., & Kutas, M. (1997). Event-related potentials elicited by spoken relative clauses. *Cognitive Brain Research*, 5(3), 193–203. doi:10.1016/S0926-6410(96)00070-5
- Münte, T. F., Heinze, H. J., & Mangun, G. R. (1993). Dissociation of brain activity related to syntactic and semantic aspects of language. *Journal of Cognitive Neuroscience*, 5(3), 335–344. doi:10.1162/jocn.1993.5.3.335

- Münter, T. F., Matzke, M., & Johannes, S. (1997). Brain activity associated with syntactic incongruencies in words and pseudo-words. *Journal of Cognitive Neuroscience*, 9(3), 318–329. doi:10.1162/jocn.1997.9.3.318
- Münter, T. F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, 395(6697), 71–73. doi:10.1038/25731
- Münter, T. F., Wieringa, B. M., Weyerts, H., Szentkuti, A., Matzke, M., & Johannes, S. (2001). Differences in brain potentials to open and closed class words: class and frequency effects. *Neuropsychologia*, 39(1), 91–102.
- Neville, H. J., Mills, D. L., & Lawson, D. S. (1992). Fractionating language: different neural subsystems with different sensitive periods. *Cerebral Cortex (New York, N.Y.: 1991)*, 2(3), 244–258.
- Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, 3(2), 151–165. doi:10.1162/jocn.1991.3.2.151
- Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, 18(1), 5–19.
- Nigam, A., Hoffman, J. E., & Simons, R. F. (1992). N400 to semantically anomalous pictures and words. *Journal of Cognitive Neuroscience*, 4(1), 15–22. doi:10.1162/jocn.1992.4.1.15
- Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 411–421. doi:10.1037/0278-7393.28.3.411
- Osterhout, L. (1990). *Event-related brain potentials elicited during sentence comprehension*. Doctoral Dissertation, Tufts University, Medford, MA.
- Osterhout, L., & Hagoort, P. (1999). A Superficial Resemblance Does Not Necessarily Mean You Are Part of the Family: Counterarguments to Coulson, King and Kutas (1998) in the P600/SPS-P300 Debate. *Language and Cognitive Processes*, 14(1), 1–14. doi:10.1080/016909699386356
- Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31(6), 785–806. doi:10.1016/0749-596X(92)90039-Z

- Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 786–803. doi:10.1037/0278-7393.20.4.786
- Osterhout, L., & Mobley, L. A. (1995). Event-Related Brain Potentials Elicited by Failure to Agree. *Journal of Memory and Language*, 34(6), 739–773. doi:10.1006/jmla.1995.1033
- Paller, K. A., Kutas, M., Shimamura, A. P., & Squire, L. R. (1987). Brain responses to concrete and abstract words reflect processes that correlate with later performance on a test of stem-completion priming. *Electroencephalography and Clinical Neurophysiology. Supplement*, 40, 360–365.
- Pesetsky, D. (1987). Wh-in-situ: movement and unselective binding. In E. Reuland & A. ter Meulen (Eds.), *The Representation of (In)Definiteness*. Cambridge, MA: MIT Press.
- Peterson, N. N., Schroeder, C. E., & Arezzo, J. C. (1995). Neural generators of early cortical somatosensory evoked potentials in the awake monkey. *Electroencephalography and Clinical Neurophysiology*, 96(3), 248–260.
- Phillips, C. (2006). The Real-Time Status of Island Phenomena. *Language*, 82(4), 795–823.
- Phillips, C. (2013a). On the nature of island constraints. I: Language processing and reductionist accounts. *Experimental Syntax and Island Effects*, 64.
- Phillips, C. (2013b). Some arguments and nonarguments for reductionist accounts of syntactic phenomena. *Language and Cognitive Processes*, 28(1-2), 156–187. doi:10.1080/01690965.2010.530960
- Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, 22(3), 407–428. doi:10.1016/j.cogbrainres.2004.09.012
- Pickering, M., Barton, S. B., & Shillcock, R. (1994). Unbounded dependencies, island constraints and processing complexity. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on Sentence Processing* (pp. 199–224). London: Erlbaum.
- Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11(3), 105–110. doi:10.1016/j.tics.2006.12.002

- Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology: Official Publication of the American Electroencephalographic Society*, 9(4), 456–479.
- Plante, E., Petten, C. V., & Senkfor, A. J. (2000). Electrophysiological dissociation between verbal and nonverbal semantic processing in learning disabled adults. *Neuropsychologia*, 38(13), 1669–1684.
- Polich, J. (1987). Task difficulty, probability, and inter-stimulus interval as determinants of P300 from auditory stimuli. *Electroencephalography and Clinical Neurophysiology*, 68(4), 311–320.
- Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 118(10), 2128–2148. doi:10.1016/j.clinph.2007.04.019
- Postal, P. M. (1998). *Three Investigations of Extraction*. MIT Press.
- Prince, A., & Smolensky, P. (2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Wiley.
- Pritchard, W. S. (1981). Psychophysiology of P300. *Psychological Bulletin*, 89(3), 506–540.
- R Development Core Team. (2009). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>
- Reinhart, R. M. G., Carlisle, N. B., Kang, M.-S., & Woodman, G. F. (2012). Event-related potentials elicited by errors during the stop-signal task. II: Human effector-specific error responses. *Journal of Neurophysiology*, 107(10), 2794–2807. doi:10.1152/jn.00803.2011
- Reinitz, M. T., Lammers, W. J., & Cochran, B. P. (1992). Memory-conjunction errors: miscombination of stored stimulus features can produce illusions of memory. *Memory & Cognition*, 20, 1–11.
- Rizzi, L. (1982). Violations of the WH-island constraint and the subjacency condition. In L. Rizzi (Ed.), *Issues in Italian syntax*. Dordrecht: Foris.
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21, 803–814.

- Rosenfelt, L., Barkley, C., Kellogg, M. K., Kluender, R., & Kutas, M. (2009). *No ERP Evidence for Automatic First-Pass Parsing*. Talk presented at the 22nd annual CUNY conference on human sentence processing, University of California, Davis.
- Ross, J. R. (1967). *Constraints on variables in syntax*. Doctoral Dissertation, MIT.
- Ross, J. R. (1987). Islands and syntactic prototypes. In B. Need, E. Schiller, & A. Bosch (Eds.), *Papers from the general session at the Twenty-Third Annual Regional Meeting of the Chicago Linguistic Society* (pp. 309–320). Chicago, IL: Chicago Linguistic Society.
- Rugg, M. D., & Nagy, M. E. (1987). Lexical contribution to nonword-repetition effects: evidence from event-related potentials. *Memory & Cognition*, *15*(6), 473–481.
- Sag, I., Hofmeister, P., & Snider, N. (2007). Processing Complexity in Subjacency Violations: The Complex Noun Phrase Constraint. *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, *43*(1), 215–229.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime User's Guide*. Pittsburgh: Psychology Software Tools, Inc.
- Sitnikova, T., Holcomb, P. J., Kiyonaga, K. A., & Kuperberg, G. R. (2008). Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *Journal of Cognitive Neuroscience*, *20*(11), 2037–2057. doi:10.1162/jocn.2008.20143
- Smith, M. E., & Halgren, E. (1987). Event-related potentials during lexical decision: effects of repetition, word frequency, pronounceability, and concreteness. *Electroencephalography and Clinical Neurophysiology. Supplement*, *40*, 417–421.
- Snyder, W. (2000). An experimental investigation of syntactic satiation effects. *Linguistic Inquiry*, *31*, 575–582.
- Soederberg Miller, L. M., Cohen, J. A., & Wingfield, A. (2006). Contextual Knowledge Reduces Demands on Working Memory during Reading. *Memory & Cognition*, *34*(6), 1355.
- Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, *15*(2), 201–292. doi:10.2307/1412107

- Sprouse, J. (2007). *A program for experimental syntax*. Doctoral Dissertation, University of Maryland.
- Sprouse, J. (2009a). *Looking for evidence in the islands debate*. Invited Talk, University of California, San Diego.
- Sprouse, J. (2009b). Revisiting Satiation: Evidence for an Equalization Response Strategy. *Linguistic Inquiry*, 40(2), 329–341. doi:10.1162/ling.2009.40.2.329
- Sprouse, J. (2011). A Test of the Cognitive Assumptions of Magnitude Estimation: Commutativity does not Hold for Acceptability Judgments. *Language*, 87(2), 274–288. doi:10.1353/lan.2011.0028
- Sprouse, J., Wagers, M., & Phillips, C. (2012). A test of the relation between working-memory capacity and syntactic island effects. *Language*, 88(1), 82–123.
- Steinhauer, K., & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language*, 120(2), 135–162. doi:10.1016/j.bandl.2011.07.001
- Stepanov, A. (2007). The End of CED? Minimalism and Extraction Domains. *Syntax*, 10(1), 80–126. doi:10.1111/j.1467-9612.2007.00094.x
- Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, 1(3), 227–245. doi:10.1080/01690968608407062
- Sussman, R. S., & Sedivy, J. (2003). The time-course of processing syntactic dependencies: Evidence from eye movements. *Language and Cognitive Processes*, 18(2), 143–163. doi:10.1080/01690960143000498
- Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science (New York, N.Y.)*, 150(3700), 1187–1188.
- Swart, H. de. (1992). Intervention Effects, Monotonicity and Scope. In C. Barker & D. Dowty (Eds.), *Proceedings of SALT II*. Columbus, OH: The Ohio State University.
- Szabolcsi, A., & den Dikken, M. (2002). Islands. In L. Cheng & R. Sybesma (Eds.), *The Second State-of-the-Article Book* (pp. 213–240). Berlin: Mouton de Gruyter.
- Szabolcsi, A., Evaraert, M., & van Riemsdijk, H. (2006). Strong vs. weak islands. In *The Blackwell companion to syntax* (Vol. 4, pp. 479–531). Wiley-Blackwell.

- Szabolcsi, A., & Zwarts, F. (1993). Weak islands and an algebraic semantics for scope taking. *Natural Language Semantics*, *1*(3), 235–284. doi:10.1007/BF00263545
- Szabolcsi, Anna, & Zwarts, Frans. (1990). Semantic Properties Of Composed Functions And The Distribution Of “wh”-phrases. ITLI: Institute for Language, Logic and Information, University of Amsterdam. Retrieved from <http://irs.ub.rug.nl/dbi/435798e84466a>
- Takami, K. (1988). Preposition stranding: Arguments against syntactic analyses and an alternative functional explanation. *Lingua*, *76*(4), 299–335. doi:10.1016/0024-3841(88)90022-8
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science (New York, N.Y.)*, *268*(5217), 1632–1634.
- Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the Processing of Unbounded Dependencies: An Eye-Tracking Study. *Journal of Memory and Language*, *35*(3), 454–475. doi:10.1006/jmla.1996.0025
- Truswell, R. (2007). Extraction from adjuncts and the structure of events. *Lingua*, *117*(8), 1355–1377. doi:10.1016/j.lingua.2006.06.003
- Ueno, M., & Garnsey, S. M. (2008). An ERP study of the processing of subject and object relative clauses in Japanese. *Language and Cognitive Processes*, *23*(5), 646–688. doi:10.1080/01690960701653501
- Ueno, M., & Kluender, R. (2003). Event-related brain indices of Japanese scrambling. *Brain and Language*, *86*(2), 243–271. doi:10.1016/S0093-934X(02)00543-6
- Ueno, M., & Kluender, R. (2009). On the processing of Japanese wh-questions: An ERP study. *Brain Research*, *1290*, 63–90. doi:10.1016/j.brainres.2009.05.084
- Van Berkum, J. J. A., Koornneef, A. W., Otten, M., & Nieuwland, M. S. (2007). Establishing reference in language comprehension: an electrophysiological perspective. *Brain Research*, *1146*, 158–171. doi:10.1016/j.brainres.2006.06.091
- Van Dyke, J. A., Johns, C. L., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition*, *131*(3), 373–403. doi:10.1016/j.cognition.2014.01.007

- Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, *55*(2), 157–166. doi:10.1016/j.jml.2006.03.007
- Van Herten, M., Chwilla, D. J., & Kolk, H. H. J. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience*, *18*(7), 1181–1197. doi:10.1162/jocn.2006.18.7.1181
- Van Petten, C. (1993). A comparison of lexical and sentence-level context effects in event-related potentials. *Language and Cognitive Processes*, *8*(4), 485–531. doi:10.1080/01690969308407586
- Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition*, *18*(4), 380–393.
- Van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context on open- and closed-class words. *Memory & Cognition*, *19*(1), 95–112.
- Van Petten, C., Kutas, M., Kluender, R., Mitchiner, M., & McIsaac, H. (1991). Fractionating the word repetition effect with event-related potentials. *Journal of Cognitive Neuroscience*, *3*(2), 131–150. doi:10.1162/jocn.1991.3.2.131
- Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language*, *97*(3), 279–293. doi:10.1016/j.bandl.2005.11.003
- Van Petten, C., & Rheinfelder, H. (1995). Conceptual relationships between spoken words and environmental sounds: event-related brain potential measures. *Neuropsychologia*, *33*(4), 485–508.
- Vissers, C. T. W. M., Chwilla, D. J., & Kolk, H. H. J. (2006). Monitoring in language perception: The effect of misspellings of words in highly constrained sentences. *Brain Research*, *1106*(1), 150–163. doi:10.1016/j.brainres.2006.05.012
- Wanner, E., & Maratsos, M. (1978). An ATN approach in comprehension. In M. Halle, J. Bresnan, & G. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 119–161). Cambridge, MA: MIT Press.
- Warrington, E. K. (1984). *Manual for Recognition Memory Tests*. Windsor, England: NFER - Nelson Publishing Company Limited.

- Waters, G. S., & Caplan, D. (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). Retrieved from <http://psycnet.apa.org/psycinfo/1996-06397-007>
- Weckerly, J., & Kutas, M. (1999). An electrophysiological analysis of animacy effects in the processing of object relative sentences. *Psychophysiology*, *36*(5), 559–570.
- Weskott, T., & Fanselow, G. (2011). On the Informativity of Different Measures of Linguistic Acceptability. *Language*, *87*(2), 249–273.
doi:10.1353/lan.2011.0041
- West, W. C., & Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: an electrophysiological investigation. *Journal of Cognitive Neuroscience*, *12*(6), 1024–1037.
- Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology*, *42*(6), 654–667.
doi:10.1111/j.1469-8986.2005.00356.x