



LUND UNIVERSITY

Reading expectations

How expectations influence our reading, eye movements, opinions, and judgments

Strukelj, Alexander

2018

Document Version:

Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (APA):

Strukelj, A. (2018). *Reading expectations: How expectations influence our reading, eye movements, opinions, and judgments*. [Doctoral Thesis (monograph), Centre for Languages and Literature]. Centre for Languages and Literature, Lund University.

Total number of authors:

1

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

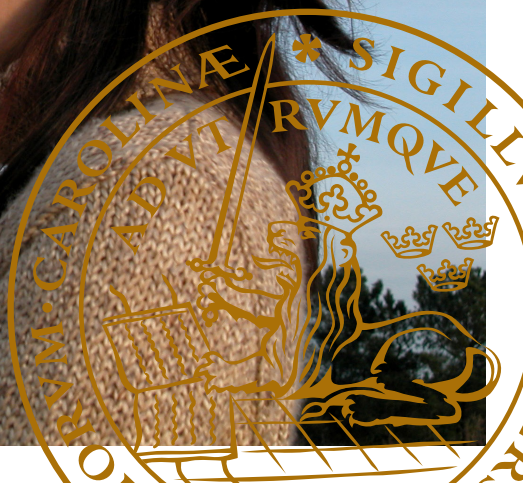
PO Box 117
221 00 Lund
+46 46-222 00 00

Reading expectations

How expectations influence our reading,
eye movements, opinions, and judgments

ALEXANDER STRUKELJ

CENTRE FOR LANGUAGES AND LITERATURE | LUND UNIVERSITY 2018



Reading is a process that frequently uses shortcuts to improve efficiency and speed. The shortcuts that help us during reading are based on things like prediction, context, biases, and previous experiences. By changing the expectations of what will be read, we change how these shortcuts operate. This book investigates the effects of expectations on reading, as evidenced by our eye movements, opinions, and judgments. In this way, we discover how the reading process works in more detail. This can help us to read more objectively, be better-informed citizens, and ultimately, become more rational people.



Reading expectations

Reading expectations

How expectations influence our reading,
eye movements, opinions, and judgments

Alexander Strukelj



LUND
UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty of Humanities and Theology, Lund University,
Lund, Sweden.

To be defended at LUX, room C121, on Friday, June 1st, 2018, 13.15-15.00.

Faculty opponent

Associate Professor Åsa Wengelin

Thesis advisors

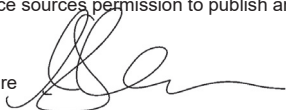
Professor Carita Paradis

Dr. Richard Andersson

Organization LUND UNIVERSITY Centre for Languages and Literature P.O. Box 201 SE-221 00 Lund SWEDEN Author(s) Alexander Strukelj	Document name DOCTORAL DISSERTATION	
	Date of issue 2018-06-01	
	Sponsoring organization	
Title and subtitle Reading expectations. How expectations influence our reading, eye movements, opinions, and judgments.		
Abstract The way we view the world is constantly affected by our expectations. The aim of this book is to determine how expectations affect the way we read. This way, we can better understand how we process information and how our expectations may change the way we read. The research is situated in experimental pragmatics and reading research. Instructions, headlines, and semantic incongruities are used to manipulate and contradict what people expect from what they read, revealing how their "default settings" operate. The methods of investigation are eye tracking to measure people's eye movements during reading, and self-reports of opinions and judgments. This book presents a series of studies. It was discovered that when people expect to read about controversial topics they tend to retain their opinions after they read the texts, but not when they expect neutral topics. Also, when people expect to read positive or negative texts, they change their opinions in line with the expected bias of the text if they do not feel strongly about the topic. Then, it was found that when people have strong opinions regarding a topic, they read the text more thoroughly, but are not that interested in the conclusions, compared to people with weak opinions. People with higher openness are more interested in reading conclusions of texts and people with lower openness are less interested in them. However, for people with weak opinions about texts they expect to agree with, the opposite pattern was found. The subsequent studies were then based on the fact that people generally assume that text and sentences are without errors. This assumption makes them sometimes fail to detect an incongruity and instead see the word they expect to see. The semantic incongruities were created by substituting words with their antonyms—for example, good with bad. It was found that both successful and failed detection of semantic incongruities affect the reading process in texts and sentences expected not to contain any errors, with people returning to the incongruent word with their eyes more frequently and spending more time on it compared to the congruent (correct) version in both cases. This shows the presence of language processing outside of awareness and highlights its importance during reading. Finally, sentences expected to originate from "less-respected" news sources result in lower acceptability judgments compared to sentences expected to originate from "well-respected" news sources, even though they in actual fact are identical. However, semantic incongruities do not affect sentence judgments. People judge the language use as being poorer when expecting sentences from "less-respected" news sources, and better when expecting sentences from "well-respected" news sources, but cannot specifically see the poor language use in incongruent sentences unless they spend some time thinking about it. In sum, the main take-home message is that expectations affect many aspects of the way people read. When reading an incongruent word in a text expected to contain no errors, people's eye movements are affected even when they do not explicitly react to the error. When people with low openness read texts expected to agree with their opinions, their eye movements are affected. When reading about topics expected to be controversial, and when reading texts expected to have evaluative bias, people's opinions are affected. Finally, when expecting sentences from less-respected or well-respected news sources, people judge the language use as poorer in the sentences from less-respected news sources.		
Key words: antonyms; cognitive linguistics; confirmation bias; expectations; experimental pragmatics; eye movements: reading; eye tracking; headlines; implicit and explicit detection of incongruities; myside bias; opinion shift; paragraph reading; processing; reading; sentence judgment; visual analog scale		
Classification system and/or index terms (if any)		
Supplementary bibliographical information		Language English
ISSN and key title		ISBN 978-91-88473-72-1
Recipient's notes	Number of pages 312	Price
	Security classification	

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature



Date 2018-04-06

Reading expectations

How expectations influence our reading,
eye movements, opinions, and judgments

Alexander Strukelj



LUND
UNIVERSITY

Front cover photo by Angel Hernandez, retrieved from Pixabay (pixabay.com), and licensed under Creative Commons Zero.

Copyright © Alexander Strukelj 2018

Humanities and Theology
Centre for Languages and Literature

ISBN 978-91-88473-72-1

Printed in Sweden by Media-Tryck, Lund University, Lund 2018



MADE IN SWEDEN 



Intertek

Media-Tryck is an environmental-
ly certified and ISO 14001 certified
provider of printed material.
Read more about our environmental
work at www.mediatryck.lu.se

Table of contents

Table of contents	7
List of tables	13
List of figures.....	16
Acknowledgements.....	19
Abstract	20
Chapter 1 Introduction.....	23
1.1 Aim and scope of the book	24
1.2 Research questions and structure of the book.....	26
Chapter 2 Background and previous research.....	29
2.1 Expectations and their relation to prediction	29
2.1.1 Expectations and reading	31
2.2 Theories on meaning.....	33
2.2.1 Implicatures, maxims, presuppositions, and the cooperative principle	33
2.2.2 Relevance.....	36
2.2.3 Linguistic competence and cognitive linguistics	37
2.3 Theories on reading.....	39
2.3.1 Word processing.....	40
2.3.2 Sentence processing	41
2.3.3 Reading models beyond sentence processing	43
2.3.4 The good enough approach to language comprehension	47
2.3.5 Context	48
2.3.6 Experimental pragmatics.....	49
2.4 Experimental methods.....	51
2.4.1 Eye-tracking methodology	51
2.4.2 Eye movements during reading.....	52
2.4.3 Statistical analyses using linear mixed-effects models.....	54
2.4.4 Questionnaire design	55

Chapter 3 The many ways of reading.....	57
3.1 Introduction.....	57
3.2 Previous research	59
3.2.1 Eye movements during reading.....	59
3.3 The study of eye movements during paragraph reading	60
3.3.1 Terms and classifications.....	61
3.3.2 Hypotheses	62
3.4 Method	64
3.4.1 Participants.....	64
3.4.2 Apparatus	64
3.4.3 Design and stimuli.....	65
3.4.4 Procedure	66
3.4.5 Data analysis.....	67
3.5 Results.....	68
3.5.1 Comprehension scores	68
3.5.2 Trial duration	69
3.5.3 Fixations and saccades.....	69
3.5.4 Total reading times per character for lines of text and for individual words.....	72
3.5.5 Total number of visits.....	75
3.6 Discussion	76
3.6.1 Thorough reading.....	77
3.6.2 Skimming.....	79
3.6.3 Spell checking.....	80
3.7 Conclusions.....	82
Chapter 4 You change your opinions (or not) when expecting controversial texts or texts with evaluative bias	85
4.1 Introduction.....	85
4.2 Previous Research	87
4.2.1 Expectations, framing, agenda setting, and affective priming.....	87
4.2.2 Newspaper headlines and click-baiting.....	88
4.2.3 Myside bias and cognitive dissonance.....	89
4.2.4 How opinions can be measured	90
4.3 The study of opinion shift after paragraph reading.....	90
4.3.1 Terms and classifications.....	91
4.3.2 Hypotheses	92

4.4 Norming data.....	96
4.4.1 Method.....	96
4.4.2 Results.....	97
4.5 Experiment 1.....	99
4.5.1 Method.....	99
4.5.2 Results.....	102
4.5.3 Discussion	104
4.6 Experiment 2.....	105
4.6.1 Method.....	105
4.6.2 Results.....	107
4.6.3 Discussion	109
4.7 Joint analysis of results from Experiments 1 and 2.....	111
4.7.1 Method.....	111
4.7.2 Results.....	112
4.7.3 Discussion	114
4.8 General discussion.....	115
4.8.1 Opinions regarding the topic: Strong or weak.....	115
4.8.2 Opinions regarding the topic: Positive or negative	116
4.9 Conclusions.....	119
Chapter 5 You change your reading (or not) when expecting controversial texts or texts with evaluative bias	121
5.1 Introduction.....	121
5.2 Previous research	122
5.2.1 Expectations and headlines, and the influence of others	122
5.2.2 Beliefs and biases	123
5.2.3 Personality and openness	124
5.2.4 Eye-tracking methodology and eye movements during reading ..	125
5.3 The study of myside bias and openness during paragraph reading.....	126
5.3.1 Terms and classifications.....	126
5.3.2 Hypotheses	127
5.4 Norming data.....	129
5.5 Experiment 1.....	129
5.5.1 Method.....	129
5.5.2 Results.....	133
5.5.3 Discussion	137
5.6 Experiment 2.....	140

5.6.1 Method.....	140
5.6.2 Results	143
5.6.3 Discussion	150
5.7 General discussion	154
5.7.1 Overall reading	154
5.7.2 Expectations, previous opinions, and openness	155
5.8 Conclusions.....	158
Chapter 6 You sometimes read the word you expect instead of an incongruent word, but you process the incongruity nevertheless Part 1: Paragraphs.....	161
6.1 Introduction	161
6.2 Previous research	163
6.2.1 Successful and failed detection of incongruities during reading ..	163
6.2.2 Lingering effects.....	167
6.2.3 Antonyms as experimental stimuli.....	167
6.3 The study of implicit detection of incongruities during paragraph reading	169
6.3.1 Terms and classifications.....	170
6.3.2 Hypotheses	174
6.4 Experiment 1.....	176
6.4.1 Method.....	176
6.4.2 Results	180
6.4.3 Discussion	182
6.5 Experiment 2.....	184
6.5.1 Method.....	184
6.5.2 Results	186
6.5.3 Discussion	190
6.6 General discussion	192
6.7 Conclusions.....	196
Chapter 7 You perceive the language in “less-respected” news sources as being poorer than in “well-respected” ones	199
7.1 Introduction	199
7.2 Previous research	200
7.2.1 Judgments of acceptability and grammaticality	200
7.2.2 Source credibility and source reliability	201
7.3 The study of source bias on sentence judgments	202

7.3.1 Terms and classifications.....	203
7.3.2 Hypotheses	203
7.4 Experiment 1.....	206
7.4.1 Method.....	206
7.4.2 Results	208
7.4.3 Discussion	210
7.5 Experiment 2.....	210
7.5.1 Method.....	211
7.5.2 Results	211
7.5.3 Discussion	213
7.6 General discussion	214
7.7 Conclusions.....	217
Chapter 8 You sometimes read the word you expect instead of an incongruent word, but you process the incongruity nevertheless Part 2: Sentences.....	219
8.1 Introduction	219
8.2 Previous research	221
8.3 The study of implicit detection of incongruities during sentence reading	221
8.3.1 Terms and classifications.....	222
8.3.2 Hypotheses	225
8.4 Experiment 1.....	227
8.4.1 Method.....	227
8.4.2 Results	231
8.4.3 Discussion	233
8.5 Experiment 2.....	235
8.5.1 Method.....	235
8.5.2 Results	236
8.5.3 Discussion	240
8.6 General discussion	241
8.7 Conclusions.....	244
Chapter 9 Further analyses.....	247
9.1 Your comprehension increases based on the time spent on the words, but the type of reading is more important.....	247
9.1.1 Introduction	247
9.1.2 Method.....	248

9.1.3 Results	248
9.1.4 Discussion	250
9.2 Your openness does not affect your opinions when expecting biased texts but actually reading neutral ones	251
9.2.1 Introduction	251
9.2.2 Method.....	252
9.2.3 Results	252
9.2.4 Discussion	252
9.3 You cannot consciously identify an incongruent sentence quickly, but your reading is instantly disrupted nevertheless.....	253
9.3.1 Introduction	253
9.3.2 Method.....	254
9.3.3 Results	254
9.3.4 Discussion	255
Chapter 10 Summary and final conclusions	257
10.1 Final conclusions	266
References.....	267
Appendices	291
Appendix 3.A Example stimuli text in English (including line breaks)	291
Appendix 3.B Remaining full statistical model outputs.....	291
Appendix 4.A Results from the joint analysis using the alternative transformation method.....	295
Appendix 4.B Full statistical model output.....	297
Appendix 6.A Stimuli texts used in Experiment 1	298
Appendix 6.B Stimuli texts used in Experiment 2	300
Appendix 6.C Example stimuli screens	302
Appendix 7.A Stimuli sentences.....	303
Appendix 8.A Figures for analyses without forgotten sentences.....	307
Appendix 8.B Figure for analyses of filler sentences in Experiment 2 (non- significant).....	308
Appendix 9.A Full statistical model outputs from Sections 9.1, 9.2, and 9.3, and remaining figure from Section 9.2	308
New analysis of data from Chapter 3	308
New analysis of data from Chapter 4 and Chapter 5	309
New analyses of data from Chapter 8.....	310

List of tables

Table	Page
Table 1.1. Overview of the chapters with the experimental studies	27
Table 3.1. Hypotheses and hypothesized effects on the experimental measures compared to regular reading	63
Table 3.2a. Mean proportions of vertical saccades and percentage point (PP) differences from baseline condition (return sweeps are removed)	70
Table 3.2b. Mean proportions of leftward saccades and percentage point (PP) differences from baseline condition (return sweeps are removed)	70
Table 3.3. Full statistical model output: Total reading times per character for lines of text as a function of line order in the text, excluding the first and last lines	74
Table 3.4. Full statistical model output: Total reading times per character for words as a function of word order in the text, when excluding skipped words, without first and last lines	75
Table 3.5. Full statistical model output: Total number of visits on words as a function of word order in the text, excluding first and last lines	76
Table 3.6. Behavior and measured effects on the experimental measures compared to regular reading	82
Table 4.1. Hypotheses and hypothesized effects on the experimental measures depending on opinion strength, Experiment 1	93
Table 4.2. Hypotheses and hypothesized effects on the experimental measures depending on positive or negative agreement score before reading, Experiment 1	94
Table 4.3. Hypotheses and hypothesized effects on the experimental measures depending on opinion strength, Experiment 2	95
Table 4.4. Hypotheses and hypothesized effects on the experimental measures depending on positive or negative agreement score before reading, Experiment 2	96
Table 4.5. Two examples of statements presented before reading and the corresponding two statements presented after reading	100
Table 4.6. Mean agreement scores after reading depending on topic, when grouping trials depending on low or high opinion strength, Experiment 1	103
Table 4.7. Mean agreement scores after reading depending on topic, when grouping trials depending on positive or negative agreement score before reading, Experiment 1	104
Table 4.8. Mean agreement scores after reading depending on type of headline, when grouping trials depending on low or high opinion strength, Experiment 2	108
Table 4.9. Mean agreement scores after reading depending on type of headline, when grouping trials depending on positive or negative agreement score before reading, Experiment 2	109
Table 4.10. Mean retention of agreement scores after reading depending on topic and headline, when grouping trials depending on positive or negative agreement score before reading, Experiment 1 and 2 combined	114
Table 4.11. Behavior and measured effects on the experimental measures depending on opinion strength, Experiment 1	117
Table 4.12. Behavior and measured effects on the experimental measures depending on positive or negative agreement score before reading, Experiment 1	117
Table 4.13. Behavior and measured effects on the experimental measures depending on opinion strength, Experiment 2	118
Table 4.14. Behavior and measured effects on the experimental measures depending on positive or negative agreement score before reading, Experiment 2	118
Table 5.1. Hypotheses and hypothesized effects on the experimental measures, Experiment 1	128
Table 5.2. Hypotheses and hypothesized effects on the experimental measures, Experiment 2	128
Table 5.3. Hypotheses and hypothesized effects on the experimental measures when accounting for openness, Experiment 2	129
Table 5.4. Behavior and measured effects on the experimental measures, Experiment 1	139
Table 5.5. Behavior and measured effects on the experimental measures, Experiment 2	153
Table 5.6. Behavior and measured effects on the experimental measures when accounting for openness, Experiment 2	154
Table 6.1. Hypotheses and hypothesized effects on the experimental measures compared to control condition, Experiment 1 (critical word) and Experiment 2 (critical word, last sentence of text, headline, and entire text)	175

Table 6.2. Hypothesis and hypothesized effects of working memory capacity on explicit detection, Experiment 2	175
Table 6.3. Implicit and explicit detection of incongruities, Experiment 1	180
Table 6.4. Mean eye movement results on the critical word, Experiment 1	181
Table 6.5. Implicit and explicit detection of incongruities, Experiment 2	187
Table 6.6. Mean eye movement results on the critical word, Experiment 2	187
Table 6.7. Mean eye movement results on last sentence, headline, and entire text, Experiment 2	187
Table 6.8. Behavior and measured effects on the experimental measures compared to control condition, Experiment 1 (critical word) and Experiment 2 (critical word, last sentence of text, headline, and entire text)	195
Table 6.9. Behavior and measured effects on explicit detection, Experiment 2	195
Table 7.1. Hypotheses and hypothesized effects on the experimental measures for incongruent sentences, compared to the control condition containing the congruent word, Experiment 1 (no source bias) and Experiment 2 (less-respected and well-respected news sources)	204
Table 7.2. Hypotheses and hypothesized effects on the experimental measures for all sentences, Experiment 2	205
Table 7.5. Behavior and measured effects on the experimental measures for incongruent sentences, compared to the control condition containing the congruent word, Experiment 1 (no source bias) and Experiment 2 (less-respected and well-respected news sources)	216
Table 7.6. Hypotheses and hypothesized effects on the experimental measures for all sentences, Experiment 2	217
Table 8.1. Hypotheses and hypothesized effects on the experimental measures on the critical word compared to the control condition, Experiment 1 and Experiment 2	227
Table 8.2. Hypothesis and hypothesized effects of source bias on explicit detection, Experiment 2	227
Table 8.3. Explicit and implicit detection of incongruities, as well as sentences that were forgotten by the participants, Experiment 1	231
Table 8.4. Mean eye movement results on the critical word, Experiment 1	232
Table 8.5. Explicit and implicit detection of incongruities, as well as sentences that were forgotten by the participants, Experiment 2	237
Table 8.6. Mean eye movement results on the critical word, Experiment 2	238
Table 8.7. Behavior and measured effects on the experimental measures on the critical word compared to the control condition, Experiment 1 and Experiment 2	244
Table 8.8. Behavior and measured effects of source bias on explicit detection, Experiment 2	244
Table 3.B1. Comprehension scores depending on instruction	291
Table 3.B2. Average fixation durations and saccade amplitudes depending on instruction	292
Table 3.B3. Proportions of leftward and vertical saccades depending on instruction	292
Table 3.B4. First fixation durations on words depending on instruction as a function of word frequency, when excluding skipped words	292
Table 3.B5. First-pass reading times on words depending on instruction as a function of word frequency, when excluding skipped words	293
Table 3.B6. Total reading times on words depending on instruction as a function of word frequency, when excluding skipped words	293
Table 3.B7. First fixation durations on words depending on instruction as a function of word length, when excluding skipped words	293
Table 3.B8. First-pass reading times on words depending on instruction as a function of word length, when excluding skipped words	294
Table 3.B9. Total reading times on words depending on instruction as a function of word length, when excluding skipped words	294
Table 3.B10. Number of visits on words depending on instruction as a function of word length	294
Table 4.A1. Mean retention of agreement scores after reading depending on topic and headline, when grouping trials depending on positive or negative agreement score before reading, using alternative transformation for agreement scores, Experiment 1 and 2 combined	295
Table 4.A2. Full statistical model output: Retention of agreement scores after reading depending on topic and headline, when grouping trials depending on positive or negative agreement score before reading, using alternative transformation for agreement scores, Experiment 1 and 2 combined	296
Table 4.B1. Retention of agreement scores after reading depending on topic and headline, when grouping trials depending on positive or negative agreement score before reading, Experiment 1 and 2 combined	297

Table 6.A1. Stimuli texts, Experiment 1	298- 299
Table 6.B1. Stimuli texts, Experiment 2	300- 301
Table 9.A1. Total reading time/character depending on comprehension scores for different reading types	308
Table 9.A2. Agreement scores after reading depending on headline (agreeing or disagreeing with opinions), when grouping trials depending on low or high opinion strength (weak or strong opinions), accounting for openness	309
Table 9.A3. Grammatical acceptability rating depending on detection, Experiment 1	310
Table 9.A4. Semantic acceptability rating depending on detection, Experiment 1	310
Table 9.A5. Grammatical acceptability rating depending on detection and expected news source, Experiment 2	310
Table 9.A6. Semantic acceptability rating depending on detection and expected news source, Experiment 2	310

List of figures

Figure	Page
Figure 3.1. Example stimulus screen	65
Figure 3.2. Mean comprehension scores depending on instruction	68
Figure 3.3. Trial duration depending on instruction	68
Figure 3.4. Average fixation durations depending on instruction	69
Figure 3.5. Mean saccade amplitudes (log-transformed) depending on instruction (return sweeps are removed)	69
Figure 3.6. First fixation durations, first-pass reading times, and total reading times on words depending on instruction as a function of word frequency per million, using a log-scale running from high to low, separated by reading type	71
Figure 3.7. First fixation durations, first-pass reading times, and total reading times on words depending on instruction as a function of word length, separated by reading type	71
Figure 3.8. Total number of visits on words depending on instruction as a function of word length, separated by reading type	71
Figure 3.9. Total reading times per character for lines of text and for individual words (excluding skipped words) depending on instruction	73
Figure 3.10. Total reading times per character for lines of text depending on instruction as a function of line order in the text	74
Figure 3.11. Total reading times per character for words as a function of word order in the text, excluding skipped words, without first and last lines, depending on instruction	75
Figure 3.12. Total number of visits on words as a function of word order in the text, without first and last lines, depending on instruction	76
Figure 4.1. Visualization of hypothesized opinion shifts after reading, Experiment 1	93
Figure 4.2. Visualization of hypothesized opinion shifts after reading, Experiment 2	95
Figure 4.3. Texts divided into two groups in Experiment 1, depending on the mean opinion strength scores for each topic from the norming data	98
Figure 4.4. Mean agreement scores before and after reading, when grouping trials depending on low or high opinion strength, Experiment 1	103
Figure 4.5. Mean agreement scores before and after reading, when grouping trials depending on positive or negative agreement score before reading, Experiment 1	104
Figure 4.6. Mean agreement scores before and after reading, when grouping trials depending on low or high opinion strength, Experiment 2	108
Figure 4.7. Mean agreement scores before and after reading, when grouping trials depending on positive or negative agreement score before reading, Experiment 2	109
Figure 4.8. Mathematical formula for transformation of agreement scores after reading (post-reading agreement scores), to remove regressions to the mean	112
Figure 4.9. Mean retention of agreement scores after reading depending on topic and headline, depending on trials with positive or negative agreement score before reading, Experiment 1 and 2 combined	113
Figure 5.1. Example stimulus, Experiment 1	132
Figure 5.2. Mean total reading times/pw (ms) on lines of text, factored by weak or strong opinion, divided into topic condition, Experiment 1	134
Figure 5.3. Mean total reading times/pw (ms) on words as a function of word number, factored by weak or strong opinion, divided into topic condition, Experiment 1	135
Figure 5.4. Mean first-pass reading times/pw (ms) on lines of text, factored by weak or strong opinion, divided into topic condition, Experiment 1	136
Figure 5.5. Mean first-pass reading times/pw (ms) on words as a function of word number, factored by weak or strong opinion, divided into topic condition, Experiment 1	136
Figure 5.6a. Example stimulus, positive bias, Experiment 2	142
Figure 5.6b. Example stimulus, negative bias, Experiment 2	142
Figure 5.7. Mean total reading times/pw (ms) on lines of text, factored by weak or strong opinion, divided into headline bias condition, Experiment 2	144
Figure 5.8. Mean total reading times/pw (ms) on words as a function of word number, factored by weak or strong opinion, divided into headline bias condition, Experiment 2	145
Figure 5.9. Mean first-pass reading times/pw (ms) on lines of text, factored by weak or strong opinion, divided into headline bias condition, Experiment 2	146

Figure 5.10. Mean first-pass reading times/pw (ms) on words as a function of word number, factored by weak or strong opinion, divided into headline bias condition, Experiment 2	146
Figure 5.11. Mean total reading times/pw (ms) on lines of text as a function of openness, factored by weak or strong opinion, divided into headline bias condition, Experiment 2	148
Figure 5.12. Mean total reading times/pw (ms) on words as a function of openness, factored by weak or strong opinion, divided into headline bias condition, Experiment 2	149
Figure 5.13. Mean first-pass reading times/pw (ms) on lines of text as a function of openness, factored by weak or strong opinion, divided into headline bias condition, Experiment 2	150
Figure 6.1a. Schematic diagram of the two different outputs created by the two parallel reading processes after fixating on an incongruent word	172
Figure 6.1b. Schematic diagram of the output similarity evaluator during normal operation	173
Figure 6.1c. Schematic diagram of the output similarity evaluator when the difference between the two outputs elicits a reaction	173
Figure 6.2. Flow chart of post-experiment interview, Experiment 1	179
Figure 6.3. Mean first fixation durations (ms) on the critical word for the control condition, implicit detection, and explicit detection, Experiment 1	181
Figure 6.4. Mean number of revisits (count) to the critical word for the control condition, implicit detection, and explicit detection, Experiment 1	182
Figure 6.5. Mean total reading times (ms) on the critical word for the control condition, implicit detection, and explicit detection, Experiment 1	182
Figure 6.6. Flow chart of post-experiment interview, Experiment 2	186
Figure 6.7. Mean first-pass reading times (ms) on the last sentence for the control condition, implicit detection, and explicit detection, Experiment 2	188
Figure 6.8. Mean number of revisits (count) to the critical word for the control condition, implicit detection, and explicit detection, Experiment 2	189
Figure 6.9. Mean total reading times (ms) on the critical word for the control condition, implicit detection, and explicit detection, Experiment 2	189
Figure 6.10. Mean total reading times (ms) on last sentence of text for the control condition, implicit detection, and explicit detection, Experiment 2	189
Figure 6.11. Mean whole-text reading times (s) for the control condition, implicit detection, and explicit detection, Experiment 2	189
Figure 6.12. Mean number of explicitly detected incongruities depending on OSPAN scores, Experiment 2	190
Figure 7.1. Mean grammatical acceptability, semantic acceptability, form-based interest, and topic-based interest ratings for congruent critical, incongruent critical, and congruent filler sentences, Experiment 1	209
Figure 7.2. Mean grammatical acceptability, semantic acceptability, form-based interest, and topic-based interest ratings for congruent critical, incongruent critical, and congruent filler sentences, divided by the expected source of the sentence, Experiment 2	212
Figure 8.1a. Schematic diagram of the two different outputs created by the two parallel reading processes after fixating on an incongruent word	224
Figure 8.1b. Schematic diagram of the output similarity evaluator during normal operation	224
Figure 8.1c. Schematic diagram of the output similarity evaluator when the difference between the two outputs elicits a reaction	225
Figure 8.2. Mean total reading times (ms) on the critical word for the control condition, implicit detection, and explicit detection, Experiment 1	232
Figure 8.3. Mean selective regression path durations (ms) on the critical word for the control condition, implicitly detected incongruent words, and explicitly detected incongruent words, Experiment 1	233
Figure 8.4. Proportion of explicit detections, implicit detections, and forgotten sentences, depending on source bias condition, Experiment 2	237
Figure 8.5. Mean total reading times (ms) on the critical word for control condition, implicitly detected incongruent words, and explicitly detected incongruent words, depending on source bias condition, Experiment 2	238
Figure 8.6. Mean selective regression path durations (ms) on the critical word for the control condition, implicitly detected incongruent words, and explicitly detected incongruent words, depending on source bias condition, Experiment 2	239
Figure 9.1. Mean comprehension scores and mean total reading times/character for regular reading, thorough reading, skimming, and spell checking, Chapter 3	249
Figure 9.2. Total reading time/character (ms) for regular reading, thorough reading, skimming, and spell checking, depending on comprehension scores, Chapter 3	249

Figure 9.3. Grammatical and semantic acceptability ratings for congruent sentences, implicitly detected incongruent sentences, explicitly detected incongruent sentences, and filler sentences, Chapter 7, Experiment 1	254
Figure 9.4. Grammatical and semantic acceptability ratings for congruent sentences, implicitly detected incongruent sentences, explicitly detected incongruent sentences, and filler sentences, divided into source bias condition, Chapter 7, Experiment 2	255
Figure 4.A1. Mathematical formula for alternative transformation of agreement scores after reading (post-reading agreement scores), to remove regressions to the mean	295
Figure 4.A2. The mathematically reduced transformation described in Figure 4.A1	295
Figure 4.A3. Mean retention of agreement scores after reading depending on topic and headline, depending on trials with positive or negative agreement score before reading, using alternative transformation for agreement scores, Experiment 1 and 2 combined	296
Figure 6.C1. Example stimuli (control condition), Experiment 1	302
Figure 6.C2. Example stimuli (control condition), Experiment 2	302
Figure 8.A1. Mean total reading times (ms) for control condition, implicitly detected incongruent words, and explicitly detected incongruent words, when removing forgotten sentences from the analysis, Experiment 1	307
Figure 8.A2. Mean selective regression path durations (ms) on the critical word for the control condition, implicit detection, and explicit detection, when removing forgotten sentences from the analysis, Experiment 1	307
Figure 8.A3. Mean total reading times (ms) for control condition, on the critical word for the control condition, implicit detection, and explicit detection, when removing forgotten sentences from the analysis, Experiment 2	307
Figure 8.A4. Mean selective regression path durations (ms) on the critical word for the control condition, implicit detection, and explicit detection, when removing forgotten sentences from the analysis, Experiment 2	307
Figure 8.B1. Mean total reading times per character (ms) for words in filler sentences, depending on source bias condition, Experiment 2	308
Figure 9.A1. Absolute difference in agreement scores before and after reading depending on openness, Chapter 4, Experiment 2, and Chapter 5, Experiment 2	309

Acknowledgements

I want to extend my most heart-felt gratitude to my two supervisors Carita Paradis and Richard Andersson. Thank you for your excellent advice, valuable input, guidance, and unwavering support during the writing of this book. This book is so much better because of you both.

I also want to thank all of my colleagues at the Lund University Humanities Lab. I have learned so much about eye tracking, experimental design, and so much more from you, and for that I am very grateful.

Thanks to all of my colleagues at the English Linguistics department at Lund University. The discussions of linguistics research and all manners of language related issues have greatly improved my understanding of the complexities of language and language research.

Finally, I want to thank all members of the Rayner Eye-tracking Lab at the University of California, San Diego for allowing me to stay for three months doing eye-tracking research. Collecting large amounts of data, talking to the researchers, spending time in the lab, and participating in the research seminar was very rewarding and made me progress as a researcher. Also, thanks the Olof Sager Foundation for the generous grant that financed my stay.

Abstract

The way we view the world is constantly affected by our expectations. The aim of this book is to determine how expectations affect the way we read. This will help us better understand how we process information, and how our expectations may change the way we read. The research is situated in experimental pragmatics and reading research. Instructions, headlines, and semantic incongruities are used to manipulate and contradict what people expect from what they read, revealing how their “default settings” operate. The methods of investigation are eye tracking, to measure people’s eye movements during reading, and self-reports of opinions and judgments.

This book presents a series of studies. It was found that when people expect to read about controversial topics, they tend to retain their opinions after they read the texts, but not when they expect neutral topics. Also, when people expect to read positive or negative texts, they change their opinions in line with the expected bias of the text if they do not feel strongly about the topic. Further, it was found that when people have strong opinions regarding a topic, they read the text more thoroughly, but are not as interested in the conclusions compared to people with weak opinions. People with a higher degree of openness are more interested in reading conclusions of texts, and people with lower openness are less interested in them. However, for people with weak opinions about texts they expect to agree with, the opposite pattern was found. The subsequent studies were then based on the fact that people generally assume that text and sentences are free of errors. This assumption makes them sometimes fail to detect an incongruity, and instead see the word they expect to see. The semantic incongruities were created by substituting words with their antonyms—for example, *good* with *bad*. It was found that both successful and failed detection of semantic incongruities affect the reading process in texts and sentences expected to be free of errors, with people returning to the incongruent word with their eyes more frequently, and spending more time on it compared to the congruent (correct) version in both cases. This demonstrates the presence of language processing independent of awareness, and highlights its importance during reading. Finally, sentences expected to originate from “less-respected” news sources result in lower acceptability judgments compared to sentences expected to originate from “well-respected” news sources, even though they in actual fact are identical. Semantic incongruities, however, do not affect sentence judgments. People judge the language usage as being poorer when expecting sentences from “less-respected” news sources, and better when expecting sentences from “well-respected” news sources, but cannot specifically see the poor language usage in incongruent sentences unless they spend some time thinking about it.

In sum, the main take-home message is that expectations affect many aspects of the way people read. When reading an incongruent word in a text expected to be free of errors, people’s eye movements are affected even when they do not explicitly react to

the error. When people with low openness read texts expected to agree with their opinions, their eye movements are affected. When reading about topics expected to be controversial, and when reading texts expected to have evaluative bias, people's opinions are affected. Finally, when expecting sentences from less-respected or well-respected news sources, people judge the language usage as poorer in the sentences from less-respected news sources.

Chapter 1

Introduction

Every day, we encounter large amounts of written information that we need to pay attention to, assess, and have an opinion about: information in newspaper articles, messages on social media platforms, op-ed pieces on internet news sites, commercial messages on billboards, angry notes about the dirty dishes in the common room, and so on. Being able to understand and evaluate information correctly is crucial in an era of internet trolls, agenda journalism, and accusations of fake news. To deal with all of this information, research has shown that our brain tends to take shortcuts, and these shortcuts can affect how we think and how we react to the information around us. Research also suggests that these shortcuts are facilitated by our expectations. We predict upcoming words and information, and we read expected words more quickly than unexpected words (Otten & Van Berkum, 2008; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005). It has also been shown that we prefer to read information that agrees with our own opinions (Klayman, 1995; Mercier, 2017; Oswald & Grosjean, 2004). Another phenomenon is that we generally assume that texts we read are correct. Accordingly, it has been shown that sometimes words that are not quite right (or not even right at all) go unnoticed, and are “replaced” with those words we expect to read, without us even noticing (Daneman, Lennertz, & Hannon, 2007; Ferreira & Patson, 2007). These shortcuts and preferences are integral parts of the reading process.

On the first page of the introduction to the *Oxford Handbook of Pragmatics*, pragmatics is defined as “the study of language use in context” (Huang, 2017, p. 1). This book focuses primarily on the context of our everyday lives, which forms the basis for our expectations. Our expectations color many aspects of behavior, perhaps even all of them; we expect, for instance, a certain level of objectivity from news sources that are considered to be well-respected, or a certain ending to a novel. As mentioned above, we know that during the reading process we regularly take shortcuts in order to improve efficiency and speed. These shortcuts are based on factors like expectations, predictions, context, biases, world knowledge, and previous experiences. Our expectations, in turn, can affect the shortcuts that we take, both consciously and unconsciously. For instance, if we expect that a text will have a positive evaluative bias in relation to a certain topic, we might be more attentive to positive points made in it. This expectation may even unconsciously affect our

opinions, making them more positive toward the topic in question. Also, if we expect poor grammar, we might be more watchful of language usage and style. This expectation might even unconsciously affect our judgment of the content negatively. This book will use experimental methods to investigate reading from a cognitive standpoint (Gärdenfors, 2014). The work is based within the field of *experimental pragmatics*, and can be described as using experimental methods to investigate the way context and expectations affect the interpretation of language (Gibbs, 2017; Noveck & Sperber, 2004; Sperber & Noveck, 2004). Experimental pragmatics thus provides an excellent framework for this book, and, arguably, an excellent framework when the research interest is to answer questions about reading as performed in language situations similar to those we encounter in our daily lives.

Knowledge about effects that expectations have on reading has important social implications. We can better understand the effectiveness of propaganda, the way people can miss egregious errors if they do not expect them (and be certain that they were not there), and how people's biases affect and change the way they perceive a text.

1.1 Aim and scope of the book

The aim of this book is to determine to what extent, and in what way, expectations regarding texts and sentences affect our opinions, judgments, and reading behavior. Reading is a well-studied field in experimental research, and eye tracking is a well-established method within the field. Also, expectations and biases have been extensively researched within fields such as media psychology and cognitive science (Kuypers, 2002; Mercier, 2017). However, not much research has combined expectations and reading. By manipulating people's expectations, we change the way people behave, the way shortcuts operate during reading, and the reading process itself.

The scope of the book is to determine how different expectations regarding texts and sentences affect the reading of these, as well as the opinions and sentence judgments after reading. The book is mainly concerned with a global view of the text reading process at the level of paragraphs, and not isolated sentences and local effects such as the word frequency effect (Bicknell & Levy, 2012; Kennedy, Pynte, Murray, & Paul, 2013). To form the basis for the studies performed in the subsequent chapters, Chapter 2 presents the relevant background information and discusses previous research.

The first study in this book, in Chapter 3, uses paragraphs of text to determine the eye movement characteristics of thorough reading, skimming, and spell checking, and how these reading types differ from regular reading. Using paragraphs of text rather than single sentences provides insights regarding different reading strategies and the

process of meaning integration at the discursive level, in which a fair amount of context is included. There has not been a great deal of research on thorough reading, as this is difficult to examine in isolated sentences (see Rayner, 1998, 2009, for overviews). The general assumptions are that increased processing is correlated with longer reading times, and therefore longer reading times signify more thorough reading. This assumption is evaluated in Chapter 3 by giving the participants specific instructions on how to read each text. In regard to the reading instructions in all subsequent studies in this book, the participants were instructed to read as they would regularly read for comprehension, which was the instruction for regular reading in Chapter 3.

Chapter 4 and Chapter 5 determine how opinions and eye movements are affected by reading a text expected to have positive or negative bias, or a text expected to be about a topic other people feel strongly about. Much research has taken an interest in opinions and opinion change (e.g., Yaniv & Milyavsky, 2007), and a large body of research has looked into cognitive biases (e.g., Stanovich, West, & Toplak, 2013). Since effects are commonly induced by manipulations of the texts given to the participants, Chapter 4 and Chapter 5 investigate whether these effects can be elicited merely by changing people's expectations regarding the texts they will read.

Chapter 6 and Chapter 8 determine how reading is affected by both successful and failed detection of semantic incongruities in texts and sentences that are expected to be free of errors. A fair amount of research has investigated the failure to detect incongruities in text (Bohan & Sanford, 2008; Daneman et al., 2007; Hannon & Daneman, 2004; Roberts, Langstein, & Galantucci, 2016; Sanford, Leuthold, Bohan, & Sanford, 2011). Previous studies, however, have only found that participants frequently miss the incongruent word in the text in favor of the word expected in the context, and have not found experimental effects stemming from the failed detection. In an attempt to produce the effects of failed detection in reading, I use stimuli better suited for eye-tracking research, namely commonly used antonym pairs. Antonyms have the advantage of being both maximally similar and maximally different at the same time (Paradis & Willners, 2011), since they express opposite properties of one and the same meaning dimension. Previous research suggests that the reading process consists of two parallel processes: one that processes the words on the page, and another that gives outputs based on expectations and predictions (e.g., Ferreira & Patson, 2007; Jackendoff, 2007). If an incongruity is not detected (according to the participant), but the participant fixates on that word with his or her eyes, it can be argued that the expected word is "read," rather than the incongruent word. If effects can be seen in eye movement measures, this can be taken as experimental evidence of these parallel processes in action.

Chapter 7 determines how sentence judgments are affected by antonymic incongruities, expectations of correctness, and the expectations that sentences originate from either well-respected or less-respected news sources. A great deal of previous research on this topic uses sentence judgment tasks (e.g., Gibson, Piantadosi,

& Fedorenko, 2011; Schütze & Sprouse, 2013), but expectations are not commonly used as manipulations. Also, in order to establish how the detection of incongruities is affected by the expectation of poor language usage in Chapter 8, Chapter 7 establishes a manipulation that induces an expectation of either good or poor language usage in the sentences, namely that the sentences were taken from either well-respected or less-respected news sources. Finally, Chapter 9 answers new questions that arise from these studies by performing new analyses of the results, and Chapter 10 contains the summary and final conclusions of the book.

Two experimental methods are employed in this book: eye tracking during reading, and participant responses on visual analog scales before and after reading. Eye tracking is an online measure that shows changes in the reading process with high sensitivity. Responses on visual analog scales is an offline measure that shows changes in people's opinions and perception, and is argued to be a somewhat subjective (Schütze, 1996, p. 120).

1.2 Research questions and structure of the book

The specific research questions in this book are as follows.

1. What are the eye movement characteristics during regular reading, thorough reading, skimming, and spell checking, and how do they differ?
2. How are opinions affected by reading a neutral text expected to be about a topic other people feel strongly about, or expected to have a positive or negative evaluative bias?
3. How are eye movements during reading affected by reading a neutral text expected to be about a topic other people feel strongly about, or expected to have a positive or negative evaluative bias?
4. How are eye movements during reading affected by failed and successful detection of semantic incongruities in *texts* that are expected to contain no errors?
5. How are sentence judgments affected by semantic incongruities and by expectations of well-respected or less-respected news sources?
6. How are eye movements during reading affected by failed and successful detection of semantic incongruities in *sentences* that are expected to contain no errors?

The book starts with a chapter covering background information and the previous research that forms the basis of the studies in the subsequent chapters. The studies are

then presented in six chapters, with discussions on how they can contribute to our understanding of expectations and reading, and how they can enrich the research field of reading in general. In order to answer the first research question, participants were given different instructions on how to read the texts. This study forms a baseline for the other studies, in which the “regular” reading behavior is changed by manipulating the participants’ expectations regarding the texts and sentences. The subsequent research questions are investigated by evoking different expectations about what the participants will read through the instructions given prior to reading or the headlines accompanying the text. Table 1.1 shows the experimental method, the size of the text stimuli given in each experimental trial, and the specific type of analysis used in each chapter. The penultimate chapter contains analyses performed by combining experiments and methods from separate chapters. The final chapter contains a summary of the answers to all research questions, and the final conclusions of the book.

Table 1.1. Overview of the chapters with the experimental studies

* Chapter 3 has also been published in *Journal of Eye Movement Research* as One page of text: Eye movements during regular and thorough reading, skimming, and spell checking (Strukelj & Niehorster, 2018). It is presented in this book with minor corrections to the language. † Visual analog scales are self-report scales similar to Likert scales, but without any visible intervals, and are capable of returning a higher number of different results (typically from 0 to 100 in steps of 1).

	Chapter 3*	Chapter 4	Chapter 5	Chapter 6	Chapter 7	Chapter 8
<i>Experimental method</i>	Eye tracking	Visual analog scales†	Eye tracking	Eye tracking	Visual analog scales†	Eye tracking
<i>Stimuli</i>	Paragraph	Paragraph	Paragraph	Paragraph	Sentence	Sentence
<i>Type of analysis</i>	Reading: All words in text	Opinion on topic	Reading: All words in text	Reading: Single word	Sentence judgment	Reading: Single word

The rationale behind the structure of the book is to first perform broad analyses, and then to narrow them down. Eye movements during paragraph reading are first investigated by giving instructions to read the text in specific ways. Eye movements during reading are then investigated without instructions for trying to read in a specific way, beginning with paragraph reading using paragraph processing analyses, continuing with paragraph reading using sentence-processing analyses, and ending with sentence reading using sentence-processing analyses. Opinion shifts after reading and sentence judgments after reading are also investigated in separate chapters.

Chapter 2

Background and previous research

In this chapter, I first present research on expectations and the related notion of prediction, followed by research on the relationship between expectations and reading. I then present theories of meaning that are relevant to experimental pragmatics, such as theories that deal with notions of implicatures and relevance. This also includes cognitively-oriented theories that are discursive and usage-based, and in which meaning in language may be described as highly dynamic and encyclopedic in nature. I then present theories on reading, beginning with reading models based on word and sentence processing, and continuing with reading models that extend beyond sentence processing by incorporating factors such as context. This is the case with the framework of experimental pragmatics, which is presented at the end of this section. Finally, I present research on different experimental methods, along with methods of data analysis used in this book.

2.1 Expectations and their relation to prediction

Expectations influence all aspects of life. Expectation, or the related term prediction, has different meanings in different fields of research. Although expectation and prediction are used interchangeably in certain fields, there is arguably a difference between the terms (Ferguson, Scheepers, & Sanford, 2010; Kuperberg & Jaeger, 2016; Levy, 2011a). According to the Oxford English Dictionary, the primary definition of expectation is “The action or fact of anticipating or foreseeing something; the belief that something will happen or be the case” (“Expectation,” 2017). The primary definition of prediction is “The action of predicting future events; an instance of this, a prophecy, a forecast” (“Prediction,” 2017). The key difference could be described as expectation being a passive process and prediction being an active process; passively expecting something versus actively predicting something. This distinction, however, is difficult to maintain in practice, as we know that both expectation and prediction occur largely without conscious involvement from the reader or listener during language processing, and, indeed, during every second of our daily lives. People make predictions based on their expectations in a

given context and situation. Expectations and predictions are intertwined in a causal relationship. They are two sides of the same coin.

It has been shown that expectation and prediction affect us in a wide variety of situations. In medicine, for instance, the mere expectation of positive effects from a drug can make a patient feel better (Moerman & Jonas, 2002; Stewart-Williams & Podd, 2004), and in experimental science, the mere expectation of a certain result in a study can produce that result (Doyen, Klein, Pichon, & Cleeremans, 2012). The anticipation of future events has also been shown to guide visual perception. In a functional magnetic resonance imaging (fMRI) study, the hypothesized location of the prediction mechanism in the brain was even identified (Ekman, Kok, & de Lange, 2017). Speech perception is also hypothesized to be guided by prediction, with the hearer actively predicting what the speaker will say (Pickering & Garrod, 2013). This is supported by the fact that noise does not hinder comprehension (Bergen & Goodman, 2015). This prediction of incoming speech has been found in the neural responses of both macaque monkeys and humans (Kikuchi et al., 2017). With regard to language, predictive language processing has been shown to be a crucial part of the language processing system (Federmeier, 2007; Kamide, 2008; Kuperberg & Jaeger, 2016; Kutas, DeLong, & Smith, 2011; Van Petten & Luka, 2012).

Previous research suggests that the brain constantly makes predictions about the world, and then compares these predictions to the actual input (A. Clark, 2013; Hohwy, 2013, 2015). This can improve the efficiency of processing, minimize surprise, and conserve cognitive resources (Friston, 2005; Friston & Kiebel, 2009; Srinivasan, Laughlin, & Dubs, 1982). Theories on predictive coding suggest that the brain predicts upcoming visual input, an idea that has been validated by eye tracking and fMRI studies (G. Edwards et al., 2017). The brain is even described as “a prediction machine,” evolved to help us successfully navigate a complex and rapidly changing environment (Van Berkum, 2010). Predictive processing is something that has been extensively researched, and is suggested to be a crucial part of language comprehension (Bonhage, Mueller, Friederici, & Fiebach, 2015). It has recently been shown to occur during both spoken (Sohoglu, Peelle, Carlyon, & Davis, 2012) and written language processing (Jakuszeit, Kotz, & Hasting, 2013). For instance, just hearing a word will drive eye movements to the corresponding object in a scene (Andersson, Ferreira, & Henderson, 2011), even when only the first syllable has been heard (Altmann & Kamide, 1999; Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995). It has been demonstrated that words can be activated even before hearing the first syllable based solely on expectations, such as anticipating a noun with a vowel sound after the indefinite article “an,” or a noun with a consonant sound after the indefinite article “a” (DeLong, Urbach, & Kutas, 2005; Foucart, Ruiz-Tada, & Costa, 2015; Gagnepain, Henson, & Davis, 2012). Predictive processing has also been found in sentence comprehension (Bonhage et al., 2015; Fine, Jaeger, Farmer, & Qian, 2013), with the prediction of the upcoming words demonstrated by measuring brain activity (Dikker & Pykkänen, 2013).

Prediction is crucial during reading, so a change in the expectations of what will be read should alter the way these predictions are manifested during reading. Importantly, these changes should be largely unconscious, as predictive processing is shown to operate mostly outside of conscious control. By introducing the expectation of the context in which a certain behavior would be activated, we can see how the behavior operates. Indeed, reading and speech processing are both highly reliant on expectations (Haberlandt, 1982; Mak & Sanders, 2012; Pickering & Garrod, 2013; Van Berkum et al., 2005). In language processing and comprehension research, prediction and expectation are terms that describe very similar concepts, and the difference between the terms is not always clearly specified. In this book, expectations are manipulated in order to manipulate people's predictions.

2.1.1 Expectations and reading

It is argued that perception in general “is guided by the anticipation of future events” (Ekman et al., 2017). In reading, effects of prediction have been found to already occur within 400 milliseconds of presenting a word, as seen in N400 effects in event-related potential (ERP) studies (Lau, Holcomb, & Kuperberg, 2013), which also suggest that meaning is more strongly predicted than form (Ito, Corley, Pickering, Martin, & Nieuwland, 2016). In reading studies, words in isolation are presented to participants, and their brain responses, or event-related potentials, are measured for differences such as that between seeing a congruent or an incongruent word. In fact, speakers often only use fragments of sentences, but are nevertheless able to convey their entire message. It is suggested that a speaker creates a model of the listener, and that this model is used to predict what the listener needs to hear in order to understand the speaker (Pickering & Garrod, 2013). In other words, speakers optimize their message based on their predictions of what the listeners need to hear in order to correctly understand the message (Levy & Jaeger, 2007).

Text comprehension can be described as a product of the information in the text, as well as the reader's knowledge of the information. The processing of discourse involves creating a mental representation of the input (Graesser, Millis, & Zwaan, 1997; Zwaan & Rapp, 2006). Based on this, the reader creates expectations for the text, which facilitates reading (Haberlandt, 1982). Accordingly, expectation is very important during comprehension (Levy, 2008; Mak & Sanders, 2012). One example is discourse markers, such as *even so*, which are shown to heavily influence what the reader expects from the subsequent text (Drenhaus, Demberg, Köhne, & Delogu, 2014; Kuperberg, Paczynski, & Ditman, 2011; Xiang & Kuperberg, 2015). Pronoun usage is affected by verbs (Hartshorne & Snedeker, 2013), showing that the verb in the main clause affects the preferred referent of a pronoun with “implicit causality,” such as in the contrasting examples “Sally frightens Mary because she...” and “Sally loves Mary because she...” The pronoun is in principle ambiguous, but most English speakers think the pronoun refers to Sally in the first example, but to Mary in the

second example (Garvey & Caramazza, 1974; Hartshorne, 2013). Online studies have used self-paced reading and eye tracking to show that pronouns inconsistent with the implicit causality in the verb disrupt reading (Koornneef & Van Berkum, 2006; Pyykkönen & Järvikivi, 2009) and integration (A. J. Stewart, Pickering, & Sanford, 2000). Bott and Solstad (2014) have proposed a novel account of causal discourse that models the expectations about upcoming discourse created by implicit causality. This model ultimately aims to create a formal theory of explanatory biases in discourse (Solstad & Bott, 2013). The predictions from the theory were confirmed in a production study, and it was also shown that the account of implicit causality in the theory was cross-linguistically valid (Bott & Solstad, 2014). The types of expectations that arise from this implicit causality of verbs, but also other features of discourse, such as relative clause processing, have been called *discourse expectations* (Roland, Mauner, O'Meara, & Yun, 2012).

During reading, the expectation of what word will come next is suggested to be automatic to a large extent, as studies show effects in early eye movements that are connected to online processes we cannot control. It is suggested that rather than being influenced by word priming, readers indeed predict the upcoming words “on the fly” (Otten & Van Berkum, 2008). This has been shown by people’s eye movements being drawn to objects by only hearing parts of words, or even when the grammatical object is just implied by a verbal construction (Altmann & Kamide, 1999; Boland, 2005). Indeed, verbs seem to have constraints that produce specific lexical preferences; that is, expectations of what will follow the specific verbal construction (Trueswell, Tanenhaus, & Kello, 1993). Word order is another factor that affects processing, highlighting the importance of expectation and prediction during reading (Arnold, Wasow, Asudeh, & Alrenga, 2004; Kaiser & Trueswell, 2004; Slioussar, 2010). It has been demonstrated in many experimental studies that a word that is highly predictable leads to faster reading, as evidenced in shorter fixation durations (Fernández, Shalom, Kliegl, & Sigman, 2014; Kennedy et al., 2013; Kliegl, Grabner, Rolfs, & Engbert, 2004; Mak & Sanders, 2012; Rayner & Duffy, 1986).

Further effects of expectation and prediction have been shown in a Dutch study using ERP, in which adjectives inconsistent with the predictions that gender inflections give rise to in the Dutch language triggered both different N400 effects and slower reading (Van Berkum et al., 2005). Prediction (or likelihood) of the subsequent word, estimated by cloze probability, and cueing from context have both been shown to produce N400 effects when using expected and unexpected words. To analyze both predictive processing and recovery from incorrect predictions, studies have used *garden path sentences*, which are sentences that intentionally mislead the reader into making an incorrect parse. An example of a garden path sentence is *When Anna dressed the baby spit up on the bed*. The sentence is usually initially analyzed as Anna dressing the baby, but after the words *spit up*, the sentence needs to be reanalyzed. These studies have shown that different discourse contexts produce different sentence reading times (Britt, Perfetti, Garrod, & Rayner, 1992), and that

recovery from incorrect analyses is easier when the target sentence is in discourse focus; that is, when referents are currently active in the context (Rayner, Garrod, & Perfetti, 1992). Expectation and prediction have thus been shown to affect reading to a very high degree (Britt et al., 1992; Rayner et al., 1992; Van Berkum et al., 2005). These expectations can be seen as micro-level processing of constructions. There are also expectations that can be seen as macro-level processing of constructions, however, which refers to, for example, expecting a certain type of text when reading a headline, or guessing the ending of a story you are reading. Accordingly, unpredictable or surprising words are indeed more difficult to process (Boston, Hale, Kliegl, & Vasishth, 2008; Hale, 2001; Levy, 2011b). When encountering surprising (or, indeed, unpredictable) words, parallel processing seems to explain much of the eye movement behavior, since many candidates need to be considered in parallel during reading; that is, when a reader moves back and forth between several different discourse referents (Boston, Hale, Vasishth, & Kliegl, 2011).

As we can see above, many effects of expectations have been shown to operate very quickly, almost instantly when a word is fixated. Many other effects of expectations, however, operate within longer time frames, and on larger areas than just one word. It is these types of effects that are of interest in this book; expectations such as reading about a controversial topic, reading a biased text, and reading sentences from well-respected or less-respected news sources. In order to investigate these effects of expectations on reading, we need to first understand meaning.

2.2 Theories on meaning

In this section, I discuss the notions of *implicatures*, *maxims*, *presuppositions*, and the *cooperative principle* as phenomena that are relevant to human communication and to how we read texts. I then discuss the notion of relevance during communication, as specified in, for instance, the maxim of relation and in relevance theory. The section ends with a presentation of the framework of cognitive linguistics, with a specific focus on the *Lexical Meaning as Ontologies and Construals* framework.

2.2.1 Implicatures, maxims, presuppositions, and the cooperative principle

The theory of implicatures attempts to disentangle, conceptualize, and understand the different types of meaning that can arise in different contexts. Grice introduced the concept of implicature as a type of implicit meaning (Baldwin, 2006; Davis, 2015; Grice, 1957, 1981). Consider these three examples from Davis (2015):

- (1) The fact that *John has a temperature* means that he is sick.
- (2) *John has a temperature* means that John's temperature is above normal.
- (3) By saying *John has a temperature*, Steve meant that John has a thermometer.

The phrase *John has a temperature* is present in all examples. In (1), it is stated that John is sick, based on the fact that he has a temperature. This is called evidential meaning, which entails that there are correlations and causal connections. In (2), it is only suggested that the phrase indicates that John's temperature is above normal. This is referred to as sentence meaning, which entails that the sentence is not drawing on evidential meaning. It refers to the sentence, rather than the speaker. However, in (3), it is indicated that when Steve used the phrase, it had a different meaning than that which is conventional. This is called speaker meaning, which entails that a speaker has a different intention, or uses a different implicature, when using a sentence (Davis, 2015).

According to the theory of *conversational implicature* (Grice, 1957, 1981), communication is governed by an underlying principle in which language is used with maximum efficiency, the cooperative principle (Huang, 2007, p. 25). The principle is divided into nine maxims of conversation in four categories: quality, quantity, relation, and manner. The maxims are speaker-oriented in the way that they encompass proper cooperation, but addressee-oriented in the way that we all assume this proper cooperation. The maxims define proper principles of cooperation from a speaker's standpoint, and are thus speaker-oriented. However, they can also be said to be addressee-oriented, because hearers take for granted that speakers adhere to them. The maxims of quality are: *Do not say what you believe is false*, and *Do not say that for which you lack adequate evidence*. As an example of conversational implicature of quality, consider the sentence *Tim Berners-Lee invented the World Wide Web in 1989*. The sentence implies that the speaker believes that Tim Berners-Lee invented the World Wide Web in 1989, and has adequate evidence that he did so. The maxims of quantity are: *Make your contribution as informative as required*, and *Do not make your contribution more informative than is required*. As an example of conversational implicature of quantity, consider the sentence *John has six credit cards*. The sentence implies that John has at most six credit cards. The maxim of relation is: *Be relevant*. As an example of conversational implicature of relation, consider the conversation between John and Mary: *John: What's the time? Mary: The museum hasn't opened yet*. The sentence spoken by Mary implies that the time is sometime before the museum normally opens. The maxims of manner are: *Avoid obscurity of expression*, *Avoid ambiguity*, *Be brief*, and *Be orderly*. As an example of the maxim of manner, consider the sentence *John went to a McDonald's and bought two hamburgers*. The utterance implies that John first went to a McDonald's, and then he bought two hamburgers.

However, it can also mean that John went to a McDonald's in order to buy two hamburgers (Grice, 1989; Huang, 2007, pp. 27-28).

A presupposition is, simply put, a proposition that is assumed to be true when uttering a sentence. This proposition operates as a precondition in order for this uttered sentence to function properly (Huang, 2007, p. 65). Implicatures are highly reliant on context, and the meaning is based on what the hearer expects the meaning to be. In other words, the hearer predicts what the speaker might mean, based on a set of rules, namely the maxims of conversation. Presuppositions, on the other hand, are not reliant on context in that sense. The meanings of presuppositions do not change depending on the context, but rather evoke predictions based on our knowledge of the world. A textbook example of presupposition is the sentence *The king of France is bald*, which presupposes that there is in fact a king of France. Similarly, the sentence *John knows that Baird invented television* presupposes that Baird invented television, and *Mary has stopped beating her boyfriend* presupposes that Mary has been beating her boyfriend (Huang, 2007, p. 65). A lexical item or linguistic construction that gives rise to a presupposition is called a *presupposition trigger*. Presuppositions can be analyzed in three ways. When using a *filtering-satisfaction* analysis, also called a *Karttunen-Stalnaker* analysis, presuppositions cannot be cancelled. Instead, any unwanted presuppositions need to be filtered out by the propositions present in the text. This can be achieved in three ways. In the example *John doesn't regret running the red light, because in fact he never did so*, a plug blocks the presupposition that John ran a red light. In the example *It's unlikely that John will play the piano again*, a hole prevents a change being made to the presupposition that John played the piano before. In the example *If the bishop promotes the politically incorrect, then he will regret doing so*, a filter screens out some of the presuppositions (Huang, 2007, pp. 76-78; Karttunen, 1973; Stalnaker, 1973). The filtering-satisfaction theory is rightfully questioned. When using a *cancellation* analysis, presupposition triggers only introduce potential presuppositions, which will only become actual ones if they are not cancelled. In the example *John hasn't discovered that Angola is in Asia*, the potential presupposition that Angola is in Asia is cancelled by our real-world knowledge that this is in fact false. Indeed, the example *The king of France is bald* above would not give rise to the presupposition that a king of France exists when using this analysis, as our real-world knowledge tells us that France is a republic.

Context is created in order of priority, and additional propositions can only be added if they do not contradict the previous propositions. Background assumptions are placed into context first, followed by contextual factors. Semantic entailments are then added, followed by conversational implicatures, with presuppositions added last, if, and only if, they do not contradict any of the previous propositions (Gazdar, 1979; Huang, 2007, pp. 81-83; Soames, 1979). When using *accommodation* analysis, which is a combination of filtering-satisfaction and cancellation analyses, the meaning of a sentence and the presuppositions have the potential to change the context. This refers to the dynamic process in which unfulfilled presuppositions can be updated, or

repaired, in discourse. In the example *I'm sorry I'm late, my car broke down*, the presupposition that the speaker arrived in his or her car is merely added to the context as if it had been there all along (Heim, 1983; Huang, 2000, p. 239, 2007, pp. 85-86).

There are two types of presuppositions: one relies on world knowledge, such as *The king of France is bald*, which relies on the world knowledge that there is a King of France (which is, as we know, inaccurate), and the other relies on lexical presupposition triggers to introduce a presupposition, such as in the statement *John regretted that he was late*, which presupposes that John was late, or the question *Have you stopped beating your wife?* which presupposes that you are, in fact, beating your wife. Importantly, both presuppositions and implications are crucial components in human communication.

2.2.2 Relevance

As mentioned in Section 2.2.1, the maxim of relation is: Be relevant. A cooperative speaker should only convey information that is relevant in the context of the utterance (Grice, 1989; Huang, 2007, pp. 27-28; Spector, 2013). The degree of similarity to relevance as defined in relevance theory is high. Relevance theory offers a way of approaching and sorting the different factors that influence comprehension, and focuses on the addressee, which is another reason it is appropriate for this book. In relevance theory, relevance is defined as “a property of inputs to cognitive processes which makes them worth processing” (Wilson & Sperber, 2012, p. 62). Relevance is not used in the regular senses of the word, but rather in a technical sense. The inputs can be both external stimuli—for instance, a smell or the sound of an utterance—and internal representations that undergo processing—for instance, a memory or the linguistic decoding of an utterance. We are constantly faced with many more inputs than we can process. Therefore, our processing resources need to be efficiently allocated to maximize cognitive benefits by “processing the inputs that offer the best expected cost/benefit ratio at the time” (Wilson & Sperber, 2012, p. 62). It has been suggested that the most important type of cognitive benefit is improvement in both theoretical and practical knowledge. The first principle of relevance, described as the cognitive principle, is that the human cognitive system favors the most relevant inputs: “the human cognitive system tends toward processing the most relevant inputs available” (Wilson & Sperber, 2012, p. 64). In communication, for instance, speakers thus strive to provide utterances that are relevant enough to be worth processing (Wilson & Sperber, 2012, p. 64). The second principle of relevance, described as the communicative principle, is that the hearer presumes that the utterance is relevant: “every utterance conveys a presumption of its own relevance” (Wilson & Sperber, 2012, p. 65). The hearer of an utterance follows a “path of least effort” when interpreting the utterance, and stops when he or she arrives at an interpretation that satisfies his or her expectations of how much meaning can (and should) be extracted from the utterance (Wilson & Sperber, 2012, p. 65).

The big difference is that in relevance theory, relevance originates from the utterance itself, while relevance in the maxim of relation originates from the speaker as an assumption that he or she should only convey relevant information in an utterance. In other words, while Grice bases the maxim of relation on communication, relevance theory also heavily relates this to cognition.

2.2.3 Linguistic competence and cognitive linguistics

In some of the pragmatics literature presented above, linguistic knowledge is considered to be the information actually present in an utterance, while non-linguistic knowledge is the information that is not present in an utterance, but nevertheless conveyed by it (Huang, 2007). For instance, in the generative framework, it is assumed that there is a clear distinction between language and thought, and that the output from different modules of language is joined into a single language of thought. This view has been questioned by scholars in cognitive linguistics and cognitive science (Gärdenfors, 2000; Langacker, 1999; Paradis, 2003; Tomasello, 2000), mainly because of its narrow view of language and lack of analysis of the complexities of authentic language in use, in favor of constructing well-formed sentences. In cognitive linguistics, an absolute boundary between linguistic and encyclopedic knowledge is shown to be very problematic, and therefore no strict dichotomy between the two is considered (Paradis, 2003, 2015). Meaning in language is dynamic and usage-based, and expression is directly mapped on a conceptual structure (Langacker, 1999; Paradis, 2003, 2015), and linguistic and non-linguistic knowledge are considered to form a continuum, since they cannot be divided on linguistic grounds (Paradis, 2003). Four tenets are central to cognitive semantics:

1. Meanings are activated conceptual patterns in the cognitive system. Meanings are based on speaker and hearer construals of situations.
2. Conceptual patterns are not independent of perceptions and bodily experience.
3. Conceptual domains are the representations on which certain modes of construal may be applied in order to structure the content in different ways.
4. Cognitive models are image-schematic and can be changed by metaphoric and metonymic operations.

(from Paradis, 2003)

These tenets all highlight the dynamics of meaning creation in language usage. The first tenet states that meaning in language is evoked by the activation of concepts in our minds and construed on the fly in human communication. The second tenet asserts that meaning in language is embodied and based on how we perceive the

world, and that we can construe different linguistic expressions based on how we perceive different situations. The third tenet states that there are different domains. There are basic domains such as color, taste, smell, and the ability to experience space and time, as well as more complex domains that are learned. When we communicate, linguistic expressions evoke sets of domains in different constellations, depending on what we want to communicate. These domains characterize various aspects of the meaning of the expression in the specific occasion of use in discourse. The fourth and final tenet asserts that we form conceptual image-based models based on our experience of the world, and that these structures organize our thoughts across more abstract domains (Paradis, 2003). In cognitive semantics, not all meaning is considered equal. In language usage, the meaning of words is defined using four centrality factors, defined on a scale based on how *conventional*, *generic*, *intrinsic*, and *characteristic* a word or concept is. All factors are separate from each other, but tend to correlate quite significantly (Langacker, 1999; Paradis, 2003). The “conventional” factor refers to how general the knowledge is about the word or concept; an example is the highly familiar word *banana*, which will be used to explain all centrality factors. The “generic” factor refers to how specific a word or a concept is, with the fact that a specific person likes eating sliced banana with their breakfast cereal being specific knowledge, and the fact that sliced banana goes well with breakfast cereal being generic knowledge. The “intrinsic” factor refers to how related to external entities the properties of a concept are, with “shape” being more intrinsic and “size” being less intrinsic. The “characteristic” factor refers to how unique or identifiable the properties of a word or concept are, with “shape” generally being more characteristic than “color,” since a banana is much easier to identify by shape than color.

Paradis (2003) discusses whether the concept of linguistic competence is viable from a cognitive linguistics perspective. Specifically, whether it is possible to pin down, isolate, or define linguistic competence, whether this presupposes a well-defined lexicon, and how (or if) it can contribute to the interpretation of lexical expression in language. The conclusion is that the notion of linguistic competence is not relevant to cognitive linguistics, since it presupposes modularity, which is something that does not exist within cognitive linguistics. In contrast to generative accounts, Paradis (2003) argues for two inseparable types of conceptual knowledge of meanings: knowledge about the word and knowledge about the world in which the word is located. Since these are inseparable, it is therefore not possible to separate linguistic and non-linguistic meaning. Finally, since the notion of linguistic competence does not reflect the nature of natural languages, language competence should be considered a non-question. By considering lexical knowledge to be something that arises from language usage in all its complexity, the cognitive linguistic framework aims to describe and explain all facets of lexical knowledge.

The framework that forms the basis of the definition of the meaning of language in this book is the Lexical Meaning as Ontologies and Construals (LOC) (Paradis, 2015), which is situated in the broad framework of cognitive linguistics. Within the

LOC framework, words used in speech and writing are not described as *having* meanings, but rather *evoking* meanings. Word forms are described as “cues to structures in conceptual space, and it is this space that is the entire use potential of a word” (Paradis, 2015, p. 274). In this view, lexical knowledge relies heavily on knowledge, understanding, and experience of the world. In the dynamic usage-based LOC framework, the meaning of words is dependent upon the context in which they are used (Paradis, 2015, p. 275). It has been argued that context can be described as a subset of a hearer’s assumptions about the world (Sperber & Wilson, 1995, p. 15), and as propositions, mental representations, and assumptions (Fetzer, 2017, p. 271; Wilson & Sperber, 2012, p. 15). In the *explicit context* of a race, the *implicit context* triggers background knowledge about racing in general, the high value generally attributed to winning, the fact that winners are normally given prizes, and so on (Wilson & Sperber, 2012, pp. 290-292) (for a thorough discussion of context, see Section 2.3.5). In this book, expectations are induced by specifying a certain explicit context, such as a text about a controversial topic, or a sentence taken from a less-respected news source. These expectations elicit a certain type of reading behavior related to the implicit context. Based on our experience of how the world works, examples of implicit context for the examples of explicit context above could be that opinions regarding controversial topics are harder to change, or that less-respected news sources use poor language in their articles, respectively. In the same way, an expectation of a text with positive evaluative bias creates a context that colors the meaning of the entire text, rather than individual words.

Having thus outlined my understanding of meaning, we can now look at theories on reading. Theories on reading create models of how reading works, based on experimental evidence from studies that examine reading in great detail. These models can be divided into three main categories: word processing models; sentence-processing models; and models with a broader approach than isolated sentences, which account for factors such as expectations and context.

2.3 Theories on reading

In this section, I will present reading models that investigate word processing and sentence processing, as well as reading models that use factors beyond single sentences to investigate reading. The good enough approach to language comprehension, a reading model based on the assumption that two parallel processes are used during reading, is then discussed. One of these processes is a syntactic analysis based on the word on the page, and the other uses fast and frugal heuristics to facilitate reading. These heuristics can be said to be based on factors like expectations, prediction, and world knowledge. Context, which has an important influence on reading, and can be

seen as a crucial factor for the heuristics analysis in the good enough approach, is then presented. This section finishes with a presentation of experimental pragmatics.

2.3.1 Word processing

Word processing can be divided into two main models: one is commonly called the feed-forward model (Schriefers, Meyer, & Levelt, 1990), and the other is commonly called the interactive model (McClelland & Rumelhart, 1981). Another model that is gaining support is named the *embodied semantics* approach (Glenberg & Robertson, 2000). All three models concern the comprehension of individual words. In the feed-forward model, lexical access to a word consists of two serial stages. Semantic properties are extracted first—that is, the meaning of the word—followed by the phonological characteristics of the word—that is, the form of the word (Schriefers et al., 1990). Importantly, if an error occurs in one of the stages, there is no way to repair the error without starting the process over again; hence the term “feed-forward.” In the interactive model, as in the feed-forward model, lexical access occurs serially, with meaning extracted before form (McClelland & Rumelhart, 1981). However, if an error occurs, it is possible in lexical access to repair the error by moving backward to a previous step. The feed-forward model is easy to empirically test, as it allows for clear, falsifiable predictions. The model seems to correctly predict results, as shown by empirical evidence in Schriefers et al. (1990). In a picture-naming task, pictures were presented and interfering words were played with a slight delay. The results showed that semantically-related words interfered at early stages of presentation, while a facilitatory effect was found for phonologically-related words at later stages of presentation. This suggests that its meaning is activated first, followed by its form (Schriefers et al., 1990). The interactive model is much more difficult to falsify, which is one of the reasons that there is much less empirical data to support it. The embodied semantics theory argues that all meaning is grounded in reality and human action (Glenberg & Robertson, 2000). To test this theory, Glenberg and Kaschak (2002) used a lexical decision task with sentences that were to be judged in terms of how sensible they were using either congruent or incongruent motion. A sentence such as “close the drawer” implies a motion away from the body. Pushing a button far from the body would thus be a congruent motion, and pushing a button near the body would be an incongruent motion. Accordingly, reaction times were shown to be longer when the motion event was incongruent. Also, merely hearing the words “lick,” “pick,” and “kick” activates the motor system connected with moving the mouth, arm, and leg, respectively (Pulvermüller, 2005).

Semantic priming refers to the observation that hearing or seeing words will prime other, related words. This is shown in lexical decision tasks, in which a word must be judged to be either a word or a non-word. Semantic priming shows that if a word related to the target word is shown before the target word, the target word is classified as a word more quickly than if the semantic prime had not been a related word

(Moss, Ostrin, Tyler, & Marslen-Wilson, 1995). It is very difficult, however, to quantify how these links are formed. Spreading activation between words show how words prime each other. In order for the entire network not to be activated, the amount of priming needs to be limited somehow. The semantic network theory investigates and classifies semantic relatedness in order to specify how and when words activate each other (Collins & Loftus, 1975). Word meaning is seen as what comes to mind when someone utters the word (Collins & Loftus, 1975). The *Latent Semantic Analysis* model (Landauer & Dumais, 1997) suggests that meaning in language is made up of abstract symbols related to what they signify in a space that is based on the co-occurrence of these words. This is also the argument in the *Hyperspace Analogue to Language* (Burgess & Lund, 1997). The meaning of a word is the relation to all other words in the space.

When investigating lexical activation, only one word is shown to participants. In lexical decision tasks, one word is shown and needs to be classified as either a word or a non-word, measuring and analyzing the time before this decision is made. This is a problem when investigating sentences using the *rapid serial visual presentation* (RSVP) paradigm (words are shown at specified intervals) and the self-paced reading paradigm (the participant presses a button after reading a word to see the next one). These paradigms do not measure reading, but rather decision-making. When words are strung together in sentences, the picture becomes increasingly complex.

2.3.2 Sentence processing

Research as to how different sentence-processing factors influence comprehension remains largely inconclusive. There are two main theories in sentence processing, which has led to the development of two types of models of reading. As with the word processing models, they can be divided into two-stage and interactive models (Ferreira & Engelhardt, 2006; Pickering & van Gompel, 2006). Two-stage models suggest that initial processing is modular; that is, completed independently, but that this is not necessarily the case for subsequent processing. As with the word processing models, when an error is made, the interactive model allows for bidirectionality in order to repair the misparse, while in the two-stage model the reader needs to “complete another loop.”

The *garden path* (GP) model is a very influential two-stage model that is modular and serial, with syntax being processed first, followed by meaning, and finally phonology (Frazier, 1979; Frazier & Rayner, 1982). If difficulties are encountered during parsing, the model requires completing another loop. Similar to the feed-forward word processing model, the GP model of sentence processing allows for clear predictions that can be examined empirically. The GP model assumes *minimal attachment*, meaning that an ambiguous phrase is interpreted based on the use of the fewest possible nodes in a phrase-structure tree, thus producing the simplest interpretation. Furthermore, *late closure* assumes that input prefers to attach to the

constituent currently being processed, rather than to constituents farther back in the sentence (Altmann, van Nice, Garnham, & Henstra, 1998). Even though empirical evidence has been found in favor of the GP model, semantically-driven expectations are difficult to reconcile with it. The *active filler strategy* is a recent addition to the GP model, explaining the processing of unbounded dependencies such as wh-questions and relative clauses (Fiebach, Schlesewsky, & Friederici, 2001). A sentence like “Which girl do you believe John loves a lot?” is argued to have a *gap* after “John loves” that should contain “which girl,” since English uses a subject-verb-object word order. The active filler strategy argues that whenever an opportunity to fill a gap is encountered, filling it is preferred. Consequently, “which girl” in the above example will fill the gap with minimal processing cost, resulting in the interpretation of the sentence as a question concerning John’s romantic interest.

Interactive models, on the other hand, assume that all sources of information can be used during sentence processing. Current interactive models are often called *constraint satisfaction* models, in which all possible analyses of a sentence are activated in parallel. Analyses are then discarded depending on the available information, until only one is chosen as the final analysis (Pickering & van Gompel, 2006). When one single analysis receives significantly more support, processing is easy, but when two analyses receive a similar amount of support, difficulties will arise. In order for a constraint-based model to correctly predict language processing, all constraints would need to be identified, which would make the model very complex. Recent computational models have successfully predicted processing, however, given various sources of information (e.g., Spivey & Tanenhaus, 1998).

Similarly, it is argued that frequency plays a major role in comprehension, since frequent analyses are favored over those that are infrequent. One influential theory suggests that parsing is dependent on the frequency of co-occurrence, that is, the constructions in which the words appear (Pickering & van Gompel, 2006, p. 464). Consider the examples “The student forgot the solution was in the book” and “The student hoped the solution was in the book.” According to large corpora, the word “forgot” is frequently followed by a direct object, producing the completed analysis that the student “forgot the solution.” The word “hoped” is frequently followed by a complement, producing an incomplete analysis in which “hoped the solution” is awaiting a complement stating where the student hoped the solution to be. Regarding the two examples above, Trueswell et al. (1993) show that without the word “that,” the sentence with “forgot” was more difficult to process.

Another factor highlighting the difference between modular or interactive sentence processing is *semantic plausibility*, which uses varying levels of plausibility to elicit different readings of a sentence. In a study, McRae, Spivey-Knowlton, and Tanenhaus (1998) showed that the sentence “the crook arrested by the detective was guilty of taking bribes” presented difficulties at “arrested” for the participants, because we know that crooks cannot arrest people, making the natural analysis of the sentence implausible (Pickering & van Gompel, 2006, p. 467, from McRae et al., 1998). In

the GP model, plausibility information is ignored during initial stages, whereas all possible analyses are activated according to theories based on constraint-satisfaction. Research into semantic plausibility has indicated the need for sentence processing to incorporate context and expectations into the models (Gibson, Bergen, & Piantadosi, 2013).

2.3.3 Reading models beyond sentence processing

Research has shown that sentence-processing models need to take into account factors such as semantic plausibility, expectations, and context (Bohan, Leuthold, Hijikata, & Sanford, 2012; Ferreira & Engelhardt, 2009). Sanford and Sturt (2002), among others, argue that predictive processing and ambiguous or vague sentences that counterintuitively pose little or no processing cost for readers indicate a number of situations that many current sentence-processing models cannot account for. Their argument is further strengthened with the presentation of errors that readers can reliably be induced to make based on sentence structure, content, and context. This is very similar to the *good enough approach* to language comprehension by Ferreira and Patson (2007), which Sanford and Sturt acknowledge.

First, Ferreira and Patson (2007) talk about ambiguities such as the pronoun *it* in “Mary bought a brand new Hitachi radio. **It** was in Selfridge’s window. Later, when Joan saw **it**, she too decided it would be a good purchase.” They claim this demonstrates underspecification in language, and conclude that this does not hinder people. When seeing the “it” in the second sentence, the pronoun could either mean the radio in the window, or that type of radio only. In the third sentence, “it” has even more interpretations. Another example is given in (4), followed by either (5a) or (5b) (Sanford & Sturt, 2002).

(4) Kelly showed every photo to a critic last month.

(5a) The critic was from a major gallery.

(5b) The critics were from a major gallery.

When participants are given either (5a) or (5b) immediately after seeing (4), no preference or advantage is found during the processing of either one, suggesting that both interpretations are considered equally likely, and therefore both are being kept active during the processing of (4). Only after seeing either (5a) or (5b) is one chosen over the other, thus resolving the ambiguity.

An eye-tracking study contrasted radically ambiguous words like “pupil” and concrete/abstract ambiguous words like “ball” with related ambiguous words like “newspaper” (Frazier & Rayner, 1990). The word “pupil” has two very different meanings: “student” and “part of eye,” as in (6a), (6b), (7a), and (7b). The word

“ball” has one concrete and one abstract meaning: “a ball that you kick” and “a dance function.” The word “newspaper” has two very similar meanings: “a printed newspaper” and “a newspaper company,” as in (8a), (8b), (9a), and (9b). The study showed that people are able to delay their decision when reading a related ambiguous word without additional processing cost, signifying that people do not “decide” which of the two ambiguous meanings is applicable. They do, however, determine the meaning immediately when reading a radically ambiguous word or a concrete/abstract ambiguous word. These consequences were found by showing disambiguating information after the occurrence of an ambiguous target, with lengthened reading times found for words with very different meanings, such as “ball” and “pupil,” but not for words with multiple senses, such as “newspaper.” Frazier and Rayner (1990) used the measure “mean reading time per character” for the entire sentence, the target word, the post-target region, and the disambiguating region.

- (6a) Deciding that the theorem was too difficult to understand, the **pupil** appeared distant and far away.
- (6b) Turning on the overhead lights abruptly and without warning, the **pupil** appeared distant and far away.
- (7a) Naturally the **pupil** appeared distant and far away, after deciding that the theorem was too difficult to understand.
- (7b) Naturally the **pupil** appeared distant and far away, after turning on the lights abruptly and without warning.
- (8a) Lying in the rain, the **newspaper** was destroyed.
- (8b) Managing advertising so poorly, the **newspaper** was destroyed.
- (9a) Unfortunately, the **newspaper** was destroyed, lying in the rain.
- (9b) Unfortunately, the **newspaper** was destroyed, managing advertising so poorly.

Sanford and Sturt (2002) talk about various semantic anomalies that are analyzed incorrectly by participants, who interpret the anomalies according to what they expect them to mean rather than what they actually mean. This again shows the role that predictive processing plays in language processing. A world knowledge violation that has been used in several studies is the *Moses illusion*, often referred to as a semantic illusion (Bottoms, Eslick, & Marsh, 2010; Büttner, 2002; Oostendorp & De Mul, 1990; Park & Reder, 2004; Reder & Kusbit, 1991). The question “How many animals of each sort did Moses put on his ark?” is posed to participants, and the spontaneous answer of *two* is incorrect, as it was Noah, not Moses, who built the ark.

Another frequently used semantic anomaly poses the question “When an airplane crashes on a border with debris on both sides, where should the survivors be buried?” to participants (Barton & Sanford, 1993; Bohan et al., 2012; Daneman et al., 2007; Hannon & Daneman, 2004). Frequently, participants do not realize that living people should not be buried, but instead misinterpret the meaning.

The term shallow processing is used to explain why general knowledge, pragmatics, and discourse affect processing (Sanford & Graesser, 2006; Sanford et al., 2011; Sanford & Sturt, 2002). Since the interpretation is erroneous, it is evidently not sufficiently processed, and becomes an underspecified parse. Shallow processing is considered to be a feature that allows for efficient parsing, and for ambiguous words to stay ambiguous during parsing without processing cost, as is the case with the underspecified pronouns mentioned above. Since no ERP effects are found for non-detected anomalies in appropriate contexts, it is assumed that semantic information is not utilized (Sanford et al., 2011). Further evidence for the existence of incomplete semantic representation is *text-change blindness* (Sturt, Sanford, Stewart, & Dawydiak, 2004). The paradigm is an extension of the *change blindness* paradigm, which shows that people are blind to changes in visual scenes (Rensink, 2002; Silverman & Mack, 2006; T. J. Smith, Lamont, & Henderson, 2012). One way of eliciting change blindness is by showing a somewhat cluttered image with several objects to participants, and after a very brief white flash, showing the same image but with one of the objects removed. This change is only noticed after a delay, and in some cases, it is not noticed at all. Text-change blindness uses sentences presented visually, changing a word in the sentence, and measuring if the participant notices the change. Sturt et al. (2004) show that meaning representations of words are not complete, since people were more likely miss the change when the words were not in linguistic focus.

- (10a) [Focus on the cider] Everyone had a good time at the pub. A group of friends had met up there for a stag night. **What Jamie really liked was the cider**, apparently.
- (10b) [Focus on Jamie] Everyone had a good time at the pub. A group of friends had met up there for a stag night. **It was Jamie who really liked the cider**, apparently.
- (11a) [Focused] **Everybody was wondering which man got into trouble**. In fact, the man with the hat was arrested.
- (11b) [Unfocused] **Everybody was wondering what was going on that night**. In fact, the man with the hat was arrested.

Change detection was measured in both experiments. In examples (10a) and (10b) from Experiment 1, *cider* was changed to *beer* (related word) or to *music* (unrelated word), with similar substitutions for *hat* in examples (11a) and (11b) from Experiment 2. In both experiments, the unrelated words were almost always detected, but related meanings out of linguistic focus were detected to a much lesser degree.

Similarly, the context can significantly change the interpretation of a sentence. A study showed that in the absence of context, when reading sentences such as (12a) and (12b) participants preferred the agentive version in (12a) to the locative version in (12b) (Pickering & van Gompel, 2006, p. 472, from Liversedge, Pickering, Branigan, & van Gompel, 1998). In other words, they spent more processing time on (12b) compared to (12a). In a follow-up experiment, however, they were given the context in (13) before reading either the locative or the agentive sentence. After introducing the expectation that a location could be specified, both sentences were shown to be easy to process (Pickering & van Gompel, 2006, p. 472, from Liversedge et al., 1998).

(12a) The shrubs were planted by the apprentice that morning.

(12b) The shrubs were planted by the greenhouse that morning.

(13) The gardener wondered where to plant the shrubs.

When analyzing ERP data, semantic anomalies and syntactic anomalies have been shown to sometimes produce similar results. Measuring the difference in brain response between the correct sentence “The cat will **eat**” and the semantically anomalous sentence “The cat will **bake**” shows an N400, a negative peak in the voltage emitted by the brain at around 400 ms (Borovsky, Elman, & Kutas, 2012). Similarly, measuring the difference in brain response between the correct sentence “The cat will **eat**” and the syntactically anomalous sentence “The cat will **eating**” shows a P600, a positive peak in voltage at around 600 ms (Hahne & Friederici, 1999; Kotz, Holcomb, & Osterhout, 2008). So, at first glance, it seems that the ERP data support the serial GP model, which claims that semantics is analyzed before syntax. The serial GP model uses garden path sentences, such as *When Anna dressed the baby spit up on the bed*, to lead the participant to perform an incorrect analysis of a sentence, and then investigate the reanalysis. N400s, however, can be elicited by incongruent case markings in German, which is a syntactic feature (Frenzel, Schlesewsky, & Bornkessel-Schlesewsky, 2011; Frisch & Schlesewsky, 2001), and P600s can be elicited by semantic anomalies (van Herten, Kolk, & Chwilla, 2005). This calls into question the “meaning comes first” account of the GP model, which is further undermined by many ERP and eye-tracking studies. It seems that the GP model is mostly supported by reaction-time studies and RSVP reading studies, which both measure decision making rather than reading. Skipping a word is impossible in

those types of studies, however, forcing participants to pay attention to words they would normally skip, such as function words and predictable words in general (Brysbaert, Drieghe, & Vitu, 2005), and the word “the” in particular (Abbott, Angele, Ahn, & Rayner, 2015).

2.3.4 The good enough approach to language comprehension

Previous research strongly suggests that separate processing routes are used in parallel during reading (Clahsen & Felser, 2006; Jackendoff, 2007; Kim & Osterhout, 2005; van Herten, Chwilla, & Kolk, 2006). This is also the case in the good enough approach to language comprehension, which assumes that reading uses one processing route that performs a syntactic analysis of each word, and another that uses fast and frugal heuristics based on semantic plausibility (Christianson, Luke, & Ferreira, 2010; Ferreira, 2003; Ferreira, Bailey, & Ferraro, 2002; Ferreira & Patson, 2007). In this approach, the heuristics analysis is considered to be favored during language processing, as it facilitates and automates reading. The heuristics can be said to be based on factors such as expectations, predictions, world knowledge, and previous experiences. If the two processing routes are different, however, such as when an incongruity or error is encountered, the heuristics analysis is often selected and the incongruity or error is not consciously noted. According to this view, the incongruity is at least partially processed, but is disregarded in favor of the expected word in the context. If this is indeed the case, the effects of this implicit detection should be present in the reading process, when compared to texts without incongruities. Ferreira and Patson (2007) suggest that the way the reading process compensates for certain incongruities, errors, and disfluencies is similar to spoken language processing, in which this type of compensation is frequent. Furthermore, it is assumed that the comprehension system only produces a meaning representation that is “good enough” for reading to continue, using less information than normally assumed (Ferreira & Patson, 2007). This would explain the many cases in which incongruities in text are not explicitly detected by participants, who instead read what they expected to see. For instance, when given sentences like *the dog was bitten by the man*, participants often incorrectly state that the dog bites the man because that is congruent with their expectations (Ferreira, 2003; Ferreira & Engelhardt, 2009). A good enough representation of the text is created that does not correctly correspond to the meaning of the text. The partial match hypothesis used to explain the Moses illusion (Kamas, Reder, & Ayers, 1996; Park & Reder, 2004) could be described as being the effect of a good enough representation.

Since the heuristics analysis is considered to be favored during language comprehension in the good enough approach, expectations are thus regarded to be a very important part of reading (Ferreira & Patson, 2007). Pickering and Garrod (2013) also show that predicting the upcoming word in a text plays a central role in language comprehension and production. Consequently, if a word is highly

incongruent with regard to the context, such as in the case of antonyms, there is a great discrepancy between what the reader expects the word to be and the actual word. Similarly, speakers use models to make predictions about certain aspects of their upcoming utterances, and listeners in turn use models based on their own potential utterances in order to predict the utterances by the speaker (Pickering & Garrod, 2013). While the comprehension systems for listening and reading are different, Ferreira and Patson (2007) argued that certain processes are shared between the two. In addition, Frazier and Clifton (2015) note that errors in written texts are repaired in the same manner as speech, with this corrected input used during interpretation.

2.3.5 Context

As mentioned, it has been argued that sentence-processing models need to take context into account in order to accurately describe the reading process (Bohan et al., 2012; Ferreira & Engelhardt, 2009). Context has been conceptualized in many diverse fields, such as natural language communication, information technology, and social-action-based analyses (Fetzer, 2017, p. 259). Because of its wide use, it is difficult to describe or formalize the concept, and robust theories for context are lacking (Meibauer, 2012). Attempts to do so have been made, however, such as at the aptly named CONTEXT interdisciplinary conferences (Blackburn, Ghidini, Turner, & Giunchiglia, 2003; Fetzer, 2017, p. 259). Context is an important concept in communication, which is seen as both context-dependent and context-creating (Fetzer, 2017, pp. 259-260). In cognitive linguistics (discussed in depth in Section 2.2.3), meaning is highly dependent on the context (Paradis, 2015, p. 275). Similarly, in relevance theory (discussed in depth in Section 2.2.2), context is considered to be a psychological construct, “a subset of the hearer’s assumptions about the world” that affects the interpretation of an utterance much more than how the world actually is (Sperber & Wilson, 1995, p. 15). Relevance theory differentiates between cognitive environment and cognitive context. The cognitive environment is described as being the set of facts manifested in the mental landscape of a listener (Sperber & Wilson, 1995, p. 39; Wilson & Sperber, 2012, p. 87), while the cognitive contexts are made up of propositions, mental representations, and contextual and factual assumptions (Fetzer, 2017, p. 271; Wilson & Sperber, 2012, p. 15). Cognitive contexts are subjective participant interpretations that are negotiated in situations such as conversations (Fetzer, 2017, p. 272). If insufficient context is given, comprehension and recall are negatively affected (Bransford & Johnson, 1972). People’s background knowledge also strongly affects comprehension, recall, and recognition (Long, Prat, Johns, Morris, & Jonathan, 2008; Long, Wilson, Hurley, & Prat, 2006; Long & Prat, 2002). Factors such as good writing quality and domain knowledge also affect comprehension, with good writing mostly improving recollection, and good domain knowledge mostly improving the correctness of these recollections (Moravcsik &

Kintsch, 1993). Wilson and Sperber (2012) differentiate between explicit and implicit context. Explicit context can be provided to participants, such as giving the two premises *A is ahead of B* and *A is ahead of C* with the explicit context *A, B and C were the top three finishers in the race last Sunday*. It has been shown that the explicit context of “a race” make more participants arrive at the conclusion *A is ahead of both B and C*. Explicit context, however, evokes implicit context. The explicit context about the race above also triggers background knowledge about racing, the value attributed to winning, the fact that winners are usually given prizes, and so on. In the example above, it is therefore likely that further conclusions are drawn because of this implicit context, such as A being happy, A being given a medal or other prize for winning, and so on. (Wilson & Sperber, 2012, pp. 290-292).

Speakers want hearers to construe certain contexts, based on the purpose of what is being communicated. The context for the language situation is based on the assumption that different ways of expressing ourselves do not occur at random, but are meaningful and functional (Fetzer, 2017, p. 265). Non-default usage, such as using informal expressions in the context of political discourse, or using formal expressions in the context of everyday conversation, signals a different type of meaning than default usage would. It is thus important to investigate not only what is said and what is meant, but also if it is said in default or non-default contexts. In the fairly new field of experimental pragmatics, the impact of context on language production and interpretation is investigated by explicitly using and adapting contextual parameters (Fetzer, 2017, p. 260). Context is reflected in a participant’s “implicit and explicit references to context and in her/his negotiation of the appropriateness of the references in order to co-construct local and global contexts” (Fetzer, 2017, p. 261). In other words, as is the case in both relevance theory and cognitive linguistics, the basis of experimental pragmatics is that context is crucial during the construction of meaning.

2.3.6 Experimental pragmatics

The field of experimental pragmatics began to develop in the mid-1970s, when certain psychologists began investigating language in social and communicative contexts rather than investigating comprehension in isolated sentences, which was the dominant method at the time (Gibbs, 2017), and, arguably, still is. With the advances in experimental methods such as eye tracking, and brain imaging techniques such as fMRI and ERP, the field of experimental pragmatics grew rapidly (Noveck & Sperber, 2012). It experienced a self-described revival in 2004 with the book titled *Experimental Pragmatics* (Noveck & Sperber, 2004), which laid down “the bases for a new field, Experimental Pragmatics, that draws on pragmatics, psycholinguistics and also on the psychology of reasoning” (Sperber & Noveck, 2004). Experimental research on pragmatic features that have traditionally been seen as semantic allows new conclusions to be drawn (e.g., Cummins & Katsos, 2010), and experimental

methods investigating pragmatic hypotheses (e.g., Noveck & Sperber, 2004) extend our understanding of language usage in more natural contexts. The question, however, of what constitutes pragmatic research, and in what ways it differs from semantic research, is a thorny one, and it has been suggested that experimental pragmatics necessarily reflects semantics as well (Meibauer & Steinbach, 2011). Research in cognitive linguistics (discussed in depth in Section 2.2.3) and experimental pragmatics (Gibbs, 2007; Giora, Balaban, Fein, & Algabets, 2004) makes the assumption that there is no principled difference between semantics and pragmatics, but rather that these concepts represent two different perspectives on the same thing. Relevance theory (discussed in depth in Section 2.2.2), deals with the cognitive aspects of pragmatics, as opposed to only basing it in communication (Wilson & Sperber, 2012). Importantly, experimental pragmatics is not a theory in the same sense as cognitive linguistics and relevance theory, but rather work on pragmatic issues is performed using experimental methods.

In experimental pragmatics, pragmatics is understood in the conventional sense as the study of language usage, and, more specifically, the study of how context and linguistic properties affect the interpretation of utterances (Sperber & Noveck, 2004). Pragmatics uses “a mixture of intuitions about interpretations and observations of behavior” as evidence, while psycholinguistics uses “sophisticated experimental methods in the study of verbal communication, but has not used them to test systematic pragmatic theories” (Sperber & Noveck, 2004). These two fields are combined in experimental pragmatics. Importantly, pragmatics not only investigates utterances and spoken communication, but also text and written communication. Experimental pragmatics should thus be described as the study of language usage and, more specifically, the use of experimental methods to study how context and linguistic properties affect the interpretation of spoken and written language.

Experimental methods are very useful in the study of pragmatics (Fein, Ben-Noked, & Giora, 2015; Gibbs, 2017; Giora, Raphaely, Fein, & Livnat, 2014; Peleg, Giora, & Fein, 2004; A. J. Stewart et al., 2000). For instance, studies investigating intentions have shown that people quickly infer the meaning of figurative language—such as the expression *Cigarettes are time bombs*, meaning that cigarettes are dangerous and can eventually kill you—and such inferences are pervasive in communication (Gibbs, 2017; Gibbs & Colston, 2012; Gibbs, Kushner, & Mills, 1991). In addition, studies of common ground in discourse have shown that speakers frequently overestimate the effectiveness of their communication (Keysar & Henly, 2002), and that readers assume communication to be cooperative and rational (Gerrig, Ohaeri, & Brennan, 2000). Common ground signifies cultural information, information shared in a common context, and information shared between two speakers.

2.4 Experimental methods

In this section, I present eye-tracking methodology, followed by a presentation of research on eye movements during reading. The section continues with a presentation of statistical analyses using mixed-effects models, which is the type of statistical analysis used in all studies of this book. This section finishes with a presentation on accurate questionnaire design.

2.4.1 Eye-tracking methodology

Eye movements are linked to visual attention and offer a window into the mind. Eye tracking is a highly applicable and highly sensitive method of research in a wide variety of fields, such as attention allocation during face recognition (Vó, Smith, Mital, & Henderson, 2012), decision making in supermarkets (Gidlöf, Anikin, Lingonblad, & Wallin, 2017), change blindness during visual search (T. J. Smith et al., 2012), and mind wandering during sentence reading (Foulsham, Farley, & Kingstone, 2013). Reflexive-like eye movements are even affected by details like disturbing sounds (Strukelj, Brännström, Holmberg, Mossberg, & Holmqvist, 2016) and other people being present in the room (Strukelj, Foulsham, & Nyström, 2016). Arguably, one of the fields that has benefitted the most by using eye tracking as an investigative method is reading (Clifton et al., 2016; Rayner, 1998, 2009).

During normal operation, the eyes make small, quick jumps called *saccades*, which last for about 30 ms; during this time, the sensitivity to visual input is significantly reduced. The saccades are interspersed with *fixations*, in which the eyes remain relatively still over a period of time, typically between 200-300 ms. During a fixation, information intake is open and visual information is processed by the brain. We use eye-tracking methodology under the assumption that it reveals something about processing, a conclusion that has been reached in many studies. In the case of reading, the cognitive processes that govern the duration of fixations and the length of saccades have received a great deal of attention (see Rayner, 1998, 2009, for overviews). Eye movements during reading are affected by, for instance, word spacing (Slattery & Rayner, 2013; Slattery, Yates, & Angele, 2016), and the perceptual span of the reader (Rayner, Slattery, & Belanger, 2010).

Eye trackers are used in order to record eye movements. The majority of current reading studies use video-based tower-mounted eye trackers, such as the EyeLink 1000 from SR Research (Ottawa, Canada) and the Hi-Speed 1250 from SensoMotoric Instruments (Teltow, Germany), both capable of recording eye movements at a sampling rate of 500 Hz and above. A combination of a camera and infrared diodes directed at the eye allows the eye to be tracked, and the position of the gaze to be mapped onto the presented stimuli. These movements are recorded, and

the results show what the participant was looking at during the entire trial (Holmqvist et al., 2011).

2.4.2 Eye movements during reading

As mentioned above, eye tracking has frequently been used to investigate the reading process (Clifton et al., 2016; Rayner, 1998, 2009). A distinction is commonly made between early and late eye-tracking measures. Early measures investigate early processing and what happens during initial reading, such as during the first encounter with a word. These measures investigate short periods of reading, and very rarely use aggregated data from the entire trial. Analyses using these measures can show immediate, even unconscious, reading behavior. Late measures contain more eye movement data and investigate longer periods of reading. Late measures frequently use aggregated data from the entire trial. Analyses using these measures can show more deliberate reading behavior. A typical early measure is *first fixation duration*, which signifies the length of the first fixation on a word when it is first visited. A typical late measure is *total reading time*, which signifies the length of time spent in an area during the entire trial. These measures are thoroughly explained below.

First fixation duration refers to the length of the first fixation on the specified text region, and is an early measure. First fixation durations increase when processing is difficult (Reichle, Rayner, & Pollatsek, 2003), and decrease for predictable words compared to unexpected ones (Rayner & Well, 1996). *First-pass reading time* (also *gaze duration*) is the length of time spent on a specified text region when it is first encountered. First-pass reading times show very similar results to first fixation durations when using words as the unit of analysis (Inhoff, 1984; Rayner, 1998). When a word is frequent, first fixation durations and first-pass reading times are shorter, compared to when a word is less frequent (Bicknell & Levy, 2012; Rayner & Duffy, 1986). Word frequency is commonly determined by finding the number of occurrences per million in a corpus. Word length is another feature that heavily determines fixation durations, since these increase with longer words (Rayner, 1998, 2009). *Regressions* are the eye movements caused by moving backward to previously-read text, and *first-pass regressions* are regressive eye movements in a region during first-pass reading. *Regression path duration* (also named *go-past time*) is the time between fixating an area for the first time and moving forward in the text. So, if previous material is reread immediately after seeing a word, all rereading is incorporated in the regression path duration (Rayner, 2009). The related measure *selective regression path duration* also starts when fixating an area for the first time and ends when moving forward in the text, but only contains the fixations and refixations in the area. This measure is preferred when the critical word has been controlled for frequency and length, but the rest of the sentence has not, and rereading therefore could differ greatly between trials. Similar to first-pass reading time, *second-pass*

reading time signifies the length of time during the second visit in an area, meaning the duration of the first time the region is reread.

As mentioned above, total reading time is the total time spent on a specified text region, meaning the sum of all fixations in the area (Mason, Tornatora, & Pluchino, 2013; Reingold & Rayner, 2006). This is a late measure that shows more deliberate stages of reading. When used on all words in a text, total reading time can give an approximation of how thoroughly a text is read, overall processing, or overall attention (Holmqvist et al., 2011; Rayner, 1998; Rayner, Smith, Malcolm, & Henderson, 2009; Strukelj, Scheiter, Nyström, & Holmqvist, 2016). Longer total reading times can show a higher amount of processing (Magliano, Graesser, Eymard, Haberlandt, & Gholson, 1993), more engagement with the text (Strukelj, Scheiter, et al., 2016), and a higher level of attention (Hyönä, 2010). Prior knowledge of the material has also been shown to facilitate processing, which is evident in reading times (Kaakinen & Hyönä, 2007; Kaakinen, Hyönä, & Keenan, 2003). It has been suggested that learning is connected to fixation times in multimedia learning (Boucheix & Lowe, 2010; Canham & Hegarty, 2010; Jarodzka, Scheiter, Gerjets, & van Gog, 2010; R. E. Mayer, 2010), while some studies find only an inconsistent link between learning and eye movements (de Koning, Tabbers, Rikers, & Paas, 2010; Schmidt-Weigand, Kohnert, & Glowalla, 2010; Strukelj, Scheiter, et al., 2016). Because of these mixed results, it has been argued that more time-locked measures are needed in addition to this global measure, to specifically isolate how learning and eye movements interact (Hyönä, 2010). Based on these results, however, total reading time appears to be a good measure of thorough reading.

When readers quickly scan the page to get an overview of the material, they are skimming the text. A study using hidden Markov models to classify reading patterns using eye movement data found that this occurred to a large extent during the beginning of reading, with behavior then transforming into more regular reading (Simola, Salojärvi, & Kojo, 2008). Another study created a “reading-skimming classifier” operating on eye movements in real time, suggesting that this type of reading is distinct from regular reading (Biedert, Hees, Dengel, & Buscher, 2012). It is assumed that so-called speed reading is similar to skimming, as the same speed-accuracy tradeoff is observed, with speed increasing at the cost of comprehension (Rayner, Schotter, Masson, Potter, & Treiman, 2016). Skimming, however, has been shown to be efficient when creating an overview of a text under time pressure (Duggan & Payne, 2009). With regard to eye movement effects, total reading times were found to be shorter compared to a more regular type of reading when only half of the text was present for participants (Duggan & Payne, 2009). Another study showed differences in total reading times on important versus unimportant sentences (Duggan & Payne, 2011). Unfortunately, all text was skimmed in the experiment, so no comparison was made with regularly-read text. The reading-skimming classifier collected reading data from participants instructed to either skim the text or read the

text regularly (Biedert et al., 2012). Specific eye movement data were not reported, but the classifier is said to use saccade lengths (saccadic amplitudes).

A large body of eye-tracking research has examined the processing of explicitly detected errors and anomalies in text, normally utilizing distinctions such as pragmatic and syntactic anomalies (Braze, Shankweiler, Ni, & Palumbo, 2002; Clifton, Staub, & Rayner, 2007; Rayner & Duffy, 1986; Reichle et al., 2003). In Braze et al. (2002), most regions of analysis contain two words, but analyses of the individual words in the region containing the incongruity and the subsequent region were also performed. The results compensated for region length (using the number of letters and spaces as regressors, and the first-pass reading time as the dependent variable in the linear regression model). For pragmatic anomalies such as *The wall will surely bite after a few years in this harsh climate*, longer length-corrected first-pass reading times were found in the critical region. When analyzing individual words, the effect was present for the word following the incongruent verb, which also revealed a significantly higher proportion of regressions. A gradual increase in regressions that reached its peak at the end of the sentence was also found for the regions following the anomaly. For syntactic anomalies such as **The wall will surely cracking after a few years in this harsh climate*, an increased proportion of regressions was found both for the two-word critical region and the individual words in the critical region. It was concluded that both types of anomaly have distinct effects on the reading process. The effects of encountering syntactic anomalies were brief, while the effects of pragmatic anomalies were prolonged and visible during a larger portion of reading. Also, these effects were more isolated to the critical region for syntactic anomalies compared to pragmatic ones.

Researchers have argued for the use of longer texts and more natural reading when investigating eye movements during reading (e.g., Jarodzka & Brand-Gruwel, 2017). This has been argued since the 1970s, as in the case of experimental pragmatics (Gibbs, 2017). As mentioned in Section 2.3.6, the field of experimental pragmatics uses experimental methods to investigate pragmatic theories. However, eye-tracking hardware has become much cheaper, more user-friendly, and more powerful. The ability to use eye tracking in research is thus better than ever, which pushes the field forward.

2.4.3 Statistical analyses using linear mixed-effects models

All statistics in this book, with very few exceptions, were performed using linear mixed-effects models (LMMs), using the R (R Development Core Team, 2015) statistical program in R Studio. LMMs provide several advantages over other statistical methods when analyzing eye-tracking data. One strength of LMMs is that models compensate for uneven data distributions, which are inevitable in eye-tracking studies (Bates, Mächler, Bolker, & Walker, 2015). Another strength is that models can use both fixed and random effects in the same model. A random effect is, in

essence, an effect that adds unwanted systematic noise into the results. This means that the model can compensate for the individual differences of participants, as well as the differences between each stimulus used in the experiment (Baayen, Davidson, & Bates, 2008). Stimuli are naturally created to be as similar as possible, but some variation is inevitable. Also, participants are often chosen within the same population, in order to be as similar as possible, but some variation is again inevitable. The effect of the manipulation is of interest, not the effect of each individual stimulus, or each individual participant. One stimulus might have a slight variation that makes reading slightly faster. Also, one participant might read all texts slightly quicker, while another might read all texts slightly slower. By modeling participant and stimuli item as random factors, the variance from these factors can be controlled, while still preserving the overall effects from the experimental manipulation. Any number of random effects can be specified, in order to remove the corresponding variance. Another strength of LMMs is that interactions are easily handled, without the need for pair-wise comparisons. How factors affect the results, as well as how factors interact with other factors to affect the results, is easily modeled.

LMMs rely on *t*-values to report significance, with a value below -1.96 and above 1.96 normally considered significant. However, the ubiquitous *p*-values can nowadays easily be approximated, for instance by Satterthwaite approximation in the *lmerTest* package.

2.4.4 Questionnaire design

When making judgments under uncertainty, several heuristics are employed in order to facilitate the process (Tversky & Kahneman, 1974). These heuristics might distort the answers given in questionnaires, resulting in inaccurate data. The heuristics also substitute a difficult question for an easier one (Denham, 2012; Kahneman, 2011). For instance, when participants are asked “How happy are you with your life these days,” the question might be unconsciously replaced with “What is my mood right now,” which is very different from the original question. Furthermore, when the choices are framed differently, they can produce very different answers (Tversky & Kahneman, 1981; Wallin, Paradis, & Katsikopoulos, 2016). Questions need to mean the same thing to every respondent, so they need to be formulated in a clear manner. The questions should also not employ vague or imprecise words like “often,” “usually,” and “frequently,” which can be interpreted very differently by respondents (Burton, 2000). Double-barreled questions combine several issues in one question, “Do you think young children should have access to cell phones and credit cards?” (Gideon, 2012, p. 96). This prevents the respondent from providing a clear answer, and makes the question vague.

Likert scales are commonly used in questionnaires, especially in social science research. Likert scales usually employ discrete scales with seven or nine available steps, and respondents are presented with a statement and asked to indicate to what degree

they agree or disagree with it (Burton, 2000; DeVellis, 2003, p. 79). Variability and sensitivity is thus fairly low, and a neutral “neither agree nor disagree” option is available in the middle of the scale. Using an ungraded continuous scale from 0 to 100 offers more sensitive results, and it also reveals very slight biases when participants respond using the neutral center of the scale. This type of scale is used in the *visual analog* scaling method (DeVellis, 2003, pp. 81-82; Mayer, 1978). Because of the many benefits of visual analog scales, they are used in all experiments that use response scales in this book (except for the mini-markers task in Chapter 5).

Chapter 3

The many ways of reading

3.1 Introduction

Research into eye movement characteristics during reading can be said to be almost as old as eye-tracking itself. Huey described the use of eye tracking to record eye movements during reading, along with extensive investigations into the underlying mechanisms of reading (Huey, 1908). In the 1920s, Tinker performed seminal studies of how people read text passages of scientific prose, elementary chemistry, algebra, and algebraic formulae (Tinker, 1927). Reading research with eye trackers took flight in the 1970s, as high quality, less invasive eye-tracking techniques emerged, along with computer systems that allowed for online manipulations of the stimuli (see Rayner, 1998, 2009, for overviews). The research on eye movements during reading, however, often studies the reading of single sentences. Studies using paragraphs or multiple pages of text are much more infrequent, leaving many open questions. This has led researchers to call for more eye-tracking research at this more natural level of reading (e.g., Jarodzka & Brand-Gruwel, 2017). This chapter investigates how eye movements during reading are affected by instructions to read the text in specific ways.

In this study, participants are presented with pages of text, along with instructions to read them using four different reading strategies: regular reading, thorough reading, skimming, and spell checking. The aim is to determine the characteristics of thorough reading, skimming, and spell checking compared to regular reading, and how the differences were manifested in eye movement measures. Some previous research has investigated the effects of instruction on reading behavior, but often only contrasting two types of reading, such as regular reading and thorough reading (Hyönä, 2010; Magliano et al., 1993) or regular reading and skimming (Biedert et al., 2012; Duggan & Payne, 2011). Using four types of reading in this study allows for a more nuanced separation between different reading behaviors, as well as the identification of similarities between these types of reading. Contrasting these types of reading provides insight into how high-level cognitive reading tasks are translated into reading behavior, and how the reading process adapts to these different task sets. Different types of reading entail different types of processing and different cognitive

demands, which should be evident in eye movements during reading (Hyönä & Niemi, 1990). By investigating the changes in eye movement parameters, we can create a blueprint of eye movement characteristics that can identify each type of reading. Being able to identify each type of reading can provide researchers with new ways of assessing how the reading process reacts to experimental manipulations. This can verify that the eye movement measures in this book accurately reflect hypothesized behavior, and can hopefully help researchers draw more accurate conclusions about reading in general.

During reading, a meaning representation is incrementally created, integrating what is currently being read with what has already been read (Kintsch, 1998; Sadoski, 1999). Using entire paragraphs of text can thus provide insight into the process of meaning integration. Research suggests that in order to more fully approximate natural reading, sentence-processing models should take factors such as semantic plausibility, expectations, and context into account (Bohan et al., 2012; Ferreira & Engelhardt, 2009; Ferreira & Patson, 2007; Sanford & Sturt, 2002). It could be argued that it is nearly impossible to control and account for all factors believed to influence reading. By using paragraphs rather than single sentences, we can investigate reading under more natural circumstances, thereby drawing more accurate conclusions of how the reading process operates in real life. Research using single sentences is crucial when investigating specific features of the reading process, but should not be used when the investigation of “natural” reading is of interest.

This study determines eye movement characteristics during the reading of one page of text on a computer screen, and how these characteristics change in response to the different instructions given to the participants. The instructions ask the participants to either read the text regularly, read the text thoroughly, skim the text, or spell check the text. Moving beyond single sentences, and instead using entire pages of text, opens up the possibility of answering questions about text reading strategies, providing a global, paragraph-level perspective to reading research. Global text processing is a subject matter that has been researched before, commonly using whole sentences as the unit of analysis (Hyönä, Lorch, & Kaakinen, 2002; Hyönä, Lorch, & Rinck, 2003). These studies have argued for the use of new measures that have been shown to identify reading styles and literacy skills (Hyönä et al., 2003; Lou, Liu, Kaakinen, & Li, 2017). Since this study does not focus on the reading of sentences or specific words, but instead on the effects of position in the text on reading behavior, lines of text are used as the unit of analysis, in addition to the words in the text. In order to establish that sufficient experimental rigor is maintained, attempts to replicate well-established effects from the sentence processing literature were performed. This shows that local effects triggered by specific and isolated words are still present when using an entire page of text as stimulus.

3.2 Previous research

3.2.1 Eye movements during reading

Early eye-tracking measures are used to investigate immediate effects on the reading process. Early measures are highly sensitive to largely unconscious effects during reading, which commonly start when a word is fixated for the first time. For instance, first fixation durations have been shown to increase when encountering an infrequent word (Rayner, 1998), and decrease when encountering a predictable word (Fernández et al., 2014). Another early measure, first-pass reading time, measures the duration of the first entry into a region, which has been shown to increase for syntactically incongruent words (Rayner, Warren, Juhasz, & Liversedge, 2004). Late measures, on the other hand, are measures that react to changes during later stages of reading, and commonly include longer periods of reading. Late measures such as total reading times increase with more rereading, and have been shown to increase with more processing (Magliano et al., 1993), more overall attention (Hyönä, 2010), and higher engagement in the text (Strukelj, Scheiter, et al., 2016).

It has been demonstrated that reading ability and comprehension can be improved via the teaching of reading strategy, which suggests that reading strategy, and awareness of reading strategy, are very important during thorough reading and learning (Gersten, Fuchs, Williams, & Baker, 2001). Indeed, it is argued that better readers exhibit better metacognitive knowledge compared to poorer readers (Baker & Beall, 2009), meaning that they are better at keeping track of their understanding of the text they are reading.

Skimming can efficiently create an overview of a text under time pressure (Duggan & Payne, 2009), and has been shown to result in shorter total reading times compared to regular reading (Simola et al., 2008). During regular reading, words are sometimes skipped and the reader does not perform a fixation on the word at all. It is plausible that skipping increases during skimming. It has been shown that proficient readers of English text skip around one-third of the words (Brysbaert et al., 2005). Function words are skipped to a larger extent than content words (Gao & Suzuki, 2005), and illegal the-previews are skipped to a much larger extent due to parafoveal processing (Abbott et al., 2015). This is related to the finding that word length influences skipping, with the likelihood of skipping a one- or two-letter word being around 76%, a four-letter word around 42%, and a nine- or ten-letter word only around 5% (Vitu, 2011). A study that compared regular reading and scanning text for specific topics, which can be considered similar to skimming, showed that early effects of frequency on first fixation durations and first-pass reading times are similar for these two reading strategies (White, Warrington, McGowan, & Paterson, 2015). Another study used a gaze-contingent eye-tracking setup that masked text that had been read, hindering participants from rereading text, simulating speedreading or

skimming (Schotter, Tran, & Rayner, 2014). The study found that comprehension decreased, concluding that comprehension is indeed supported by regressive eye movements. The study questioned the viability of speedreading, which could be seen as a process similar to skimming the text. This is also argued by Rayner et al. (2016), who suggest that the only way of increasing reading speed while maintaining high comprehension is to increase language skill. The measure *number of visits* to a word can show skipping from a global perspective.

Normally, studies use the term *proofreading* when participants search for spelling errors. This term, however, is problematic, since proofreading can also signify searching for other mistakes such as subject-verb agreement errors, logical mistakes, and possibly even incorrect facts. The term *spell checking* is therefore used here. A previous study investigated spell checking by giving people sentences without spelling errors to be read for comprehension, followed by sentences with an occasional spelling error (Kaakinen & Hyönä, 2010). In this second group of sentences, instructions specified that spelling errors might occur, and that the participant should indicate if this were the case after the sentence. Many differences in eye movements were observed. For instance, *saccade amplitudes* (the size of a saccade) were shorter, and refixation probabilities were higher during spell checking. Moreover, both the word length and word frequency effects in first-pass reading times were greater during spell checking. The study concluded that task effects significantly modulate attentional resources. The increased frequency effect for spell checking was replicated in another study, further suggesting that readers change their reading strategies depending on the task demands (Schotter, Bicknell, Howard, Levy, & Rayner, 2014). It is possible that spell checking increases fixation durations compared to regular reading because of the need for higher amounts of activation in order to verify the correct spelling of each word.

3.3 The study of eye movements during paragraph reading

This study used four different types of instructions to encourage participants to read a text in a certain manner. The four commonly used types of reading studied here were: regular reading, thorough reading, skimming, and spell checking. Regular reading was used as the baseline condition, and compared to the remaining three types. Previous research, as described above, has shown differences in these types of reading, making in depth study and comparison of them using eye tracking very interesting. The tasks of regular and thorough reading can be described as studying the content and information of the text, while the task of spell checking can be described as investigating the details of the individual words. The task of skimming can be described as quickly getting an overview of the information without dwelling on the specifics of the text. Specific reading styles have been identified when performing

cluster analyses of eye movement patterns in long expository text (Hyönä et al., 2002; Hyönä & Nurminen, 2006). This shows that reading styles can be identified in eye movement measures during paragraph reading. Repeated reading has been shown to facilitate processing (Hyönä & Niemi, 1990), possibly because it is less cognitively demanding to integrate new information into an existing meaning representation, than into a meaning representation that is less complete (Kintsch, 1998; Sadoski, 1999). Hyönä and Niemi (1990) showed that this was observed in shorter total reading times. It is possible that these effects could be seen in different eye movement characteristics in the beginning and ends of texts. Accordingly, Hyönä and Niemi (1990) showed that average fixation durations got shorter as text was read, ending with the shortest average fixation durations on sentences in the conclusions.

3.3.1 Terms and classifications

Four different types of reading were used in this study: regular reading, thorough reading, skimming, and spell checking. A short description of how the terms were understood in this study follows below.

The term regular reading signifies what has been referred to as *reading for comprehension* in previous research (Rayner, 2009; Schotter, Bicknell, et al., 2014; Vasishth, Suckow, Lewis, & Kern, 2010), in which participants are instructed to read the text as they would normally (Foraker & Murphy, 2012; Kim & Osterhout, 2005). This type of reading was used as the baseline condition, to which the remaining three types of reading were compared.

The term thorough reading signifies what has been referred to as *reading to learn* (Sanders & Gernsbacher, 2004) in previous research, in which participants are instructed to read the text thoroughly in order to learn the material it contains (Strukelj, Scheiter, et al., 2016), or are told to learn the material, since there is to be a test afterward (Eitel & Kühl, 2016). When participants are thoroughly reading a text, it can be assumed that the comprehension and, most importantly, retention of the material will be higher than in regular reading, and that the total time spent on the text will be longer than in regular reading.

The term skimming signifies what has been referred to as *reading for gist* (Rayner, 2009) in previous research; that is, trying to read a text as quickly as possible while still understanding it (Duggan & Payne, 2009; Rayner et al., 2016; Wilkinson, Reader, & Payne, 2012). When participants are skimming a text, it can be assumed that the comprehension and most importantly retention of the material will be lower than in regular reading, and that the total time spent on the text will be shorter than during regular reading.

The term spell checking signifies the way participants read when they try to find words that are misspelled, something that has been referred to as *proofreading* in previous research (Kaakinen & Hyönä, 2010; Schotter, Bicknell, et al., 2014), in which participants were instructed to pay attention “to the letters and the spelling of

the words without paying attention to the meaning” (Brybaert et al., 2005) and “are expected to attend strongly to the orthography of words” (Wotschack, 2009). When participants are spell checking a text, it can be assumed that the comprehension and retention of the material will be lower than in regular reading, and that the total time spent on the text will be longer than during regular reading.

3.3.2 Hypotheses

Several hypotheses were formed in terms of the differences between regular reading and the other three reading types. A summary of the hypotheses and the hypothesized effects is given in Table 3.1.

Better comprehension, as demonstrated by higher comprehension scores, was expected during thorough reading, since participants engaged more profoundly with the material. A larger distance between two subsequent fixations, shown in an increased saccade amplitude, indicates a higher frequency of skips during reading. This was not expected during thorough reading, but more rereading of previously read text up and down the page was hypothesized to be shown in a higher proportion of vertical saccades. A higher proportion of leftward, (regressive) saccades, indicating a higher number of regressions, were also expected, since this might indicate thorough reading of individual lines of text. Processing as defined by average fixation durations indicates higher processing during *every* fixation. This was not expected for thorough reading, but more time was expected to be spent on each word during the entire trial, demonstrated by increases in total reading times per character and the total number of visits. The pattern for reading one page of text was expected to be similar to regular reading, demonstrated in longer total reading times per character and a higher total number of visits to words early in the text compared to the words toward the end. Participants were expected to spend more time in total on the beginning of the text compared to the end of the text, as was also the case in regular reading, but all durations were expected to be longer than those for regular reading.

Decreased comprehension was expected during skimming, as demonstrated by lower comprehension scores, since participants only formed an overview of the material. It was expected that processing would be lower for each word during every fixation, overall reading would be faster and more spread out over the page, and more word skips would be found, demonstrated in shorter average fixation durations, and larger saccade amplitudes. A lower proportion of vertical saccades was expected, indicating less rereading of previously-read text up and down the page, and a lower proportion of leftward (regressive) saccades, indicating a lower number of regressions. Less time was expected to be spent on each word during the entire trial, as demonstrated by decreases in total reading times per character and total number of visits. The pattern for reading one page of text was expected to be uniform regardless of which word was fixated, as shown by similar total reading times per character and a similar total number of visits regardless of the location of the words in the text. In

other words, eye movement characteristics were expected to be manifested in a uniform manner during reading, irrespective of the word's location in the text. Participants were expected to form only an overview of the material, and not use more processing than needed.

Table 3.1. Hypotheses and hypothesized effects on the experimental measures compared to regular reading
The hypotheses and effects supported by the experimental data are reported in **bold** in Table 3.6 in Section 3.6.

Reading type	Hypothesis compared to regular reading	Hypothesized effects compared to regular reading
Thorough reading	Better comprehension of material	Higher comprehension scores
	No difference in processing of words for all fixations	Average fixation durations similar to regular reading
	No difference in distances between two subsequent fixations during reading	Saccade amplitudes similar to regular reading
	More rereading of previously read text	Higher proportion of vertical saccades
	More regressions	Higher proportion of leftward saccades
	More deliberate reading during entire trial	Longer total reading times, higher number of visits to words
	Reading behavior is very similar to regular reading when investigating fixation durations on words in entire trial	Reading pattern is similar to regular reading during entire trial
Skimming	Reduced comprehension of material	Lower comprehension scores
	Lower amounts of processing on words for all fixations	Shorter average fixation durations
	Greater distances between two subsequent fixations during reading	Larger saccade amplitudes
	Less rereading of previously read text	Lower proportion of vertical saccades
	Fewer regressions	Lower proportion of leftward saccades
	Less deliberate reading during entire trial	Shorter total reading times, lower number of visits to words
	Uniform reading pattern across the entire text	Total reading times and number of visits are similar over the entire text
Spell checking	Reduced comprehension of material	Lower comprehension scores
	Higher amounts of processing of words for all fixations	Longer average fixation durations
	Smaller distances between two subsequent fixations during reading	Smaller saccade amplitudes
	More pronounced frequency effect	Even longer first fixation durations and first-pass reading times for less frequent compared to more frequent words
	More pronounced word length effect	Even longer total reading times for longer compared to shorter words
	No differences in rereading of previously read text	Proportion of vertical saccades similar to regular reading
	More regressions	Higher proportion of leftward saccades
	More deliberate reading during entire trial	Longer total reading times, higher number of visits to words
	Uniform reading pattern across the entire text	Total reading times and number of visits are similar over the entire text

Spell checking is more similar to a search task than reading. It was expected to produce lower comprehension scores, since participants were not reading the text in order to understand the content, but were instead searching for errors. Average

fixation durations were expected to be longer, and saccade amplitudes shorter, since participants were expected to perform eye movements that were slower and less spread out over the page, and skip fewer words during reading. The frequency effect on first fixation durations and first-pass reading times was expected to be greater, since participants would be less certain of the spelling of infrequent words and thus need more processing when initially encountering them. The effect of word length on total reading times was expected to be greater, since long words contain more letters and therefore there are more possibilities for spelling errors. Total reading times per character and total number of visits was expected to increase. The pattern of reading one page of text was expected to be uniform regardless of which word was fixated, as shown by similar total reading times per character and similar total number of visits, regardless of the location of the words in the text. Participants were expected to search for errors rather than read regularly.

3.4 Method

3.4.1 Participants

Sixty-four native speakers of Swedish (34 female; 18 to 29 years of age) participated in the experiment. The participants were recruited through student lists and advertisements at the Lund University campus. All participants had normal or corrected-to-normal vision. Four participants were excluded from the results due to technical problems with the eye-tracking equipment or poor recordings. This resulted in 60 participants in total (32 female; $M=22.9$ years, $SD=3.69$). The participants were naïve to the purpose of the experiment.

3.4.2 Apparatus

Eye movement data were recorded binocularly at a sampling rate of 120 Hz using RED-m remote video-based eye-trackers (SensoMotoric Instruments, Teltow, Germany). The experiment took place in the Digital Classroom at the Humanities Laboratory, Lund University, Sweden. The distance from a participant's eyes to the stimulus monitor was approximately 60 cm. Stimuli were displayed on a Dell P2210 22" widescreen LCD display with a refresh rate of 60 Hz at a resolution of $1,680 \times 1,050$ pixels (47.5×30.0 cm, 43.2×28.1 degrees of visual angle at a distance of 60 cm). The eye-tracking system was controlled by SMI iView RED-m (3.2.177), while stimuli presentation, a 5-point calibration, and a 4-point validation of the calibration accuracy were controlled by SMI Experiment Center (3.5.281).

3.4.3 Design and stimuli

The stimuli consisted of eight texts taken from the International Reading Speed Texts (IReST) (Trauzettel-Klosinski & Dietz, 2012). One text was also used as an example stimulus. The IReST texts are balanced for text length, word length, and linguistic complexity. The IReST texts are available in several languages, and the Swedish version was used. All Swedish texts contained 146 words and 16 lines, for 684 characters in total, with an identical readability index (LIX 35). The readability index corresponds approximately to the readability of typical works of fiction. The mean word length is 4.61 ($SD=0.01$) characters, and is balanced across the texts. Each text was adapted to on-screen reading by using a monospaced ClearType font, Consolas 26-point (approximately 4.5×6.5 mm, 0.35×0.50 degrees of visual angle at a distance of 60 cm). Each individual line of text was separated by one blank line in order to facilitate accurate eye-tracking measures. An example stimulus screen can be found in Figure 3.1. The original English version of the text can be found in Appendix 3.A. The translations were verified to correspond accurately to the original texts. None of the texts contained any spelling errors, even though one of the instructions was to inspect a text for spelling errors. Adding spelling errors into the text was considered, but in order to compare identical texts, this was not carried out.

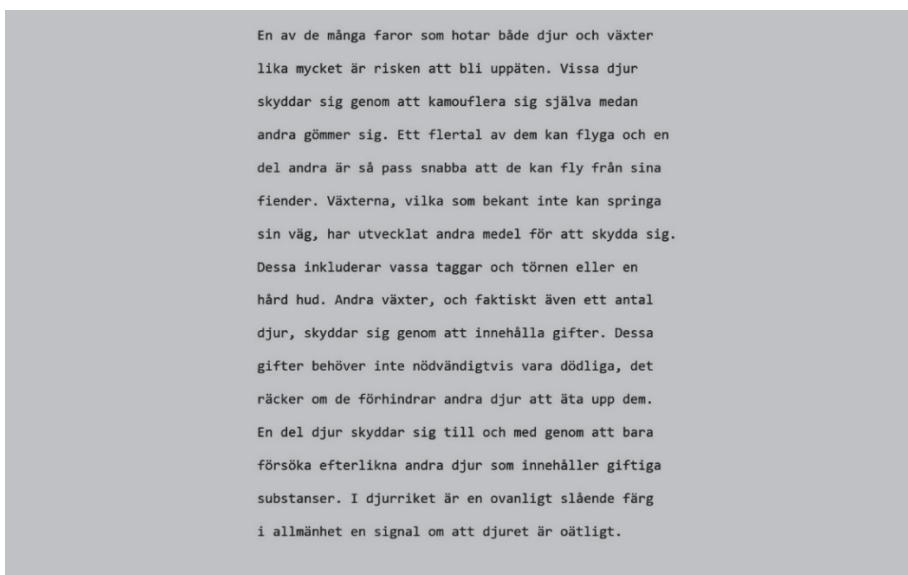


Figure 3.1. Example stimulus screen

The example stimulus is shown using the identical proportions of the experimental setup. The original English version can be found in Appendix 3.A.

Each text was followed by five multiple-choice questions. Each question had five possible answers, only one of which was correct. An example of a question for the text

in Figure 3.1 can be found in (14) below, translated into English. The questions were asked in Swedish during the experiment.

- (14) What is described as a danger for animals and plants alike?
- A. To be killed.
 - B. To be poisoned.
 - C. To be eaten.
 - D. To be injured.
 - E. None of the above.

The texts were distributed over four lists using a Latin square design, with each list containing two texts for each one of the four conditions: thorough reading, skimming, spell checking, and regular reading (the baseline condition). The distribution was counterbalanced so that the instructions were paired with different texts for each participant, and the order of appearance of the texts was randomized. For regular reading, the written instructions were to read the text in order to understand it. For thorough reading, the written instructions were to read the text very thoroughly. For skimming, the written instructions were to “read the text quickly, that is, skim the text,” but to try to still understand it. For spell checking, the written instructions were to look for spelling errors in the text. The texts in the spell checking condition were accompanied by a question regarding the number of spelling errors found.

3.4.4 Procedure

The experimental session lasted around 30 minutes, but no time restrictions were imposed during the experiment. The participants were welcomed to the study and told to sit at specific places in the Digital Classroom so that they could not see the screens of any other participants. Participant groups ranged from one to five, with three being the most common. The participants were first informed that their participation was voluntary, and that they could stop the experiment at any time. They were then told that the study was a reading study in which they would read texts and answer five multiple-choice comprehension questions afterwards, and that they would be given one of four different instructions prior to each text. They were specifically told that it was very important that these instructions were followed during each trial, and were given thorough descriptions of each instruction.

They were then seated in front of the eye tracker and calibration was begun. The calibration was repeated until the offset was less than 1.0 degrees of visual angle in

both the horizontal and the vertical directions. The experiment began with the participants completing a questionnaire about their age and gender. They were then given a sample trial, using the baseline condition instructions of regular reading. The text was also taken from the IReST set of texts, and was accompanied by five questions. After the example trial, they were instructed to call for the experimenter if anything remained unclear, otherwise the experiment would commence. The eye-tracking portion then started, with instructions being shown before each text, and five multiple-choice comprehension questions following each text.

After the eye-tracking portion of the experiment was completed, they were taken outside the Digital Classroom and told about the specifics of the study, signed a consent form, were thanked again, and were given a movie ticket as compensation for participating in the study.

3.4.5 Data analysis

Velocity-based high-speed event detection was performed using SMI BeGaze 3.5 with default settings (peak velocity threshold=40°/s, minimum fixation duration=50 ms), which transformed the raw data into fixations and saccades. The recordings from the right eye were used since the right eye is dominant for the majority of people. Fixations above 1,000 ms were removed.

The data processing, figures, and statistical analyses were made using R (3.2.2) (R Development Core Team, 2015) and the lme4 package (Bates et al., 2015) in R Studio (0.99.473). Models used random intercepts and slopes of condition within group for participant and item, with intercept-slope correlation factors for participants and items. In the cases in which models did not converge, the intercept-slope correlation factors were dropped. This was done on the models for first fixation durations and first-pass reading times as a function of word frequency, and first fixation durations and first-pass reading times as a function of word length. All statistical models used log-transformed data to better fit a Gaussian distribution except where indicated. In order to improve readability, no log-transformation was performed on the data in the figures, except for saccade amplitudes. The lmerTest package (2.0-29) was used to estimate *p*-values with Satterthwaite approximation.

In all saccade analyses, all *return sweeps* (the saccade from the end of one line of text to the beginning of the next line) that were part of initial reading were removed from the results by excluding all saccades directed downward and right-to-left that were over 500 pixels in horizontal width (the width of one line of text was around 700-800 pixels) and under 100 pixels in vertical width (the distance between the middle of two lines of text was around 60 pixels). This resulted in the removal of 5,580 saccades from the data (from a total of 93,865). Saccades of larger vertical widths and/or smaller horizontal widths were considered to be part of patterns of rereading as opposed to the regularity of initial reading patterns. The thresholds were chosen in order to under-remove rather than over-remove data. Without the large return

sweeps, the analyses of saccade amplitudes should be more accurate, and the analyses of saccade directions should better reflect the special characteristics of each type of reading without the return sweeps that are common to all types and add unnecessary noise to the data.

The Swedish corpus Korp (<https://spraakbanken.gu.se/korp/>), consisting of around 9 million words, was used when investigating frequency effects.

3.5 Results

Full statistical model outputs that are not given in the corresponding results section below can be found in Appendix 3.B. All models except comprehension scores and total number of visits on words used log-transformed data to better fit a Gaussian distribution.

3.5.1 Comprehension scores

To obtain data for compression, each text was accompanied by five multiple-choice comprehension questions, for a maximum of five points per text. Figure 3.2 shows the comprehension scores. Compared to regular reading, a significant increase was found for the comprehension scores in the condition with thorough reading (Estimate=0.349, $SE=0.135$, $t=2.59$, $p<.02$), and a significant decrease was found in the remaining two conditions (Skimming: Estimate=-0.687, $SE=0.165$, $t=-4.16$, $p<.001$; Spell checking: Estimate=-0.719, $SE=0.152$, $t=-4.72$, $p<.001$).

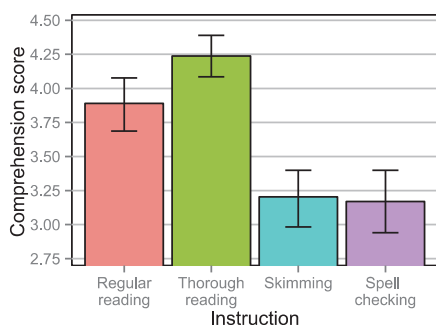


Figure 3.2. Mean comprehension scores depending on instruction
Error bars denote standard errors of the mean.

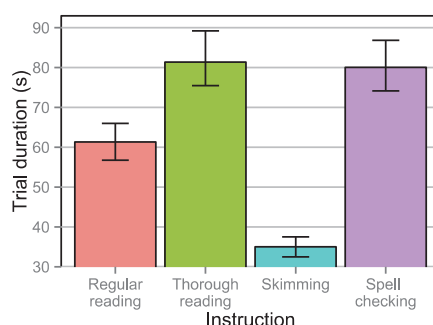


Figure 3.3. Trial duration depending on instruction
Error bars denote standard errors of the mean.

3.5.2 Trial duration

Figure 3.3 shows trial duration, i.e., the total time a text was shown on screen before the participant moved on to the comprehension questions. For trial durations, a significant decrease was found for skimming (Estimate=-0.560, $SE=0.052$, $t=-10.80$, $p<.0001$) and significant increases were found for thorough reading (Estimate=0.275, $SE=0.030$, $t=9.22$, $p<.0001$) and spell checking (Estimate=0.253, $SE=0.067$, $t=3.77$, $p<.001$), compared to regular reading.

3.5.3 Fixations and saccades

Figure 3.4 shows average fixation durations and Figure 3.5 shows log-transformed saccade amplitudes (due to a large variance in saccade amplitudes throughout the trial, a figure with log-transformed means is much more informative.) For average fixation durations, a significant decrease was found for skimming (Estimate=-0.031, $SE=0.011$, $t=-2.74$, $p<.01$) and a significant increase was found for spell checking (Estimate=0.084, $SE=0.015$, $t=5.60$, $p<.0001$), compared to regular reading. No differences were found for thorough reading (Estimate=-0.005, $SE=0.010$, $t=-0.51$, $p=.62$). For saccade amplitudes, a significant increase was found for thorough reading (Estimate=0.059, $SE=0.015$, $t=3.92$, $p<.0001$) and skimming (Estimate=0.090, $SE=0.018$, $t=4.92$, $p<.0001$) and a significant decrease was found for spell checking (Estimate=-0.167, $SE=0.015$, $t=-10.90$, $p<.0001$), compared to regular reading.

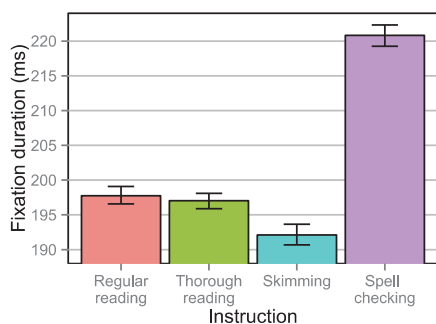


Figure 3.4. Average fixation durations depending on instruction
Error bars denote standard errors of the mean.

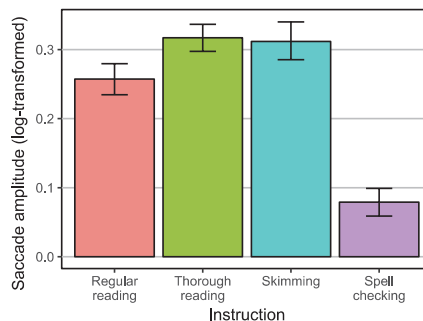


Figure 3.5. Mean saccade amplitudes (log-transformed) depending on instruction (return sweeps are removed)
Error bars denote standard errors of the mean.

Table 3.2a shows proportions of vertical saccades. For proportions of vertical saccades, significant increases were found for thorough reading (Estimate=0.063, $SE=0.026$, $t=2.44$, $p<.01$) and spell checking (Estimate=0.149, $SE=0.026$, $t=5.77$, $p<.0001$), and a significant decrease was found for skimming (Estimate=-0.176, $SE=0.033$, $t=-5.38$, $p<.0001$), compared to regular reading. Table 3.2b shows proportions of leftward saccades. For proportions of leftward, i.e., regressive, saccades,

significant increases were found for spell checking (Estimate=0.063, $SE=0.026$, $t=2.44$, $p<.01$), and significant decreases were found for skimming (Estimate=-0.123, $SE=0.026$, $t=-4.68$, $p<.0001$), compared to regular reading. No significant results were found for thorough reading (Estimate=0.016, $SE=0.020$, $t=0.82$, $p=.41$).

Table 3.2a. Mean proportions of vertical saccades and percentage point (PP) differences from baseline condition (return sweeps are removed)

Statistically significant results in **bold**. * indicates that the binomial generalized linear mixed model reported $p<.05$ for the comparison with regular reading, *** indicates $p<.001$.

Instruction	Vertical saccades	
	%	PP +/-
Regular reading	14.92	-
Thorough reading	15.53*	0.61*
Skimming	12.79***	-2.13***
Spell checking	16.94***	2.02***

Table 3.2b. Mean proportions of leftward saccades and percentage point (PP) differences from baseline condition (return sweeps are removed)

Statistically significant results in **bold**. * indicates that the binomial generalized linear mixed model reported $p<.05$ for the comparison with regular reading, *** indicates $p<.001$.

Instruction	Leftward saccades	
	%	PP +/-
Regular reading	23.37	-
Thorough reading	24.17	0.80
Skimming	21.25***	-2.02***
Spell checking	24.45*	1.08*

Figure 3.6 shows first fixation durations, first-pass reading times, and total reading times on words as a function of word frequency. The frequency effect for first fixation durations, well-known from previous sentence reading research (Bicknell & Levy, 2012; Rayner & Duffy, 1986), was replicated, with first fixation durations increasing as a function of decreasing frequency (Estimate=0.014, $SE=0.003$, $t=4.67$, $p<.0001$). Overall, first fixation durations were also significantly longer for spell checking (Estimate=0.075, $SE=0.015$, $t=4.84$, $p<.0001$), compared to regular reading. No other significant results were found (all $ps>.15$). This effect was also found for first-pass reading times, which increased as a function of decreasing frequency (Estimate=0.040, $SE=0.003$, $t=11.7$, $p<.0001$). Overall, first-pass reading times were also significantly longer for spell checking (Estimate=0.075, $SE=0.015$, $t=4.84$, $p<.0001$), compared to regular reading. An interaction was also found between spell checking and frequency (Estimate=0.024, $SE=0.005$, $t=5.02$, $p<.0001$), showing that the frequency effect was larger for spell checking than for regular reading. This effect was also found for total reading times, which increased as a function of decreasing frequency (Estimate=0.101, $SE=0.005$, $t=21.7$, $p<.0001$). All conditions and all interactions were also significantly different compared to regular reading (for condition and interaction data, see the full statistical model outputs in Appendix 3.B).

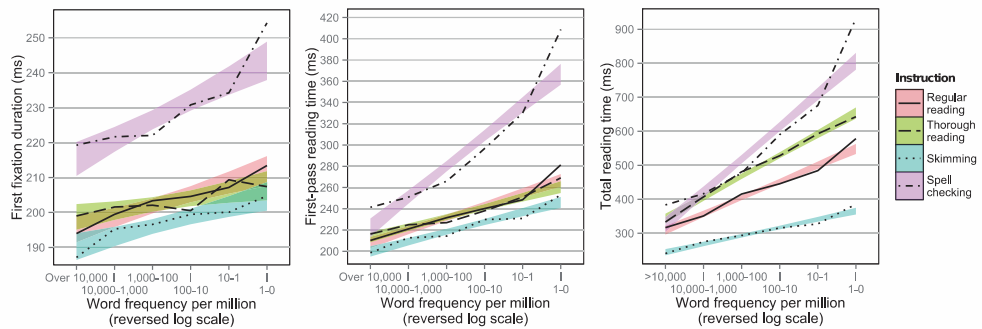


Figure 3.6. First fixation durations, first-pass reading times, and total reading times on words depending on instruction as a function of word frequency per million, using a log-scale running from high to low, separated by reading type
Means are shown by lines, only 95% confidence intervals are shown for regression lines.

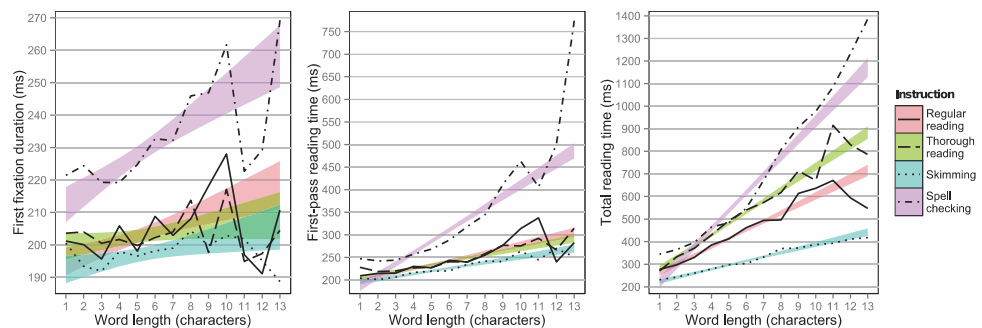


Figure 3.7. First fixation durations, first-pass reading times, and total reading times on words depending on instruction as a function of word length, separated by reading type
Means are shown by lines, only 95% confidence intervals are shown for regression lines.

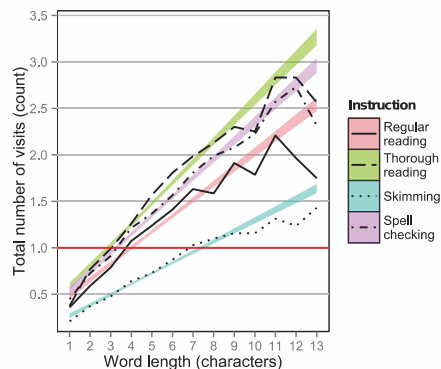


Figure 3.8. Total number of visits on words depending on instruction as a function of word length, separated by reading type
Means are shown by lines, only 95% confidence intervals are shown for regression lines.

Figure 3.7 shows first fixation durations, first-pass reading times, and total reading times on words as a function of word length. Figure 3.8 shows total number of visits as a function of word length. The word length effect, well-established by previous sentence reading research (Rayner, 1998, 2009), was replicated, with significant increases as a function of word length for first fixation durations (Estimate=0.009, $SE=0.002$, $t=4.29$, $p<.0001$), first-pass reading times (Estimate=0.030, $SE=0.002$, $t=13.6$, $p<.0001$), total reading times (Estimate=0.078, $SE=0.003$, $t=26.4$, $p<.0001$), and number of visits (Estimate=0.171, $SE=0.004$, $t=42.6$, $p<.0001$). Compared to regular reading, spell checking showed longer first fixation durations (Estimate=0.061, $SE=0.016$, $t=3.92$, $p<.0001$), and skimming showed shorter total reading times (Estimate=-0.144, $SE=0.040$, $t=-3.63$, $p<.0001$) and fewer number of visits (Estimate=-0.166, $SE=0.070$, $t=-2.36$, $p<.03$). Significant interactions with word length were also found, compared to regular reading. First-pass reading times increased for spell checking (Estimate=0.023, $SE=0.003$, $t=7.44$, $p<.0001$), and both total reading times and number of visits increased for thorough reading and spell checking, but decreased for skimming (for interaction data, see the full statistical model outputs in Appendix 3.B).

In the subsequent analyses in this study, interesting results were found when using the position of the word or line in the text. However, the results found for word frequency or word length should be unaffected by this, as the IReST stimuli texts are balanced for word length, lexical complexity, number of words, and number of lines.

3.5.4 Total reading times per character for lines of text and for individual words

Figure 3.9 shows total reading times per character for lines of text and for individual words (excluding skipped words) depending on reading type. For total reading times per character for lines of text, significant increases were found for thorough reading (Estimate=0.226, $SE=0.062$, $t=3.66$, $p<.01$) and spell checking (Estimate=0.301, $SE=0.076$, $t=3.97$, $p<.001$), and significant decreases were found for skimming (Estimate=-0.496, $SE=0.072$, $t=-6.87$, $p<.0001$), compared to regular reading. When using individual words, all skipped words were removed from the analyses and figures. As when using lines of text, significant increases were found for thorough reading (Estimate=0.145, $SE=0.029$, $t=4.94$, $p<.0001$) and spell checking (Estimate=0.188, $SE=0.046$, $t=4.09$, $p<.0005$), and significant decreases were found for skimming (Estimate=-0.288, $SE=0.030$, $t=-9.58$, $p<.0001$), compared to regular reading.

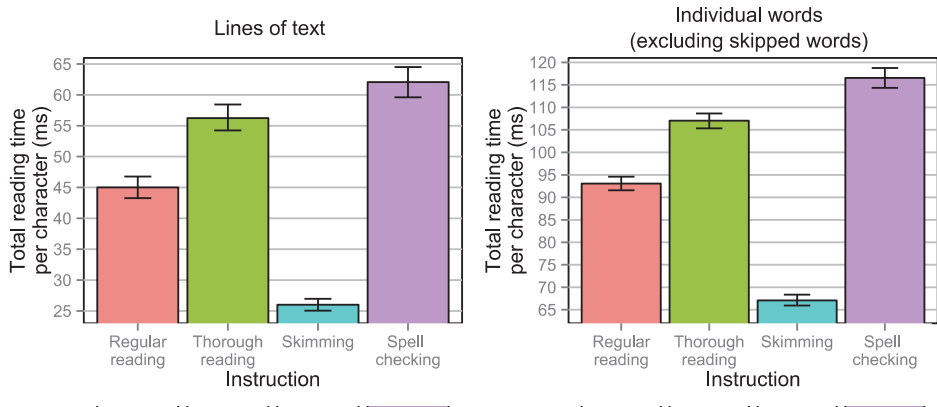


Figure 3.9. Total reading times per character for lines of text and for individual words (excluding skipped words) depending on instruction
Error bars denote standard errors of the mean.

Figure 3.10 shows total reading times per character on individual lines, on each individual line, along with a regression line for all lines except the first and last. This was done because the first and last lines were found to differ greatly from the rest of the data. All subsequent analyses and figures were made without the words in the first and last line of text. When using the relative position of the individual lines, total reading times per character were longer as a function of line number for spell checking only, compared to regular reading. The full statistical model output can be found in Table 3.3.

Figure 3.11 shows regression lines for total reading times per character for words as a function of the relative position of each word in the text, excluding skipped words and the words in the first and last lines. When using the relative position of the words in the text, total reading times per character decreased less as a function of word number for skimming and spell checking, compared to regular reading. The full statistical model output can be found in Table 3.4.

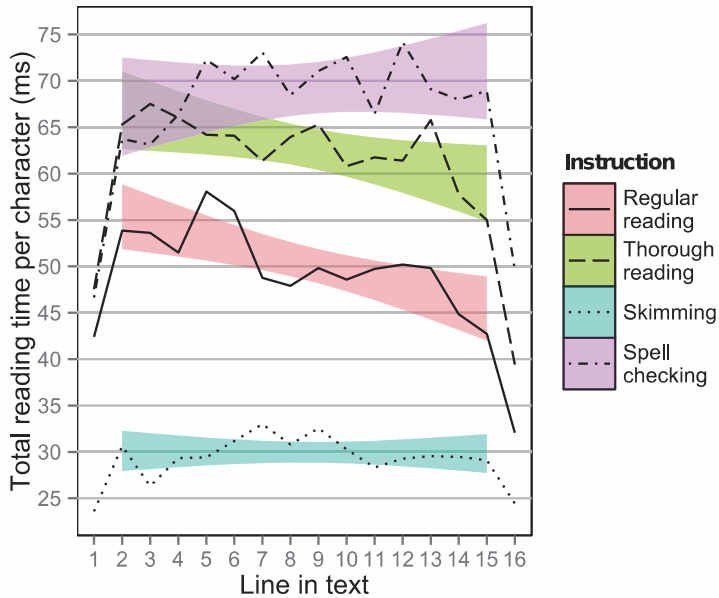


Figure 3.10. Total reading times per character for lines of text depending on instruction as a function of line order in the text

Means are shown by lines, 95% confidence intervals are shown for regression lines when excluding first and last lines.

Table 3.3. Full statistical model output: Total reading times per character for lines of text as a function of line order in the text, excluding the first and last lines

Statistically significant results in **bold**. A colon denotes an interaction.

	Estimate	SE	df	t-value	p-value	
Thorough reading	0.2334	0.0729	100	3.202	.00183	**
Skimming	-0.6218	0.0868	49	-7.162	.00000	***
Spell checking	0.1789	0.0879	75	2.034	.04550	*
Line number	0.0037	0.0042	6415	0.865	.38717	
Thorough reading: Line number	0.0003	0.0060	6422	0.047	.96261	
Skimming: Line number	0.0117	0.0061	6431	1.920	.05489	.
Spell checking: Line number	0.0148	0.0060	6413	2.472	.01347	*

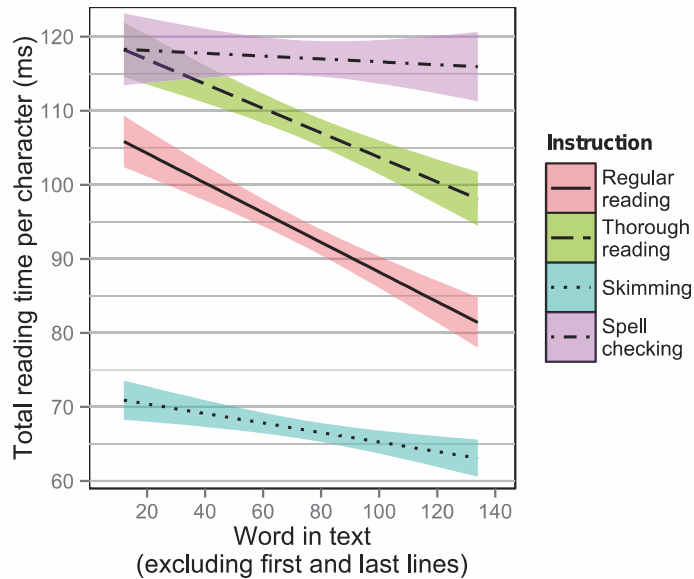


Figure 3.11. Total reading times per character for words as a function of word order in the text, excluding skipped words, without first and last lines, depending on instruction
Regression lines include 95% confidence interval.

Table 3.4. Full statistical model output: Total reading times per character for words as a function of word order in the text, when excluding skipped words, without first and last lines
Statistically significant results in **bold**. A colon denotes an interaction.

	Estimate	SE	df	t-value	p-value	
Thorough reading	0.1181	0.0390	35	3.026	.00461	**
Skimming	-0.3571	0.0394	74	-9.050	.00000	***
Spell checking	0.0669	0.0550	33	1.215	.23317	
Word number	-0.0011	0.0002	36490	-5.249	.00000	***
Thorough reading: Word number	0.0004	0.0003	36440	1.511	.13082	
Skimming: Word number	0.0008	0.0003	36120	2.532	.01134	*
Spell checking: Word number	0.0016	0.0003	36720	5.583	.00000	***

3.5.5 Total number of visits

Figure 3.12 shows regression lines for total number of visits on words as a function of the relative position of each word in the text, excluding the words in the first and last lines. For the total number of visits on individual words, they were found to be higher as a function of word number for skimming and spell checking, compared to regular reading. These results indicate that the decrease found for regular reading is not present for skimming and spell checking. The full statistical model output can be found in Table 3.5.

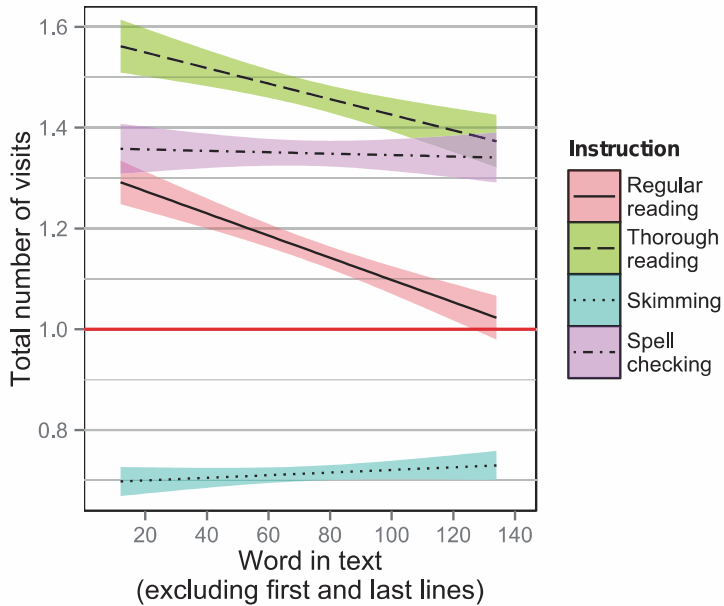


Figure 3.12. Total number of visits on words as a function of word order in the text, without first and last lines, depending on instruction
Regression lines include 95% confidence interval.

Table 3.5. Full statistical model output: Total number of visits on words as a function of word order in the text, excluding first and last lines
Statistically significant results in **bold**. A colon denotes an interaction.

	Estimate	SE	df	t-value	p-value	
Thorough reading	0.2536	0.0755	26	3.358	.00247	**
Skimming	-0.6153	0.0730	34	-8.431	.00000	***
Spell checking	0.0661	0.0986	37	0.671	.50675	
Word number	-0.0021	0.0003	59640	-7.460	.00000	***
Thorough reading: Word number	0.0007	0.0004	59640	1.780	.07502	.
Skimming: Word number	0.0024	0.0004	59640	5.812	.00000	***
Spell checking: Word number	0.0018	0.0004	59640	4.362	.00001	***

3.6 Discussion

This eye-tracking study investigated the eye movement characteristics of different types of reading using an entire page of text instead of single sentences. Both entire lines of text and isolated words were used in the analyses. Participants were reminded of the importance of using the indicated reading strategy, both during the verbal instructions before the experiment and the written instructions before each trial. To

verify that participants were actually performing the instructed reading, the results from the comprehension questions were used. Significant increases in comprehension scores were found for thorough reading, and significant decreases for skimming and spell checking, compared to regular reading. Also, trial durations (the total time a text was present on screen before the participant continued to the comprehension questions) were significantly shorter for skimming, and significantly longer for thorough reading and spell checking, as would be expected. This suggests that participants followed the instructions, and that the eye-tracking data would thus reflect the reading characteristics for the different reading types.

The word frequency effect found in previous eye-tracking research was replicated (Bicknell & Levy, 2012; Rayner & Duffy, 1986), with significant increases found in first fixation durations and first-pass reading times as a function of decreasing frequency. The effect for word length found in previous eye-tracking research was also replicated (Rayner, 1998, 2009), with significant increases in first fixation durations, first-pass reading times, total reading times, and total number of visits as a function of word length. The replication of frequency and word length effects using a page of text as opposed to the single sentences used in previous studies demonstrated that local effects were still evident in the data. Using an entire page of text as stimuli offers the additional advantage of allowing for the analysis of reading from a global perspective, answering questions about the reading process when it operates in a more natural way.

In this study, rather than using skipping rate or total number of refixations, the total number of visits was used as a global measure for examining rereading operating on an entire page of text. Word skipping has been shown to increase for short words (Vitu, 2011), function words (Gao & Suzuki, 2005), and illegal the-previews (Abbott et al., 2015), and it has been argued that proficient readers of English normally skip around one-third of the words (Brysbaert et al., 2005). In this study, examining the number of visits to a word throughout the text revealed that skipping also depends on the word's location in the text. Thus, the total number of visits to a word was used to show general reading strategies rather than specific skipping effects.

Each section below begins with replications of previous research on different reading types, followed by the additional conclusions that can be drawn from the data in this study.

3.6.1 Thorough reading

Trial durations were significantly longer compared to regular reading. This means that people spent more time reading the texts before moving to the next one.

3.6.1.1 Replication of previous research

Thorough reading had significantly longer total reading times per character compared to regular reading. Since comprehension also increased, as shown in higher

comprehension scores, this suggests a link between learning and eye movements, replicating previous research in multimedia learning (Boucheix & Lowe, 2010; Canham & Hegarty, 2010; Jarodzka et al., 2010; R. E. Mayer, 2010). Longer total reading time can show more processing (Magliano et al., 1993), more engagement (Strukelj, Scheiter, et al., 2016), and more attention (Hyönä, 2010), all of which can be said to reflect thorough reading.

3.6.1.2 Discussion of new findings

The profiles for thorough reading and regular reading were very similar. No differences in saccade amplitudes were expected, but these turned out to be significantly larger than in regular reading, with average fixation durations that were not significantly different. The increases in saccade amplitudes might be explained by the significantly longer trials, in combination with the higher proportion of vertical saccades, which are long regressive saccades to much earlier parts of the text. As these were fairly frequent, the average saccade amplitudes were affected. The finding that the proportion of leftward (regressive) saccades was not significantly larger shows that rereading was not primarily of the current line of text but rather of earlier material. This suggests that overall reading strategies were fairly similar to those used in regular reading, but with more rereading of previously-read text rather than more frequent rereading on the current line. Leftward, regressive saccades are performed when rereading the information currently being processed. Regressive saccades are important, since completely blocking the ability to reread previously-read text has been shown to negatively affect comprehension (Schotter, Tran, et al., 2014). Regressive saccades, however, seem to play a minor role in increasing comprehension, with vertical saccades being much more important in this case and possibly also that of learning. Vertical saccades can be viewed as a type of global rereading focused on “filling in the blanks” of knowledge. This potentially reflects a reading strategy that leads to better comprehension (Baker & Beall, 2009; Gersten et al., 2001). Furthermore, total reading times were longer, but not significantly different over the page. The downward slope found in total reading times from the beginning to the end of the text was similar to the one found for regular reading. This slope, combined with the higher proportion of vertical saccades, shows the rereading behavior of previously-read material. As a result, reading of the initial parts of the text was more thorough when considering reading during the entire trial. It is likely in these cases that the participant needed to recheck a detail that was forgotten, or perhaps establish a context for the rest of the text. It is also possible that the text was reread to a larger extent simply because the text was available for rereading for a longer time. It is possible that the pattern reflected the high demands of integrating new information and building a meaning representation of the text during initial stages of reading, when this representation is very sparse (Kintsch, 1998; Sadoski, 1999). These demands decrease as the memory representation becomes more complete during later stages of reading. This is similar to how total reading times become shorter after

repeated reading and how average fixation durations become shorter as more of a text is read (Hyönä & Niemi, 1990).

Total reading times per character were the longest for spell checking, both overall and as a function of word length, but very long for thorough reading as well. The total number of visits to words as a function of word length, however, was highest for thorough reading. In combination, these results suggest that words were reread and revisited to a larger extent during thorough reading, and that the results for spell checking signify increased processing and less rereading. This was also found in the reading pattern for total number of visits, which exhibits the same downward slope as for total reading times for thorough reading, and an even pattern across the entire page for spell checking.

3.6.2 Skimming

Trial durations were significantly shorter compared to regular reading. This means that people spent less time reading the texts before moving to the next one.

3.6.2.1 *Replication of previous research*

Skimming resulted in significantly larger saccade amplitudes compared to regular reading. This replicates the theory described for the real-time reading-skimming classifier in Biedert et al. (2012). Total reading times per character were significantly shorter compared to regular reading. This replicates the results by Simola et al. (2008) and Duggan and Payne (2009). The amount of rereading was very low compared to regular reading, with significantly smaller proportions of vertical and leftward (regressive) saccades. The correlation between less rereading and lower comprehension scores replicates Schotter, Tran, et al. (2014), in which participants were not able to reread text. Frequency effects in first fixation durations and first-pass reading times consistent with regular reading were found, replicating White et al. (2015).

3.6.2.2 *Discussion of new findings*

With regard to the reading pattern across the page, similarities were found with spell checking, with both total reading times and total number of visits being very even across the entire text (albeit much longer and higher for spell checking, and much shorter and lower for skimming.) These results show that, compared to regular reading, processing was lower but more evenly distributed across the entire page of text, longer saccades were performed, and reliance on rereading—both of previously read text and regressions to previous words on the current line of text—was lower. It is possible that the pattern reflects that a lower level of effort is used when building the meaning representation of the text (Kintsch, 1998; Sadoski, 1999). Furthermore, no difference in these demands was found in total reading times toward the ends of texts, as meaning never becomes adequately constructed during reading, which also

results in lower comprehension scores. Word skipping is an automatic part of reading (Abbott et al., 2015; Brysbaert et al., 2005; Gao & Suzuki, 2005; Vitu, 2011), but is also consciously affected by the type of reading, as was evidenced by the lower number of visits to words during skimming. The data showing that increased reading speed and fewer visits to words leads to lower comprehension supports the contention of Rayner et al. (2016) in their critique of speed-reading, in which they argue that reading speed can only be increased while maintaining high comprehension by increasing language skill.

3.6.3 Spell checking

Trial durations were significantly longer compared to regular reading. This means that people spent more time reading the texts before moving to the next one.

3.6.3.1 *Replication of previous research*

Spell checking was the only reading instruction in which participants were not reading normally, but rather searching for errors. This task effect was shown in lower comprehension scores and, more importantly, in significantly larger frequency effects compared to regular reading in first-pass reading times and total reading times. This replicates the work of both Kaakinen and Hyönä (2010) and Schotter, Bicknell, et al. (2014), and demonstrates that participants were accessing the words to a larger extent than when reading regularly in order to determine whether or not the spelling was correct. In the case of infrequent words—words with which the participants are not very familiar—this process is more difficult, resulting in longer fixation durations. Significantly greater word length effects compared to regular reading were also found in first-pass reading times and total reading times, replicating research by Kaakinen and Hyönä (2010). This further suggests that words are inspected rather than read during spell checking. Spell checking resulted in significantly shorter saccade amplitudes compared to regular reading. This replicates the results in Kaakinen and Hyönä (2010). In Kaakinen and Hyönä (2010), refixation probability was also found to be higher. In the current study, the total number of visits increased significantly compared to regular reading, which can be considered similar to Kaakinen and Hyönä's (2010) finding.

3.6.3.2 *Discussion of new findings*

Compared to regular reading, spell checking showed significantly longer first fixation durations and first-pass reading times. These early effects show that participants changed their early processing because of the task. Shorter saccade amplitudes combined with longer average fixation durations during the entire trial show that the reading pattern was much more deliberate, with higher amounts of processing during each fixation on a word. A significant difference in overall reading across the page was

found, with regular and thorough reading showing a consistent decrease in total reading times per character and total number of visits, while spell checking showed a striking evenness across the page. As with skimming, it is possible that the pattern reflects that a lower level of effort is used when building a meaning representation of the text (Kintsch, 1998; Sadoski, 1999). While total reading times are longer, no decrease in these demands are found in total reading times toward the ends of texts, since meaning never becomes adequately constructed during reading, which also results in lower comprehension scores. A higher number of regressions and a higher proportion of vertical saccades were also found compared to regular reading, showing that the pattern of reading was very different. This further indicates that people use a different reading strategy when spell checking, increasing their processing, and distributing it much more evenly across the page. It is plausible that a higher amount of activation is needed to correctly identify the spelling of the word compared to identifying sufficient amounts of meaning to continue reading. The larger effects of both frequency and word length indicate that the processing of long and infrequent words was increasingly difficult when correct spelling was required. The increases in total reading times led to increased comprehension for thorough reading, but a decrease in comprehension was seen for spell checking. This suggests that global eye movements are highly dependent on task demands, and that the measure in isolation is not sufficient when investigating learning (Hyönä, 2010). This is further suggested by previous research in multimedia learning showing only inconsistent links between eye movements and learning (de Koning et al., 2010; Schmidt-Weigand et al., 2010; Strukelj, Scheiter, et al., 2016).

Table 3.6 shows a summary of the behavior and measured effects, and how they correlate with the hypotheses and hypothesized effects presented in Section 3.3.2 above.

Table 3.6. Behavior and measured effects on the experimental measures compared to regular reading

Hypotheses presented in Section 3.3.2 supported by the data in **bold**. All behavior and effects in the table are supported by the data.

Reading type	<i>Behavior compared to regular reading</i>	<i>Measured effects compared to regular reading</i>
Thorough reading	Better comprehension of material	Higher comprehension scores
	No difference in processing of words for all fixations	Average fixation durations similar to regular reading
	Larger distances between two subsequent fixations during reading	Larger saccade amplitudes
	More rereading of previously read text	Higher proportion of vertical saccades
	No difference in regressions on the current line of text	No difference in proportion of leftward saccades
	More deliberate reading during entire trial	Longer total reading times, higher number of visits to words
	Reading behavior is similar to regular reading when investigating fixation durations on words in entire trial	Reading pattern is similar to regular reading during entire trial
Skimming	Worse comprehension of material	Lower comprehension scores
	Lower amounts of processing on words for all fixations	Shorter average fixation durations
	Greater distances between two subsequent fixations during reading	Larger saccade amplitudes
	Less rereading of previously read text	Lower proportion of vertical saccades
	Fewer regressions	Lower proportion of leftward saccades
	Less deliberate reading during entire trial	Shorter total reading times, lower number of visits to words
	Uniform reading pattern across the entire text	Total reading times and number of visits are similar over the entire text
Spell checking	Worse comprehension of material	Lower comprehension scores
	Higher amounts of processing of words for all fixations	Longer average fixation durations
	Smaller distances between two subsequent fixations during reading	Smaller saccade amplitudes
	More pronounced frequency effect	Even longer first fixation durations and first-pass reading times for less frequent compared to more frequent words
	More pronounced word length effect	Even longer total reading times for longer compared to shorter words
	More rereading of previously read text	Higher proportion of vertical saccades
	More regressions	Higher proportion of leftward saccades
	More deliberate reading during entire trial	Longer total reading times, higher number of visits to words
	Uniform reading pattern across the entire text	Total reading times and number of visits are similar over the entire text

3.7 Conclusions

The data gathered when using one paragraph of text have been shown to accurately reflect reading behavior. Early effects found in eye-tracking research on single sentences were replicated in this study—namely, the word frequency effect showing increased first fixation durations on less frequent words, as well as the word length

effect showing increased first fixation durations on longer words. The “usual” sentence-level effects can be found, but conclusions about reading an entire page of text can also be drawn. Using paragraphs as stimuli in eye-tracking studies can therefore be seen as viable and very useful, which should encourage the field to move beyond the traditional sentence-level reading research (also see Jarodzka & Brand-Gruwel, 2017).

Compared to regular reading, all three types of reading investigated in this study were notably different. During thorough reading, longer total reading times and a higher number of visits to each word were found, along with higher comprehension scores. This suggests that the material had been read more thoroughly. With regard to the reading pattern over the entire page, similarities were found with regular reading, with total reading times and total number of visits steadily decreasing down the page. This suggests that material found earlier in the text is reread to a higher extent. There might be a need to recheck forgotten details so that the context can be firmly established during initial reading, or the effect may be present simply because it is possible to reread the earlier material, but not possible to reread text that has not yet been seen. During skimming, longer saccades and shorter average fixation durations were combined with fewer visits to each individual word, as well as lower comprehension scores. This is interpreted as more evidence of word skipping, less processing, and less rereading, respectively, and led to less thorough reading being reflected in lower comprehension scores. Total reading times were also shorter and more evenly distributed across the entire page compared to regular reading. This suggests that reading during skimming was deliberately faster and more uniform. During spell checking, shorter saccades and longer average fixation durations were found, along with lower comprehension scores. This is interpreted as evidence of less word skipping and more processing, but these factors do not signify thorough reading, since comprehension was drastically worse. Total reading times were longer and evenly distributed across the entire page. This suggests that reading was deliberately slower and more uniform, but that these results cannot be considered “more thorough” with respect to regular reading, since comprehension decreased.

The study of reading continues in Chapter 4, with participants being given single pages of text to read regularly. To change the participants’ reading behavior, their expectations of what they would be reading were manipulated. In the first experiment, instructions were given about the topics; participants were told either that they would be reading about controversial topics that others have strong opinions about, or neutral topics. In the second experiment, various headlines were used that were either positively or negatively biased toward the topic. The effect that these expectations had on the way people read and processed the information in the texts was analyzed by measuring people’s opinions regarding the topics before and after reading.

Chapter 4

You change your opinions (or not) when expecting controversial texts or texts with evaluative bias

4.1 Introduction

Are our opinions affected to a higher degree when we read a text that we expect to be considered “controversial” by others than if we are told the same text is not considered “controversial”? What about if we expect to read a text that agrees or disagrees with our own opinions? Controversial issues can be defined as “issues on which our society is clearly divided and significant groups within society advocate conflicting explanations or solutions based on alternative values” (Stradling, 1985, p. 9) and “significant numbers of people argue about them without reaching a conclusion” (Oulton, Dillon, & Grace, 2004, p. 2). When reading about controversial topics such as gun control and affirmative action, the participants often disregarded or came up with alternative explanations for the arguments that contradicted their own beliefs, while they uncritically accepted the arguments that agreed with their own beliefs (Taber & Lodge, 2006). It has also been suggested that we prefer information that agrees with our opinions and are more likely to disregard information that disagrees with them (Hergovich, Schott, & Burger, 2010; Kardash & Scholes, 1996; Lord, Ross, & Lepper, 1979; Taber & Lodge, 2006). Similarly, it is suggested that we actively try to find support for our own opinions (Mercier, 2017; Stanovich et al., 2013). Our opinions seem to be “guarded” or conserved by our cognitive processes and biases, and opinion shifts are quite difficult to achieve. Previous research suggests that mere expectations are able to produce cognitive biases (Marsh & Hanlon, 2007; Nickerson, 1998). In this chapter, I investigate if it is enough to *expect* that a text will agree or disagree with our opinions to change how the information in the text is perceived, and produce effects of these opinion-conserving biases.

It seems reasonable to assume that opinion shifts are more likely to occur if people have weak opinions about a topic than if they feel strongly about a topic. If peoples’

opinions are strong, they should be more likely to adhere to these opinions, regardless of whether they are positive or negative. Also, if people state that they do not have strong opinions about a topic, but learn that *other people* feel strongly about it, this may have an effect on their own opinion. In such circumstances, people with weak opinions might in fact start behaving like people whose opinions are strong. When expecting to read a text that agrees or disagrees with our own opinions, it again seems reasonable to assume that opinion shifts are more likely to occur if we have weak opinions about a topic than if we feel strongly about it. This chapter therefore investigates how opinions about a topic are affected by the expectation that other people have strong opinions about it, and the expectation that the text will have a positive or negative bias.

One way of testing biased reading of a text would be to change the headline. Headlines are argued to create a context that optimizes processing (Ifantidou, 2009), and help the reader extract as much information as possible with the least amount of effort (Dor, 2003). It can thus be argued that the perception of a text might change with different headlines. If people read a headline that agrees with their opinions on the topic, they should expect that the text will be biased in favor of their opinions. They should thus become even more certain about their opinions, or at least retain them. Conversely, if people read a headline that disagrees with their opinions, they expect that the text will be biased against their opinions. They should then retain their opinions, as they develop a resistance against being coerced into changing their mind, and their responses after reading should be very similar to the responses given before reading.

In spite of the fact that the effect of headlines on reading and opinions is extremely relevant, very little research has been done on this topic. It has been demonstrated that headlines play an important role in drawing attention to the text in news media (Molek-Kozakowska, 2013, 2014; Schaffer, 1995). In online news media in particular, *click-baiting* is becoming increasingly common. Click-baiting is the practice in which online media uses eye-catching or biased headlines in order to elicit clicks to their articles (Blom & Hansen, 2015; Chen, Conroy, & Rubin, 2015). The possible impact of these click-baiting headlines on opinions has neither been investigated nor problematized.

In order to investigate opinion shifts contingent upon other people's opinions, and on positive and negative headlines accompanying neutral texts, two experiments were conducted. The specific opinions of the participants were measured both before and after reading the texts, and the strength of their opinions were measured before reading. The results were analyzed in two ways in each experiment. In one analysis, trials were grouped in relation to whether people had positive or negative opinions on each topic before reading the texts. The other analysis grouped trials depending on if the participants stated they had a weak or strong opinion on each topic before reading. In Experiment 1, all headlines were neutral. Participants' expectations were instead changed by giving different descriptions of the topics; they were told either

that the text covered “various topics,” or topics that “people say they feel very strongly about” and were “considered highly controversial according to survey data.” This description was meant to increase how important the participants expected the topic to be, and thereby increase how important they considered their own opinion. In Experiment 2, participants’ expectations were manipulated by using different headlines that were either positively or negatively biased toward the topic. This changed the participants’ expectations as to whether a positively or negatively biased text would be read.

4.2 Previous Research

4.2.1 Expectations, framing, agenda setting, and affective priming

Many aspects of our behavior is affected by our expectations, such as expecting a certain result in an experiment or expecting medicine to work (Doyen et al., 2012; Stewart-Williams & Podd, 2004). *Framing* is a concept in which information is presented in different ways, thereby achieving different results (Druckman, 2001; Tversky & Kahneman, 1981; Wallin et al., 2016). The term framing has been used in increasingly diverse ways, becoming more of a universal term for describing communication in general and media effects in particular (Benford & Snow, 2000; Carragee & Roefs, 2004; Scheufele, 1999, 2000; Van Gorp, 2007). Indeed, language even affects factors like economic rationality (Jullien, 2016). The slightly different but related term *agenda setting* investigates mass media communication, and how news coverage can change the opinions and behavior of the public (McCombs & Shaw, 1972); the suggestion is that there is a strong correlation between the strong emphasis placed on an issue by mass media and how important the issue is perceived to be by mass audiences (Scheufele & Tewksbury, 2007). A review paper analyzing fifty years of mass communication research showed that agenda-setting and framing were the two most frequently used theories in several high-profile journals (Bryant & Miron, 2004).

In the context of political communication and mass media, *priming* is considered to be an extension of agenda setting. When the importance of an issue is highlighted and elevated by agenda setting, people’s judgments regarding this issue are affected by priming (Scheufele & Tewksbury, 2007). *Affective priming* is argued to be the affective-evaluative component of priming. The media influences affective priming by using compelling arguments in combination with affect, by using an evaluative tone when writing about an issue (Kiousis, 2005; Sheaffer, 2007). It is argued that the media helps people to attach affective attributes to these issues. A study using data from five Israeli elections showed that more positive media presentations resulted in more positive evaluations, and more negative media presentations resulted in more

negative evaluations (Sheafer, 2007). Using the *hot cognition* hypothesis, which states that every sociopolitical concept that a person evaluates becomes affectively charged (Abelson, 1963; Sheafer, 2007), it is argued that all sociopolitical concepts have affective loadings (Bargh, 1994; Lodge & Stroh, 1993; Lodge & Taber, 2005). In other words, it is suggested that people have positive and negative opinions regarding all sociopolitical concepts. A study indicated that implicit evaluations were made very quickly—within 400 milliseconds of seeing an affective statement—and that the emotional valence of a political issue is stored with the concept itself in our minds (J. P. Morris, Squires, Taber, & Lodge, 2003). Another study showed faster reaction times for congruent pairings and slower reaction times for incongruent pairings, as well as that people with strong prior attitudes seemed to be affected the most (Lodge & Taber, 2005). This demonstrates that people have opinions on most topics, and that they are affected by information that contradicts these opinions.

4.2.2 Newspaper headlines and click-baiting

Dor (2003) argues that newspaper headlines operate as relevance optimizers, helping readers to minimize processing while also maximizing information uptake. Headlines are also said to create ad hoc contexts that optimize processing (Ifantidou, 2009). Similarly, it has been suggested that headlines create a schema that is used during subsequent reading, with appropriate schemas improving comprehension and recall (Bransford & Johnson, 1972), as well as a decrease in rereading behavior and quicker processing of ambiguous words (E. E. Smith & Swinney, 1992; Wiley & Rayner, 2000). In other words, headlines can be seen as framing devices for newspaper articles. In news contexts, however, headlines cannot give comprehensive information, since this could potentially cause readers to ignore the main story and move on to the next headline instead (Kronrod & Engel, 2001). Headlines are more often than not syntactically impoverished, ambiguous, vague, or even misleading (Marquez, 1980; Perfetti, Faux, Beverly, Bell, & Rodgers, 1987). Newspapers have a long tradition of using headlines to attract attention to articles using sensationalism (Molek-Kozakowska, 2013), lexical and syntactic ambiguity (Bucaria, 2004), and plain exaggeration.

The online news media are heavily reliant on headlines to generate traffic for their content, since this is the only bit a reader sees before deciding to click the hyperlink to read the story. The term click-baiting describes the practice of creating headlines in which the purpose is to entice the reader to click on it. When a link is clicked, this results in more activity on the website, as well as more commercial revenue (Blom & Hansen, 2015).

When the online news media use headlines that are misleading, as is often the case with click-baiting headlines, the expectations do not match the content of the text. Headlines, descriptions, and information affect opinions based on the way they are framed (Levin & Gaeth, 1988; Levin, Gaeth, Schreiber, & Lauriola, 2002; Levin,

Schneider, & Gaeth, 1998; Pan & Kosicki, 1993; Van Gorp, 2007). It is possible that misleading and click-baiting headlines could have unintended or unexpected consequences. Since opinions are generally not very stable, consistent or well-informed (Chong & Druckman, 2007; Converse, 1962, 2006; Zaller, 1992), it is possible that the expectations from the headline alone can change opinions, regardless of what the text contains. Research has been performed on opinion shift in regard to specific topics (e.g., Bettinghaus & Baseheart, 1969; Bettinghaus, 1986; Etgar & Goodwin, 1982) and over time (e.g., Keasey, 1973). As mentioned above, the affective priming hypothesis states that more positive and negative media content lead to more positive and negative evaluations (Bargh, 1994; Kioussis, 2005; Lodge & Taber, 2005; J. P. Morris et al., 2003; Sheaffer, 2007). These effects, however, were induced by actually biased content, and not merely the expectation of biased content.

4.2.3 Myside bias and cognitive dissonance

Information that supports people's opinions and beliefs is very often favored, while information that argues against people's opinions and beliefs is disregarded (Bastardi, Uhlmann, & Ross, 2011; J. K. Clark, Wegener, Habashi, & Evans, 2012; Hergovich et al., 2010; Jonas, Schulz-Hardt, Frey, & Thelen, 2001; Klayman, 1995; Nickerson, 1998). The term commonly used to describe this phenomenon is *confirmation bias* (Hergovich et al., 2010; Jonas et al., 2001; Klayman, 1995; Nickerson, 1998). Confirmation bias is considered to be an automatic rather than a conscious process. It functions as a heuristic, or shortcut, for cognitively complex tasks (Friedrich, 1993; Oswald & Grosjean, 2004); in other words, it is a process that helps us process information in a more efficient way, albeit at a cost. Expectations were shown to affect observers during a task in which animal behavior was classified using qualitative judgments, but confirmation bias accounted for less than 20% of the mean values in reporting (Marsh & Hanlon, 2007). This, along with other studies, suggests that confirmation bias might not be as strong as previously thought (Mercier, 2017). It has therefore been argued that the more recent theory of *myside bias* should instead be used. Myside bias shows that people evaluate evidence and test hypotheses with a bias toward their own opinions and beliefs (Mercier, 2017; Stanovich et al., 2013), and that people prefer their own opinions over an advisor's opinion (Yaniv, 2004). Myside bias proposes a more limited set of predictions using stricter experimental control and more rigorous methods to verify the claims. With the knowledge that the two theories are very similar, the term myside bias will be used exclusively from now on. In addition, the way controversial issues are read mimics the effects of myside bias, with participants coming up with alternative explanations to arguments in texts that contradict their beliefs and uncritically accepting arguments that agree with their beliefs (Taber & Lodge, 2006). This suggests that controversial issues might produce more myside bias than "regular" issues.

When people encounter information that does not correspond to their current opinions or knowledge, a conflict arises. This conflict is called *cognitive dissonance*, and it is resolved by altering or adapting one of the two conflicting pieces of information (Bem, 1967; Festinger, 1962). Since it may be painful to change one's opinion, it has been argued that people avoid information that causes cognitive dissonance (Festinger, 1962; Harmon-Jones & Mills, 1999; Kardash & Scholes, 1996).

4.2.4 How opinions can be measured

Agreement bias shows that respondents tend to agree with a statement regardless of its content (DeVellis, 2003, p. 69). Participants are also more likely to disagree with negative questions—for example, “this issue is bad; agree-disagree”—than to agree with positive questions—for example, “this issue is good; agree-disagree” (Kamoen, Holleman, Mak, Sanders, & van den Bergh, 2011; Kamoen, Holleman, & van den Bergh, 2013). Since statements about opinions are described in either positive or negative ways—for example, agreeing or disagreeing with statements such as “I believe that this issue is good” and “I believe that this issue is bad”—these biases need to be taken into account when creating survey statements so that results are as accurate and robust as possible. Changes in opinions can be identified by measuring responses before and after receiving advice, as suggested in the literature review by Bonaccio and Dalal (2006). This should also be the case after other tasks such as reading texts.

When performing two separate measurements, if the first result deviates far from the mean, the second result will be closer to it, on average. In other words, if one measurement is extreme, the other measurement will be less extreme, on average. This statistical phenomenon is called *regression to the mean* (Barnett, van der Pols, & Dobson, 2005; Bland & Altman, 1994). It is commonly used in medical research, in which the nature of measurements in intervention studies need to be compensated for a regression to the mean, so that results show the true effects of the intervention (Shepard & Finison, 1983; Whitney & Von Korff, 1992; Yudkin & Stratton, 1996). On the second measurement, extreme results will be drawn toward the middle of the distribution, where the probability density is the highest. Therefore, if a manipulation removes or decreases this regression to the mean, the manipulation could be said to have an effect.

4.3 The study of opinion shift after paragraph reading

In two experiments, I investigate the effect of framing on text perception, and, more specifically, whether expectations alone can cause opinion shifts after reading a text. The critical texts in both experiments were congruent with positive, negative, and neutral headlines. Apart from the headlines, the texts were identical in the two

experiments except for very minor corrections. The texts were all descriptive in nature, and highlighted both the positive and negative sides of the topic in a balanced way. In this study, the term *neutral* is used to refer to these texts. No manipulations were performed in the actual texts in either experiment.

In Experiment 1, the headlines were neutral and uninformative. One half of the texts, however, were introduced as being about “various topics,” and the other half were introduced as texts “on various topics that people say they feel very strongly about, and that are considered highly controversial according to survey data.” The hypothesis was that by hearing that others felt strongly about a topic, participants would feel more strongly about their own opinions. The headline was shown at the same time as the text. The two groups of texts were created based on opinion data from a norming experiment in which participants stated how controversial they considered the topics to be. In order to ensure experimental rigor, the two groups of texts needed to be as similar as possible with regard to how controversial the topics were considered to be overall, but also needed to contain topics that were considered to be both controversial and non-controversial.

In Experiment 2, the headlines were either positively or negatively worded. This was intended to cause participants to expect biased texts, and thus create biased frames that would affect the reading of the texts. Headlines were shown in isolation for two seconds at the top of the page before the texts appeared below, and the participants were informed that this would happen. Showing the headline for two seconds before the entire text is presented emulates an online news article, in which only the headline is visible until the reader clicks the link. The participant was thus implicitly guided to read the headline and not skip it during the experiment.

4.3.1 Terms and classifications

In order to evaluate participants’ opinions regarding the topics, they responded to two statements about their specific opinions on each, both before and after reading; responses were recorded on continuous scales ranging from 0.00 to 1.00. One statement was worded positively and the other was worded negatively; for instance, “Women’s financial situations are more positively affected by divorce than men’s” or “Women’s financial situations are more negatively affected by divorce than men’s.” The response to the negatively worded statement (where a higher value signifies higher disagreement) was subtracted from the response on the positively worded statement (where a higher value signifies higher agreement), producing two differences ranging from -1.00 to 1.00 for each trial, one before and one after reading. I refer to this as *agreement score*. Participants also stated how strongly they felt about each topic on a continuous scale ranging from 0.00 to 1.00. I refer to this as *opinion strength*.

Trials were classified into two groups in two different ways. One analysis grouped the trials based on the participant having positive or negative opinions about the topic

before reading; that is, having either positive or negative agreement scores before reading. If the agreement score before reading was positive (equal to or above 0.00), the trial was put into the group in which participants had positive prior opinions on the topic. If the agreement score before reading was negative (below 0.00), the trial was put into the group in which participants had negative prior opinions on the topic. The other analysis grouped the trials based on the participant having weak or strong opinions about the topic; that is, having low or high opinion strength. If the opinion strength was below 0.50, the participant was classified as having a weak opinion on the topic. If the opinion strength was equal to or above 0.50, the participant was classified as having a strong opinion on the topic.

In Experiment 1, the texts specified as being on “various topics” will be referred to as *neutral*. The texts specified as being on “various topics that people say they feel very strongly about, and that are considered highly controversial according to survey data” will be referred to as *controversial*.

4.3.2 Hypotheses

In Experiment 1, neutral texts were paired with neutral headlines, and the topics were introduced as being either controversial or neutral. If readers would expect to read a controversial text, they might increase their own opinion strength in order to protect themselves from having to re-evaluate their current position, which they now considered to be more important. Experiment 1 thus examined whether increasing the perceived importance of the topic would produce resistance to opinion shifts.

When participants demonstrate high opinion strength, they are sure of their opinion and are unwilling to change it. Resistance to opinion shifts was therefore expected for all texts. When participants demonstrate low opinion strength, they are not sure of their opinion and might be more likely to change it. Resistance to opinion shifts was therefore expected only for the controversial topics, in which they might behave similarly to participants with high opinion strength. Changes in opinion were expected for the neutral topics only. The mean agreement scores, however, were expected to be very close to zero before reading, and no hypothesis was made as to the particular direction of the measurements. The texts were carefully tailored to be neutral and balanced. If the participants perceived the texts to be slightly positively biased, however, agreement scores would increase slightly, and if the participants perceived the texts to be slightly negatively biased, agreement scores would decrease slightly. These changes were not expected to be significant.

When participants had either a positive or negative prior opinion about the topics, resistance to opinion shifts was expected for the controversial topics. For the neutral topics, regressions to the mean were expected instead. Regression to the mean occurs for repeated measurements when one group of measurements is very different (Barnett et al., 2005; Bland & Altman, 1994). This is the case in this study, as the trials are divided into either only positive agreement scores before reading, or only

negative agreement scores before reading. A visualization of the hypothesized opinion shifts in Experiment 1 is given in Figure 4.1. Note that these are not the actual results gathered from participants. A summary of the hypotheses and the hypothesized effects in Experiment 1 are given in Table 4.1 and Table 4.2.

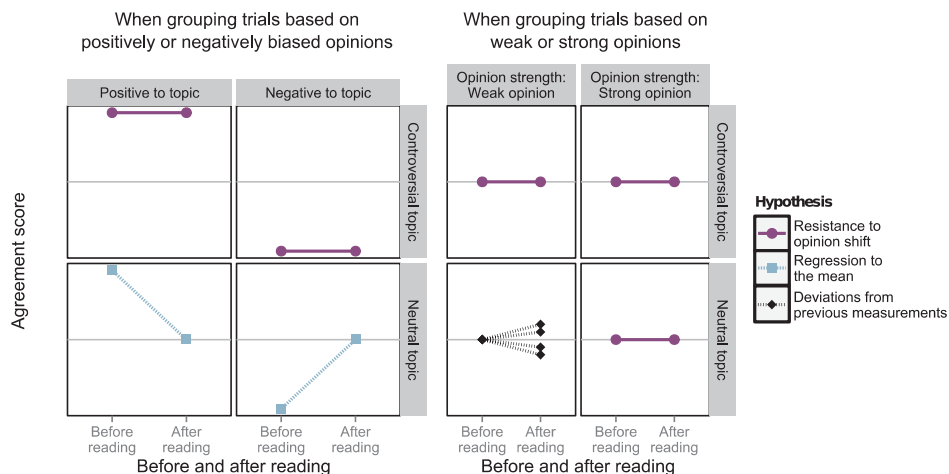


Figure 4.1. Visualization of hypothesized opinion shifts after reading, Experiment 1
Figure does not show actual experimental data.

Table 4.1. Hypotheses and hypothesized effects on the experimental measures depending on opinion strength, Experiment 1

The hypotheses and effects supported by the experimental data are reported in **bold** in Table 4.11 in Section 4.8.

Topic	Participant opinion strength on the topic	
	Weak opinion	Strong opinion
Framed as a controversial topic	Hypothesis: The participants do not feel strongly about the topic. Hearing that others feel strongly about the topic makes them feel more strongly about the topic and retain their opinions.	Hypothesis: Since the participants feel strongly about the topic, the manipulation is not necessary for their opinions to be retained.
	Hypothesized effects: Resistance to opinion shift.	Hypothesized effects: Resistance to opinion shift.
Framed as a neutral topic	Hypothesis: Since the participants do not feel strongly about the topic, their opinions change.	Hypothesis: Since the participants feel strongly about the topic, their opinions are retained.
	Hypothesized effects: Opinion shift, seen in deviations from previous measurements.	Hypothesized effects: Resistance to opinion shift.

Table 4.2. Hypotheses and hypothesized effects on the experimental measures depending on positive or negative agreement score before reading, Experiment 1

The hypotheses and effects supported by the experimental data are reported in **bold** in Table 4.12 in Section 4.8.

Topic	Participant agreement score on the topic before reading	
	Positive to topic	Negative to topic
Framed as a controversial topic	<i>Hypothesis:</i> The participants will retain their currently held opinions because they hear that others feel strongly about the topic, causing them to feel more strongly about it themselves.	<i>Hypothesis:</i> The participants will retain their currently held opinions because they hear that others feel strongly about the topic, causing them to feel more strongly about it themselves.
	<i>Hypothesized effects:</i> Resistance to opinion shift.	<i>Hypothesized effects:</i> Resistance to opinion shift.
Framed as a neutral topic	<i>Hypothesis:</i> Agreement scores exhibit the regular behavior of subsequent measurements, when one group is very different.	<i>Hypothesis:</i> Agreement scores exhibit the regular behavior of subsequent measurements, when one group is very different.
	<i>Hypothesized effects:</i> Opinion shift, seen in regressions to the mean.	<i>Hypothesized effects:</i> Opinion shift, seen in regressions to the mean.

In Experiment 2, neutral texts were paired with either positively or negatively biased headlines, producing expectations about the content, and causing participants to read the texts in light of the introduced bias.

When participants have high opinion strength, they are sure of their opinion and are unwilling to change it. Strong opinions on a topic also cause participants to be affected by incongruent affective priming (Lodge & Taber, 2005). Resistance to opinion shifts was therefore expected for all texts. When participants have low opinion strength, they are not sure of their opinion and might be more likely to change it. The affective priming induced by the biased headlines should therefore influence their opinions. Therefore, agreement scores were expected to increase for texts with positive headlines, and decrease for texts with negative headlines.

When the participants have either a positive or negative prior opinion about the topics, this should produce resistance to opinion shift, since the participants are affected by myside bias and cognitive dissonance. Again, regression to the mean occurs for repeated measurements when one group of measurements is very different, as it is in this study. So deviations from this, that is, only very small regressions to the mean, would signify that the manipulation affected opinions. A visualization of the hypothesized opinion shifts in Experiment 2 is given in Figure 4.2. Note that these are not the actual results gathered from participants. A summary of the hypotheses and the hypothesized effects in Experiment 2 are given in Table 4.3 and Table 4.4.

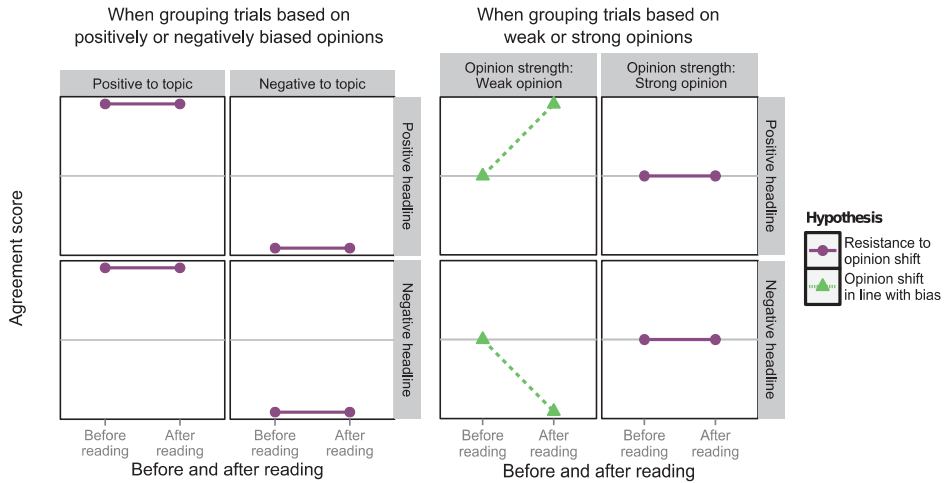


Figure 4.2. Visualization of hypothesized opinion shifts after reading, Experiment 2
Figure does not show actual experimental data.

Table 4.3. Hypotheses and hypothesized effects on the experimental measures depending on opinion strength, Experiment 2

The hypotheses and effects supported by the experimental data are reported in **bold** in Table 4.13 in Section 4.8.

Headline	Participant opinion strength on the topic	
	Weak opinion	Strong opinion
Positive headline	Hypothesis: Since the participants do not feel strongly about the topic, the expectation-induced affective priming from the headline influences the reading of neutral text.	Hypothesis: Since the participants feel strongly about the topic, they will not change their opinions.
	Hypothesized effects: Opinion shift in line with bias, seen in higher agreement scores.	Hypothesized effects: Resistance to opinion shift.
Negative headline	Hypothesis: Since the participants do not feel strongly about the topic, the expectation-induced affective priming from the headline influences the reading of neutral text.	Hypothesis: Since the participants feel strongly about the topic, they will not change their opinions.
	Hypothesized effects: Opinion shift in line with bias, seen in lower agreement scores.	Hypothesized effects: Resistance to opinion shift.

Table 4.4. Hypotheses and hypothesized effects on the experimental measures depending on positive or negative agreement score before reading, Experiment 2

The hypotheses and effects supported by the experimental data are reported in **bold** in Table 4.14 in Section 4.8.

Headline	Participant agreement score on the topic before reading	
	Positive to topic	Negative to topic
Positive headline	<i>Hypothesis:</i> The participants will retain their currently held opinions because they expect the text to agree with them, due to myside bias.	<i>Hypothesis:</i> The participants will retain their currently held opinions because they expect the text to disagree with them, and therefore pay less attention to it, due to myside bias and cognitive dissonance.
	<i>Hypothesized effects:</i> Resistance to opinion shift.	<i>Hypothesized effects:</i> Resistance to opinion shift.
Negative headline	<i>Hypothesis:</i> The participants will retain their currently held opinions because they expect the text to disagree with them, and therefore pay less attention to it, due to myside bias and cognitive dissonance.	<i>Hypothesis:</i> The participants will retain their currently held opinions because they expect the text to agree with them, due to myside bias.
	<i>Hypothesized effects:</i> Resistance to opinion shift.	<i>Hypothesized effects:</i> Resistance to opinion shift.

4.4 Norming data

The texts in Experiment 1 were described as being either about “various topics” or about topics that “people say they feel very strongly about” that were “considered highly controversial according to survey data.” These two groups were created based on previously gathered opinion data so that the groups were as similar as possible with regard to mean agreement scores, while also having the largest possible range of topics considered both controversial and neutral.

4.4.1 Method

4.4.1.1 Participants

Sixty-four native speakers of English (38 female; 18 to 41 years of age; $M=27.6$ years, $SD=5.92$) participated in the experiment. The participants were recruited through student lists and advertisements at Lund University, Lund, Sweden.

4.4.1.2 Apparatus

The experiment took place at the Lund University Humanities Lab at the Centre for Languages and Literature. The questionnaire was completed on a computer using PsychoPy (1.80.00).

4.4.1.3 Design and stimuli

The opinion questionnaire contained two groups of 20 statements, each regarding the strength of the participants’ opinions on each topic. This resulted in 40 statements in total, with two statements for each topic. The first group of statements addressed how

strongly the participant felt about each topic. Using the mouse, the participants stated to what degree they agreed with the statement on an ungraded visual analog scale ranging from “completely disagree” to “completely agree.” An example statement is “I have a strong opinion on vaccination.” The continuous scale ranged from 0.00 to 1.00, and measured the participant’s opinion strength. The responses equal to or lesser than 0.50 were classified as not having strong opinions on the topic (hereafter referred to as weak opinions, for brevity), and responses greater than 0.50 were classified as having a strong opinions. The statements given to participants, however, were specifically created so that the extremes did not signify “a weak opinion” and “a strong opinion” respectively, which would make the middle of the scale a neutral response. The participants were also specifically instructed to interpret the scale as how strongly they felt about the topic. Marking one extreme of the scale, labeled *completely agree*, would mean that they indeed felt very strongly about the topic, and marking the other extreme, labeled *completely disagree*, would mean that they did *not* have strong feelings about the topic. This was done so that participants would not interpret marking the middle of the scale to mean not having strong feelings or being neutral about the topic. This justified the division into two groups. Also, the difficult self-evaluation “how strong or weak is my opinion on this topic?” was turned into the simpler “do I feel strongly about this topic?”

4.4.1.4 Procedure

Each participant was introduced to the experiment and recorded individually. The experimental session lasted for around 5 minutes, but no time restrictions were imposed during the experiment. The participants were welcomed to the study, and were told that they were going to answer a questionnaire about their opinions on several different topics. They were informed that their participation was completely voluntary and that they were free to end the experiment at any time. The experiment started with the participants completing a questionnaire about their age and gender, followed by the opinion questionnaire. The participants were instructed to answer spontaneously, and not to think too long about their answers. After completing the questionnaire, the participants were thanked for participating in the study.

4.4.1.5 Data analysis

The data processing and figure were made using R (3.2.2) (R Development Core Team, 2015) in R Studio (0.99.473).

4.4.2 Results

Two groups of texts were created by sorting the mean opinion strength scores about the topics in descending order, and then distributing the texts between the two groups of texts in an alternating fashion. Figure 4.3 shows the distribution of the texts

and their respective mean opinion strength scores. As can be seen in Figure 4.3, the range of the mean opinion strength scores was as large as possible within each group, while also resulting in very similar mean opinion strength scores for all texts in the group (Group 1: $M=0.568$, $SD=0.141$; Group 2: $M=0.568$, $SD=0.128$).

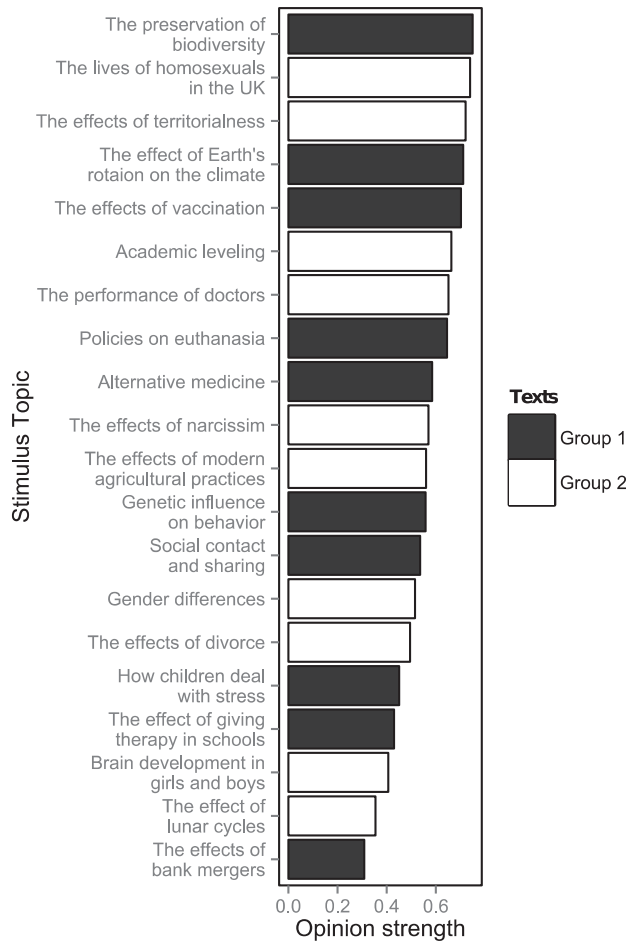


Figure 4.3. Texts divided into two groups in Experiment 1, depending on the mean opinion strength scores for each topic from the norming data
Mean opinion strength for all texts is identical in both groups (0.568), with different standard deviations (Group 1 $SD=0.141$; Group 2 $SD=0.128$).

4.5 Experiment 1

4.5.1 Method

4.5.1.1 Participants

Seventy-two native speakers of English (51 female; 18 to 32 years of age; $M=20.3$ years, $SD=2.57$) participated in the experiment. The participants were recruited with SONA Experiment Scheduling at the Department of Psychology at the University of California, San Diego, United States.

4.5.1.2 Apparatus

The experiment took place at the Rayner Eye Tracking Lab, University of California, San Diego, United States. Stimuli were displayed on a ViewSonic VX922 19" 4:3 LCD display at a resolution of $1,280 \times 1,024$ pixels with a refresh rate of 60 Hz. Questionnaires before and after reading were presented on a separate computer using PsychoPy (1.80.00).

4.5.1.3 Design and stimuli

A total of 20 texts were taken from the English reading comprehension part of the Swedish Scholastic Aptitude Test, and adapted to the current study. Neutral and uninformative headlines were created for the texts. The texts were similar in style to the short articles found in mainstream online media. These texts highlighted both negative and positive aspects of a topic, were descriptive in nature, and were kept neutral so that both versions of the headlines were congruent with the corresponding texts. The texts were slightly modified so that all of them were 90 words in length, excluding headlines. Half of the texts were presented with the instruction that they dealt with various topics, coded as neutral in this study, or topics that "people say they feel very strongly about" that were "considered highly controversial according to survey data," coded as controversial in this study. For the other group, the controversial texts were given as neutral, and vice versa. A total of 12 texts were critical texts that were part of the analysis. In order to move from the text to the comprehension question, participants pressed the spacebar. No time limits were set for any task during the experiment.

Texts were given in groups of 10, with instructions regarding whether the topics were controversial or not given before each group of texts. The order of the instructions counterbalanced, and the order of the texts in each group was randomized in Experiment Builder. The headline was presented at the same time as the rest of the text during each trial. An example stimulus is given in (15).

(15) Divorce.

Statistics show that women usually get the short end of the stick when their marriages end, but it is men who feel less satisfied with their divorce settlements. Psychologist Virgil Sheets asked 372 ex-couples how they felt about the terms of their divorce, including custody arrangements, child support, and division of property. Crucially, he found that women were far happier than their former mates despite the fact that the average woman's standard of living plummets 27 percent the year after her marriage dissolves, while a man's instead increases 10 percent.

Two questionnaires were given to each participant, one before and one after reading. The pre-reading questionnaire examined opinion strength, that is, the strength of the participant's opinion on each topic. This was evaluated by the responses on 20 statements formulated as "I have a strong opinion on," followed by each specific topic. The pre- and post-reading questionnaires examined agreement scores. When comparing these scores, opinion shift after reading was observed. Both the pre- and the post-reading questionnaire contained 40 statements regarding specific opinions on the 20 topics. This resulted in 80 statements in total, with four statements for each topic. In the current study, agreement bias was possible in the questionnaire, since neutral statements could not be used (DeVellis, 2003, p. 69). Opposing questions with only a yes/no answer have been shown to be unreliable (Kamoen et al., 2011, 2013), and this might also be the case for 7-point Likert scale responses (Holleman, 2000; Kamoen et al., 2011). Since statements were used instead of questions in this study, this possible limitation should be weaker. To further offset this, two separate statements were given about specific opinions on each topic, with one being positive and the other negative. In addition, since questions prime each other (Strack, Martin, & Schwarz, 1988), the two pairs were reformulated in order to minimize recognition, and they were given in separate questionnaires. Examples of the questions can be found in Table 4.5. For half of the participants, the specific opinion statements in the pre-reading questionnaire were substituted with the post-reading questionnaire and vice versa, for counterbalancing purposes. Participants stated to what degree they agreed with the statement. The order of the statements was randomized, and all statements were presented one by one in isolation.

Table 4.5. Two examples of statements presented before reading and the corresponding two statements presented after reading

Bias highlighted in **bold** (this was not present in the statements given to the participants.) The example statements correspond to the text in (15) above.

Before reading	After reading
Divorce improves an individual's financial situation.	Divorce does not improve an individual's financial situation.
Women's financial situations are more positively affected by divorce than men's.	Women's financial situations are more negatively affected by divorce than men's.

4.5.1.4 Procedure

Each participant was introduced to the experiment and recorded individually. The experimental session lasted around 40 minutes, but no time restrictions were imposed during the experiment. The participants were welcomed to the study, and were told that they would participate in a reading study consisting of several different sections, and that written instructions would be given during each part of the experiment. They were informed that their participation was completely voluntary, and that they could stop the experiment at any time.

The experiment began with the participants completing a questionnaire about their age and gender, followed by the pre-reading questionnaire. The participants were instructed to answer spontaneously, and not think too long about their answers. The participants were then told that they would read several texts, each followed by a comprehension question. When finished with a text, they should press the spacebar in order to answer the comprehension question. After the reading phase, a filler task was performed for 20 minutes, after which the post-experiment questionnaire was completed. The participants were then told about the real purpose of the experiment, and shown an example of a manipulation.

4.5.1.5 Data analysis

The data processing, figures, and statistical analyses were made using R (3.2.2) (R Development Core Team, 2015) and the lme4 package (Bates et al., 2015) in R Studio (0.99.473). The responses to the pre- and post-reading questionnaire were coded into agreement scores by subtracting the answer for the negatively worded statement (where a higher value signifies higher disagreement) from the positively worded statement (where a higher value signifies higher agreement), resulting in agreement scores ranging from -1.00 to 1.00. The responses in the pre-reading questionnaire were used to create two groups with either positive or negative prior opinions. The statements for how strongly the participants felt about the topic were also used to form two groups, with a weak or a strong opinion about a topic. Visual analog scales are also sensitive to slight positive and negative biases for the responses at the center of the scale, since they contain no visual anchor at the center. Linear mixed effect models were used to examine opinion shift, manifested in the difference in agreement score between before and after reading, as a function of opinion strength on the topic, and of positive or negative prior opinions about the topic. The distribution was Gaussian with the exception of peaks at the two extremes: 1.00 and -1.00. One of the factors in both designs was text that addressed either controversial or neutral topics. In order to produce a Gaussian distribution suitable for regression analysis, 25 agreement scores of exactly -1.00 or 1.00 were excluded, resulting in a total of 1,703 trials. Models used random intercepts and slopes of condition within group for participant and item, with intercept-slope correlation factors for participants and items. All models converged when these parameters were used. The

lmerTest package (2.0-29) was used to estimate p -values with Satterthwaite approximation.

4.5.2 Results

All statistical models used log-transformed data, but in order to improve readability, no log-transformation was performed on the data in the figures. Before reading the texts, the weak opinion group had significantly lower agreement scores compared to the strong opinion group (Estimate=-0.116, $SE=0.021$, $t=-5.40$, $p<.001$). The scores in the weak opinion group were closer to 0.00 than the strong opinion group.

Figure 4.4 shows the agreement scores before and after reading, depending on whether the participants had low or high opinion strength on the topic before reading. No significant changes in the agreement scores before and after reading were found (controversial texts: Estimate=-0.035, $SE=0.045$, $t=-0.78$, $p=.45$; neutral texts: Estimate=-0.031, $SE=0.048$, $t=-0.65$, $p=.53$). All scores decreased after reading texts. No significant interactions were found between opinion strength and agreement scores (controversial texts: Estimate=0.034, $SE=0.042$, $t=0.80$, $p=.43$; neutral texts: Estimate=0.023, $SE=0.043$, $t=0.55$, $p=.58$). To further examine the differences between controversial and neutral, a chi-square test of goodness of fit compared the two models. One model used condition (either controversial or neutral text) and one did not. The models were not significantly different (Chi df=4, Chisq=1.96, $p=.74$). No significant interaction was found between the grouping of positive or negative opinion before reading and weak or strong opinions on the topics (controversial texts: Estimate=0.001, $SE=0.071$, $t=0.01$, $p=.99$; neutral texts: Estimate=-0.024, $SE=0.071$, $t=-0.32$, $p=.75$). Table 4.6 shows the mean agreement scores, depending on whether participants had low or high opinion strength on the topic before reading.

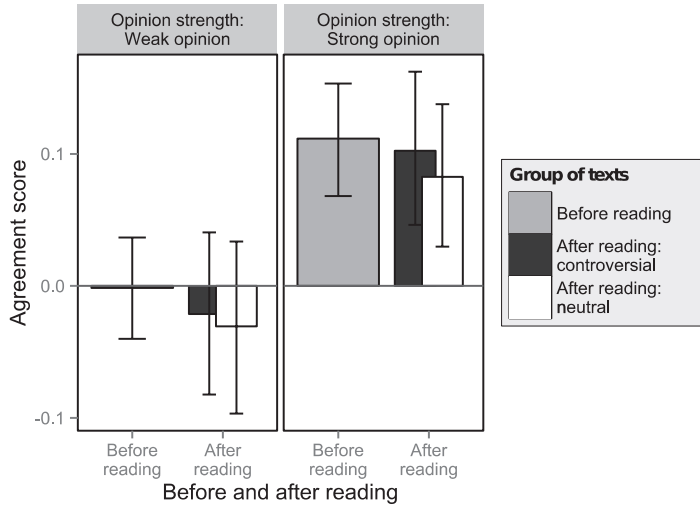


Figure 4.4. Mean agreement scores before and after reading, when grouping trials depending on low or high opinion strength, Experiment 1
Error bars denote standard errors of the mean.

Table 4.6. Mean agreement scores after reading depending on topic, when grouping trials depending on low or high opinion strength, Experiment 1
Standard deviations in parentheses.

Topic	Group with weak opinions (opinion strength<0.50)	Group with strong opinions (opinion strength>=0.50)
Controversial topic	-0.021 (0.422)	0.101 (0.471)
Neutral topic	-0.031 (0.443)	0.082 (0.438)

Figure 4.5 shows the agreement scores before and after reading, depending on whether participants were positive or negative to the topic before reading. The mean agreement score before reading for all participants was 0.064 ($SD=0.463$). After participants had read the texts, agreement scores approached the mean for all participants for both the positive opinion group and the negative opinion group. A significant decrease was found for the positive opinion group for both controversial (Estimate=-0.140, $SE=0.033$, $t=-4.26$, $p<.001$) and neutral texts (Estimate=-0.176, $SE=0.048$, $t=-3.65$, $p<.005$). A significant increase was found for the negative opinion group for both controversial (Estimate=0.146, $SE=0.063$, $t=2.31$, $p<.05$) and neutral texts (Estimate=0.164, $SE=0.058$, $t=2.86$, $p<.02$). Table 4.7 shows the mean agreement scores, depending on whether participants were positive or negative to the topic before reading.

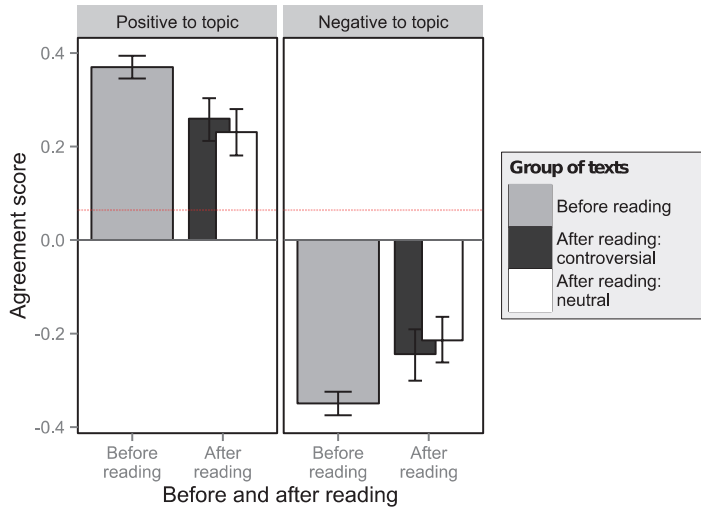


Figure 4.5. Mean agreement scores before and after reading, when grouping trials depending on positive or negative agreement score before reading, Experiment 1
Mean agreement score before reading for all participants is marked with a dotted line in the figure. Error bars denote standard errors of the mean.

Table 4.7. Mean agreement scores after reading depending on topic, when grouping trials depending on positive or negative agreement score before reading, Experiment 1
Standard deviations in parentheses.

Topic	Group with positive agreement scores before reading ($M=0.370$, $SD=0.282$)	Group with negative agreement scores before reading ($M=-0.349$, $SD=0.256$)
Controversial topic	0.260 (0.370)	-0.244 (0.396)
Neutral topic	0.231 (0.410)	-0.215 (0.350)

4.5.3 Discussion

When dividing trials into two groups depending on high or low opinion scores, the two groups represented trials in which participants had either a strong or a weak opinion on the topic. A resistance to opinion shift was hypothesized for the trials with high opinion scores (strong opinions), in both conditions. For the trials in which opinion was stated as being weak, this result was expected under the controversial condition. Under the neutral condition, however, results were only expected to differ from the measurements obtained before reading. Instead, all agreement scores decreased slightly, but no significant effects were found (all $ps>.43$). Also, the model comparison between the controversial and neutral condition was not significantly different ($p=.74$). The fact that all agreement scores decreased is surprising, since the texts, along with the headlines, were neutral. This could mean that people generally expect texts to be negative rather than positive. When given unbiased and neutral headlines, there is a tendency to expect a negative text, which might make opinions

change accordingly. Since the results did not decrease significantly, however, the results from the strong opinion group were interpreted as resistance to opinion shift.

When dividing trials depending on participants having either positive or negative agreement scores about the topic before reading, a similar resistance to opinion shift was expected for the topics introduced as “controversial topics” that “other people felt strongly about.” The introduction would indicate that the opinion held by the participant was more important, and the participant would therefore be more likely to retain it. Results should therefore show a smaller regression to the mean. In previous research, controversial issues were described as issues about which society is clearly divided, and that people argue about without reaching a conclusion (Oulton et al., 2004, p. 2; Stradling, 1985, p. 9). However, in both conditions, the mean agreement scores decreased significantly and approached zero for the positive opinion group, and increased significantly and approached zero for the negative opinion group.

Because of the limited results in Experiment 1, a stronger manipulation was needed. An expectation of an evaluative bias was introduced for the texts in Experiment 2. Since a great deal of research has been done about how people read texts that either agree or disagree with their opinions, the manipulation was expected to produce a much stronger resistance to opinion shifts in those with strong opinions, and produce opinion shifts in those who did not have strong opinions.

4.6 Experiment 2

Except for the differences explained in the Method section below, Experiment 2 was identical to Experiment 1.

4.6.1 Method

4.6.1.1 *Participants*

Sixty-four native speakers of English (38 female; 18 to 41 years of age; $M=27.6$ years, $SD=5.92$) participated in the experiment. The participants were recruited through student lists and advertisements at Lund University, Lund, Sweden. The participants were naïve to the purpose of the experiment.

4.6.1.2 *Apparatus*

The experiment took place at the Lund University Humanities Lab at the Centre for Languages and Literature. Stimuli were displayed on a Samsung SyncMaster 245T 24" widescreen LCD display at a resolution of $1,920 \times 1,200$ pixels with a refresh rate of 60 Hz.

4.6.1.3 *Design and stimuli*

The 20 texts (12 critical texts and eight filler texts) from Experiment 1 were used, except for very minor corrections. The neutral and uninformative headlines used in Experiment 1 were replaced with headlines that were either positively or negatively biased regarding the overall subject of the texts. A total of eight texts were filler texts, since they were not completely congruent with both of the versions of the headlines. This decreased the risk of participants understanding what was being tested. The order of the texts was randomized. To ensure that participants read the headline and also emulated how a link is seen in online media, the headline was presented for two seconds before the rest of the text was shown during each trial.

An example stimulus is given in (16). The headline read “The good effects of divorce” for the positive headline condition and “The bad effects of divorce” for the negative headline condition. The alternatives are indicated in **bold** in (16), with only one alternative shown to participants. No words were displayed in bold during the experiment.

- (16) The **good/bad** effects of divorce.

Statistics show that women usually get the short end of the stick when their marriages end, but it is men who feel more dissatisfied with their divorce settlements. Psychologist Virgil Sheets asked 372 ex-couples how they felt about the terms of their divorce, including custody arrangements, child support, and division of property. He found that women were far more pleased than their former mates despite the fact that the average woman’s standard of living plummets 27 percent the year after her marriage dissolves, while a man’s instead increases 10 percent.

4.6.1.4 *Procedure*

Identical to Experiment 1.

4.6.1.5 *Data analysis*

In order to produce a Gaussian distribution suitable for regression analysis, 36 agreement scores of exactly -1.00 or 1.00 were excluded, resulting in a total of 1,356 trials. Two separate two-by-two designs were used. The headline was one of the factors in both designs. For one design, positive or negative agreement score before reading was the second factor. For the other design, opinion strength on the topic was the second factor. Models used random intercepts and slopes of condition within group for participant and item, with intercept–slope correlation factors for participants and items. All models converged when using these parameters.

4.6.2 Results

All statistical models used log-transformed data, but in order to improve readability, no log-transformation was performed on the data in the figures. Before reading the texts, the weak opinion group had significantly lower agreement scores compared to the strong opinion group (Estimate=-0.121, $SE=0.038$, $t=-3.19$, $p<.005$). The scores in the weak opinion group were closer to 0.00 than the strong opinion group.

Figure 4.6 shows the agreement scores before and after reading, depending on whether participants had low or high opinion strength on the topic before reading. When investigating opinion shifts after reading for the two groups, no significant changes in agreement scores were found for texts with positive (Estimate=-0.037, $SE=0.031$, $t=-1.22$, $p=.23$) or negative headlines (Estimate=-0.013, $SE=0.030$, $t=-0.43$, $p=.67$). However, an interaction was found between opinion strength and agreement score for texts with positive headlines (Estimate=0.115, $SE=0.048$, $t=2.38$, $p<.02$), with a significant increase after reading for the weak opinion group. Texts with negative headlines showed a decrease in agreement scores, but this decrease was not significant (Estimate=-0.010, $SE=0.048$, $t=-0.22$, $p=.83$). In order to confirm these results, the analyses were remade using opinion strength as a continuous variable rather than two groups. The results were confirmed, increasing opinion strength produced decreasing changes in attitudes after reading for positive headlines (positive headline: Estimate=-0.155, $SE=0.077$, $t=-2.00$, $p<.05$; negative headline: Estimate=-0.0006, $SE=0.079$, $t=-0.08$, $p=.94$). No significant interaction was found between the grouping of positive or negative opinion before reading, and strong or weak opinions on the topics for texts with positive or negative headlines (positive headline: Estimate=0.140, $SE=0.084$, $t=1.66$, $p=.10$; negative headline: Estimate=-0.038, $SE=0.083$, $t=-0.45$, $p=.65$). Table 4.8 shows the mean agreement scores, depending on whether participants had low or high opinion strength on the topic before reading.

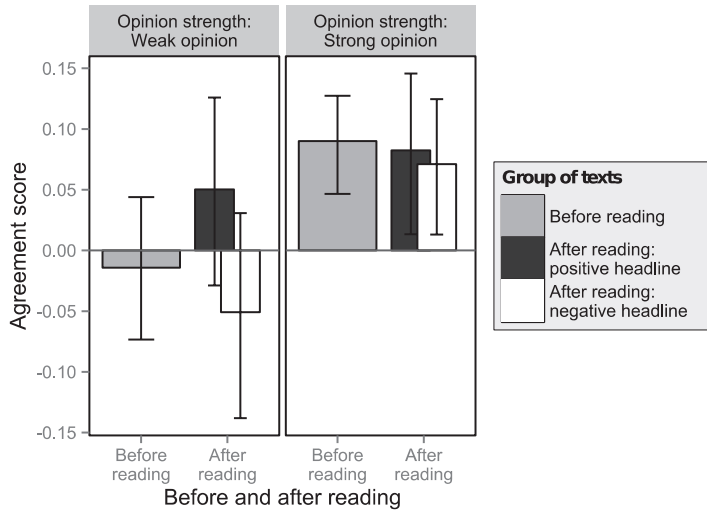


Figure 4.6. Mean agreement scores before and after reading, when grouping trials depending on low or high opinion strength, Experiment 2
Error bars denote standard errors of the mean.

Table 4.8. Mean agreement scores after reading depending on type of headline, when grouping trials depending on low or high opinion strength, Experiment 2
Standard deviations in parentheses.

Headline	Group with positive agreement scores before reading ($M=0.370$, $SD=0.282$)	Group with negative agreement scores before reading ($M=-0.349$, $SD=0.256$)
Positive headline	0.063 (0.421)	0.082 (0.464)
Negative headline	-0.052 (0.460)	0.068 (0.414)

Figure 4.7 shows the agreement scores before and after reading, depending on whether participants were positive or negative to the topic before reading. The mean agreement score before reading for all participants was 0.051 ($SD=0.448$). After participants had read the texts, agreement scores approached the mean for all participants for both the positive opinion group and the negative opinion group. A significant decrease was found for the positive opinion group in both the positive (Estimate=-0.097, $SE=0.024$, $t=-3.99$, $p<.001$) and the negative headline conditions (Estimate=-0.123, $SE=0.025$, $t=-4.85$, $p<.001$). A significant increase was found for the negative opinion group in both the positive (Estimate=0.124, $SE=0.030$, $t=4.16$, $p<.001$) and the negative headline condition (Estimate=0.153, $SE=0.028$, $t=5.43$, $p<.001$). Table 4.9 shows the mean agreement scores, depending on whether participants were positive or negative to the topic before reading.

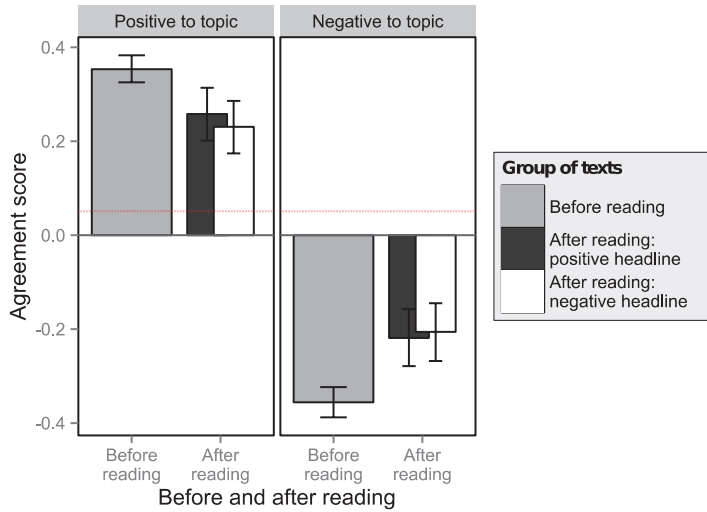


Figure 4.7. Mean agreement scores before and after reading, when grouping trials depending on positive or negative agreement score before reading, Experiment 2

Mean agreement score before reading for all participants is marked with a dotted line in the figure. Error bars denote standard errors of the mean.

Table 4.9. Mean agreement scores after reading depending on type of headline, when grouping trials depending on positive or negative agreement score before reading, Experiment 2

Standard deviations in parentheses.

Headline	Group with positive agreement scores before reading ($M=0.353$, $SD=0.271$)	Group with negative agreement scores before reading ($M=-0.356$, $SD=0.276$)
Positive headline	0.258 (0.408)	-0.219 (0.339)
Negative headline	0.231 (0.370)	-0.206 (0.384)

4.6.3 Discussion

The affective priming hypothesis states that media information presented in a biased way produces evaluations in line with the bias (Kiousis, 2005; Sheafer, 2007), and that information framed in a biased way can affect opinions (Levin & Gaeth, 1988; Levin et al., 2002, 1998; Pan & Kosicki, 1993; Van Gorp, 2007). In this study, the mere expectation of affective priming was shown to affect opinions, but only for participants who stated their opinions to be weak. All texts shown to the participants were descriptive, and highlighted both positive and negative sides of the topic in a balanced way, with no manipulation made to the actual texts. The participants expected the texts to be biased due to reading headlines that contained either positively or negatively biased descriptions of the topics. By introducing the mere expectation of bias using headlines, an effect on opinions was found, but only for participants stating they had weak opinions on the topic. For these participants, opinions became significantly more positive after reading texts with positive

headlines, and more negative after reading texts with negative headlines (albeit not significantly). It has been argued that affect regarding sociopolitical concepts is automatic (Bargh, 1994; Lodge & Stroh, 1993; J. P. Morris et al., 2003), but also that people with strong opinions are more strongly affected when the priming is incongruent with the currently held opinion (Lodge & Taber, 2005). It is possible that this is a result of myside bias and cognitive dissonance, which state that people are generally biased toward information that agrees with them and avoid information that does not (Bastardi et al., 2011; Bem, 1967; J. K. Clark et al., 2012; Festinger, 1962; Harmon-Jones & Mills, 1999; Hergovich et al., 2010; Jonas et al., 2001; Kardash & Scholes, 1996; Klayman, 1995; Mercier, 2017; Nickerson, 1998; Stanovich et al., 2013; Taber & Lodge, 2006; Yaniv, 2004). A more straightforward reason, however, could be the fact that the participants had what they claimed to be strong opinions. They were therefore much less likely to change their opinions. In this case, the expectation of a biased text was not sufficient to prompt myside bias and cognitive dissonance for participants with strong opinions.

The results for participants with weak opinions could have large implications for how we see news media and media communication. Research has shown that confirmation bias and myside bias can be induced by expectations (Marsh & Hanlon, 2007; Nickerson, 1998), but to the best of my knowledge, affective priming and resulting opinion shifts have not been induced by expectations alone. Headlines raise interest in news media articles to entice people to read the rest of the text (Blom & Hansen, 2015; Bucaria, 2004; Molek-Kozakowska, 2013; Schaffer, 1995), and do not necessarily correctly reflect what the articles are actually about (Chen et al., 2015; Kronrod & Engel, 2001; Marquez, 1980; Perfetti et al., 1987); this is similar to how the headlines were used in this study. Importantly, headlines have been argued to trigger expectations and schemas that are held during subsequent reading (Bransford & Johnson, 1972; Dor, 2003; Ifantidou, 2009; E. E. Smith & Swinney, 1992; Wiley & Rayner, 2000), further suggesting that affective priming and resulting opinion shifts could be induced by headlines alone. To simulate the effect of click-baiting in this study, which is the use of enticing headlines in online news media, the headlines were shown at the top of the page without the text for two seconds. This is similar to how a link on an online news page is seen in isolation, clicked, and how the reader is then taken to the text after a short delay.

As already mentioned, studies investigating myside bias and cognitive dissonance suggest that people are actively biased toward information that agrees with their opinions, and that information not corresponding to the current belief is largely ignored. This would result in resistance to opinion shifts for headlines that were both congruent and incongruent with the currently held opinions. This was not observed, however, since all mean results moved closer to zero. People's opinions have been argued to be fairly unstable (Chong & Druckman, 2007; Converse, 1962, 2006; Zaller, 1992), but the results were interpreted as regressions to the mean, masking the effects of the headlines (Barnett et al., 2005; Bland & Altman, 1994). In order to

examine these results accurately, and determine if the expectation of a biased text would induce myside bias and cognitive dissonance, the regression to the mean needed to be removed from the results. This was done in the next analysis.

4.7 Joint analysis of results from Experiments 1 and 2

When the participants were divided depending on having positive or negative agreement scores before reading, the mean agreement scores after reading were significantly closer to zero. It was suspected that the significant results in Experiment 1 were regressions to the mean, which prompted the stronger manipulation in Experiment 2. The results in Experiment 2, however, showed the same patterns. This prompted another analysis of these results, in which the regressions to the mean were removed in order to isolate the effects from the manipulations.

4.7.1 Method

4.7.1.1 *Participants*

The data from the participants from both experiments were used, for a total of 136 native speakers of English. Experiment 1: 72 participants (51 female; 18 to 32 years of age; $M=20.3$ years, $SD=2.57$), Experiment 2: 64 participants (38 female; 18 to 41 years of age; $M=27.6$ years, $SD=5.92$). For full descriptions, see Section 4.5.1.1 and Section 4.6.1.1.

4.7.1.2 *Apparatus*

See Section 4.5.1.2 and Section 4.6.1.2.

4.7.1.3 *Design and stimuli*

See Section 4.5.1.3 and Section 4.6.1.3.

4.7.1.4 *Procedure*

See Section 4.5.1.4 and Section 4.6.1.4.

4.7.1.5 *Data analysis*

Transformations were performed on the results in order to remove the regression to the mean and show only the effects of resistance to opinion shift. In this section, the term *pre-reading* is used for values before reading, and the term *post-reading* is used for values after reading. The agreement scores were transformed independently for the two groups. In both transformations, pre-reading agreement scores were transformed to zero by subtracting the mean agreement score from each agreement score value.

This was performed separately for both groups, since the means were different (positive opinions: $M=0.371$, $SD=0.295$; negative opinions: $M=-0.361$, $SD=0.272$). The transformation procedure for the post-reading agreement scores is described in Figure 4.8. All post-reading agreement scores were first multiplied by the quotient of the mean pre-reading agreement score divided by the mean agreement score after reading neutral texts. The mean pre-reading agreement score was then subtracted, resulting in agreement scores without the regression to the mean found between the pre-reading agreement scores and the post-reading agreement scores, using the neutral texts with neutral headlines in Experiment 1 as a baseline. Transformed pre-reading agreement scores were used as baselines in the statistical models. This resulted in the post-reading *retention of agreement score*. The data processing and figure were made using R (3.2.2) (R Development Core Team, 2015) in R Studio (0.99.473).

An alternative transformation procedure is to subtract the mean post-reading agreement score from the control condition from the other three conditions after reading. This procedure was found after mathematically reducing a longer formula. The transformation described in Figure 4.8 was used, since it also used mean pre-reading agreement scores. It therefore compensates better for the differences between the two measurements, and consequently for the post-reading regression to the mean. When performing the alternative transformation, however, only the scale of the results changed, with the distributions being identical. The figures describing the alternative transformation procedure, the results, as well as the full statistical model output of the results when using the alternative transformation, can be found in Appendix 4.A.

$$\text{Agreement score} \times \frac{\text{Mean postreading agreement score}}{\text{Mean postreading agreement score for noncontroversial texts}} - \text{Mean prereading agreement score}$$

Figure 4.8. Mathematical formula for transformation of agreement scores after reading (post-reading agreement scores), to remove regressions to the mean

Agreement score in **bold** signifies every individual agreement score; all other values are constant. Pre-reading and post-reading are referred to as prereading and postreading, in order to avoid confusion with minus signs.

4.7.2 Results

The statistical model did not use log-transformed data, as the data contained a large number of zero values. However, the distribution was Gaussian, and thus suitable for statistical analysis. Figure 4.9 and Table 4.10 show the mean retention of agreement scores after reading depending on topic and headline for Experiments 1 and 2 combined, when grouping trials depending on positive or negative agreement scores before reading. Overall, a significant change was found for the controversial texts (Estimate=0.074, $SE=0.035$, $t=2.14$, $p<.03$). Note that this result only trended towards significance in the analysis using the alternative transformation method

(Estimate=0.046, $SE=0.024$, $t=1.91$, $p=.06$), albeit being very close to significance. The full statistical model output when using the alternative method for transformation can be found in Appendix 4.A. No significant changes were found for texts with positive (Estimate=0.040, $SE=0.039$, $t=1.04$, $p=.30$) or negative headlines (Estimate=0.021, $SE=0.041$, $t=0.51$, $p=.61$). The group of texts with negative agreement scores before reading was not significantly different to the group with positive agreement scores before reading (Estimate=-0.002, $SE=0.025$, $t=-0.09$, $p=.99$). An interaction was found between retention of agreement scores after reading and positive or negative agreement scores before reading for controversial texts (Estimate=-0.119, $SE=0.053$, $t=-2.23$, $p<.03$), with a significant retention of agreement scores after reading for the group of trials with negative agreement scores before reading. The corresponding retention of agreement scores was also seen for the group of trials with positive agreement scores before reading. No other significant interactions were found (all $ps>.46$). The interaction for the controversial texts was also significant when using the alternative transformation method described in Appendix 4.A (Estimate=-0.074, $SE=0.037$, $t=-2.01$, $p<.04$). The full statistical model output with all interactions can be found in Appendix 4.B.

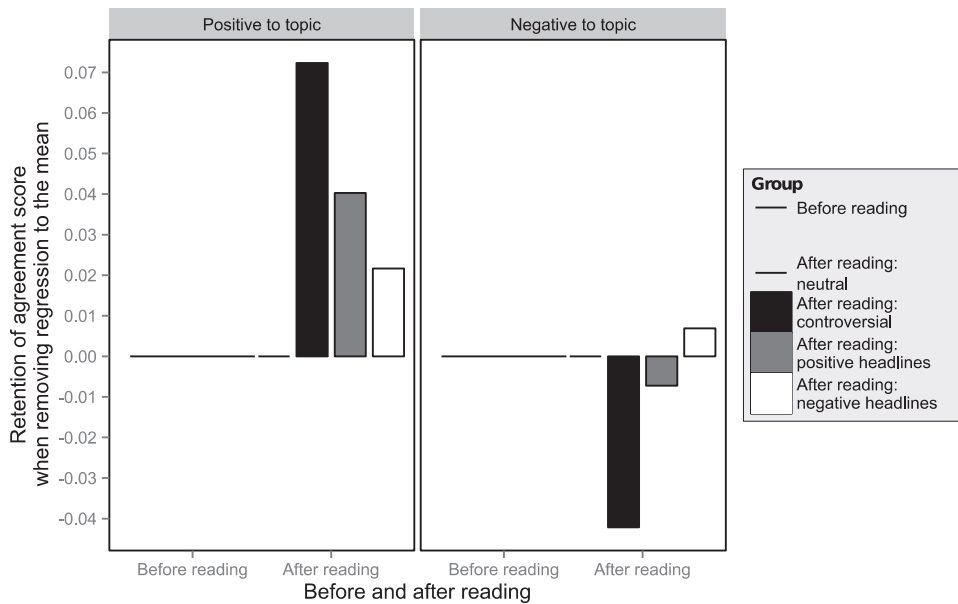


Figure 4.9. Mean retention of agreement scores after reading depending on topic and headline, depending on trials with positive or negative agreement score before reading, Experiment 1 and 2 combined
The agreement scores after reading the neutral texts in Experiment 1 were used as the baseline (see Figure 4.8), which results in the zero values in the figure. A larger value in the figure signify a higher retention of agreement scores.

Table 4.10. Mean retention of agreement scores after reading depending on topic and headline, when grouping trials depending on positive or negative agreement score before reading, Experiment 1 and 2 combined

Standard deviations in parentheses.

Topic, Headline	Group with positive agreement scores before reading	Group with negative agreement scores before reading
Controversial topic	0.0723 (0.626)	-0.0422 (0.653)
Positive headline	0.0403 (0.665)	-0.0072 (0.552)
Negative headline	0.0217 (0.601)	0.0069 (0.634)

4.7.3 Discussion

The results from Experiment 1 and Experiment 2 were combined, and two groups of trials were made, one with positive agreement scores before reading, and one with negative agreement scores before reading. This signified that the participant was either positive or negative to the topic before reading the text. The regression to the mean was then removed separately for the two groups of trials, using the neutral texts with neutral headlines as a baseline. Importantly, two different methods were used—one that multiplied every acceptability score after reading with a baseline value, and one that subtracted a baseline value from every acceptability score. Both methods produced identical distributions of the results, with the results obtained from the second method being slightly smaller overall. The first method is described in Figure 4.8 and the results can be found in 4.7.2. The alternative method is described in Appendix 4.A, which also contains the results. Both methods produced results that showed the retention of agreement scores after the regressions to the mean had been removed. In other words, when results were positive for the group of positive trials, or negative for the group of negative trials, this showed how much of the positive or negative opinion was retained.

Higher retention of opinions was found when topics were introduced as being controversial, compared to when they were introduced as neutral, both for participants who had positive opinions on the topic before reading, and for participants who had negative opinions. When hearing that the topics of the upcoming texts were considered controversial and that others feel strongly about them, people seemed to value their own opinions much more than when reading texts on neutral topics. They were consequently much less likely to change their own opinions. This provides another reason why people are more likely to disagree on controversial issues, as shown by Oulton et al. (2004, p. 2) and Stradling (1985, p. 9), and why controversial issues seem to produce more myside bias, as shown by Taber and Lodge (2006).

Surprisingly, no significant retention of opinions was seen when the headlines were positive or negative. This suggests that the expectation of a text with evaluative bias is not sufficient to produce myside bias and cognitive dissonance, and that the texts themselves need evaluative bias in order for this to occur. It is also possible that

expectations did produce myside bias and cognitive dissonance, but not to a sufficient degree to produce effects of resistance to opinion shifts in this study.

4.8 General discussion

Two experiments investigated how expectations about the topic and the specific bias of the text affect opinions after reading.

4.8.1 Opinions regarding the topic: Strong or weak

Previous research has shown that expectations can affect the outcomes of such things as experimental results and medical treatment (Doyen et al., 2012; Stewart-Williams & Podd, 2004). In this study, mere expectations affected people's opinions after reading a news text, using predictions based in affective priming and the effects of affective compelling arguments. In Experiment 1, it was hypothesized that people's personal opinions would become stronger after hearing that the topic was controversial, since controversial topics divide people and are argued about extensively (Oulton et al., 2004, p. 2; Stradling, 1985, p. 9). Therefore, this should indicate to them that their opinion was important, and strengthen their opinion on the topic. This would result in resistance to opinion shifts in all cases but one, namely for the participants with weak opinion strength reading neutral topics. This was not the case, since slight decreases in agreement scores and statistical null results were observed in all cases. Instead of interpreting these results further, another experiment with a stronger manipulation was conducted.

All sociopolitical concepts are argued to become affectively charged when read, something that has been shown to happen very quickly, and can affect opinions on both the current and related issues (Abelson, 1963; Bargh, 1994; Lodge & Stroh, 1993; Lodge & Taber, 2005; J. P. Morris et al., 2003). This is referred to as affective priming, and the media can influence this by using affective compelling arguments when writing about an issue (Kioussis, 2005; Sheaffer, 2007). Experiment 2 in this study investigated if affective priming could be influenced by people *expecting* that affective compelling arguments would be used in a news text. This was achieved by combining neutral texts on various news topics with headlines that appeared to be biased summaries of the topic, such as "the good effects of divorce" or "the bad effects of divorce" combined with a text that described the effects of divorce on men and women in a balanced and neutral way. Opinion change in the direction of the headline was observed, but only for participants stating they had weak opinions on a topic before reading. When comparing opinions before and after reading the texts, their opinions became more positive when reading texts with positive headlines, and more negative when reading texts with negative headlines.

The main purpose of a headline is to summarize the article and to provide context, but media has a long history of using headlines to raise interest in the article. With the online news media, however, this practice is now bigger than ever with the use of click-baiting headlines. Here, the main purpose is to entice the reader to click the link, thereby generating more internet traffic to the page. These headlines might have very little to do with the article, or might even be misleading. This practice has been questioned, but only in terms of whether it is ethical to “trick” users to click on articles they might not be interested in. That the text could be read differently when introduced with a bias is generally not a concern for publishers, and the fact that people may actually change their opinions after reading a text solely based on the headline is an issue that is not discussed at all.

Importantly, in Experiment 2, all stimuli were presented using only the headlines for the first two seconds of presentation, after which the rest of the text was also shown. This ensured that participants paid attention to the headline, but also emulated the process of browsing online media. We see clickable headlines, and if one is interesting enough, we click it and see the rest of the text after a brief delay when the new page loads. The results from this study have implications for the click-baiting practices that the online news media commonly employ, in which the headline or link is often primarily designed to elevate the number of clicks and not to accurately portray the content of an article. It seems that people with weak opinions will perceive the text differently based on what the headline says, and even change their opinions as a result.

4.8.2 Opinions regarding the topic: Positive or negative

When trials were divided depending on whether participants had positive or negative prior opinions about a topic, all agreement scores regressed toward the mean in both experiments; that is, they became less extreme. The joint analysis of the results from both Experiment 1 and Experiment 2 that removed the regression to the mean, however, revealed an interesting result. For the texts introduced as describing controversial topics that “other people feel strongly about,” more resistance to opinion shift was found compared to reading the texts introduced as being on various topics. In Experiment 1, it was hypothesized that the introduction would make the participant feel more strongly about his or her own opinion, and therefore be more likely to resist any opinion shifts. This would result in smaller regressions to the mean, which was seen when combining the results from both experiments and removing the effects of regressions to the mean. It appears that the mere expectation of an issue that is described as controversial makes people more likely to retain their opinions on the matter.

A resistance to opinion shift was also found for the texts with positive and negative headlines, though none of the results were significant. This seems to confirm the conclusion in Experiment 2 about the participants with strong opinions, and

indicates that this conclusion extends to all participants. The mere expectation of a biased text, prompted by positive or negative headlines, seemed to prompt only very little myside bias and cognitive dissonance. Instead, as Section 4.8.1 shows, strong opinions create a high amount of resistance to opinion shift.

Table 4.11, Table 4.12, Table 4.13, and Table 4.14 show a summary of the behavior and measured effects, and how they correlate with the hypotheses and hypothesized effects presented in Section 4.3.2.

Table 4.11. Behavior and measured effects on the experimental measures depending on opinion strength, Experiment 1

Hypotheses presented in Section 4.3.2 supported by the data in **bold**. All behavior and effects in the table are supported by the data.

Topic	Participant opinion strength on the topic	
	Weak opinion	Strong opinion
Framed as a controversial topic	<i>Behavior:</i> The participants' opinions become slightly more negative.	<i>Behavior:</i> The participants' opinions become slightly more negative.
	<i>Measured effects:</i> Small, non-significant decrease in agreement scores.	<i>Measured effects:</i> Small, non-significant decrease in agreement scores.
Framed as a neutral topic	<i>Behavior:</i> The participants' opinions become slightly more negative.	<i>Behavior:</i> The participants' opinions become slightly more negative.
	<i>Measured effects:</i> Small, non-significant decrease in agreement scores.	<i>Measured effects:</i> Small, non-significant decrease in agreement scores.

Table 4.12. Behavior and measured effects on the experimental measures depending on positive or negative agreement score before reading, Experiment 1

The joint analysis of the data from both Experiment 1 and Experiment 2 is also used. Hypotheses presented in Section 4.3.2 supported by the data in **bold**. All behavior and effects in the table are supported by the data.

Topic	Participant agreement score on the topic before reading	
	Positive to topic	Negative to topic
Framed as a controversial topic	<i>Behavior:</i> The participants retain their currently held opinions because they hear that others feel strongly about the topic, causing them to feel more strongly about it themselves.	<i>Behavior:</i> The participants retain their currently held opinions because they hear that others feel strongly about the topic, causing them to feel more strongly about it themselves.
	<i>Measured effects:</i> Resistance to opinion shift.	<i>Measured effects:</i> Resistance to opinion shift.
Framed as a neutral topic	<i>Behavior:</i> Agreement scores exhibit the regular behavior of a subsequent measurement, when one group is very different.	<i>Behavior:</i> Agreement scores exhibit the regular behavior of a subsequent measurement, when one group is very different.
	<i>Measured effects:</i> Opinion shift, seen in regressions to the mean.	<i>Measured effects:</i> Opinion shift, seen in regressions to the mean.

Table 4.13. Behavior and measured effects on the experimental measures depending on opinion strength, Experiment 2

Hypotheses presented in Section 4.3.2 supported by the data in **bold**. All behavior and effects in the table are supported by the data.

Headline	Participant opinion strength on the topic	
	Weak opinion	Strong opinion
Positive headline	<i>Behavior:</i> Since the participants do not feel strongly about the topic, the expectation-induced affective priming from the headline influences the reading of neutral text.	<i>Behavior:</i> Since the participants feel strongly about the topic, they do not change their opinions.
	<i>Measured effects:</i> Opinion shift in line with bias, seen in higher agreement scores.	<i>Measured effects:</i> Resistance to opinion shift.
Negative headline	<i>Behavior:</i> Since the participants do not feel strongly about the topic, the expectation-induced affective priming from the headline influences the reading of neutral text.	<i>Behavior:</i> Since the participants feel strongly about the topic, they do not change their opinions.
	<i>Measured effects:</i> Opinion shift in line with bias, seen in lower agreement scores.	<i>Measured effects:</i> Resistance to opinion shift.

Table 4.14. Behavior and measured effects on the experimental measures depending on positive or negative agreement score before reading, Experiment 2

The joint analysis of the data from both Experiment 1 and Experiment 2 is also used. Hypotheses presented in Section 4.3.2 supported by the data in **bold**. All behavior and effects in the table are supported by the data.

Headline	Participant agreement score on the topic before reading	
	Positive to topic	Negative to topic
Positive headline	<i>Behavior:</i> The participants are not sufficiently influenced by myside bias to maintain their opinions. Agreement scores therefore exhibit the regular behavior of a subsequent measurement, when one group is very different.	<i>Behavior:</i> The participants are not sufficiently influenced by myside bias and cognitive dissonance to maintain their opinions. Agreement scores therefore exhibit the regular behavior of a subsequent measurement, when one group is very different.
	<i>Measured effects:</i> Regression to the mean.	<i>Measured effects:</i> Regression to the mean.
Negative headline	<i>Behavior:</i> The participants are not sufficiently influenced by myside bias and cognitive dissonance to maintain their opinions. Agreement scores therefore exhibit the regular behavior of a subsequent measurement, when one group is very different.	<i>Behavior:</i> The participants are not sufficiently influenced by myside bias to maintain their opinions. Agreement scores therefore exhibit the regular behavior of a subsequent measurement, when one group is very different.
	<i>Measured effects:</i> Regression to the mean.	<i>Measured effects:</i> Regression to the mean.

4.9 Conclusions

This chapter shows that people retain their opinions after reading about topics presented as controversial, but not after reading topics presented as neutral. It is possible that controversial issues create much more myside bias than neutral issues. This would explain why controversial issues, as shown in previous research, so often create opinions that people stubbornly stick to, in spite of evidence to the contrary, which is the very definition of myside bias. Also, the results suggest that the mere expectation of reading about a controversial topic creates enough myside bias for people to refuse to change their opinions.

Media presented in a biased way—that is, media with affective priming—is shown to affect opinions and produce evaluations in line with the bias. This chapter shows that mere expectations of affective priming make people change their opinions regarding a topic in line with the bias after reading a text that was in fact neutral, but only if they do not feel strongly about the topic to begin with. It seems that the expectation of evaluative bias makes people read the text as if it were *actually* biased, and produces the associated effects. If people felt strongly about the topic, however, the texts had no effect on their opinions. It appears that feeling strongly about a topic serves as a defense against affective priming, and people refuse to change their opinions. Since myside bias was not found when grouping participants according to either positive or negative prior opinions, it appears that myside bias only arises from the combination of expectations of affective priming and strong opinions on the topic.

Substantial changes in opinions after reading were found in this chapter. Chapter 5 investigates whether effects are present in eye movements during reading as well.

Chapter 5

You change your reading (or not) when expecting controversial texts or texts with evaluative bias

5.1 Introduction

The previous chapter investigated opinion shifts depending on other people's opinions, and on positive and negative headlines accompanying the neutral texts. Effects on opinions were found. It is therefore plausible to think that this should be evident in eye movements during reading. This chapter therefore investigates how reading is affected by the expectation that other people have strong opinions about the topics, and the expectation that the texts will either agree or disagree with the participant's currently held opinion.

We read texts of many kinds in our daily lives, but we cannot engage with all texts we encounter in the same way, and at the same level of thoroughness and open-mindedness. We know that influence from peers can affect our opinions (Gaertner & Dovidio, 2012; Hewstone, Rubin, & Willis, 2002) and the decisions we make (Cialdini & Goldstein, 2004). While we prefer information from our peers, it seems that we can be influenced in a similar way by anyone (Guazzini, Yoneki, & Gronchi, 2014). It therefore seems reasonable to assume that the opinions of other people affect the way people read neutral texts. As mentioned in Section 4.2.3, *myside bias* shows that people tend to find support for their own opinions (Mercier, 2017; Stanovich et al., 2013). *Myside bias* has been argued to be a desirable cognitive feature (Mercier, 2017) and expectations alone seem sufficient to change our evaluation of information (Marsh & Hanlon, 2007; Nickerson, 1998). Previous research suggests that text is processed differently when it agrees or disagrees with the reader's opinions. But could the mere expectations of a text that either agrees or disagrees with our opinions affect the way the text is read, even if the text is neutral? In addition, how strongly we feel about a topic is likely to affect our reading. But, how does this interact with our expectations and opinions? Finally, with regard to personality, it can be hypothesized that people who are more open to other opinions

are by definition more inclined to evaluate the validity of arguments that conflict with their own, and should therefore read texts more thoroughly. Another possibility is that higher openness corresponds to lower *epistemic vigilance*, which is argued to be the resistance and vigilance against misinformation from others (Sperber et al., 2010). If this is the case, high openness would result in less cross-examination of the text, and thus less thorough reading. But how does personality interact with people's preference for information that agrees with them, and their bias against information that disagrees with them?

In order to investigate reading depending on other people's opinions, and on either agreeing or disagreeing with the text, two eye-tracking experiments were conducted. The specific opinions of the participants and the strength of their opinions were measured before reading. The results were analyzed in two ways in each experiment. In one analysis, trials were grouped in relation to whether people had positive or negative opinions on each topic before reading the texts. The other analysis grouped trials depending on if the participants stated they had a weak or strong opinion on each topic before reading. In Experiment 1, all headlines were neutral. Participants' expectations were instead changed by giving different descriptions of the topics, which were described as either "various topics," or topics that "people say they feel very strongly about" and were "considered highly controversial according to survey data." This description was meant to increase how important the participants expected the topic to be, and thereby increase how important they considered their own opinion. In Experiment 2, participants' expectations were manipulated by using different headlines that were either positively or negatively biased toward the topic. This changed the participants' expectations regarding if a positively or negatively biased text would be read, and consequently whether the text agreed or disagreed with their own opinions. The participants' openness to other opinions was also investigated, since these factors should affect reading, when considering whether they agree with the expected bias or not.

5.2 Previous research

5.2.1 Expectations and headlines, and the influence of others

A study measuring sentence reading times with key presses found faster reading when the participant's expectations matched the causal relations in the text (Haberlandt, 1982). This demonstrates facilitation when expectations are met. Similarly, causal expectations were shown to facilitate processing of causally-related sentences in an eye-tracking study (Mak & Sanders, 2012). Headlines, however, can also be used to create expectations. Within the framework of relevance theory (Sperber & Wilson, 1995b), Dor (2003) has argued that headlines optimize the relevance of the text, and

help readers extract the maximal amount of context with a minimal amount of processing. Ifantidou (2009) has suggested that headlines create contexts in order to optimize processing effort. Appropriate headlines or titles have been shown to activate the appropriate schema for the text, leading to several changes in reading behavior, such as less rereading, quicker processing of ambiguous words, and shorter overall reading times per sentence (E. E. Smith & Swinney, 1992; Wiley & Rayner, 2000), as well as increased comprehension ratings and recall scores (Bransford & Johnson, 1972). Much research has investigated how the news media use headlines to draw attention to the articles (Bucaria, 2004; Molek-Kozakowska, 2013, 2014; Schaffer, 1995), but little research has investigated how the attention-grabbing headlines in news media might affect the reading of the corresponding articles.

Social influence has a large effect on decision making and opinions, showing that we are vulnerable to peer influence, and thus affected by what others think (Cialdini & Goldstein, 2004). *In-group bias* (also called *intergroup bias* and *in-out group bias*) shows that we are more biased toward the opinions of our peers than the opinions of people in groups that are unrelated to us (Gaertner & Dovidio, 2012; Hewstone et al., 2002). An eye-tracking study investigated how the amount of social information available to participants affected decision making in a picture judgment task (Guazzini et al., 2014). Participants were given either no information, information about the average responses of previous participants, or information about the response of a “friend.” Judgments were unaffected by what type of information was given, but eye movements differed in terms of the duration of picture inspection and number of looks between the “no information” and the two other information conditions. Eye movements were very similar, however, for both the in-group (friend) and out-group (previous participants). This suggests that people are influenced similarly when responses come from a peer and from an unknown, previous participant. But the way people process information is also affected by cognitive biases, as described in the next section.

5.2.2 Beliefs and biases

Confirmation bias, and the more recent and refined theory of myside bias, describe the tendency for people to be actively biased toward information that agrees with their opinions and to ignore information that disagrees with their opinions (Hergovich et al., 2010; Jonas et al., 2001; Klayman, 1995; Mercier, 2017; Nickerson, 1998; Stanovich et al., 2013). Myside bias, which proposes similar predictions but is a more refined and well-researched theory than confirmation bias, is used in this study and this entire book. For a thorough explanation of confirmation bias and myside bias, see Section 4.2.3.

Several studies have examined how information that is consistent or inconsistent with current beliefs is processed. By using questionnaires measuring beliefs before and after reading, it was found that beliefs were generally weaker after reading, but that

some participants' beliefs were strengthened, while other people's beliefs were weakened (McCrudden & Sparks, 2014). Two reader profiles were identified based on interviews: *belief-protection* (defending the current opinion and rejecting inconsistent arguments), and *belief-reflection* (questioning the current opinion and incorporating inconsistent arguments). An eye-tracking study showed how the strength of prior beliefs, as well as whether these beliefs were consistent or inconsistent with the presented information, affected processing and evaluation on the internet (van Strien, Kammerer, Brand-Gruwel, & Boshuizen, 2016). Interestingly, no significant differences in reading were found between texts that agreed or disagreed with opinions for participants with strong feelings about a topic. This suggests that high and low strengths of opinion affect reading in different ways. A self-paced sentence reading study found that arguments were read more slowly in texts inconsistent with beliefs regarding evolution, and summary sentences were read more quickly in texts inconsistent with beliefs regarding TV violence (Wolfe, Tanner, & Taylor, 2013). The study also used titles that were consistent with the text, such as, "TV Violence Causes Real Violence."

Cognitive dissonance has been shown to arise when people are confronted with new information that conflicts with their current opinions or knowledge (Stone & Cooper, 2001). People actively try to reduce this dissonance by reconciling the conflicting information to create consonance. Several theories have been proposed about the underlying reasons for cognitive dissonance, with the *self-standards model* reconciling three different perspectives and suggesting that self-confidence plays a large role when mediating dissonance (Harmon-Jones & Mills, 1999; Stone & Cooper, 2001). Personality affects how cognitive dissonance is mediated, and it is possible that this can be seen in the reading processes.

5.2.3 Personality and openness

Myside bias, presented in Section 4.2.3 and briefly in Section 5.2.2, is argued to be correlated with rational thinking (Mercier, 2017). Stanovich et al. (2013) propose a model to evaluate rational thinking, with dimensions encompassing aspects such as "resistance to miserly processing," "absence of irrelevant context effects in decision making," and "open-minded thinking." This would therefore suggest a connection between myside bias, rational thinking, and the *openness* personality trait, which is also referred to as *intellect or openness* and *openness to experience* (DeYoung, Quilty, & Peterson, 2007; Thompson, 2008). Eye movement behavior seems to be consistent with this (Stanovich et al., 2013). The term openness will be used in this book.

When examining personality, the *big-five trait taxonomy* (also called the *five-factor model*) is by far the most commonly used (Costa, 1991; John & Srivastava, 1999). Five distinct personality traits are examined, including the previously mentioned openness personality trait. The remaining four traits are *agreeableness*, *extraversion*, *conscientiousness*, and *neuroticism*. In a study with university students, Roccas, Sagiv,

Schwartz, and Knafo (2002) showed that openness correlated most positively with self-direction (autonomy of thought, openness to new ideas) as well as with universalism (understanding and tolerance for all people and ideas), agreeableness with benevolence, extraversion with achievement, and conscientiousness with conformity. Openness was specifically investigated by Dollinger, Leong, and Ulicni (1996), who found that openness correlated positively with self-direction values, and with being broadminded and imaginative. It is plausible that the two reader profiles suggested by McCrudden and Sparks (2014)—belief-protection and belief-rejection—are connected with openness.

Several tests have been developed in order to quantify the big-five trait taxonomy, such as the *mini-markers task* (Saucier, 1994). The task consists of 40 personality characteristics (referred to as markers) such as “imaginative” and “sympathetic,” with a participant reporting how well each marker describes their personality. A nine-point scale is used, ranging from “extremely inaccurate” to “extremely accurate.” The responses are divided into five groups, with eight specific markers in each group, and a mean is calculated, producing the big-five personality traits for the participant. Personality traits and learning styles have long been used to evaluate students’ educational preferences and measure their scholastic performance (Dunn, Beaudry, & Klavas, 2002; Schutte & Malouff, 2004). While extraversion was found to affect reading when listening to preferred and non-preferred music (Johansson, Holmqvist, Mossberg, & Lindgren, 2011), however, little research has been performed on how personality affects the reading process and reading comprehension (Sadeghi, Kasim, Tan, & Abdullah, 2012).

5.2.4 Eye-tracking methodology and eye movements during reading

Section 2.3.1 gives a thorough explanation of eye-tracking methodology, and Section 2.3.2 gives a thorough review of the research on eye movements during reading. Please note that Chapter 3 in this book investigates the eye movements during reading of pages of text. Section 2.3.4 gives an explanation of good research practice when creating surveys, along with visual analog scales.

When investigating eye movement behavior on survey responses, participants were found to spend more time on the first few options and less time on the last ones (Galesic, Tourangeau, Couper, & Conrad, 2008). During reading, the ends of paragraphs are particularly important when creating an accurate memory representation of the material (Hyönä et al., 2002).

5.3 The study of myside bias and openness during paragraph reading

In two eye-tracking experiments, I investigated how the reading process is affected by expectations and how strongly a participant feels about a topic. Prior to reading, participants stated how strongly they felt about each topic, as well as their specific opinions on each topic. These responses were used to classify participants into having either a weak or strong opinion on the topic, as well as agreeing or disagreeing with the headlines. No manipulation was performed on the texts in either experiment.

In Experiment 1, neutral texts were combined with neutral and fairly uninformative headlines such as *Vaccination*. The texts were framed by instructing participants that they would either read texts “on various topics,” or texts “on various topics that people say they feel very strongly about, and that are considered highly controversial according to survey data.” These cues created expectations of text topics that other people feel strongly about. Since participants can be influenced by social pressures (Guazzini et al., 2014), this framing was expected to increase how strongly the participants felt about the topics. In Chapter 4, it was indeed found that the participants expecting to read texts about these “controversial” topics retained their opinions to a significantly larger extent compared to the texts about various topics, when their opinions were measured 20 minutes after reading. This retention of opinion was also found for people who had strong opinions on a topic, while people who did not have a strong opinion on a topic changed their opinion in line with the expected bias of the text. In combination, these results suggest that people feel more strongly about a topic when they hear that *other* people feel strongly about a topic, and that the topic is considered controversial.

In Experiment 2, neutral texts were combined with either positively or negatively biased headlines, such as *The good current vaccination practices* or *The bad current vaccination practices*. This created expectations that the texts would either agree or disagree with the participants’ own opinions. The activated schema was, in addition to the general topic of the text, either one of agreement or disagreement with the information expressed in it. Since the texts were in fact neutral, however, the effects of the two different schemas could be compared. In addition, a personality test was used to evaluate openness since it was expected to influence the way participants would read the texts.

5.3.1 Terms and classifications

All stimuli texts were given on individual single screens, divided up into six to eight lines of text. Entire lines of text or individual words were divided by the line width (in pixels) for all analyses in order to compensate for word lengths. This resulted in *total reading times per pixel-width* (hereafter referred to as *total reading times/pw*) and *first-*

pass reading times per pixel-width (hereafter referred to as *first-pass reading times/pw*). In the per-line analyses, texts were divided into three sections consisting of the first two lines, the middle lines, and the last two lines, which were used as factors. When words were used as areas of interest in the reading analyses, the words in the text were numbered from 1-90 and used as variables. How strongly the participant feels about a topic, as specified in the questionnaire response, is referred to as *opinion strength*.

Note that per-character measures are more common in reading studies (e.g., Hyönä et al., 2002), in which monospaced fonts are used. In this study, a non-monospaced font was used to create texts that looked more similar to texts encountered in online and written media. A per-character analysis therefore only gives an approximation of the width of the word or line, while per-pixel-width measures provide exact widths. Analyses using per-character measures were conducted, and showed highly similar results to the per-pixel-width measures. A significant result changed to a non-significant one, or vice versa in two cases in Experiment 2, while no results changed in Experiment 1. First-pass reading times per character on the middle lines as a function of opinion strength changed from $p=.058$ to $p=.047$, and on the last two lines as a function of opinion strength and openness changed from $p=.048$ to $p=.051$. These effects can be considered marginal, and were not addressed in this study. The full statistical model outputs for the per-character analyses can be found in Supplementary Materials 5.B and 5.D (the per-pixel-width analyses can be found in Supplementary Materials 5.A and 5.C). Due to space constraints, these Supplementary Materials are found online rather than in this book.

5.3.2 Hypotheses

In Chapter 3, eye movement characteristics for regular reading and thorough reading showed that the beginnings of texts were read more thoroughly, and the ends of texts were read less thoroughly, as demonstrated in total reading times on words. This was also expected in the current chapter, and deviations from this pattern are analyzed.

To facilitate the task of the reader, the hypotheses and the hypothesized effects are shown in Table 5.1, Table 5.2, and Table 5.3. Differences in eye movement patterns are hypothesized to be most observable at the ends of texts, where conclusions or summaries were usually located.

Table 5.1. Hypotheses and hypothesized effects on the experimental measures, Experiment 1

Topic condition is predicted to impact processing of end-section and overall reading times, and opinion strength is predicted to impact overall reading times. The hypotheses and effects supported by the experimental data are reported in **bold** in Table 5.4 in Section 5.5.3.

Topic	Participant opinion strength on the topic	
	Weak opinion	Strong opinion
Framed as a controversial topic	<i>Hypothesis:</i> The participants expect a controversial and important topic, and therefore read the concluding lines thoroughly. Their overall reading is also more thorough, similar to participants that have strong opinions regarding the topic.	<i>Hypothesis:</i> The participants feel strongly about the topic and are certain about the conclusions they have drawn about it. Therefore, the concluding lines of text are not read thoroughly.
	<i>Hypothesized effects:</i> - End-section read more thoroughly. - Overall reading similar to participants with strong opinions.	<i>Hypothesized effects:</i> End-section read less thoroughly.
Framed as a neutral topic	<i>Hypothesis:</i> The participants do not read the text or the concluding lines thoroughly, because they do not feel strongly about the topic. Their overall reading is also less thorough than participants with strong opinions.	<i>Hypothesis:</i> The participants feel strongly about the topic and are certain about the conclusions they have drawn about it. Therefore, the concluding lines of text are not read thoroughly.
	<i>Hypothesized effects:</i> - End-section read less thoroughly. - Overall reading not as thorough as participants with strong opinions.	<i>Hypothesized effects:</i> End-section read less thoroughly.
Overall reading patterns	After the initial lines, total reading times on lines will decrease and first-pass reading times on lines will stay similar during all reading. Total reading times/pw on words will decrease as a function of word number. First-pass reading times/pw will remain similar as a function of word number.	

Table 5.2. Hypotheses and hypothesized effects on the experimental measures, Experiment 2

Bias condition is predicted to impact processing of end-section and overall reading times, and opinion strength is predicted to impact overall reading times. The hypotheses and effects supported by the experimental data are reported in **bold** in Table 5.5 in Section 5.6.3.

Bias	Participant opinion strength on the topic	
	Weak opinion	Strong opinion
Headline agrees with participants' opinions	<i>Hypothesis:</i> The participants expect their opinions to agree with the text. This makes them read the concluding lines of text more thoroughly since they are biased toward information that agrees with them, even though they do not feel strongly about the topic.	<i>Hypothesis:</i> The participants expect their opinions to agree with the text. This makes them read the concluding lines of text more thoroughly since they are biased toward information that agrees with them.
	<i>Hypothesized effects:</i> - End-section read more thoroughly. - Overall reading is less thorough than participants with strong opinions.	<i>Hypothesized effects:</i> - End-section read more thoroughly. - Overall reading is more thorough compared to participants with weak opinions.
Headline disagrees with participants' opinions	<i>Hypothesis:</i> The participants expect their opinions to disagree with the text. This makes them read the concluding lines of text less thoroughly since they are biased against information that disagrees with them.	<i>Hypothesis:</i> The participants expect their opinions to disagree with the text. This makes them read the concluding lines of text less thoroughly since they are biased against information that disagrees with them.
	<i>Hypothesized effects:</i> - End-section read less thoroughly. - Overall reading is less thorough than participants with strong opinions.	<i>Hypothesized effects:</i> - End-section read less thoroughly. - Overall reading is more thorough compared to participants with weak opinions.
Overall reading patterns	After the initial lines, total reading times on lines will decrease and first-pass reading times on lines will remain similar during all reading. Total reading times/pw on words will decrease as a function of word number. First-pass reading times/pw will stay similar as a function of word number.	

Table 5.3. Hypotheses and hypothesized effects on the experimental measures when accounting for openness, Experiment 2

Openness is predicted to impact processing of end-sections. The hypotheses and effects supported by the experimental data are reported in **bold** in Table 5.6 in Section 5.6.3.

Bias	Participant opinion strength on the topic			
	Weak opinion	Strong opinion	Weak opinion	Strong opinion
Headline agrees with participants' opinions	<i>Hypothesis:</i> The participants do not care about other possible conclusions.	<i>Hypothesis:</i> The participants do not care about other possible conclusions.	<i>Hypothesis:</i> The participants read the conclusions thoroughly to confirm their opinions.	<i>Hypothesis:</i> The participants read the conclusions thoroughly to confirm their opinions.
	<i>Hypothesized effects:</i> • End-section read less thoroughly.	<i>Hypothesized effects:</i> • End-section read less thoroughly.	<i>Hypothesized effects:</i> • End-section read more thoroughly.	<i>Hypothesized effects:</i> • End-section read more thoroughly.
Headline disagrees with participants' opinions	<i>Hypothesis:</i> The participants do not care about other possible conclusions.	<i>Hypothesis:</i> The participants do not care about other possible conclusions.	<i>Hypothesis:</i> The participants read the conclusions thoroughly to see how their own opinions are wrong.	<i>Hypothesis:</i> The participants read the conclusions thoroughly to see how their own opinions are wrong.
	<i>Hypothesized effects:</i> • End-section read less thoroughly.	<i>Hypothesized effects:</i> • End-section read less thoroughly.	<i>Hypothesized effects:</i> • End-section read more thoroughly.	<i>Hypothesized effects:</i> • End-section read more thoroughly.
Low openness		High openness		
Participant openness				

5.4 Norming data

The two groups of texts were identical to the grouping in Chapter 4. The norming data used when creating the two groups can be found in Section 4.4.

5.5 Experiment 1

5.5.1 Method

5.5.1.1 Participants

Sixty-four native speakers of English (47 female; 18 to 31 years of age) participated in the experiment. The participants were recruited with SONA Experiment Scheduling at the Rayner Lab, University of California, San Diego, United States. All participants reported having normal or corrected-to-normal vision. Due to technical problems with the recording equipment or bad recordings, six participants were excluded from

the results. This resulted in 58 participants in total (42 female; $M=20.1$ years, $SD=2.24$). The participants were naïve to the purpose of the experiment.

5.5.1.2 Apparatus

Eye-tracking data were recorded binocularly at a sampling rate of 500 Hz using an EyeLink 1000 tower-mounted eye tracker from SR Research (Ottawa, Canada). The experiment took place at the Rayner Lab, University of California, San Diego, United States. The head of the participant was stabilized using a chin and forehead rest during recording. The distance from each participant's eyes to the stimulus monitor was approximately 60 cm. Stimuli were displayed on a ViewSonic VX922 19" 4:3 LCD display at a resolution of $1,280 \times 1,024$ pixels (39×29 cm, 36.0×27.2 degrees of visual angle at a distance of 60 cm), with a refresh rate of 60 Hz. The eye-tracking system was controlled with the host software for EyeLink 1000 running on a data recording computer. Stimulus presentation, questionnaire responses during eye tracking, a 13-point calibration and 13-point validation of the calibration accuracy were handled using Experiment Builder (1.10.165) running on a separate stimulus computer. Pre-test questionnaires were completed on a separate computer using PsychoPy (1.80.00).

5.5.1.3 Design and stimuli

The pre-reading questionnaire contained 20 statements followed by 40 statements. The first group of statements addressed how strongly the participant felt about each topic. Using the mouse, the participants stated to what degree they agreed with the statement on an ungraded visual analog scale ranging from *completely disagree* to *completely agree*. An example statement is *I have a strong opinion on vaccination*. The continuous scale ranged between 0.00 and 1.00. The responses equal to or lesser than 0.50 were classified as not having strong opinions on the topic (hereafter referred to as weak opinions, for brevity), and responses greater than 0.50 were classified as having a strong opinions. The statements given to participants, however, were specifically created so that the extremes did not signify "a weak opinion" and "a strong opinion" respectively, which would make the middle of the scale a neutral response. The participants were also specifically instructed to interpret the scale as how strongly they felt about the topic. Marking one extreme of the scale, labeled *completely agree*, would mean that they indeed felt very strongly about the topic, and marking the other extreme, labeled *completely disagree*, would mean that they did *not* have strong feelings about the topic. This was done so that participants would not interpret marking the middle of the scale to mean not having strong feelings or being neutral about the topic. This justified the division into two groups. Also, the difficult self-evaluation "how strong or weak is my opinion on this topic?" was turned into the simpler "do I feel strongly about this topic?" These responses were used to classify trials into two

groups: texts on topics that the participants felt strongly about, and texts on topics that the participants did not feel strongly about.

The second group of statements addressed specific opinions on each of the 20 topics, with one positively worded statement and one negatively worded statement for each topic. Participants once again used the mouse to indicate to what degree they agreed with the statement on an ungraded visual analog scale ranging from *completely disagree* to *completely agree*. An example statement pair is *Vaccinating children affects society positively* and *Vaccination is not an important part of public health policy*. The order of both sets of statements was randomized by PsychoPy. Importantly, the data on the positivity or negativity of the participants' opinions were not used in Experiment 1. They were included to create an identical situation to Experiment 2, where the data were used.

The stimuli used for the eye tracking consisted of 20 texts (eight fillers and 12 critical texts) taken from the English reading comprehension portion of the Swedish Scholastic Aptitude Tests, which uses a wide variety of online and written sources as texts. The texts were adapted to the current study to have exactly 90 words excluding headlines. The texts were also chosen to have a similar structure and style to the short articles that are commonly found in mainstream online media. Text was adapted to on-screen reading by using a large font, Arial 24-point (5×10 mm, 0.4×0.8 degrees of visual angle at a distance of 60 cm), and with the individual lines of text were separated by line spacing equaling approximately two blank lines in order to facilitate accurate eye tracking. Line breaks were automatically created by Experiment Builder. The order of the texts was randomized by Experiment Builder. Two groups of texts were given to participants in two instruction conditions. The texts referred to as being on "various topics" are hereby referred to as *neutral*. The texts referred to as being on "various topics that people say they feel very strongly about, and that are considered highly controversial according to survey data" are hereby referred to as *controversial*. The instructions as well as the order of the instructions were counterbalanced. An example stimulus screen is shown in Figure 5.1. After each text, the participants answered one multiple-choice comprehension question. The comprehension question for the text in Figure 5.1 was *What does the text imply?* The four possible alternatives were: *Vaccination is necessary to maintain a low death rate among children*; *Vaccination has had relatively little effect on public health*; *Vaccination is a waste of resources in the case of most diseases*; and *Vaccination no longer gives protection against tuberculosis*. The correct answer in this example is the first alternative, but the results from the comprehension questions were not used in any analysis, and were present only to ensure that the participants read all texts.

Vaccination.

According to a recent study, doctors observe increasing numbers of people denouncing vaccination as a part of a public health policy. This view is argued to be fashionable amongst the privileged, who take for granted the relative freedom from lethal infectious diseases as a result of successful immunization. The idea that improved living standards are solely responsible for the decline of infectious disease is possibly true in the case of immunization for tuberculosis, which has been successfully eradicated in several countries. Usually, this does not apply to other illnesses, however.

Figure 5.1. Example stimulus, Experiment 1

The example stimulus is shown using the identical proportions of the experimental setup.

5.5.1.4 Procedure

Each participant was introduced to the experiment and recorded individually. The experimental session lasted around 45 minutes, but no time restrictions were imposed during the experiment. The participants were welcomed to the study, and were told that they were going to participate in a reading study consisting of several parts, and that written instructions would be given during each part of the experiment. They were informed that their participation was completely voluntary, and that they were free to end the experiment at any time.

The experiment started with the participants completing a questionnaire about their age and gender, followed by the pre-reading questionnaire. The participants were instructed to answer spontaneously, and not think too long about their answers.

The participants were then told that they would read several texts, each followed by a comprehension question. They were then seated in front of the eye tracker and calibration was begun. The calibration was repeated until the offset was less than 0.75 degrees of visual angle in both the horizontal and the vertical directions.

After the eye-tracking portion of the experiment, participants were told about the real purpose of the study. They were thanked again, and given one course credit for their participation.

5.5.1.5 Data analysis

Velocity-based high-speed event detection was performed using SR Research Data Viewer with default settings (acceleration threshold=8,000°/s², velocity threshold=30°/s), which transformed the raw data into fixations and saccades. The recordings from the right eye were used, since the right eye is dominant for the majority of people. Data from the headlines were not used in any analyses.

The data processing, figures, and statistical analyses were made using R (3.2.2) (R Development Core Team, 2015) and the lme4 package (Bates et al., 2015) in R Studio (0.99.473). Linear mixed effect models were used to examine the effect of framing on total reading times/pw and first-pass reading times/pw, using either lines of text or individual words. Words without any fixations were removed from the analysis. All data were log-transformed to better fit a Gaussian distribution. Two fixed effects with two levels each were used, namely texts introduced as controversial or neutral, and topics that participants had a strong or weak opinion about. Models used random intercepts and slopes of condition within group for participant and text, with intercept-slope correlation factors for participants and texts. All models converged using intercept-slope correlation factors. The lmerTest package (2.0-29) was used to estimate *p*-values with Satterthwaite approximation.

Figures for the word-by-word analyses show the linear regression lines used in the linear mixed models as well as bars for three parts of the text (the first 30 words, the middle 30 words, and the final 30 words).

5.5.2 Results

All statistical models used log-transformed data, but in order to improve readability, no log-transformation was performed on the data in the figures. Full statistical model outputs can be found in Supplementary Material 5.A. The term *word number* refers to the order of the words in the text from the first word to the 90th word of a text (all texts were 90 words). Full statistical model outputs using per character measures can be found in Supplementary Material 5.B. No results changed from significant to non-significant or vice versa when using per character measures.

5.5.2.1 Total reading times per pixel-width

Figure 5.2 shows observed mean total reading times/pw on lines of text by opinion, divided into topic condition. If compared to the reading of the first two lines, a significant decrease was found for both the middle lines (Estimate=-0.078, *SE*=0.032, *t*=-2.45, *p*<.01) and the last two lines (Estimate=-0.129, *SE*=0.037, *t*=-3.49, *p*<.001).

In addition, participants with high opinion strengths had significantly shorter total reading times/pw on the last two lines, compared to participants with low opinion strengths (Estimate=-0.168, $SE=0.047$, $t=-3.59$, $p<.001$). No other significant results were found (all $ps>.10$). Figure 5.3 shows observed mean total reading times/pw on words as a function of word number by opinion, divided into topic condition. Bars (for illustration purposes only) show means for the first 30 words, middle 30 words, and last 30 words. A significant decrease was found between the beginnings and the ends of texts (Estimate=-0.0014, $SE=0.0003$, $t=-5.04$, $p<.0001$). An interaction was found between topic condition, opinion strength, and word number (Estimate=-0.0013, $SE=0.0005$, $t=2.65$, $p<.01$). Participants with high opinion strengths had longer total reading times/pw at the beginnings of neutral texts compared to participants with low opinion strength, with all participants in the controversial condition having very similar total reading times/pw.

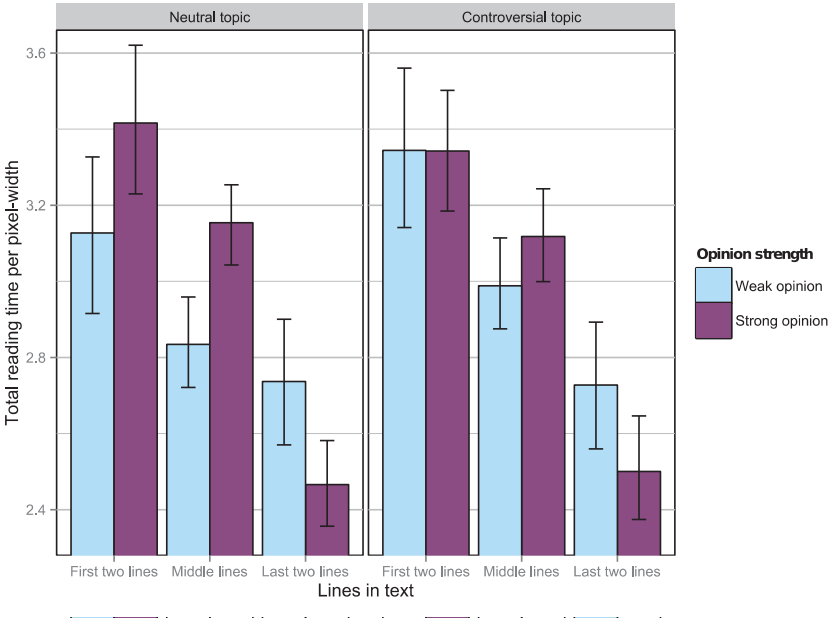


Figure 5.2. Mean total reading times/pw (ms) on lines of text, factored by weak or strong opinion, divided into topic condition, Experiment 1
Error bars denote standard errors of the mean.

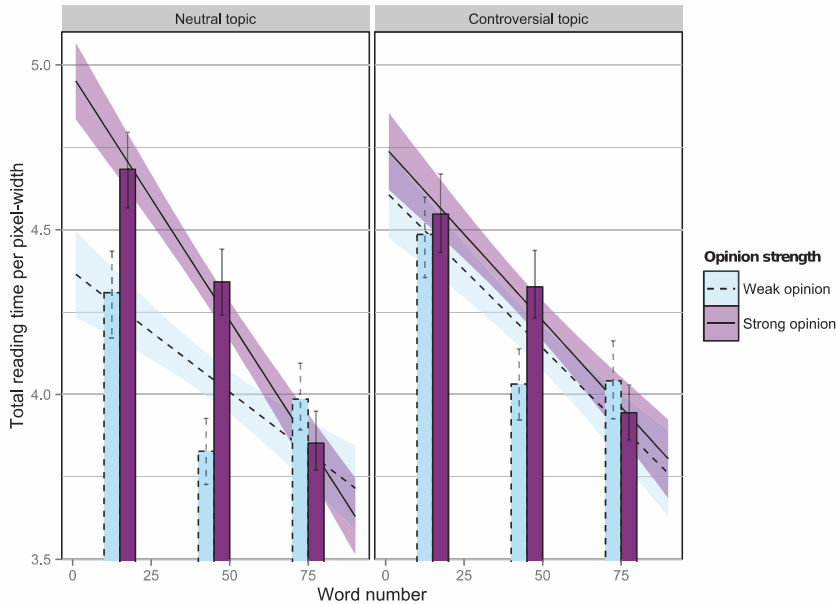


Figure 5.3. Mean total reading times/pw (ms) on words as a function of word number, factored by weak or strong opinion, divided into topic condition, Experiment 1

For illustration purposes, bars show means for the first 30 words, middle 30 words, and last 30 words. Error bars denote standard errors of the mean, and regression lines include 95% confidence intervals.

5.5.2.2 First-pass reading times per pixel-width

Figure 5.4 shows observed mean first-pass reading times/pw on lines of text by opinion, divided into topic condition. If compared to the reading of the first two lines, a significant increase was found for both the middle lines (Estimate=7.43, $SE=0.089$, $t=8.37$, $p<.0001$) and the last two lines (Estimate=6.96, $SE=0.103$, $t=6.75$, $p<.0001$). No other significant results were found (all $ps>.29$).

Figure 5.5 shows observed mean first-pass reading times/pw on words as a function of word number by opinion, divided into topic condition. Bars (for illustration purposes only) show means for the first 30 words, middle 30 words, and last 30 words of each text. A significant decrease was found between the beginnings and the ends of text (Estimate=-0.0014, $SE=0.0003$, $t=-5.44$, $p<.0001$). No other significant results were found (all $ps>.26$).

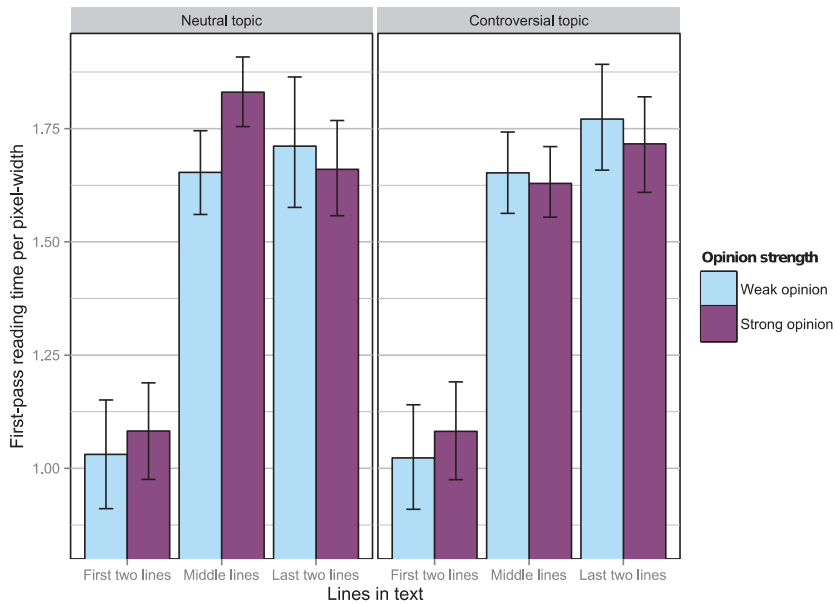


Figure 5.4. Mean first-pass reading times/pw (ms) on lines of text, factored by weak or strong opinion, divided into topic condition, Experiment 1
Error bars denote standard errors of the mean.

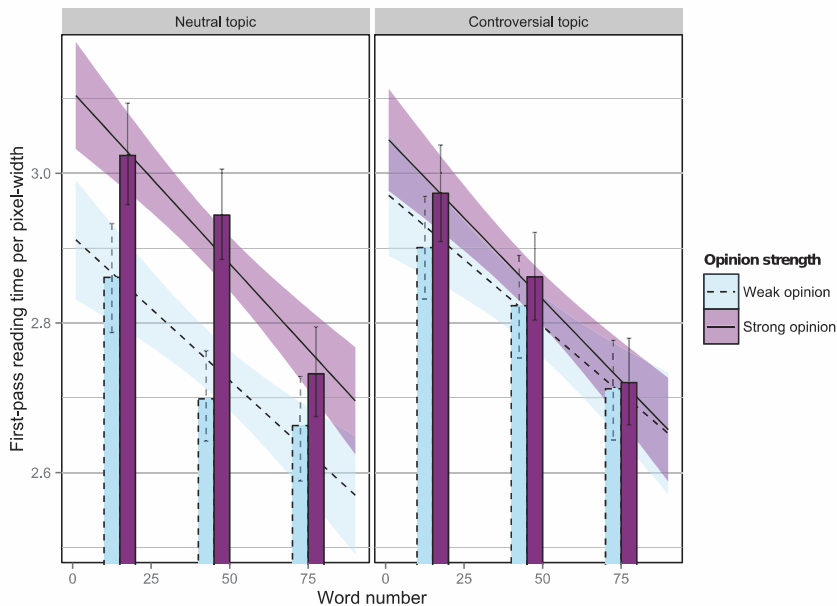


Figure 5.5. Mean first-pass reading times/pw (ms) on words as a function of word number, factored by weak or strong opinion, divided into topic condition, Experiment 1
For illustration purposes, bars show means for the first 30 words, middle 30 words, and last 30 words. Error bars denote standard errors of the mean, and regression lines include 95% confidence intervals.

5.5.3 Discussion

Headlines are believed to help readers minimize or optimize text processing by creating an occasion-specific context (Dor, 2003; Ifantidou, 2009). The headlines accompanying the texts in Experiment 1, however, were not very informative and were neutral in nature, often using only one word to describe the overall topic (for example, *Vaccination*). Thus, participants instead needed to create a context using the beginning of the text, and started reading by skimming the first few lines. This was demonstrated by significantly shorter first-pass reading times on the first two lines of the text compared to the other lines. First-pass reading times/pw on words indicates early processes (Inhoff, 1984; Rayner & Well, 1996; Reichle et al., 2003), while first-pass reading times/pw on entire lines of text indicate the conscious interruption in the reading of a single line of text. The measure aggregates all fixations during the first visit to the line, and ends when the line is left for the first time, which could happen very early during reading. The measure is then divided by the width of the line to compensate for size differences between the areas of interests (AOIs). This suggests that participants skimmed the beginning of the text to get an overview before continuing to read at a slower pace, since the headline did not give a great deal of information.

All participants spent more time on the initial parts of the texts compared to the final parts. This was shown in total reading times/pw on lines and words. This is similar to how participants read survey responses, focusing more on the first few responses compared to the ones at the end of the list (Galesic et al., 2008). In Chapter 3 of this book this pattern was also observed, with longer total reading times per character in the beginnings of texts compared to the ends during regular and thorough reading. This might reflect the fact that rereading is possible for a longer time at the beginnings of texts, since the end of the text has not yet been read. First-pass reading times/pw on words, however, were also longer at the beginnings of texts, compared to the ends. This reflects early processing, indicating that participants found the beginnings of texts more difficult. In combination, this suggests that effort is high at the beginning, when little text has been read. Effort then decreases over time, since more information has been read and becomes available in the mental representation of the text, which was also the conclusion in Chapter 3. It should be mentioned, however, that it is possible that our attention simply begins to fade toward the end of texts.

Participants with weak opinions regarding the topic read less thoroughly compared to participants with strong opinions. This was seen in significantly shorter total reading times/pw on words, which can be used to evaluate attention and processing (Holmqvist et al., 2011; Mason et al., 2013; Rayner, 1998, 2009; Reingold & Rayner, 2006; Strukelj, Scheiter, et al., 2016). When topics were presented as ones that “people say they feel very strongly about” which were “considered highly controversial according to survey data,” however, no significant difference was found

between participants with weak and strong opinions, with the results being very similar in the two groups. It would be expected to be the case that, in general, the participants had stronger opinions regarding some of the topics they read about, and weaker opinions regarding others. Since van Strien et al. (2016) found differences in reading between participants with strong and weak opinions, differences were also expected in the current study. When hearing that others felt strongly about the topics, people with weak opinions started reading more thoroughly and in a similar way to the group stating they had strong opinions on the topics. This result suggests that we are influenced by other people, regardless of who these people are. This might be in contrast to findings that say we are influenced by the opinions of our peers rather than people with whom we share no relation (Gaertner & Dovidio, 2012; Hewstone et al., 2002). An eye movement study, however, showed that both in-group and out-group information produced similar eye movement patterns (Guazzini et al., 2014). Thus, hearing that members of an out-group feel strongly about a topic seemed to be sufficient to change reading behaviors in the current experiment.

Interestingly, participants stating that they had strong opinions on a topic spent less time at the ends of the texts compared to participants with weak opinions, as demonstrated by shorter total reading times/pw on the last two lines. In addition, total reading times/pw on words became shorter toward the end of texts. Spending less time on the end of texts suggests that the overall processing of the text decreased. It is possible that participants who have strong opinions on a topic have already decided that they are right, making them less interested in changing their opinions. Paragraph ends have been found to be important when constructing a memory representation of the information (Hyönä et al., 2002), with large differences between different reading strategies. This could further suggest that participants with strong opinions in this study were helped by the information they already had. It is possible, however, that strength of opinion is related to prior knowledge. If this is the case, reading might be less thorough because of the higher familiarity with the material (Kaakinen et al., 2003).

Table 5.4 shows a summary of the behavior and measured effects, and how they correlate with the hypotheses and hypothesized effects presented in Section 5.3.2.

Table 5.4. Behavior and measured effects on the experimental measures, Experiment 1

Topic condition is predicted to impact processing of end-section and overall reading times, and opinion strength is predicted to impact overall reading times. Hypotheses presented in Section 5.3.2 supported by the data in **bold**. All behavior and effects in the table are supported by the data.

Topic	Participant opinion strength on the topic	
	Weak opinion	Strong opinion
Framed as a controversial topic	<i>Behavior:</i> The participants expect a controversial and important topic, and therefore read the concluding lines thoroughly. However, the participants did not change their reading of the conclusions depending on how the topic was framed. Their overall reading is also more thorough, similar to participants that have strong opinions regarding the topic.	<i>Behavior:</i> The participants feel strongly about the topic and are certain about the conclusions they have drawn about it. Therefore, the concluding lines of text are not read thoroughly.
	<i>Measured effects:</i> • End-section read more thoroughly, but with no interaction with condition. • Overall reading similar to participants with strong opinions.	<i>Measured effects:</i> • End-section read less thoroughly.
Framed as a neutral topic	<i>Behavior:</i> The participants read the concluding lines thoroughly, not changing their reading of the conclusions depending on how the topic is framed. Their overall reading is less thorough than participants with strong opinions.	<i>Behavior:</i> The participants feel strongly about the topic and are certain about the conclusions they have drawn about it. Therefore, the concluding lines of text are not read thoroughly.
	<i>Measured effects:</i> • End-section read more thoroughly, but with no interaction with condition. • Overall reading not as thorough as participants with strong opinions.	<i>Measured effects:</i> • End-section read less thoroughly.
Overall reading patterns	After the initial lines, total reading times on lines decreased , and first-pass reading times on lines increased. Total reading times/pw and first-pass reading times/pw on words decreased as a function of word number.	

The manipulation in Experiment 1 was intended to affect the strength of opinion by adding a social component. Very limited effects were found with regard to the manipulation, however. A stronger manipulation was therefore used in Experiment 2, and the strength of opinion, which affected the reading process significantly in Experiment 1, was more thoroughly investigated. The manipulation in Experiment 2 created expectations of either a positive or negatively biased text, to examine if myside bias and cognitive dissonance would arise in neutral texts. Openness was also measured, and was used to see if the strength of opinion was affected by this personality trait. While the neutral headlines in Experiment 1 activated background knowledge and theoretically caused slight myside bias, it was believed that the biased headlines in Experiment 2 would cause much higher levels of myside bias. High levels of cognitive dissonance would also be expected if the bias explicitly disagrees with the participant's opinion. This would theoretically give rise to more thorough reading of texts expected to agree with the participant's opinion, and less thorough reading of texts expected to disagree with their opinions. It was also considered likely that

reading would be affected by how strongly participants felt about a topic and that their openness would interact with how strongly they felt about the topic. It is plausible that the two reader profiles identified by McCrudden and Sparks (2014)—belief-protection and belief-reflection—are consistent with low and high openness, respectively. Strong opinions on the topic, as well as higher levels of openness, were expected to result in more thorough reading of the conclusions of the text in particular. Another experiment was conducted to investigate these assumptions.

5.6 Experiment 2

In Experiment 2, the texts from Experiment 1 were used, with very minor corrections. In order to also examine the effects of openness, a mini-markers task was added in Experiment 2 (Saucier, 1994). The task evaluated personality using the big-five trait taxonomy (Costa, 1991; John & Srivastava, 1999), and openness was the personality trait assumed to affect reading in this study. Except for the differences explained in the Method section below, Experiment 2 was identical to Experiment 1.

5.6.1 Method

5.6.1.1 *Participants*

Sixty-four native speakers of English (42 female; 18 to 41 years of age) participated in the experiment. The participants were recruited through student lists and advertisements at Lund University. All participants reported having normal or corrected-to-normal vision. Due to technical problems with the recording equipment or bad recordings, six participants were excluded from the results. This resulted in 58 participants in total (40 female; $M=27.6$ years, $SD=5.92$). The participants were naïve to the purpose of the experiment.

5.6.1.2 *Apparatus*

Eye-tracking data were recorded binocularly at a sampling rate of 500 Hz using a Hi-Speed 1250 tower-mounted eye tracker from SensoMotoric Instruments (Teltow, Germany). The experiment took place at the Humanities Lab, Lund University, Sweden. The head of the participant was stabilized using a chin and forehead rest during recording. The distance from each participant's eyes to the stimulus monitor was approximately 70 cm. Stimuli were displayed on a Samsung SyncMaster 245T 24" widescreen LCD display at a resolution of $1,920 \times 1,200$ pixels (52×32 cm, 40.8×26.1 degrees of visual angle at a distance of 70 cm), with a refresh rate of 60 Hz. The eye-tracking system was controlled with SMI iView X 2.8 running on a data recording computer. Stimulus presentation, questionnaire responses during eye tracking, a 9-point calibration, and a 4-point validation of the calibration accuracy

were handled using MATLAB (2012a) with the Psychophysics toolbox (3.0.11) running on a separate stimulus computer. The mini-markers task was completed on a different computer using PsychoPy (1.80.00).

5.6.1.3 Design and stimuli

The 20 texts (12 critical texts and eight filler texts) from Experiment 1 were used, with very minor corrections. The neutral and uninformative headlines used in Experiment 1 were replaced with headlines that were either positively or negatively biased in terms of the overall subject of the texts.

As in Experiment 1, the strength of opinion was evaluated via the pre-reading questionnaire, with the responses regarding the specific opinion also used in Experiment 2. The statements addressing specific opinions on each of the 20 topics, with one positively worded statement and one negatively worded statement for each topic, resulted in responses ranging from 0.00 to 1.00. The response to the negatively worded statement (where a higher value signifies higher disagreement) was subtracted from the response on the positively worded statement (where a higher value signifies higher agreement), producing two differences ranging from -1.00 to 1.00 for each trial, one before and one after reading. Trials with a difference greater than 0.00 in the trials were classified as participant agreement with a positive headline and disagreement with a negative headline. Trials with a difference equal to or lesser than 0.00 were classified as the participant disagreeing with a positive headline, and agreeing with a negative headline.

Text was adapted to on-screen reading using a large font, Arial 36-point (5.0×10 mm, 0.40×0.80 degrees of visual angle at a distance of 70 cm), and with the individual lines of text separated by two blank lines in order to facilitate accurate eye tracking. At the beginning of each trial, the headline was shown in isolation for two seconds, in order to implicitly make sure the participant would not skip the headline. After this, the whole text appeared below the headline. Two different headlines were created for all texts, either with a positive or negative bias on the topic. The order of the texts was randomized in MATLAB. An example stimulus pair is shown in Figures 5.6a and 5.6b.

The mini-markers task contained 40 different personality characteristics, shown one at a time. The participant stated how accurately each characteristic described their own personality on a scale ranging from one (signifying *extremely inaccurate*) to nine (signifying *extremely accurate*). This resulted in five means for the big-five trait taxonomy.

The good current vaccination practices.

According to a recent study, doctors observe an increasing number of people
denouncing vaccination as a part of a public health policy.

This view is argued to be fashionable amongst the privileged, who take for granted
the relative freedom from lethal childhood diseases as a result of successful immunization.

The idea that improved living standards are solely responsible for the decline of infectious disease
is possibly true in the case of immunization for tuberculosis, which has been successfully eradicated
in several countries, but it does not apply to other illnesses.

Figure 5.6a. Example stimulus, positive bias, Experiment 2

The example stimulus is shown using the identical proportions of the experimental setup.

The bad current vaccination practices.

According to a recent study, doctors observe an increasing number of people
denouncing vaccination as a part of a public health policy.

This view is argued to be fashionable amongst the privileged, who take for granted
the relative freedom from lethal childhood diseases as a result of successful immunization.

The idea that improved living standards are solely responsible for the decline of infectious disease
is possibly true in the case of immunization for tuberculosis, which has been successfully eradicated
in several countries, but it does not apply to other illnesses.

Figure 5.6b. Example stimulus, negative bias, Experiment 2

The example stimulus is shown using the identical proportions of the experimental setup.

5.6.1.4 Procedure

The participants were instructed that the headline would be shown in isolation for two seconds, after which the text would appear automatically.

After the eye-tracking part of the experiment, the mini-markers task was completed. The participants were given a movie ticket for participating in the study.

5.6.1.5 Data analysis

Velocity-based high-speed event detection was performed using SMI BeGaze 3.5 with default settings (peak velocity threshold=40°/s, minimum fixation duration=50 ms), which transformed the raw data into fixations and saccades.

Two fixed effects with two levels each were used: a headline either agreeing or disagreeing with the participant's opinion prior to reading, and topics about which participants had either a strong or weak opinion. All models converged using intercept-slope correlation factors.

Openness was used as a continuous variable in all statistical models, ranging from 1.00 to 9.00.

5.6.2 Results

All statistical models used log-transformed data, but in order to improve readability, no log-transformation was performed on the data in the figures. Full statistical model outputs can be found in Supplementary Material 5.C. The term word number refers to the order of the words in the text from the first word to the 90th word of a text (all texts were 90 words). Full statistical model outputs using per character measures can be found in Supplementary Material 5.D. When using per character measures, first-pass reading times per character on the middle lines as a function of opinion strength changed from non-significant ($p=.058$) to significant ($p=.047$), and on the last two lines as a function of opinion strength and openness changed from significant ($p=.048$) to non-significant ($p=.051$). Therefore, these effects were not addressed in this study.

5.6.2.1 Total reading times per pixel-width

Figure 5.7 shows observed mean total reading times/pw on lines of text by opinion, divided into bias condition. If compared to the reading of the first two lines, a significant decrease was found for the last two lines (Estimate=-0.273, $SE=0.056$, $t=-4.87$, $p<.0001$). No other significant results were found (all $ps>.05$). Figure 5.8 shows observed mean total reading times/pw on words as a function of word number by opinion, divided into bias condition. Bars (for illustration purposes only) show means for the first 30 words, middle 30 words, and last 30 words. A significant decrease was found between the beginnings and the ends of text (Estimate=-0.002, $SE=0.0004$, $t=-5.36$, $p<.0001$). A significant decrease was also found for participants with high

opinion strengths compared to participants with low opinion strengths (Estimate=-0.089, $SE=0.026$, $t=-3.39$, $p<.001$). An interaction between these results was found (Estimate=-0.001, $SE=0.0005$, $t=2.17$, $p<.05$). Reading for participants with high opinion strengths was relatively uniform, while participants with low opinion strengths had higher total reading times/pw at the beginnings of texts, and lower total reading times/pw at the ends of texts. Notably, no interaction was found with bias condition (Estimate=-0.0002, $SE=0.0007$, $t=-0.32$, $p=.75$). No other significant results were found (all $ps>.83$).

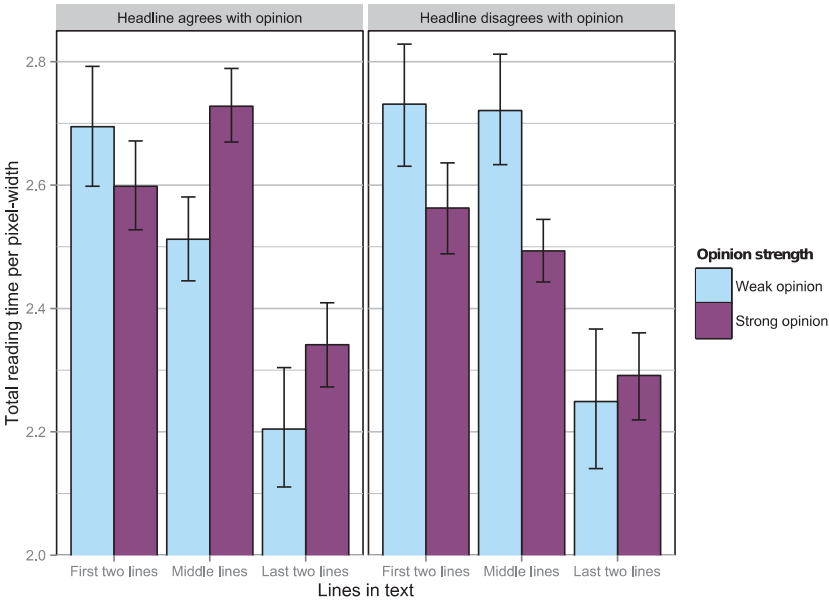


Figure 5.7. Mean total reading times/pw (ms) on lines of text, factored by weak or strong opinion, divided into headline bias condition, Experiment 2
 Error bars denote standard errors of the mean.

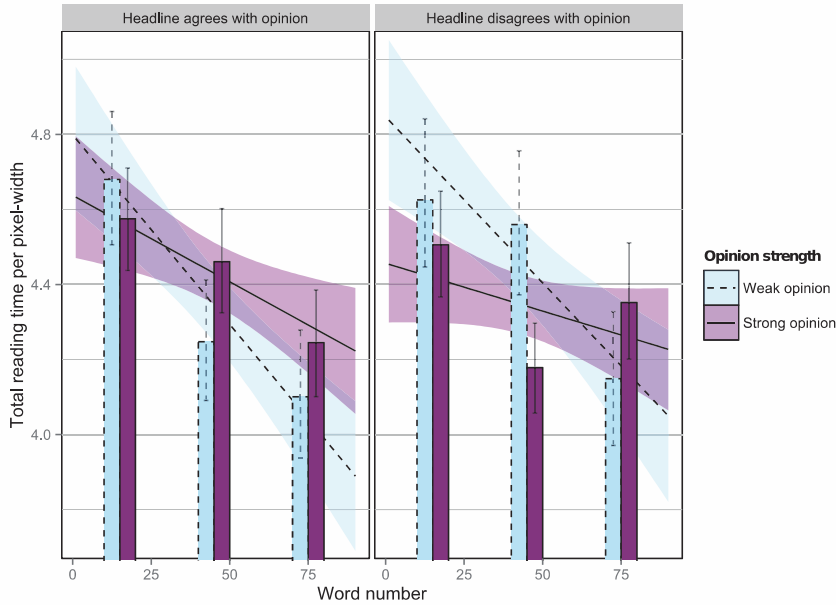


Figure 5.8. Mean total reading times/pw (ms) on words as a function of word number, factored by weak or strong opinion, divided into headline bias condition, Experiment 2

For illustration purposes, bars show means for the first 30 words, middle 30 words, and last 30 words. Error bars denote standard errors of the mean, and regression lines include 95% confidence intervals.

5.6.2.2 First-pass reading times per pixel-width

Figure 5.9 shows mean first-pass reading times/pw on lines of text by opinion, divided into bias condition. If compared to the reading of the first two lines, a significant increase was found for the middle lines (Estimate=0.211, $SE=0.058$, $t=3.59$, $p<.001$). No other significant results were found (all $ps>.06$). Figure 5.10 shows mean first-pass reading times/pw on words by opinion, divided into bias condition. Bars (for illustration purposes only) show means for the first 30 words, middle 30 words, and last 30 words. A significant decrease was found between the beginnings and the ends of text (Estimate=-0.0072, $SE=0.0003$, $t=-2.13$, $p<.05$). No other significant results were found (all $ps>.26$).

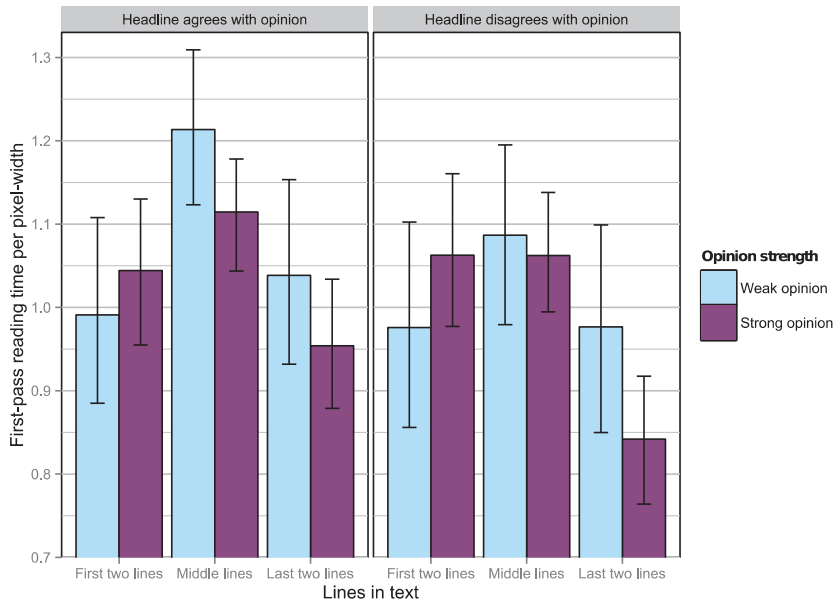


Figure 5.9. Mean first-pass reading times/pw (ms) on lines of text, factored by weak or strong opinion, divided into headline bias condition, Experiment 2
Error bars denote standard errors of the mean.

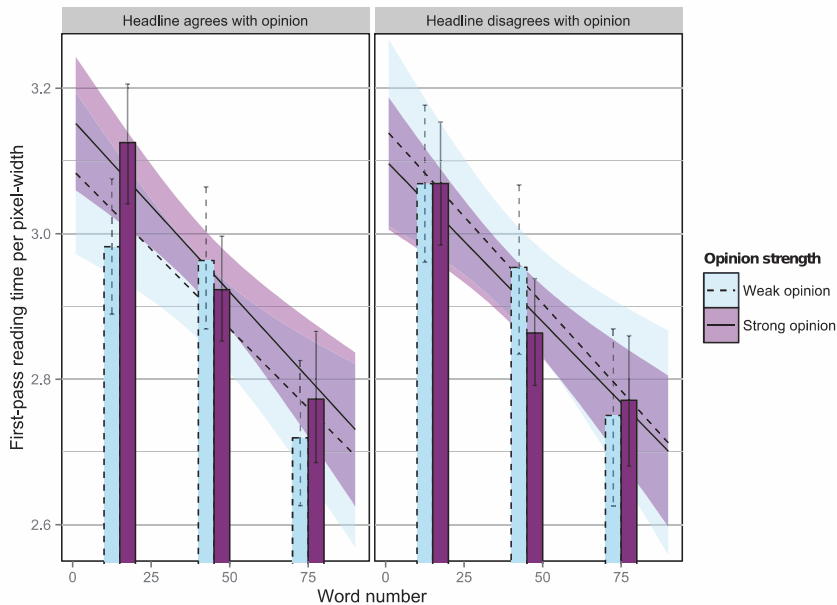


Figure 5.10. Mean first-pass reading times/pw (ms) on words as a function of word number, factored by weak or strong opinion, divided into headline bias condition, Experiment 2
For illustration purposes, bars show means for the first 30 words, middle 30 words, and last 30 words. Error bars denote standard errors of the mean, and regression lines include 95% confidence intervals.

5.6.2.3 *The effect of openness on reading*

Figure 5.11 shows observed mean total reading times/pw on lines of text as a function of openness, by opinion, divided into bias condition. If compared to the reading of the first two lines, an interaction was found on the last two lines for bias condition, opinion strength, and openness (Estimate=-0.337, $SE=0.111$, $t=-3.05$, $p<.01$). Total reading times/pw increased when openness increased for all participants except when headlines agreed with opinions of participants with low opinion strengths. In those cases, total reading times/pw *decreased* when openness increased. This interaction is also found on words, using word number (Estimate=-0.003, $SE=0.0007$, $t=-3.86$, $p<.0001$). Figure 5.12 shows observed mean total reading times/pw on words as a function of openness, by opinion, divided into bias condition. The location of the interaction is confirmed when using this categorization, with a significant result only on the last 30 words (Estimate=-0.179, $SE=0.046$, $t=-3.87$, $p<.0001$). Figure 5.13 shows observed mean first-pass reading times/pw on lines of text as a function of openness, by opinion, divided into bias condition. If compared to the reading of the first two lines, an interaction was found on both the middle lines and the last two lines for bias condition, opinion strength, and openness (middle lines: Estimate=-0.262, $SE=0.124$, $t=-2.12$, $p<.05$; last two lines: Estimate=-0.321, $SE=0.144$, $t=-2.23$, $p<.05$). First-pass reading times/pw was fairly uniform as a function of openness for participants with high opinion strengths, for all parts of the text. However, large changes were found for participants with low opinion strengths. When headlines disagreed with their opinions, first-pass reading times increased when openness increased, and when headlines agreed with opinions first-pass reading times/pw decreased when openness increased. No significant results were found for first-pass reading times/pw on words when using word number (all $ps>.18$) or when dividing a text into three equal groups (all $ps>.17$).

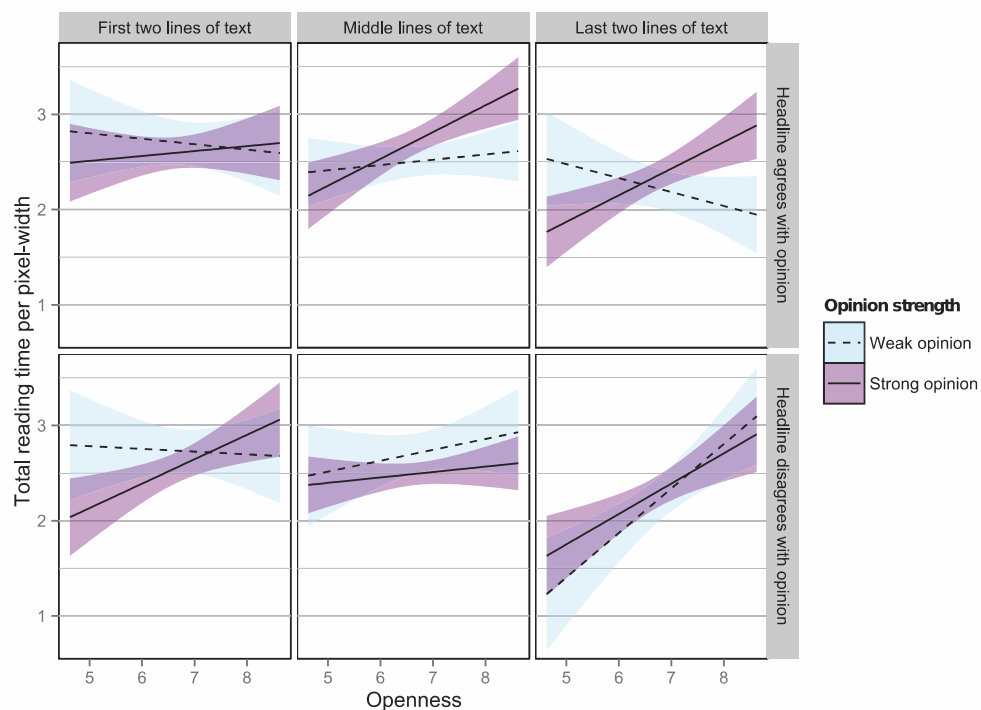


Figure 5.11. Mean total reading times/pw (ms) on lines of text as a function of openness, factored by weak or strong opinion, divided into headline bias condition, Experiment 2
Regression lines include 95% confidence intervals.

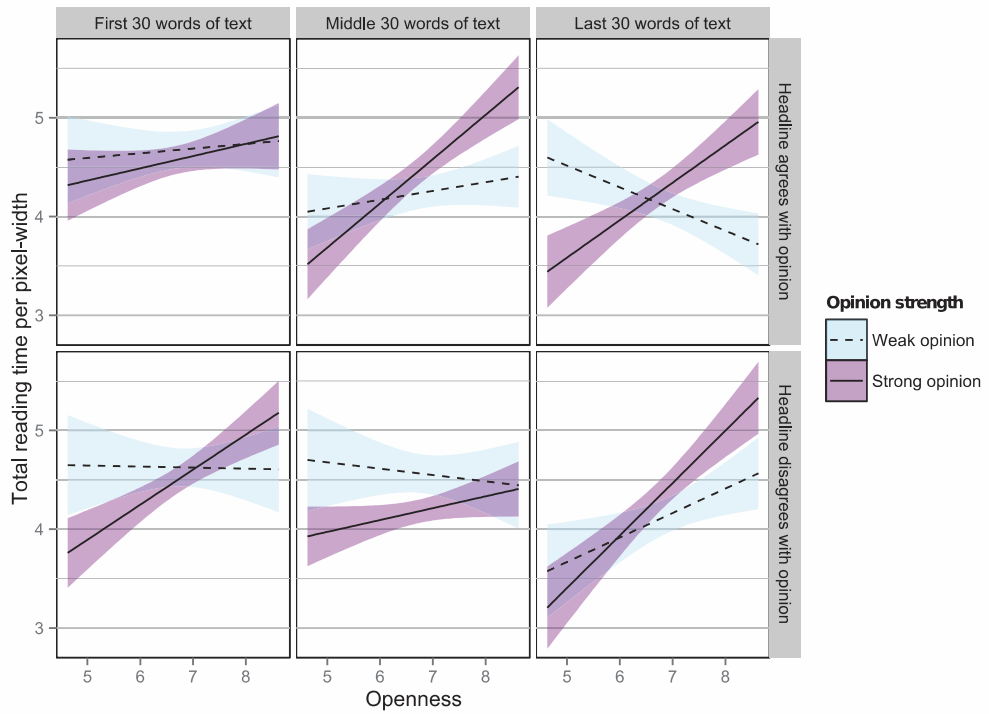


Figure 5.12. Mean total reading times/pw (ms) on words as a function of openness, factored by weak or strong opinion, divided into headline bias condition, Experiment 2
Regression lines include 95% confidence intervals.

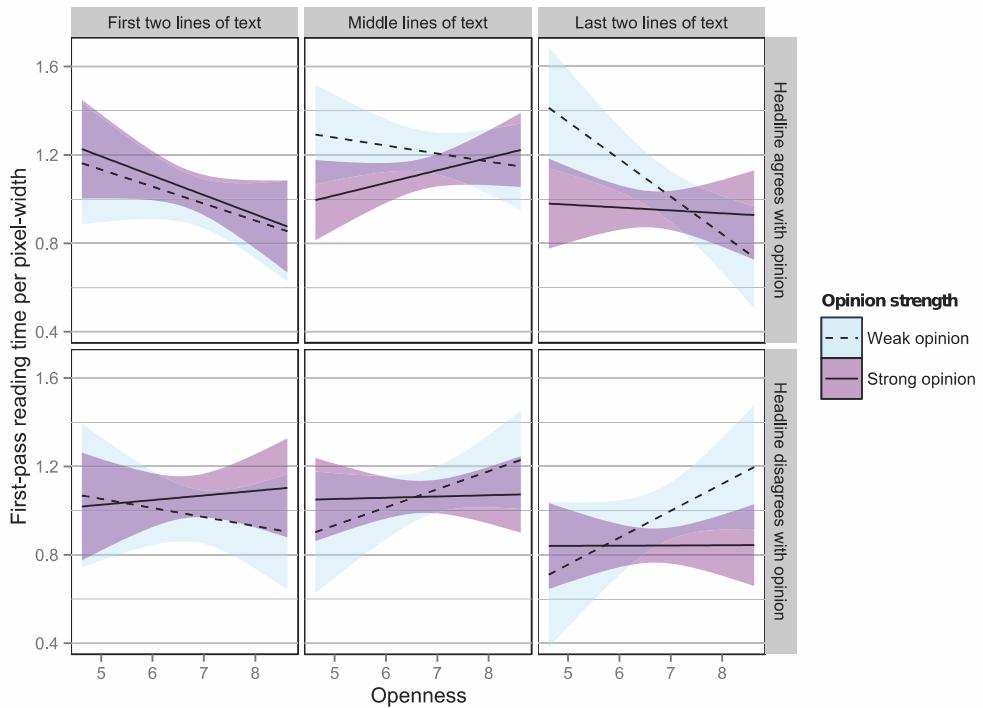


Figure 5.13. Mean first-pass reading times/pw (ms) on lines of text as a function of openness, factored by weak or strong opinion, divided into headline bias condition, Experiment 2
Regression lines include 95% confidence intervals.

5.6.3 Discussion

In Experiment 2, biased headlines were used in conjunction with neutral texts in order to create schemas in which participants expected a text that would either agree or disagree with the information. This is similar to how biased headlines were used in Wolfe et al. (2013), but these were consistent with the position of the text. In the current study, the sentences were not classified in detail, as was the case in Wolfe et al. (2013), in favor of using multiple stimuli texts in order to find more generalizable tendencies. Total reading times/pw on words and first-pass reading times/pw on words were longer at the beginning of the texts compared to the ends; this was also demonstrated by total reading times/pw on the last two lines being shorter compared to the first two lines. This suggests that beginnings of texts were read more thoroughly by all participants compared to the ends. It is also possible that this suggests initial delays caused by the need to establish contextual connections at the beginning of reading. Both explanations would suggest that reading is difficult initially and becomes easier as more of the text is read. While total reading times/pw on lines were not significantly different for participants with strong or weak opinions, total reading times/pw on words were lower overall for participants with strong

opinions. The ends of the texts were read more thoroughly by participants with weak opinions. No effects on the reading process were found for expecting to agree or disagree with the text. This supports the results in van Strien et al. (2016), in which reading times were not affected by agreement or disagreement with the opinion expressed in the text when opinions were strong, but does not support the results found for participants with weak opinions, who spent longer reading texts disagreeing with their opinions than texts agreeing with their opinions. The results do not coincide with those of Wolfe et al. (2013), in which unitizing sentences (summary sentences, wrap-up sentences) were read more quickly by participants who disagreed with the text. When openness was incorporated in the analysis of the results in the current study, however, interesting results appeared.

Openness is the personality trait in the big-five trait taxonomy (Costa, 1991; John & Srivastava, 1999) that has been shown to correlate most positively with openness to new ideas and tolerance for all people and ideas (Roccas et al., 2002), and broadmindedness and imaginativeness (Dollinger et al., 1996), among other things. In both conditions, participants with strong opinions and low openness read the last two lines for a shorter amount of time compared to participants with strong opinions and high openness. Participants with high openness spent longer on the conclusion, suggesting that they were open to changing their opinions by thoroughly reading the conclusions, regardless of the bias of the text. The ends of paragraphs seem to be especially important for the creation of a memory representation of the material (Hyönä et al., 2002), which could explain the thorough reading. This was also found for participants with weak opinions, but only when the headline disagreed with their opinions. When the headline agreed with the participants' opinions, the trend was instead reversed. Higher openness resulted in less thorough reading, and lower openness resulted in more thorough reading. This result was also found for words as the area of interest in the analysis, as well as in first-pass reading/pw per line. These results are surprising, but tentative explanations can be offered on the basis of the structure of the text.

While headlines always either agreed or disagreed with the opinion of the participant, the texts themselves always contained neutral and unbiased descriptions of the issue. Participants therefore expected to read a biased text, but they were likely to assume that at least some other information would be present in the text. Therefore, participants with weak opinions but high openness actively refrained from processing the text too thoroughly. They see themselves as being open to other opinions, so they value other perspectives. But it could also be the case that they did not feel very strongly about the topic in question, and they thought they would agree with the conclusion of the text. To conserve cognitive resources, they implicitly evaluated other perspectives less thoroughly by decreasing processing of the conclusion. With weak opinions and low openness, processing was instead high. It is possible that participants spent a longer amount of time on texts that agreed with them merely because of the fact that they agreed with them. This would be predicted

by confirmation bias (Hergovich et al., 2010; Jonas et al., 2001; Klayman, 1995; Nickerson, 1998), which states that participants are biased toward texts that confirm their opinions, and are biased against texts that disagree with them, and the more refined theory of myside bias (Mercier, 2017; Stanovich et al., 2013), which states that participants will have a tendency to find arguments that support their own opinions. Note that longer reading times were interpreted as representing more thorough integration of information, and that retention tasks after reading could strengthen this claim. Indeed, processing times for texts with headlines participants disagreed with were low for both participants with weak and strong opinions of the topic when openness was low. This provides evidence that openness affects the reading process more than disconfirmation bias (K. Edwards & Smith, 1996). The alternative hypothesis is that higher openness correlates with lower epistemic vigilance, which would predict that less cross-examination of information would occur when openness is higher. Thus, reading would become less thorough, and information would be accepted at face value. This pattern is only seen in the one case outlined above. While this explanation is certainly possible, it is considered less likely than the explanation given here. High openness might indeed signify low epistemic vigilance, but with regard to the reading process and the results seen in this study, I argue that this is expressed not by less cross-examination but by more thorough processing of the information.

This study shows that shorter processing time indicates higher resistance to information in the text. Another interpretation would be that shorter processing times indicate more immediate integration of the information. Two main factors serve to rebut these interpretations. Most importantly, the suggestion that higher openness increases processing was largely consistent with the results. In addition, the texts did not contain biased information to integrate, but only the headline. To properly examine this claim, however, it would have been necessary to test retention and comprehension after all reading was concluded.

Table 5.5 and Table 5.6 show a summary of the behavior and measured effects, and how they correlate with the hypotheses and hypothesized effects presented in Section 5.3.2.

Table 5.5. Behavior and measured effects on the experimental measures, Experiment 2

Bias condition is predicted to impact processing of end-section and overall reading times, and opinion strength is predicted to impact overall reading times. Hypotheses presented in Section 5.3.2 supported by the data in **bold**. All behavior and effects in the table are supported by the data.

Bias	Participant opinion strength on the topic	
	Weak opinion	Strong opinion
Headline agrees with participants' opinions	<i>Behavior:</i> As the participants do not feel strongly about the topic, they are losing interest in the text by the word. Also, expecting a text they agree with does not make them behave differently compared to expecting one they disagree with.	<i>Behavior:</i> The participants read the concluding lines of text more thoroughly. This is not due to myside bias, however, since expecting a text they agree with does not make them behave differently compared to expecting one they disagree with.
	<i>Measured effects:</i> Word-by-word reading goes from thorough at the beginning to skimming at the end.	<i>Measured effects:</i> Word-by-word reading is more uniform.
Headline disagrees with participants' opinions	<i>Behavior:</i> As the participants do not feel strongly about the topic, they are losing interest in the text by the word. Also, expecting a text that disagrees with them does not make them behave differently compared to expecting one that agrees with them.	<i>Behavior:</i> As the participants feel strongly about the topic, their interest is still high toward the end of the text. They do not care, however, whether the text agrees with them or not. Also, expecting a text that disagrees with them does not make them behave differently compared to expecting one that agrees with them.
	<i>Measured effects:</i> Word-by-word reading goes from thorough at the beginning to skimming at the end (hypothesized: end-section read less thoroughly).	<i>Measured effects:</i> Word-by-word reading is more uniform.
Overall reading patterns	After the initial lines, total reading times on lines decreased , but only for the last lines in a text. First-pass reading times on lines increased, as in Experiment 1. Total reading times/pw on words decreased as a function of word number. First-pass reading times/pw decreased as a function of word number, as in Experiment 1.	

Table 5.6. Behavior and measured effects on the experimental measures when accounting for openness, Experiment 2

Openness is predicted to impact processing of end-sections. Hypotheses presented in Section 5.3.2 supported by the data in **bold**. All behavior and effects in the table are supported by the data.

Bias	Participant opinion strength on the topic			
	Weak opinion	Strong opinion	Weak opinion	Strong opinion
Headline agrees with participants' opinions	<i>Behavior:</i> The participants are not open to other opinions, so reading convincing conclusions is not likely to change their opinions. They are therefore biased toward the concluding lines, which they expect to reaffirm their opinions. Myside bias rather than openness affects reading when opinions are weak and the text is expected to agree with opinions.	<i>Behavior:</i> The participants do not care about other possible conclusions.	<i>Behavior:</i> The participants are open to other opinions, so reading convincing conclusions might change their opinions. They therefore avoid the concluding lines, so as not to spend unnecessary effort on the topic. Myside bias rather than openness affects reading when opinions are weak and the text is expected to agree with opinions.	<i>Behavior:</i> The participants read the conclusions thoroughly to confirm their opinions.
	<i>Measured effects:</i> End-section read more thoroughly. Middle and concluding lines are thoroughly read at first entry.	<i>Measured effects:</i> End-section read less thoroughly.	<i>Measured effects:</i> End-section read less thoroughly. Middle and concluding lines are skimmed at first entry.	<i>Measured effects:</i> End-section read more thoroughly.
Headline disagrees with participants' opinions	<i>Behavior:</i> The participants do not care about other possible conclusions.	<i>Behavior:</i> The participants do not care about other possible conclusions.	<i>Behavior:</i> The participants read the conclusions thoroughly to see how their own opinions are wrong.	<i>Behavior:</i> The participants read the conclusions thoroughly to see how their own opinions are wrong.
	<i>Measured effects:</i> End-section read less thoroughly.	<i>Measured effects:</i> End-section read less thoroughly.	<i>Measured effects:</i> End-section read more thoroughly.	<i>Measured effects:</i> End-section read more thoroughly.
Low openness		High openness		
Participant openness				

5.7 General discussion

Two experiments investigated how expectations about the topic and the specific bias of the text affect reading, and how this is manifested in the reading process.

5.7.1 Overall reading

The beginnings of texts were read more thoroughly at compared to the ends by all participants in both experiments. This was seen in total reading times/pw in both

Experiment 1 and Experiment 2. In addition, total reading times/pw on lines were shorter for both the middle and the last two lines in Experiment 1, and shorter for the last two lines in Experiment 2. As longer total reading times show overall processing (Holmqvist et al., 2011; Mason et al., 2013; Rayner, 1998, 2009; Strukelj, Scheiter, et al., 2016), this suggests that the beginnings of texts are read more thoroughly compared to the ends.

The first two lines of text were consciously left behind earlier when a text was first encountered. This was shown by shorter first-pass reading times/pw on the first two lines compared to the rest of the text in Experiment 1. Thus, rereading was indeed higher at the beginnings of texts, but was caused in part by this initial skimming. This was interpreted as being caused by the low amount of information in the headlines. More informative headlines in Experiment 2 produced more uniform reading, but first-pass reading times on the first two lines were still significantly shorter compared to the middle lines. The results could be interpreted as the first couple of lines of short, informative texts being skimmed in order to get an overview of the topic before more regular reading begins, with informative headlines eliminating some of this need. This is in line with research that suggests that the role of headlines is to optimize text processing (Dor, 2003; Ifantidou, 2009), and that headlines are used to create expectations and interest in the text (Bucaria, 2004; Molek-Kozakowska, 2013, 2014; Schaffer, 1995). First-pass reading times/pw on words suggest that early processing was high at the beginnings of texts, and that processing decreased toward the end. Longer first fixation durations show higher levels of early processing (Reichle et al., 2003), with first-pass reading times showing very similar results to first fixation durations with words as the unit of analysis (Inhoff, 1984; Rayner, 1998). This suggests that the process of building a mental representation of the text seems to be more difficult at the beginnings of texts, and that difficulties decrease as the mental representation becomes more complete. That the first two lines are skimmed initially indicates an effort to quickly create an initial context, which is then used as a framework for incorporating the remaining information. The fact that the uninformative headlines in Experiment 1 resulted in more skimming than the more informative headlines in Experiment 2 supports this notion.

5.7.2 Expectations, previous opinions, and openness

Participants expected unbiased texts in Experiment 1, since headlines were neutral, short, and uninformative. The texts in Experiment 2 had biased headlines, however, creating an expectation of texts that either agreed or disagreed with the participants' previous opinions. The neutral headlines in Experiment 1 activated schemas that were not biased, but only contained the topic of the text. In Experiment 2, in addition to the topic, the schemas also contained the expectation of either agreeing or disagreeing with the text. As a result, a much higher level ofmyside bias and cognitive dissonance could be expected in Experiment 2. This hypothesized increase inmyside bias did not

create any significant resistance to opinion shift in Chapter 4, which was surprising. Eye movements are a much more sensitive measure, however, than measuring opinions in a questionnaire, and are also an online measure that studies reading in real-time. Effects of myside bias could thus arguably be present in eye movement measures. It should be mentioned that the texts were neutral, and that only expectations were producing these effects. It is indeed suggested, however, that confirmation bias is affected by our beliefs and expectations (Nickerson, 1998), and it is considered to be an automatic process (Friedrich, 1993; Oswald & Grosjean, 2004), which might extend to myside bias as well. It is therefore plausible that the expectations from the biased headlines induce substantial levels of myside bias, and change the way a neutral text is read. Expectations were indeed sufficient to induce myside bias in an observation task, albeit to a fairly limited extent (Marsh & Hanlon, 2007).

In Experiment 1, when reading texts framed as neutral, total reading times/pw on words were shorter for participants with weak opinions compared to participants with strong opinions. Differences between the two groups of participants disappeared, however, when texts were framed as texts that other people felt strongly about. This suggests that expecting topics that others feel strongly about is sufficient to raise interest in the topic and produce a reading similar to participants with strong opinions about the topic. Also, eye movements seem to react similarly for information given by an in-group (for example, friends) and an out-group (for example, the “other people” in the instructions for Experiment 1). Our sensitivity to influence from others (Cialdini & Goldstein, 2004) thus seems to be visible in our reading behavior. Furthermore, total reading times/pw on the last two lines were shorter for participants with strong opinions on the topic, regardless of whether or not the text was introduced as a controversial topic. Conclusions and summaries are often found at the ends of texts, and participants possibly avoid them because they are uninterested in other opinions.

Experiment 2, however, produced reading patterns that were distinctly different. Apart from decreases toward the ends of texts (explained in Section 5.7.1, above), no significant differences were found between texts that agreed or disagreed with previous opinions. This mirrors the results for participants with strong opinions in van Strien et al. (2016). It has been demonstrated that reading is facilitated when the reader’s expectations are met (Haberlandt, 1982; Mak & Sanders, 2012). Since the texts were neutral in this case, this did not happen. But differences in openness have been shown to significantly affect the results, which could have masked the predicted effects of myside bias and confirmation bias (Friedrich, 1993; Hergovich et al., 2010; Jonas et al., 2001; Klayman, 1995; Mercier, 2017; Nickerson, 1998; Oswald & Grosjean, 2004; Stanovich et al., 2013), and cognitive dissonance (Harmon-Jones & Mills, 1999; Stone & Cooper, 2001). These biases predict that participants read texts they expect to agree with their opinions more thoroughly than texts they expect to disagree with them.

In this study, openness affected reading similarly when headlines agreed and disagreed with the participant's opinion in three out of four cases. Higher openness led to more thorough reading of the ends of texts, as shown by significantly longer total reading times/pw. This suggests that openness is more important than myside bias when resolving cognitive dissonance, possibly in a similar way as in-out-group bias seems to protect from other cognitive biases (Guazzini et al., 2014). The results are also largely consistent with the prediction that the two reader profiles identified by McCrudden and Sparks (2014)—belief-protection and belief-reflection—are closely linked to openness.. The self-standards model proposes that self-confidence and self-esteem determine how cognitive dissonance is resolved (Harmon-Jones & Mills, 1999; Stone & Cooper, 2001). In other words, personality seems to play a major role. Since openness is correlated with autonomy of thought (Roccas et al., 2002), it might be related to self-confidence and self-esteem.

Myside bias is the tendency for people to evaluate evidence based on their own opinions (Stanovich et al., 2013). Since the texts in this study were neutral, it is possible that myside bias was not activated by the mere expectation of reading a text with an evaluative bias. This would suggest that openness constantly operates during reading, while myside bias only operates when the text contains clear evaluative bias. Myside bias is argued to be related to rational thinking, with a model evaluating rational thinking encompassing open-minded thinking, again suggesting that openness affects the resolution of cognitive dissonance. Finally, openness produced differences between participants who expected to agree or disagree with the text only when participants did not have strong opinions on the topic. These results support the conclusions of van Strien et al. (2016), who found differences in reading texts that agreed or disagreed with opinions only for participants who did not feel strongly about the topic.

In Experiment 2, participants with strong opinions and low openness read the last two lines less thoroughly compared to participants with strong opinions and high openness. This effect, however, did not change according to bias condition. Since the texts were neutral, the ends of the texts were the only remaining place in which the bias introduced by the headline could be resolved. More thorough reading toward the end of a text is thus seen as a combination of increased interest in the conclusion of a text and the participants' attempt to find the information pertaining to the bias introduced by the headline, which is also more likely to be located at the end of the text. This means that, when participants have strong opinions on a topic, higher levels of openness seem to show a high interest in other opinions, regardless of whether or not these opinions are in conflict with previous opinions. First-pass reading times were uniform regardless of openness for the beginnings and middles of the texts, suggesting that participants read the texts similarly at first. At the ends of the texts, higher levels of openness led the participants with weak opinions about texts with headlines that they agreed with to initiate rereading more quickly, and led them to initiate rereading more slowly for texts with headlines that they disagreed with. This

suggests that the participants are confused by a biased headline accompanying a neutral text. Rather than interpreting these results as contradicting theories of myside bias and cognitive dissonance, it is suggested that the effects are overridden when opinions on a topic are strong. Since openness correlates positively with openness to new ideas and tolerance for all people and ideas (Roccas et al., 2002), as well as being broadminded and imaginative (Dollinger et al., 1996), openness seems to be able to override cognitive biases such as myside bias. This does not hold, however, when participants have weak opinions on a topic.

When participants with weak opinions on a topic read texts they thought they would disagree with, the patterns are the same as for the participants with strong opinions. Openness overrides myside bias, just as it did for participants with strong opinions on the topic. But when texts agreed with participants that stated they had a weak opinion on the topic, the opposite trend was found. Participants with lower levels of openness read the texts more thoroughly than participants with higher levels of openness. As suggested by myside bias, they might have been seeking support for their own views, and were biased toward opinions that they thought would agree with their own, respectively. Because their openness was low, it was less likely that their opinions would change as a result of reading contradictory information that might be present in the text. The fact that participants with higher levels of openness were found to read the texts less thoroughly is also thought to maintain myside bias. A text that agrees with someone's opinions can still contain other arguments or reasoning. By reading less thoroughly, less attention is given to information that could contradict the currently held opinion. A weak opinion on a topic is likely considered less important than a strong opinion. Therefore, considering the validity of the currently held opinion or evaluating other opinions might be seen as unnecessary effort. Since the likelihood of changing an opinion increases when openness is high, avoiding any contradictory information helps maintaining the opinion.

5.8 Conclusions

This chapter shows that people read short informational text less thoroughly when they do not feel strongly about a topic. Reading became more thorough, however, when they were informed that "people say they feel very strongly about" a topic. This supports research suggesting that influence from others can affect reading. In addition, when activating schemas that not only activate the general topic, but also the expectation of either agreeing or disagreeing with the text, the level of openness seems to affect the reading process much more than myside bias. Low openness resulted in less thorough reading of the ends of the texts (which commonly contains conclusions and summaries), and high openness resulted in more thorough reading. This is true except in one case: when people do not feel strongly about a topic but

expect the text to agree with them. In this case, openness instead *reinforces* myside bias. When openness is low, opposing information is unlikely to change opinions, and the thorough reading predicted by myside bias occurred in the conclusions of the texts. When openness was high, opposing information is likely to change opinions, and reading of the conclusions therefore became less thorough. A possible explanation is that by using less effort when reading a text that potentially contains opposing information, the currently held opinion can be kept safe. It seems as if myside bias has little effect on the reading process when induced by expectations alone, unless people do not feel strongly about the topic. When they do not feel strongly about the topic and they expect the text to agree with them, myside bias is activated and their opinions are maintained. This finding suggests that understanding and bias can be conditioned by expectations. People read what they want to see, thereby missing information that does not support their view.

In Chapter 6, one specific feature of the reading process is investigated: the fact that people sometimes do not consciously notice incongruities in text. It is plausible to assume that people do not expect texts to contain incongruities that are substantial enough to warrant an interruption to their reading. Previous research indeed shows that people sometimes do not consciously notice incongruities in text, and instead read what they expect. There are two possible explanations to why this occurs. One explanation is that the incongruity is insufficiently processed, and therefore does not affect the reading process. As a consequence, the incongruity can be said to be missed. Another explanation is that the incongruity is sufficiently processed, and affects the reading process. Since the participant does not consciously notice the incongruity, however, the reading process can be said to replace the incongruent word with the expected word. As a consequence, the incongruity is not missed, but can be said to be *disregarded*. Chapter 6 investigates this by giving people texts with semantic incongruities, and then evaluating the detection of the incongruities after reading.

Chapter 6

You sometimes read the word you expect instead of an incongruent word, but you process the incongruity nevertheless

Part 1: Paragraphs

6.1 Introduction

Previous chapters have shown that people tend to maintain their opinions to a larger extent when expecting controversial topics. People also change their reading if they expect that a text will agree with them (if they do not feel strongly about the topic). This suggests that people misread, or do not note, statements that do not support their own opinion in certain cases. Since these effects arise from mere expectations, with no changes being made to the text itself, it becomes interesting to investigate how reading changes when an actual incongruity is encountered. In cases in which the reader detects an incongruity, it is natural to assume that his or her reading is affected in some way or another. The reader may return to it in order to reread the relevant bit and make sense of the text. When an incongruent word is read but the error is not detected, however, is the word missed, or is it in fact processed and then disregarded in favor of the word that is expected in the context? This is an important question, since it would indicate the presence of parallel processing routes during reading (Clahsen & Felser, 2006; Ferreira & Patson, 2007; Jackendoff, 2007). This chapter therefore investigates how reading is affected by successful and failed detection of semantic incongruities in text. This chapter also investigates how detection is affected by the working memory capacity of the participant.

As mentioned above, previous research suggests that reading uses parallel processing routes: one based on the actual word on the page, and the other based on factors such as context, previous experiences, and expectations (Clahsen & Felser, 2006; Ferreira & Patson, 2007; Jackendoff, 2007). Consequently, when reading an incongruent word, one processing route would contain the incongruent word while another would contain the word expected in the context. In the case of failed

detection, the expected word meaning would be used in the reader's meaning representation of the text (what the text is about). While the effects of different types of incongruities on the reading process have been extensively researched using eye tracking (Clifton et al., 2007; R. K. Morris, 2006), much less attention has been given to *failures* of detection. Such studies are important, since they provide detailed evidence of the reading process. To the best of my knowledge, no consequences for not noticing words that are incongruent with the text have been found (Bohan & Sanford, 2008; Daneman et al., 2007). The reason for this may be that the stimuli have not been clear or strong enough to elicit these small effects. Clearly, the choice of critical words for a task of this kind is of importance for the result. Antonym pairs such as *high* and *low* were used as critical words in the stimuli texts. One word in the pair created a direct contradiction of the previous context and the other one was the word in the correct control condition. This was assumed to be obvious enough to reliably trigger detection so that an error is identified on unambiguous grounds by a reader. Antonym pairs serve as excellent as experimental manipulations, since they express polar opposites within the same meaning dimension. They are thus both maximally different and maximally similar at the same time (see Section 6.2.3).

A distinction is made between *explicit detection*—reading and reacting to an incongruity—and *implicit detection*—reading but not explicitly reacting to an incongruity. In other words, explicit detection is a conscious reaction to the incongruity, while implicit detection could be described as an unconscious reaction. Explicit detection is immediate and is expected to produce a strong disruption to the reading process, since the reader rereads the text. Implicit detection is expected to give rise to smaller disruptions, since the incongruity does not elicit a conscious reaction from the reader. Nevertheless, implicit detection should be evident in lingering effects during subsequent reading. This would suggest that the incongruity was at least partially processed, but that this output was disregarded in favor of an interpretation based on what was expected in the context.

In both cases, outputs using the incongruity are created, along with outputs using the word that is expected in the context, as suggested, for instance, by the good enough approach to language comprehension (Ferreira & Patson, 2007). Consequently, one of these two conflicting interpretations needs to be disregarded during the creation of a consistent meaning interpretation of the text. During explicit detection, the violation is deemed to be large enough to require attention from the reader. When the incongruity is not detected, however, the expected word meaning is used during meaning integration without the participant's involvement—allowing heuristic usage to continue in subsequent reading—and the output with the incongruity is disregarded. Effects of implicit detection should be found in eye-tracking measures, when compared to a control condition in which the text is congruent, due to disruptions in the reading process arising from the disregarded output. This would indicate the presence of a selection mechanism dealing with two contrasting interpretations from parallel reading processes, with the incorrect one

selected to be the most appropriate for the text currently being read. A contrasting hypothesis would state that when not explicitly detected, an incongruity is missed due to insufficient processing, rather than a parallel reading process taking precedence. If this is the case, effects should only be found in eye-tracking measures for explicit detection.

In order to investigate the reading process during successful and failed detection of incongruities in texts, two eye-tracking experiments were conducted. In Experiment 1, short texts were read by participants, with some texts containing an incongruent word. The incongruity was created by replacing a correct word—for example, *save*—with its corresponding antonym—for example, *lose*. Detection was investigated in a post-experiment interview by asking about each incongruent text in turn. In Experiment 2, the main difference was that the classification of detection was much more geared toward implicit detection. Eye movement differences were investigated between the incongruities that had been successfully detected, the incongruities that were not explicitly detected, and the texts with the correct word.

6.2 Previous research

Expectations have been shown to play a major role in reading, as Chapter 4 and Chapter 5 have shown. When expecting texts with a positive or negative evaluative bias, people with weak opinions changed their opinions to become more positive or negative after reading the texts, depending on the type of bias. In other words, they read the text as if it was biased, and they actually changed their opinions accordingly. Openness was shown to strongly affect the reading of the ends of texts: lower openness led to less thorough reading and higher openness led to more thorough reading. This pattern, however, was the opposite for people with weak opinions on a topic, who expected to read a text that they agreed with.

6.2.1 Successful and failed detection of incongruities during reading

Studies specifically investigating failures to explicitly detect incongruities have used specific cultural knowledge incongruities such as the *Moses illusion* to decrease explicit detection. The question *How many animals of each kind did Moses take on the ark* is given to participants, and around 80% incorrectly answer *two* (it was Noah, not Moses, who built the ark) (Bredart & Modolo, 1988; Oostendorp & De Mul, 1990; Reder & Kusbit, 1991). Skimming is not believed to be a reason for not reporting an incongruity, since reading times for manipulated words are in fact shorter when participants explicitly detect the manipulation, compared to incongruities that were not reported as detected by participants. Instead, the *partial match* hypothesis is proposed to be the cause (Kamas et al., 1996; Park & Reder, 2004), in which linguistic input only needs to pass a level of sufficient match to continue. The

inherent problem with the use of this type of stimuli, referred to as *semantic illusions*, is the reliance on pragmatic knowledge in the sense of world and cultural knowledge, such as knowledge of the Bible, history, geography, fairytales, etc. Also, Mata, Ferreira, and Reis (2013) propose that semantic illusions are resolved by two largely independent groups of processes: controlled processes checking the information in the sentence with the information present in memory, and automatic processes evaluating the impression of truth in the semantic associations between the elements in the sentence. Mata et al. (2013) showed that an increase in cognitive load increased the number of semantic illusions that were not explicitly detected by participants.

Incongruities known as *semantic anomalies* operate on a more general level than more culturally-dependent illusions (Croft & Cruse, 2004; Paradis, 2003, 2015). General in this sense means that they are independent of cultural information about society, which has nothing to do with the general world knowledge that is crucial for lexical knowledge. Instead of using cultural knowledge like *Moses building the ark*, semantic anomalies rely on concepts like *burying survivors* and *negotiating with hostages* (Barton & Sanford, 1993; Bohan & Sanford, 2008; Park & Reder, 2004). Bohan and Sanford (2008) performed an eye-tracking study on anomalies designed to be difficult to detect, and divided the results into detected anomalies, nondetected anomalies, and nonanomalies. The stimuli consisted of experimental items from a bank of pretested semantic anomalies and non-anomalous filler passages, such as *The authorities negotiated/communicated with the scared and desperate **hostages** and calmed them down*. The critical word (bold) became anomalous or nonanomalous depending on the context introduced by the manipulation (underlined). No significant effects were found when comparing the nonanomalies with the anomalies that were not explicitly detected by participants. When participants consciously detected the anomalies, late measures showed significantly longer *total reading times*; that is, the total time spent on the specified text region compared to nondetected trials. An increase in the number of fixations on the anomalous region was also found. The number of first-pass regressions on the region following the critical region increased, as did increases in regression-path durations. No effects on early eye-tracking measures were found.

Rayner et al. (2004) investigated plausibility during reading. Three conditions were used—plausible (likely theme), implausible (unlikely theme), and anomalous (inappropriate theme)—in sentences such as *John used a **knife** to chop/an **axe** to chop/a **pump** to inflate the large carrots for dinner*. The anomalous condition can be considered a semantic incongruity (it is impossible to inflate real carrots; the sentence would only be acceptable in a fantasy situation or, indeed, if the carrots in question were balloons) and the implausible condition can be seen as a pragmatic incongruity (we know that axes are not commonly used to chop carrots). The differences between the conditions are based on knowledge about the entities and the frequency of encounters with expressions of this kind in language (Paradis, 2015). Some of these actions are common, while others are very strange or even unheard of. Immediate

effects of a disruption to the reading process were found for anomalous sentences, as shown by increased gaze durations on the word, and delayed effects were found for implausible sentences, as shown in increased go-past times. Other studies have used manipulations such as *he took a sip from the waterfall/transmitter* (Kutas & Hillyard, 1980) and *the woman told the tourist/suitcase* (Nieuwland & Van Berkum, 2005). These types of manipulations, however, are problematic. The type of knowledge needed to correctly identify the incongruities, while certainly very similar across participants, cannot be assumed to be identical.

Rayner et al. (2004) found effects from reading both implausible and inappropriate sentences, but did not investigate detection. This was controlled for in research by Barton and Sanford (1993), who used a text with five different alternatives to measure the effects of detecting semantic anomalies during reading. The text from their first experiment is given in (17) below, with the five alternatives in italics.

- (17) There was a tourist flight traveling from Vienna to Barcelona. On the last leg of the journey, it developed engine trouble. Over the Pyrenees, the pilot started to lose control. The plane eventually crashed right on the border. Wreckage was equally strewn in France and Spain. The authorities were trying to decide where to bury the *survivors* / *injured* / *surviving injured* / *surviving maimed* / *surviving dead*.

After reading the text in (17), participants were asked to answer a question: *What should the authorities do?* Answers such as “let the relatives of the deceased decide where they want the bodies to be buried” were considered to be evidence of not explicitly detecting a manipulation. This type of stimuli offers advantages compared to the Moses illusion paradigm, since it can be analyzed from a linguistic angle by specifying the linguistic feature containing the manipulation. Furthermore, the manipulation is located in running text, as opposed to the questions posed in the Moses illusion paradigm. A replication and extension of the study included two additional texts (Hannon & Daneman, 2004), and was further expanded upon using eye tracking (Daneman et al., 2007). Daneman et al. (2007) used two-word constructions based on the stimuli in Barton and Sanford (1993) in three conditions: internally coherent (*surviving injured*); internally incoherent (*surviving dead*); and an unambiguous control (*unfortunate dead*). The findings in Daneman et al. (2007) showed effects of detection of incongruities in late measures, but not early ones, for both internally coherent and incoherent conditions. A significant increase was found in the number of regressions as well as look-back time on explicitly detected manipulations compared to the control condition. Since detection was delayed and only occurred after the reader had moved away from the manipulation, a fair amount of processing is needed to interpret this type of incongruity. No effects were found on the critical word, however, for manipulations that were not explicitly detected. The authors argue that this could be caused by participants seeing the manipulation as an

unintentional error in the stimuli. They did not report this during the post-experiment interview, and it was therefore classified as non-detected. The findings suggest that although semantic anomalies introduced in larger texts are frequently missed by participants, there are limited effects on the reading process as a consequence of the failed detection. It is possible, however, that semantic anomalies are still not strong enough to elicit effects in the reading process, or that the manipulation needs to be more evident but still not noted by participants.

The tolerance for incoherence has been shown to be surprisingly high. In a study conducting conversations between two participants about what celebrity a participant “would want to spend a day with” via instant messaging, more than a third of participants did not notice a message that was clearly incoherent (Roberts et al., 2016). These messages were introduced by the experimenters unbeknownst to the participants during the experiment, and were either incoherent with regard to the task-relevant information about a celebrity not present on the screen (Experiment 1), or with regard to the gender of the interlocutor (a message that the celebrity in question was “an icon for women like me” was inserted by male interlocutors and “an icon for men like me” by female interlocutors) (Experiment 2) (Roberts et al., 2016). With regard to spoken language, “um”-disfluencies (filled pauses), are actually argued to help processing (Bosker, Quené, Sanders, & de Jong, 2014), by improving word recognition (Corley & Hartsuiker, 2011) and memory (Collard, Corley, & MacGregor, 2008), and silent pauses are shown to limit context-driven expectations for the upcoming words (MacGregor, Corley, & Donaldson, 2010).

With regard to failure of detection in visual scenes, a study examined how *scene semantics* (the context and the composition of visual scenes), affected the identification of objects and visual search (Cornelissen & Võ, 2016). One of the crucial findings was that participants looked longer at incongruent items, even when they stated that nothing was wrong with the scene, suggesting that the incongruities had been implicitly detected. Another study showed that people generally are not aware of where they look, further suggesting that eye movements are connected with unconscious processes (Võ, Aizenman, & Wolfe, 2016). A similar case of looking but not seeing is *inattention blindness*, which is the failure to notice an unexpected but salient stimulus. In a study, participants saw a video clip of people playing basketball, and were asked to count the number of passes a team made. In the middle of the game, however, a man dressed in a gorilla suit walked through the two teams passing basketballs to each other (Simons & Chabris, 1999). Not only did participants fail to see the man in the gorilla suit, but a subsequent study showed that participants were equally likely to fixate on the incongruity, regardless of whether they consciously noticed it or not (Beanland & Pammer, 2010). This was shown to be the case also for experts performing a familiar task (Drew, Võ, & Wolfe, 2013). In the eye-tracking study, a familiar type of search task was performed by radiologists, with a gorilla inserted in one of the images. It was found that eighty-three percent of the

radiologists did not see the incongruity, but that the majority of these in fact looked directly at it.

6.2.2 Lingering effects

If expectations are violated and a sentence is reanalyzed, the initial analysis has been shown to leave a memory trace, such as in the case of garden path sentences. Garden path sentences are commonly used to show effects of prediction during reading, but also to illustrate the problematic situations that can arise from reanalyzing an already performed analysis (Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Ferreira, Christianson, & Hollingworth, 2001; Lau & Ferreira, 2005). When reading the garden path sentence *While Anna dressed the baby spit up on the bed*, participants are likely to initially assume that the baby is being dressed by Anna. But as they reach *spit up*, they discover that their initial analysis was incorrect and reread the sentence. As expected, post-experiment, participants report that the baby spit up on the bed. A significant number of participants, however, also incorrectly report that Anna dressed the baby. This error occurs to a higher degree the longer the participant is committed to the incorrect syntactic structure, with regard to phrase length and the position of the manipulation (Christianson et al., 2001; Ferreira et al., 2001; Lau & Ferreira, 2005). The phenomenon in which an initially incorrectly formed analysis leads to incorrect responses to follow-up questions shows that this initial sentence analysis is committed to memory extremely quickly. An eye-tracking study by Slattery, Sturt, Christianson, Yoshida, and Ferreira (2013) showed that the lingering misinterpretations caused by reading garden path sentences are likely caused by the initial output not being completely erased. This is similar to the *digging-in effect*, also examined using sentence structures with different lengths in order to keep an output active for different durations before a reanalysis is made. Reading times become longer toward the end of the sentence the longer an output is kept active (Tabor & Hutchins, 2004). Furthermore, participants are more likely to use the incongruent information in subsequent comprehension questions when not explicitly detecting Moses illusion-like incongruities (Bottoms et al., 2010). Since participants do have sufficient previous knowledge, the error must be disregarded by the parsing process for reading to continue unhindered. In accordance with the parallel processing hypothesis, the erroneous information is at least partially processed, since it does affect subsequent comprehension checks. Again, it seems that an incorrect analysis is corrected without the conscious involvement of the reader.

6.2.3 Antonyms as experimental stimuli

Since prediction is central to language comprehension (Pickering & Garrod, 2013), any lingering effects of implicit detection should be more likely with a large

discrepancy between the actual word and the word expected in the context. Antonyms are therefore well-suited for use in stimuli investigating implicit detection.

Antonym pairs in language usage are semantically and discursively similar in the sense that they share the same minimal conceptual dimension and occur in the same types of structural environment, but are also at the same time maximally different since they express opposite properties of the shared *meaning dimension* (Paradis, Löhndorf, van de Weijer, & Willners, 2015; Paradis & Willners, 2011). Some antonyms are strongly conventionalized as pairs along a meaning dimension—for example *high* and *low* for height (Paradis, Willners, & Jones, 2009)—while others require a substantial amount of context to be understood as antonyms, such as *tall* and *mowed* in the sentence *fireflies prefer tall grass to mowed lawns* (Jones, Murphy, Paradis, & Willners, 2012, p. 132). Binary opposition is central to human perception and cognition (Bianchi, Savardi, & Kubovy, 2011; Paradis, 2001, 2005, 2013), and several studies have shown that native speakers of a language—even small children—seem intuitively aware of antonymy (Bentin, 1987; Bianchi et al., 2011; Kretchmar, Bornkessel-Schlesewsky, & Schlesewsky, 2009; Murphy & Jones, 2008; Paradis et al., 2009; Roehm, Bornkessel-Schlesewsky, Rösler, & Schlesewsky, 2007; Tribushinina, 2013; van de Weijer, Paradis, Willners, & Lindgren, 2012, 2014). A number of studies have used antonym relations to examine the reading process (Barton & Sanford, 1993; Daneman et al., 2007). As shown in Section 6.2.1, Barton and Sanford (1993) used self-paced reading and investigated detection of a semantic anomaly that was antonymic in nature (*survivors* and *dead*), and found that participants consistently missed the incongruities in the given texts. Daneman et al. (2007) used eye tracking on stimuli adapted from Barton and Sanford (1993), and found that detection affected late measures but not early ones. This suggests that incongruities are detected online, but no eye-movement differences between detectors and non-detectors were found.

While these studies used antonym relations, however, they did not make use of common core antonym pairs such as *high* and *low* in a context in which they would be immediately identifiable as incorrect. Instead, their stimuli used a more complex context, in which the anomaly required a fairly high amount of cognitive processing in order to be correctly resolved. Ideally, an incongruity should be easily identifiable, as is the case with syntactic errors, but still be syntactically correct. Because antonyms are similar in all respects but one—the polar opposition—they make excellent stimuli for investigating successful and failed detection. Antonyms also allow for syntactically correct sentences to be used.

6.3 The study of implicit detection of incongruities during paragraph reading

In two eye-tracking experiments, I investigated the effects of successful and failed detection of antonymic incongruities when participants were reading a paragraph of text. To ensure that the incongruities were clearly incongruent in their various contexts, while still having a high chance of not being detected, antonyms were used. Antonym pairs operate in one meaning dimension only, namely the polar opposites within a meaning space. All other features are identical, making antonyms maximally similar, while being maximally different at the same time (Jones et al., 2012; Paradis et al., 2015, 2009; Paradis & Willners, 2011). This semantic relation is fundamental for native speakers (Bentin, 1987; Bianchi et al., 2011; Kretchmar et al., 2009; Murphy & Jones, 2008; Paradis et al., 2009; Roehm et al., 2007; Tribushinina, 2013; van de Weijer et al., 2012, 2014). The manipulations are very closely linked to semantic properties, and they rely very little on world knowledge, which even obvious cases such as *the woman told the **suitcase*** (Nieuwland & Van Berkum, 2005) and *John used a **pump** to inflate the large carrots for dinner* (Rayner et al., 2004) can be said to do to some extent. The onset of detection is also clear, and the binary meaning relation is identical in all manipulations. All of these advantages make antonyms very well-suited for experimental manipulations.

The experiment was designed so that participants could indicate detection of an incongruity after each text by choosing an alternative that stated that the text was ambiguous. Detection was then definitely determined in a post-experiment interview, in which participants were asked questions about two congruent filler texts in order to lower suspicion, then about each manipulated text in turn, and finally about the texts indicated by the participant as being ambiguous. An incongruity was considered to be successfully detected when a participant referenced the antonymic incongruity (Experiment 1) or responded that something was strange with the text (Experiment 2). It is important to note that the incongruities were highly conspicuous in the overall context, and clearly erroneous. In Experiment 1, the only region of analysis was the critical word. In Experiment 2, the analysis also included the headline and the last sentence of the text to further analyze the effects of both the explicit and implicit detection of the incongruent word. The stimuli were slightly altered so that the sentence immediately following the incongruent sentence was the last sentence in all texts.

An example text can be found in (18) below. All stimuli texts and follow-up questions can be found in Appendices 6.A and 6.B, and example stimulus screens can be found in Appendix 6.C. In (18), the antonym pair is indicated in italics, with the incongruent word also in bold. Only one word was given to each participant, with *save* being the word given in the correct version of the text, and *lose* being the word

given in the incongruent version of the text. No words were displayed in bold or italics during the experiment.

- (18) Petrol stations profit by getting rid of cash payment machines

Many gas stations are removing the option to pay with cash at petrol kiosks. Cash requires additional handling and it makes the stations more vulnerable to robberies. The petrol companies are expected to *save/lose* money by removing the cash machines entirely. The companies plan to implement this during the next few years.

6.3.1 Terms and classifications

Previous research suggests that two parallel processes operate during reading (Jackendoff, 2007; Kim & Osterhout, 2005; van Herten et al., 2006): one process that syntactically analyzes the text, and a second that uses heuristics based on semantic plausibility and a “surface-order-to-thematic-template” to create an output that is “good enough” (Christianson et al., 2010; Ferreira, 2003; Ferreira et al., 2002; Ferreira & Engelhardt, 2009; Ferreira & Patson, 2007). The output from the good enough analysis is the one commonly used by the reading process, and is based on factors such as expectations, prediction, context, and previous knowledge. Context and prediction have indeed been shown to be important during reading in several studies (Britt et al., 1992; Rayner et al., 1992; Van Berkum et al., 2005). Ferreira and Patson (2007) used garden path sentences, which are syntactic and structural, when arguing for the good enough approach to language comprehension. They also used several studies in which semantic incongruities are missed to argue for the usage of heuristics during reading. It thus seems as though the focus in much previous research has been on the usage of heuristics during reading, rather than the specificity of the syntactic analysis. In my study, the manipulation is purely semantic in nature, and the good enough approach to language comprehension is used as the reading model. In order to avoid confusion, the parallel process that is referred to as a syntactic analysis in the good enough approach is here referred to as *word-meaning retrieval*. Note that I do not argue for the validity of the term syntactic analysis, but merely extend the notion to also incorporate semantic details. The term word-meaning retrieval does not differentiate between semantic and syntactic details. It can plausibly be said that erroneous syntax is discovered by retrieving the meaning of the word and noticing that it does not fit.

6.3.1.1 Proposed model for when an incongruent word is fixated

This study introduces two mechanisms—*output similarity evaluator* and *expectation-based output selection*—to explain the effects that the explicit and implicit detection of incongruities has on the reading process. The proposed model is visualized in Figure

6.1a, Figure 6.1b, and Figure 6.1c. The two critical words used in the figures are taken from the incongruent version of (18) above. This means that *save* is the word expected in the context, but *lose* is the actual word that the reader encounters. The two parallel processes that operate during reading create two outputs for every word that is fixated. Of these two outputs, only one is used during the creation of the cumulative meaning representation of the text. Figure 6.1a shows a schematic diagram of the output similarity evaluator when fixating on an incongruent word and comparing the two outputs. The two outputs are normally very similar when reading, but differ when fixating a semantic incongruity such as the type used in this study, in which the expected word and the word on the page are antonymic opposites. The heuristics analysis is assumed to finish slightly more quickly than the word-meaning retrieval. The fact that the process of word-meaning retrieval is not fully completed is represented as a hypothetical area labeled *not completed* in the figure.

When participants in this study encounter an antonymic incongruity, the two parallel reading processes deliver contrasting outputs. When the output similarity evaluator compares the two outputs, the difference is either deemed small enough to not require further attention from the reader, or sufficiently large or incongruent to elicit a response. Importantly, the output similarity evaluator prefers to use the heuristics analysis; in other words, the expected word is commonly favored. It is possible that better reading proficiency might not necessarily decrease this preference, but could in fact increase it, since reading is automated to a higher degree. Figure 6.1b shows a schematic diagram of expectation-based output selection under normal circumstances, that is, when the two outputs are considered to be sufficiently similar for regular reading to continue. Expectation-based output selection uses the output from the good enough analysis when updating and adding to the meaning representation of the text. The reader remains unaware of the error, and heuristics usage continues uninterrupted. The output from the word-meaning retrieval is left incomplete and is disregarded by the reading process. This is also the case when there is only a slight contrast between the two outputs, such as with spelling or grammatical errors. Indeed, the good enough analysis seems to be used frequently during reading, as evidenced by the many cases in which semantic incongruities, anomalies, and illusions remain undetected by readers (Barton & Sanford, 1993; Daneman et al., 2007; Ferreira & Engelhardt, 2009; Ferreira & Patson, 2007; Hannon & Daneman, 2004; Park & Reder, 2004). This is similar to how prediction and context is used during speech processing (Pickering & Garrod, 2013) and reading (Britt et al., 1992; Rayner et al., 1992; Van Berkum et al., 2005).

Figure 6.1c shows a schematic diagram of expectation-based output selection during explicit detection of an incongruity; that is, when the two outputs are considered to be different enough to warrant further inspection by the reader. Expectation-based output selection is not used, and the word-meaning retrieval is added to the meaning representation. The incomplete analysis, which contains the incongruent word, makes the reader aware of the incongruity. The fact that

expectation-based output selection is not used further signals the presence of an incongruity, and the fact that a more thorough inspection of the text is needed. Heuristics usage during reading is halted, and subsequent reading relies on a word-meaning retrieval of the incongruent word and the rest of the text. As mentioned in the introduction, a successfully detected incongruity is considered to be a case of explicit detection, and signifies fixating on an incongruity and consciously reacting to it. A failure to notice any issues in the text despite having fixated the incongruity is considered to be a case of implicit detection. The participant makes a fixation on the word, but no conscious attention is drawn to the incongruity in order for heuristics usage to continue during reading.

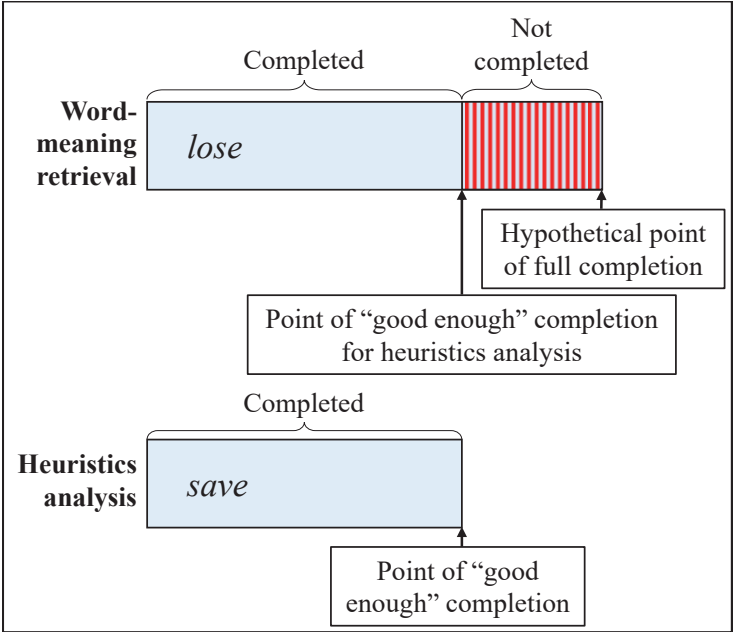


Figure 6.1a. Schematic diagram of the two different outputs created by the two parallel reading processes after fixating on an incongruent word
The figure represents a schematic diagram of a fixation on the incongruent critical word in the example text (18) above.

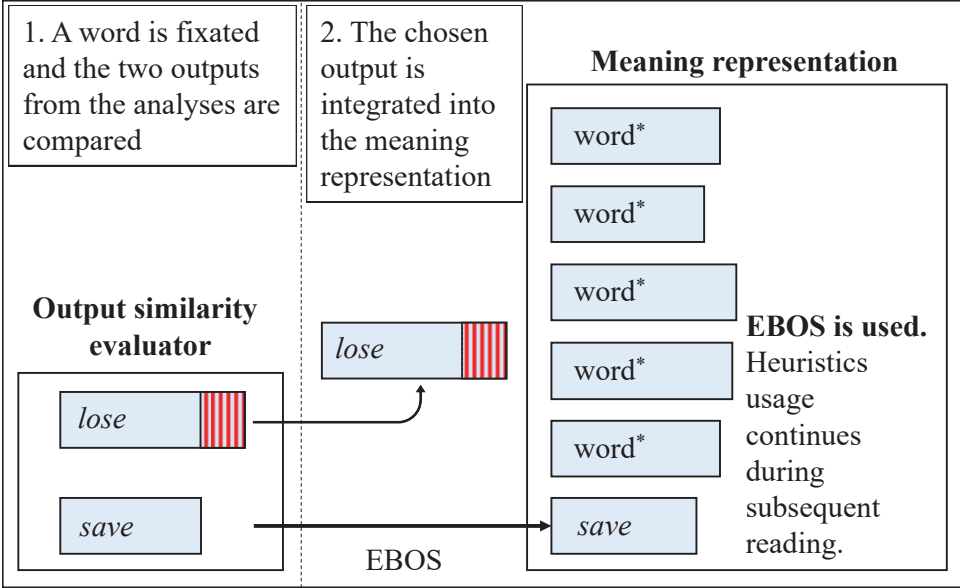


Figure 6.1b. Schematic diagram of the output similarity evaluator during normal operation
 The difference between the two outputs does not elicit a reaction, and expectation-based output selection (EBOS) is therefore used. The * denotes one small piece of the parsing process that is entered into the ongoing meaning representation. The example word in the figure is the incongruent critical word in the example text (18) above.

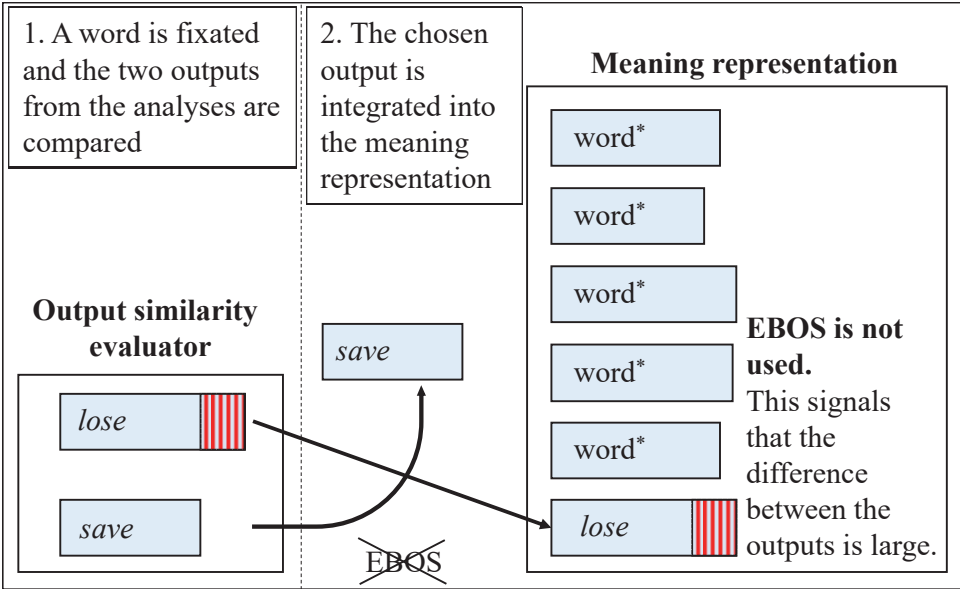


Figure 6.1c. Schematic diagram of the output similarity evaluator when the difference between the two outputs elicits a reaction
 The difference between the two outputs elicits a reaction, and expectation-based output selection (EBOS) is therefore not used. The * denotes one small piece of the parsing process that is entered into the ongoing meaning representation. The example word in the figure is the incongruent critical word in the example text (18) above.

Note that the model is simplified, describing only the selection process for whether or not the heuristics analysis is used during reading on a fixation-by-fixation basis. As mentioned, lingering effects in subsequent reading are assumed to be caused by the disregarded word-meaning retrieval. Concepts such as parafoveal processing (Hyönä, 2011), word skipping (Drieghe, Desmet, & Brysbaert, 2007), and lexical activation (Moss et al., 1995) all influence reading, but are disregarded in the model.

6.3.2 Hypotheses

In this study, disruptions to the reading process were expected both during explicit detection of the incongruity, and also during failures to explicitly detect the incongruity, which is referred to as implicit detection. The disruptions during explicit detection were hypothesized to be the result of the participant noticing the incongruity, and rereading the incongruent word and the rest of the text as a result. The disruptions during implicit detection were hypothesized to be the result of the disregarded output from the parallel reading process that contained the incongruent word. The experimental manipulation could be said to be large, since antonym pairs are maximally different within the same meaning dimension. This creates a large contrast between the two parallel outputs, which should create visible effects of both detection and non-detection of the incongruity, compared to when the correct word is read.

Implicit detections were classified much more conservatively in Experiment 2, thus favoring these results and ensuring that they accurately measured actual failures to detect the incongruent word. Evidence of implicit detection would be expected to be seen in eye movement effects similar to explicit detection, as lingering effects arising from the incongruent output from the parallel analysis (Bottoms et al., 2010; Christianson et al., 2001; Lau & Ferreira, 2005; Slattery et al., 2013; Tabor & Hutchins, 2004). A higher number of *revisits* to the incongruent word were expected (meaning the total number of times the specified text region was reentered), as well as longer total reading times compared to the congruent, control condition. In Experiment 2, it was expected that explicit detection would be seen in the headline and the last sentence of the text as well, due to rereading and reanalyzing the text. Commonly, reading disruptions result in longer first-pass reading times. However, after having read the incongruent sentence, an interruption would be evident in *shorter* first-pass reading times on the subsequent sentence, which is again congruent to the text, compared to the control condition. In addition, total reading times on the headline and last sentence, as well as total text presentation times, would be expected to be longer. This would signify increased rereading. During implicit detection, the partial, parallel output with the incongruent word was expected to produce lingering effects similar to explicitly detected incongruities on the incongruent word, but not on the headline and last sentence. All effects of implicit detection should be smaller but still significant when compared to the control condition. This is why total reading

times on the headline and the last sentence were not expected to change enough to produce significant effects.

To attempt to influence the total number of explicitly detected incongruities, the working memory capacity of the participant was measured in Experiment 2. A larger working memory capacity has been linked to better comprehension in several studies (see Daneman & Merikle, 1996, for an overview, but also see Van Dyke, Johns, & Kukona, 2014). Higher working memory capacity is therefore hypothesized to result in a higher number of explicitly detected incongruities.

A summary of the hypotheses and the hypothesized effects are given in Table 6.1 and Table 6.2.

Table 6.1. Hypotheses and hypothesized effects on the experimental measures compared to control condition, Experiment 1 (critical word) and Experiment 2 (critical word, last sentence of text, headline, and entire text)

Hypotheses are given both for how the proposed model reacts and for the changes in the participant's reading behavior. The hypotheses and effects supported by the experimental data are reported in **bold** in Table 6.8 in Section 6.6.

Detection type	Hypothesis compared to the control condition	Hypothesized effects compared to the control condition
Implicit detection	Model: The output similarity evaluator judges the two outputs from the two analyses as sufficiently similar, and uses expectation-based output selection to insert the output from the heuristics analysis into the meaning representation.	<u>On critical word:</u> - Longer total reading times and more revisits, but not to the extent that is found for explicit detection. - No difference in first fixation durations.
	Behavior: The participants continue their reading without consciously reacting to the incongruity. The output with the incongruent word, however, results in more rereading of the incongruent word. The lingering effects should also stop the reading of the subsequent sentence more quickly, since it is incongruent with the previous one. Also, reading of the entire text should take longer.	<u>On last sentence, headline, and entire text (Experiment 2):</u> - Shorter first-pass reading times on last sentence, but not to the extent that is found for explicit detection. - Longer whole-text reading times on entire text, but not to the extent that is found for explicit detection.
Explicit detection	Model: The output similarity evaluator judges the two outputs from the two analyses as different enough to require reanalysis. Expectation-based output selection is not used, and heuristics usage is halted in favor of conscious rereading of the sentence.	<u>On critical word:</u> - Longer total reading times and more revisits. - No difference in first fixation durations.
	Behavior: The participants halt their reading in favor of rereading the incongruent word. The effects should be observable in late measures only. The participants interrupt their reading of the last sentence more quickly, since it is incongruent with the previous one, and begin rereading the previous material. The participants reread the last sentence, headline, and entire text to try to make sense of the incongruity.	<u>On last sentence, headline, and entire text (Experiment 2):</u> - Shorter first-pass reading times and longer total reading times on last sentence. - Longer total reading times on headline. - Longer whole-text reading times on entire text.

Table 6.2. Hypothesis and hypothesized effects of working memory capacity on explicit detection, Experiment 2

Working memory capacity is represented by OSPAN score. The hypotheses and effects supported by the experimental data are reported in **bold** in Table 6.9 in Section 6.6.

	Hypothesis	Hypothesized effects
Working memory capacity	Higher working memory capacity is correlated with better text comprehension of text. Consequently, the participants should be better at identifying irregularities and errors in the text.	A higher number of explicitly detected incongruities for participants with high working memory capacity, as measured by OSPAN absolute score.

6.4 Experiment 1

Experiment 1 investigated the effects of both explicit and implicit detection of incongruities on the reading process for the critical word, using both early and late measures.

6.4.1 Method

6.4.1.1 Participants

Thirty-five native speakers of English (20 female; 18 to 45 years of age) participated in the experiment. The participants were recruited through student lists and advertisements at Lund University. All participants reported having normal or corrected-to-normal vision. One participant was excluded from the results due to technical problems with the eye-tracking equipment. Another participant was removed since the participant began searching for incongruities at an early stage of the experiment, as stated in the post-experiment interview. This resulted in 33 participants in total (19 female; $M=28.2$ years, $SD=6.73$). The participants were naïve to the purpose of the experiment.

6.4.1.2 Apparatus

Eye movement data were recorded binocularly at a sampling rate of 500 Hz using a Hi-Speed 1250 tower-mounted eye-tracker from SensoMotoric Instruments (Teltow, Germany). The experiment took place at the Lund University Humanities Laboratory at the Centre for Languages and Literature. The head of the participant was stabilized using a chin and forehead rest during recording. The distance from each participant's eyes to the stimulus monitor was approximately 70 cm. Stimuli were displayed on a Samsung SyncMaster 245T 24" widescreen LCD display at a resolution of $1,920 \times 1,200$ pixels (52×32.5 cm, 40.8×26.1 degrees of visual angle at a distance of 70 cm), with a refresh rate of 60 Hz. The eye-tracking system was controlled with SMI iView X 2.7 running on a data recording computer. Stimulus presentation, a 9-point calibration, and a 4-point validation of the calibration accuracy were handled using SMI Experiment Center 3.0 running on a separate stimulus computer.

6.4.1.3 Design and stimuli

The stimuli consisted of 16 short news texts in English, each followed by an accompanying comprehension question (see Appendix 6.A). All texts were short and simple (number of words $M=53.9$; $SD=9.76$). The antonymic incongruities used were common antonymic partners such as *high* and *low*. An example stimulus text can be found in (18) in Section 6.3.

Each text was adapted to on-screen reading by using a large font, Times New Roman 36-point (approximately 5.0×10 mm, approximately 0.40×0.80 degrees of visual angle at a distance of 70 cm), with the individual lines of text separated by two blank lines, in order to facilitate accurate eye-tracking measures (see Appendix 6.C). The headline and the first half of the text always established the context prior to the incongruity. In order to emulate antonym usage in natural language as closely as possible, word class distribution of antonyms in the stimuli reflected the distribution found in a dictionary and in corpus data (Paradis et al., 2009). The critical word was always located in the last third of the text in order for the context to be established prior to the incongruity. The critical word was also preceded and followed by at least two words, in order to measure fixation durations on the word preceding the critical word and the word directly following it. To control for word length, antonyms were matched for length within pairs, given the constraints of needing to be clear antonyms. The Corpus of Contemporary American English (COCA) (Davies, 2008) was used to control for frequency, ensuring that the two groups were as similar as possible, again, given the constraints of needing to be clear antonyms.

Each text was accompanied by a statement about the text, and the answer options *true*, *false*, *true – ambiguous text* and *false – ambiguous text*. The statement was designed to be easy to answer based on the text and addressed the overall topic of the text. The statement given for (18) is *It is getting harder to pay with cash at petrol stations nowadays*, and the correct answer is *true*.

The data were divided into three conditions post hoc: explicit detection, implicit detection, and control. All texts in which participants indicated that they noticed the antonymic incongruity when prompted during the post-experiment interview were classified as explicitly detected trials.

The texts were distributed over two experimental lists, and the order of appearance of the texts was randomized with the constraint that two incongruities could never occur in sequence. Participants were assigned to the lists in an alternating manner. For each list, six of the texts contained an antonym incongruity, and six belonged to the control condition using the word correctly corresponding to the context. The six manipulated texts in one list were used in an undistorted form in the other, thus functioning as controls for the manipulated texts, and vice versa. Both lists also contained four filler texts.

6.4.1.4 Procedure

Each participant was introduced to the experiment and recorded individually. The experimental session lasted around 20 minutes, but no time restrictions were imposed during the experiment. The participants were welcomed to the study, and completed a pre-experiment questionnaire in which their age and gender was noted. The participants were informed that they would participate in a reading study examining how pupil dilation is affected by the responses to the questions. They were also informed that their participation was completely voluntary and that they could

interrupt the experiment at any time. The procedure of the experiment was then explained by the experimenter, but written instructions were also provided during the eye-tracking recording. The participants were told that the experiment consisted of small news items, each followed by a simple true or false question. They were instructed to read the texts as quickly and as accurately as possible, press the space-bar when they were finished, and answer the subsequent question using the mouse. Memorizing the texts was not necessary, since the subsequent questions were straightforward and did not focus on specific details in the text. In addition, they were instructed to tick the *ambiguous* option if the text felt ambiguous to them. The idea was to not reveal the real research objective, while still giving the participants the ability to indicate explicit detection on-line. Nevertheless, the responses to the post-experiment questions formed the most important basis for classification into explicit and implicit detection.

Having confirmed that they understood the instructions, participants were seated in front of the eye-tracker and calibration was started. The calibration was repeated until the offset reported by Experiment Center was less than 0.5 degrees of visual angle in both the horizontal and the vertical directions.

The experiment began with the participants completing a questionnaire about their age, gender, and educational background. The instructions were repeated, followed by a practice trial consisting of one text and one question before the experimental texts and questions were displayed. No eye tracking was performed during the responses to questions.

After the eye-tracking portion was concluded, an interview was conducted with the participants. First, the participants were asked if they had “Any particular problems with any of the texts and questions off the top of your head?” Any text mentioned was recorded. Specific questions were then asked about each of the critical texts, for example “What did you think about the text about the petrol station situation? Any problems?” The underlined topic in the example above was replaced accordingly. Cases in which a text had been classified as *ambiguous* by participants during reading were also specifically addressed, with the question “You stated that the text about the petrol station situation was ambiguous. Why?” The interview was conducted in order to find out whether the incongruities had been explicitly detected. Importantly, the risk that participants incorrectly recalled noticing incongruities post hoc was minimized since the probing questions largely used the topics, and not the incongruities themselves. If the reason given by a participant was unrelated to the incongruity, the trial was classified as implicitly detected (this term is used to mean that effects of implicit detection were found). In the last question, the participants were asked what they thought was the purpose of the experiment. The classification procedure can be seen in Figure 6.2.

After the interview, the participants were informed of the real purpose of the experiment, and were shown an example incongruity. They were given a debriefing

form to take home, thanked again, and given an instant lottery ticket for participating in the study.

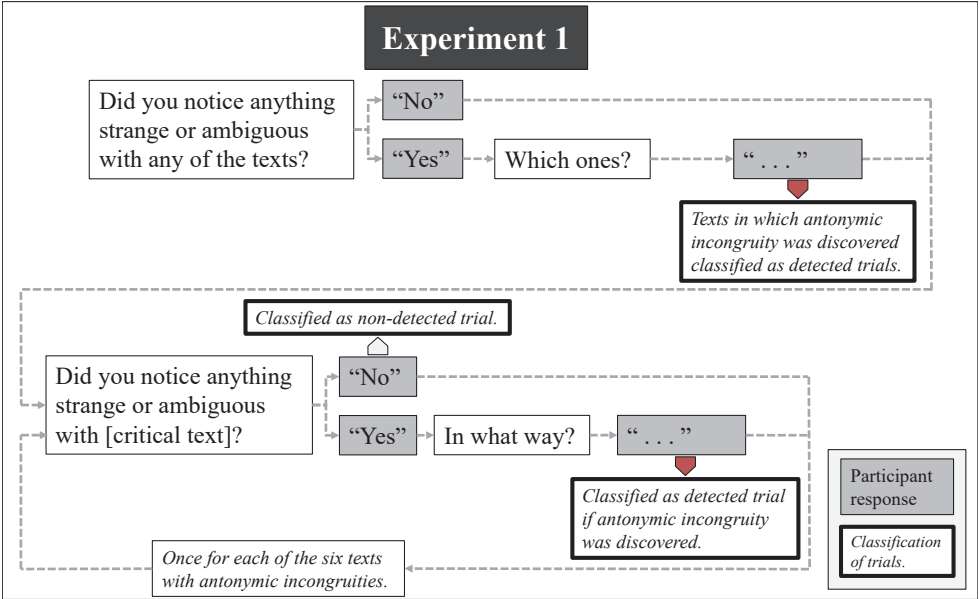


Figure 6.2. Flow chart of post-experiment interview, Experiment 1
The figure explains the classification procedure for explicitly and implicitly detected trials.

6.4.1.5 Data analysis

Velocity-based high-speed event detection was performed using SMI BeGaze 2.7 with default settings (peak velocity threshold=40°/s, minimum fixation duration=50 ms), which transformed the raw data into fixations and saccades. The recordings from the right eye were prioritized, since the right eye is dominant for the majority of people; the left eye, however, was used in trials in which data from the right eye were unavailable (26% of the cases). Only trials in which the critical word was fixated at least once were included in the models.

The data processing, figures, and statistical analyses were made using R (R Development Core Team, 2015) and the lme4 package (Bates et al., 2015) in R Studio (0.98.504). First fixation durations were log-transformed to better fit a Gaussian distribution. Since revisits could not be coerced into a Gaussian distribution, they were coded into a binomial distribution, with 0 signifying *no revisit to the target area*, and 1 signifying *one or more revisits*. A generalized linear mixed effects model was used to examine revisits to the critical word, while linear mixed effect models were used to examine the effect of sentence manipulation on first fixation durations and total reading times. A fixed factor with three levels (explicit detection, implicit detection, and control) was used, using treatment coding with the

control condition as the baseline. Models used random slopes of condition for participants and items, with intercept-slope correlations for participants and items. In cases in which the model did not converge, the model was gradually reduced. This was the case for revisits, which were therefore examined using only random intercept for participants and items. The lmerTest package (2.0-29) was used to estimate p -values with Satterthwaite approximation.

6.4.2 Results

All statistical models used log-transformed data, but in order to improve readability, no log-transformation was performed on the data in the figures.

6.4.2.1 Performance

Table 6.3 shows detection of incongruities for Experiment 1, excluding the trials where the critical word was not fixated. Results are shown as both the total number of trials, and the proportion of detection type (implicit or explicit), with the distribution between implicit and explicit detection being fairly evenly distributed.

Table 6.3. Implicit and explicit detection of incongruities, Experiment 1

Only trials where the critical word was fixated at least once are reported, with both the total number of trials (count) and the proportion of trials in the critical condition (percent) shown.

	Trials	Proportion
Control condition	158	-
Implicit detection	70	45.2
Explicit detection	85	54.8

6.4.2.2 First fixation durations

Figure 6.3 shows mean first fixation durations on the critical word (log-transformed). First fixation durations were significantly shorter on explicitly detected incongruities (Estimate=-0.114, $SE=0.055$, $t=-2.08$, $p<.05$) compared to the control condition. No significant difference was found between implicitly detected incongruities and the control condition (Estimate=-0.018, $SE=0.064$, $t=-0.28$, $p=.78$). All participant eye movement results can be found in Table 6.4.

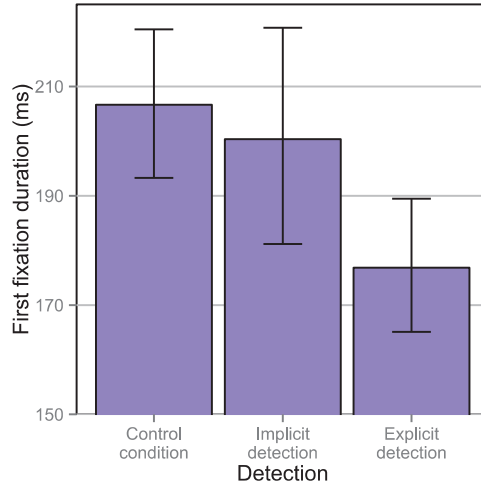


Figure 6.3. Mean first fixation durations (ms) on the critical word for the control condition, implicit detection, and explicit detection, Experiment 1
Error bars denote standard errors of the mean.

Table 6.4. Mean eye movement results on the critical word, Experiment 1

All measures are given with standard errors in parentheses. Statistically significant differences when comparing with the control condition are presented in **bold**; * indicates that the LMM reported $p < .05$, ** indicates $p < .01$, *** indicates $p < .001$.

	Control condition	Implicit detection	Explicit detection
<i>First fixation duration (ms)</i>	207 (6.66)	200 (9.65)	177 (6.38)*
<i>Revisits (count)</i>	0.268 (0.046)	0.493 (0.106)*	1.52 (0.135)***
<i>Revisits (binomial)</i>	0.204 (.032)	0.310 (.055)*	0.759 (.46)***
<i>Total reading time (ms)</i>	265 (13.2)	297 (18.5)	505 (30.7)***

6.4.2.3 Revisits and total reading times

Figure 6.4 shows mean number of revisits to the critical word. The number of revisits were significantly larger to both explicitly detected (Unaltered distribution: Estimate=1.22, $SE=0.117$, $t=10.4$, $p<.0001$; Binomial distribution: Estimate=2.87, $SE=0.398$, $t=7.23$, $p<.0001$) and implicitly detected incongruities (Unaltered distribution: Estimate=0.260, $SE=0.125$, $t=2.08$, $p<.05$; Binomial distribution: Estimate=0.866, $SE=0.387$, $t=2.24$, $p<.03$), compared to the control condition.

Figure 6.5 shows mean total reading times on the critical word. Explicitly detected incongruities were significantly longer (Estimate=0.573, $SE=0.116$, $t=4.95$, $p<.001$), compared to the control condition. No significant differences were found for implicit detection (Estimate=0.096, $SE=0.104$, $t=0.92$, $p=.38$).

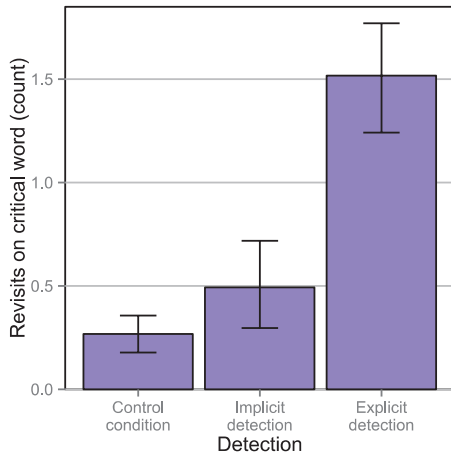


Figure 6.4. Mean number of revisits (count) to the critical word for the control condition, implicit detection, and explicit detection, Experiment 1. Error bars denote standard errors of the mean.

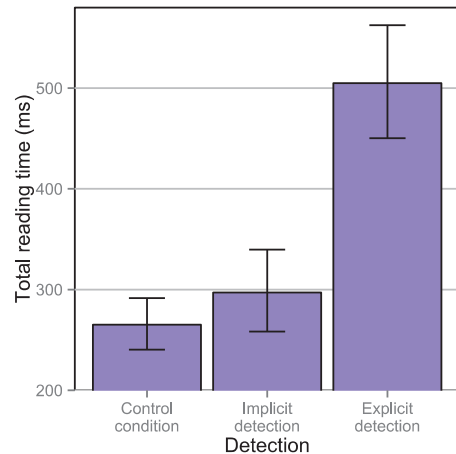


Figure 6.5. Mean total reading times (ms) on the critical word for the control condition, implicit detection, and explicit detection, Experiment 1. Error bars denote standard errors of the mean.

6.4.3 Discussion

Incongruities not explicitly detected by participants did not result in significant differences in first fixation durations on the critical word compared to the control condition. This suggests that participants did not react to the incongruity when they first saw it. The input was instead interpreted in line with what the readers expected from the text, resulting in what the participants interpreted as a correct text. This is not surprising, since the same result has been found in a large number of studies investigating failure to explicitly detect incongruities in text (Barton & Sanford, 1993; Bohan & Sanford, 2008; Ferreira et al., 2002; Hannon & Daneman, 2004; Kamas et al., 1996; Park & Reder, 2004; Reder & Kusbit, 1991). When analyzing the effects in late measures for the incongruities that were not explicitly detected by participants in this study, however, very interesting findings appeared. Significantly higher numbers of revisits to the critical word were found compared to the control condition, suggesting that the incongruities were in fact implicitly detected. The parallel processes that govern reading, as hypothesized by Ferreira and Patson (2007) among others, deliver conflicting outputs that have to be integrated into the meaning representation of the text. The preferred output comes from the heuristics based on context and expectations, while the other process uses a word-meaning retrieval. Expectation-based output selection uses the interpretation from the heuristics analysis when creating a meaning representation of the text, since the incongruity is seen as involuntary or irrelevant. The word expected in the context is used instead of the

incongruent word, producing an interpretation that does not correctly correspond to the input, and reading continues. The effects in late measures can be compared to the digging-in effect, which shows that reading times increase when an incorrect analysis is allowed to stay uncorrected for longer periods of time (Tabor & Hutchins, 2004). This could suggest the presence of an incorrect analysis when participants in the current study failed to explicitly detect an incongruity. The results can also be related to the lingering effects of misinterpreting garden path sentences, which show that incorrect parsing of a text produces an increase in incorrect answers to content questions, even if the text is immediately reanalyzed (Christianson et al., 2001; Ferreira et al., 2001). It is conceivable to assume that these effects arise from an incorrect analysis of the word, which is outside of the explicit awareness of the reader. It should be pointed out that not detecting Moses illusion-type incongruities (Reder & Kusbit, 1991) also increases the usage of the incorrect information in comprehension questions (Bottoms et al., 2010). This suggests that the incorrect information is partially processed, but corrected without the conscious involvement of the reader, so that he or she does not react to the error. In other words, the output containing the incorrect information is disregarded in favor of the expected output. The incongruity used in this study can be considered to be a fairly substantial error, not a mere spelling error. Because of this, the disregarded output containing this error therefore produces lingering effects largely similar to the effects found for explicit detection in late measures.

Explicitly detected incongruities showed significantly shorter first fixation durations on the critical word compared to the control condition. The results mirror studies using the Moses illusion, in which explicitly detected anomalies are read more quickly than ones that are not explicitly detected (Oostendorp & De Mul, 1990; Reder & Kusbit, 1991). Nevertheless, these results are somewhat surprising, since they are not in line with Braze et al. (2002). Braze et al. (2002) found no significant eye movement effects for the critical word for pragmatic anomalies, but did for syntactic anomalies, for which an increased number of first-pass regressions were found. The decrease in first fixation durations when detecting an incongruity can be interpreted as the effect of the interruption of the use of the heuristics suggested by the good enough approach, and a switch to a thorough rereading of the text. This switch occurs when the encountered incongruity is seen as an error that requires conscious involvement from the reader during the reading process. That this is evident in early measures is surprising, however, since both Bohan and Sanford (2008) and Daneman et al. (2007) found that the effects of detection were only evident in late eye-tracking measures for semantic incongruities.

In late measures, explicitly detected manipulations showed significantly higher numbers of revisits and increased total reading times on the critical word compared to the control condition. This demonstrates that participants reread the text in order to make sense of it. Normally, the output similarity evaluator uses expectation-based output selection when a conflict arises between two outputs, resulting in the heuristics

analysis being used in the meaning representation of the text (for a thorough explanation of these terms, see Section 6.3.2). If the conflict is deemed too great, however, the word-meaning retrieval is instead used, the incongruity is explicitly detected, and more thorough inspection is initiated without reliance on heuristics. In Braze et al. (2002), longer first-pass reading times were found on the word after a pragmatic anomaly, showing that reading was affected by detection of an anomaly, but that this detection was only visible during subsequent reading.

In order to evaluate and extend the results in Experiment 1, the texts were further controlled so that analyses could be conducted on two additional areas: the headline and the last sentence. Since the effects for implicit detection were smaller than the effects for explicit detection, the analysis of implicit detection was favored in Experiment 2. This was achieved by making classification of the implicitly detected trials more conservative and stringent, with all trials in which participants expressed that they noted something strange in a text classified as explicitly detected, regardless of the reason. With the use of this system of classification, the results for implicit detection were very unlikely to be contaminated by the inclusion of behavior from explicit detection. The classification of explicit detection can thus be considered more liberal, since the precise details of the detections are largely unknown, only that there was “something wrong” with the text. As a consequence, results for implicit detections can be considered somewhat more accurate, while results for explicit detection can be considered somewhat less accurate. To evaluate the detection of incongruities, a working memory task was also incorporated into Experiment 2. A larger working memory capacity has been linked to better comprehension in several studies (see Daneman & Merikle, 1996, for an overview, but also see Van Dyke et al., 2014). Thus, the memory trace of the context should be stronger with a larger working memory capacity, and participants with larger working memory capacity should therefore be more likely to explicitly detect incongruities.

6.5 Experiment 2

Except for the differences explained in the Method section below, Experiment 2 was identical to Experiment 1.

6.5.1 Method

6.5.1.1 *Participants*

Sixty native speakers of English (40 female; 18 to 41 years of age) participated in the experiment. Two participants were removed since they started searching for incongruities at an early stage of the experiment, as stated during the post-experiment

interview. This resulted in 58 participants in total (39 female; $M=27.4$ years, $SD=5.92$).

6.5.1.2 Apparatus

Updated software was used for the eye-tracking recordings (SMI iView X 2.8, SMI Experiment Center 3.4, SMI BeGaze 3.4). An automated operation span (OSPAN) working memory task was presented on a separate computer using E-Prime 2.

6.5.1.3 Design and stimuli

All texts in which participants indicated that they noticed something strange or ambiguous when prompted during the post-experiment interview were classified as explicitly detected trials, regardless of what the participant reacted to.

The stimuli consisted of 16 short news texts in English, each followed by an accompanying comprehension question (see Appendix 6.B). All texts were short and simple (number of words $M=62.7$; $SD=5.39$). Each text was adapted to on-screen reading by using a large font, Arial 36-point (approximately 5.0×10 mm, approximately 0.40×0.80 degrees of visual angle at a distance of 70 cm), with the individual lines of text separated by two blank lines, in order to facilitate accurate eye-tracking measures (see Appendix 6.C). The critical word was always located in a separate sentence, making the sentence with the critical word internally congruent, but immediately identifiable as incongruent with regard to the entire text. Each sentence containing a critical word was always followed by exactly one sentence.

An automated OSPAN working memory task (Unsworth, Heitz, Schrock, & Engle, 2005) was completed by the participants on a separate computer. Participants were presented with a single letter followed by a math problem. After solving the math problem and continuing to the next screen, the participants were asked to indicate whether a presented solution to the math problem was correct or not. After the set size ranging from three to seven single letters followed by math problems and presented solutions, they were asked to correctly recall all letters in order of presentation. Working memory capacity is reflected in the measure *OSPAN absolute score*, signifying the number of perfectly recalled letter sets (max 75).

6.5.1.4 Procedure

The experimental session lasted around 45 minutes, but no time restrictions were imposed during the experiment.

In the post-experiment interview, incongruities were classified as explicitly detected when a participant stated that something was strange with the specific text, regardless of what the participant reacted to, making classification of implicit detection very conservative. The classification process can be seen in Figure 6.6. Cases in which a text had been specified as *ambiguous* by participants in a comprehension question during reading were also classified as explicitly detected.

The final task for participants was the OSPAN task, completed on a separate computer.

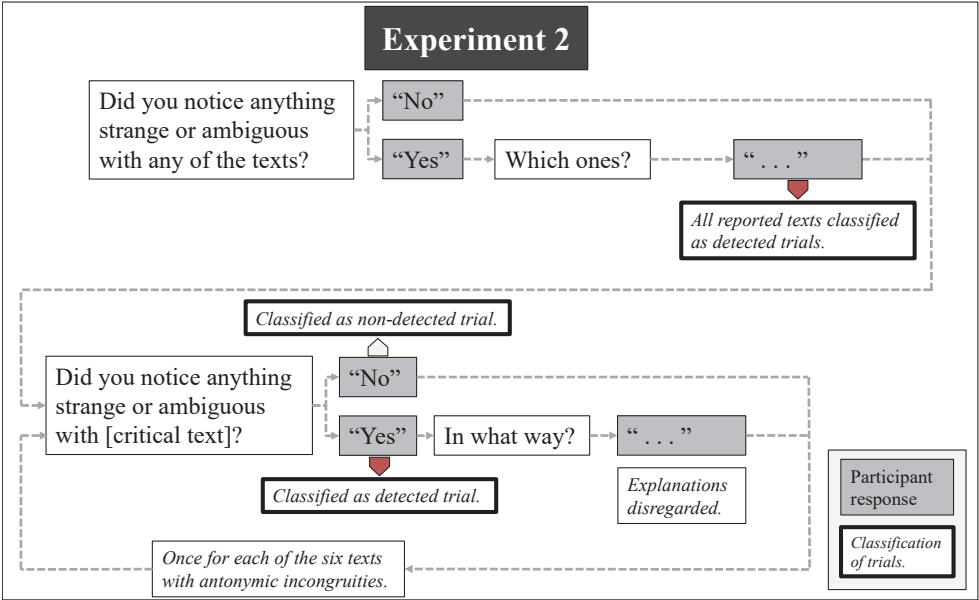


Figure 6.6. Flow chart of post-experiment interview, Experiment 2
The figure explains the classification procedure for explicitly and implicitly detected trials.

6.5.1.5 Data analysis

Only the recordings from the right eye were used. Only trials in which the critical word was fixated at least once were included in the models. First-pass reading times and whole-text reading times were log-transformed to better fit a Gaussian distribution. Linear mixed effect models were used to examine the effect of the sentence manipulation on total reading times and first-pass reading times on the last sentence, total reading times on the headline, and whole-text reading times. A linear model was used to examine the number of explicit detections of incongruities using OSPAN absolute scores as a predictor. All models converged using random slopes of condition for participants and items, with intercept-slope correlations for participants and items.

6.5.2 Results

6.5.2.1 Performance

Table 6.5 shows detection of incongruities for Experiment 2, excluding the trials where the critical word was not fixated. Results are shown as both the total number of trials, and the proportion of detection type (implicit or explicit), with the distribution

between implicit and explicit detection being fairly evenly distributed. Note that the classification of implicit detection was more conservative and stringent in Experiment 2 compared to Experiment 1.

Table 6.5. Implicit and explicit detection of incongruities, Experiment 2

Only trials where the critical word was fixated at least once are reported, with both the total number of trials (count) and the proportion of trials in the critical condition (percent) shown.

	Trials	Proportion
Control condition	216	-
Implicit detection	123	48.4
Explicit detection	131	51.6

6.5.2.2 First fixation durations on words and first-pass reading times on sentences

All participant eye movement results for the critical word can be found in Table 6.6, and results for all other regions can be found in Table 6.7. No significant differences were found for mean first fixation durations on explicitly (Estimate=0.053, $SE=0.051$, $t=1.05$, $p=.30$) or implicitly detected incongruities (Estimate=0.025, $SE=0.052$, $t=0.49$, $p=.63$), compared to the control condition.

Figure 6.7 shows mean first-pass reading times on the last sentence of the text, i.e., the sentence after the sentence containing the critical word. All plots show the control condition, explicitly detected incongruities, and implicitly detected incongruities. First-pass reading times were significantly shorter in the condition with explicitly detected (Estimate=-0.329, $SE=0.100$, $t=-3.28$, $p<.01$) and implicitly detected incongruities (Estimate=-0.316, $SE=0.091$, $t=-3.48$, $p<.005$), compared to the control condition.

Table 6.6. Mean eye movement results on the critical word, Experiment 2

All measures are given with standard errors in parentheses. Statistically significant differences when comparing with the control condition are presented in **bold**; * indicates that the LMM reported $p<.05$, *** indicates $p<.001$.

	Control condition	Implicit detection	Explicit detection
First fixation duration (ms)	202 (5.75)	200 (7.42)	208 (7.69)
Revisits (count)	0.322 (0.037)	0.992 (0.111)***	1.65 (0.151)***
Revisits (binomial)	0.227 (.023)	0.496 (.039)***	0.630 (.038)***
Total reading time (ms)	288 (14.9)	433 (36.6)**	579 (45.6)**

Table 6.7. Mean eye movement results on last sentence, headline, and entire text, Experiment 2

All measures are given in milliseconds with standard errors in parentheses. Statistically significant differences when comparing with the control condition are presented in **bold**; * indicates that the LMM reported $p<.05$, ** indicates $p<.01$, *** indicates $p<.001$.

	Control condition	Implicit detection	Explicit detection
Last sentence			
First-pass reading time	1,149 (71.3)	906 (91.9)**	743 (80.0)**
Total reading time	2,325 (92.4)	2,670 (164)	2,783 (154)**
Headline			
Total reading time	1,997 (89.3)	1,925 (108)	1,813 (111)
Entire text			
Whole-text reading time	20.4 (0.430)	22.0 (0.635)*	26.7 (0.829)***

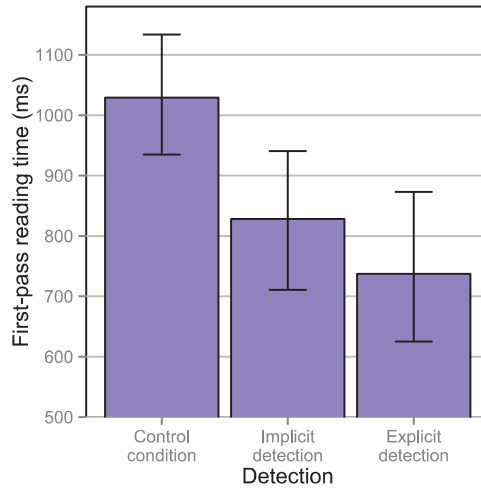


Figure 6.7. Mean first-pass reading times (ms) on the last sentence for the control condition, implicit detection, and explicit detection, Experiment 2
Error bars denote standard errors of the mean.

6.5.2.3 Revisits and total reading times

Figure 6.8 shows mean number of revisits to the critical word (in the binomial coding explained in Section 6.5.1.5, above). The number of revisits were significantly larger to both explicitly (Unaltered distribution: Estimate=1.25, $SE=0.263$, $t=4.76$, $p<.0001$; Binomial distribution: Estimate=2.16, $SE=0.441$, $z=4.90$, $p<.0001$) and implicitly detected incongruities (Unaltered distribution: Estimate=0.605, $SE=0.173$, $t=3.50$, $p<.005$; Binomial distribution Estimate=1.47, $SE=0.333$, $z=4.41$, $p<.0001$) compared to the control condition.

Figure 6.9 shows mean total reading times on the critical word. Total reading times were significantly longer on both explicitly detected (Estimate=0.531, $SE=0.133$, $t=4.00$, $p<.005$) and implicitly detected incongruities (Estimate=0.277, $SE=0.092$, $t=3.02$, $p<.01$) compared to the control condition.

No significant differences were found in total reading times on the headline in the conditions with explicitly (Estimate=0.030, $SE=0.100$, $t=0.37$, $p=.72$) or implicitly detected incongruities (Estimate=-0.069, $SE=0.091$, $t=-1.02$, $p=.32$), compared to the control condition.

Figure 6.10 shows mean total reading times on the last sentence of the text, i.e., the sentence after the sentence containing the critical word. Total reading times were significantly longer in the condition with explicitly detected incongruities (Estimate=0.258, $SE=0.087$, $t=2.99$, $p<.01$), compared to the control condition. No significant differences were found in the condition with implicitly detected incongruities (Estimate=0.142, $SE=0.083$, $t=1.71$, $p=.13$).

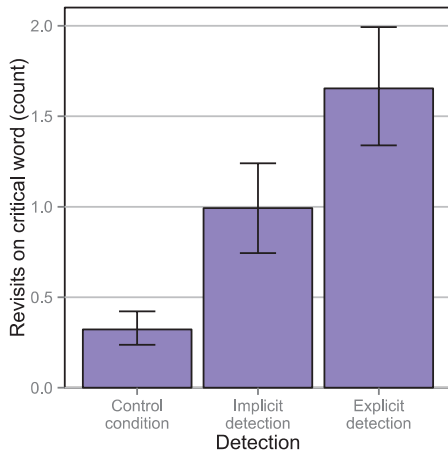


Figure 6.8. Mean number of revisits (count) to the critical word for the control condition, implicit detection, and explicit detection, Experiment 2
Error bars denote standard errors of the mean.

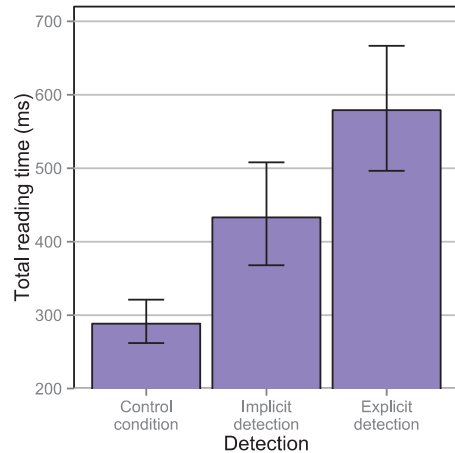


Figure 6.9. Mean total reading times (ms) on the critical word for the control condition, implicit detection, and explicit detection, Experiment 2
Error bars denote standard errors of the mean.

6.5.2.4 Whole-text reading times

Figure 6.11 shows mean whole-text reading times. Whole-text reading times were significantly longer in the conditions with both explicitly detected (Estimate=0.267, $SE=0.045$, $t=5.90$, $p<.0001$) and implicitly detected incongruities (Estimate=0.065, $SE=0.028$, $t=2.34$, $p<.03$), compared to the control condition.

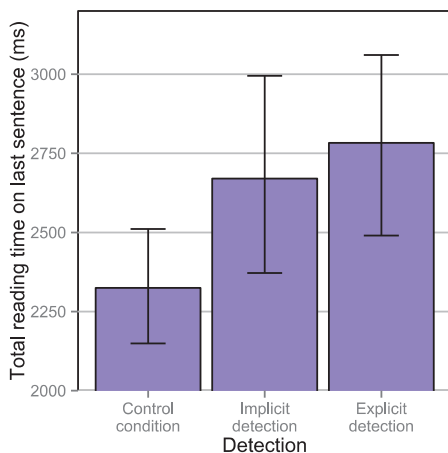


Figure 6.10. Mean total reading times (ms) on last sentence of text for the control condition, implicit detection, and explicit detection, Experiment 2
Error bars denote standard errors of the mean.

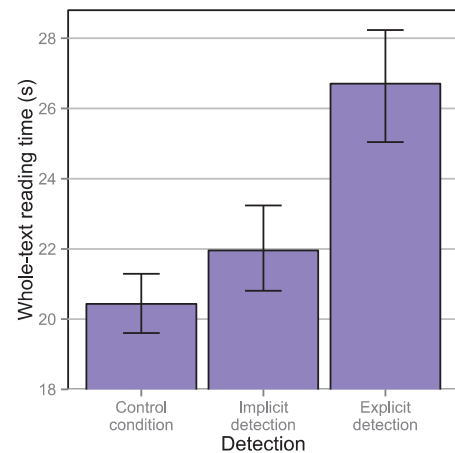


Figure 6.11. Mean whole-text reading times (s) for the control condition, implicit detection, and explicit detection, Experiment 2
Error bars denote standard errors of the mean.

6.5.2.5 Working memory capacity

Figure 6.12 shows the number of explicitly detected incongruities depending on the participants' OSPAN absolute scores. No significant differences were found on the total number of explicit detections depending on OSPAN absolute score (Estimate=0.003, $SE=0.009$, $t=0.33$, $p=.74$). OSPAN absolute score did not significantly affect first fixation durations (explicit detection: Estimate=0.004, $SE=0.002$, $t=1.77$, $p=.08$; implicit detection: Estimate=0.003, $SE=0.002$, $t=1.05$, $p=.29$), total reading times (explicit detection: Estimate=0.001, $SE=0.004$, $t=0.38$, $p=.71$; implicit detection: Estimate=-0.006, $SE=0.004$, $t=-1.48$, $p=.14$), or revisits (explicit detection: Estimate=0.004, $SE=0.014$, $t=0.27$, $p=.79$; implicit detection: Estimate=-0.002, $SE=0.015$, $t=-0.12$, $p=.90$) on the critical word.

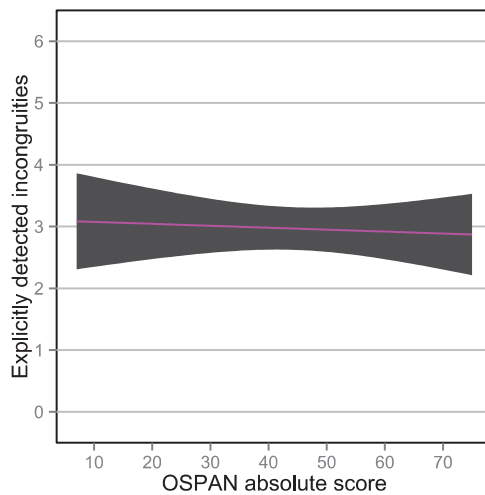


Figure 6.12. Mean number of explicitly detected incongruities depending on OSPAN scores, Experiment 2
Regression line includes 95% confidence interval.

6.5.3 Discussion

Implicitly detected critical words caused no significant differences in first fixation durations on the critical word compared to the control condition, reflecting the results in Braze et al. (2002), Bohan and Sanford (2008), and Daneman et al. (2007). In late measures, however, implicitly detected incongruities showed a significantly higher numbers of revisits and longer total reading times on the critical word compared to the control condition. This mirrors the results from explicitly detected incongruities in the current study, and contradicts Bohan and Sanford (2008) and Daneman et al. (2007), who found no effects during failure to explicitly detect incongruities. In Experiment 2, classification of implicit detection was performed as

conservatively as possible in order to make sure the results were caused by actual failures to detect the incongruities during normal reading. Thus, initial text comprehension may compensate for erroneous items during the initial stages of the analysis without drawing attention to the error. An accrued cost is shown in later measures, however, strongly suggesting that the incongruities were implicitly detected. In addition, first-pass reading times on the sentence following the sentence with the incongruity are significantly shorter for texts with both explicitly and implicitly detected incongruities compared to the control conditions. This shows that participants moved to other areas from the last sentence more quickly, regardless of whether they were aware of the error or not; this further suggests a similar effect on the reading process for both explicitly and implicitly detected incongruities. Also, texts with both explicitly and implicitly detected incongruities showed significantly longer whole-text reading times, when compared to the control condition. This shows that they were read for longer amounts of time.

Explicitly detected incongruities showed no significant effects in first fixation durations on the critical word, but several effects in late measures; specifically, significantly higher numbers of revisits, and longer total reading times compared to the control condition. Furthermore, texts with explicitly detected incongruities showed significantly longer total reading times on the last sentence of the text, but not on the headline. This suggests an interruption to the reading process that is visible in late measures. This mirrors the results in Braze et al. (2002), in which reading pragmatically anomalous words did not result in early effects, but first-pass reading times in the region *following* the incongruity increased compared to the control condition. This is further strengthened by the findings in Bohan and Sanford (2008) and Daneman et al. (2007), in which effects of detection were also only evident in late eye-tracking measures. When participants in the current study reread the text after explicit detection, however, no effects were found for the headline. This is somewhat surprising, since information in headlines usually gives a succinct summary of the text. Thus, when making sense of an incongruity, it would seem logical to return to the headline. The stimuli in Experiment 2 were constructed so that the incongruities were present in grammatically correct sentences that always clearly contradicted the information given in the headline and the sentence following the incongruent sentence, in addition to the general context.

Evaluating comprehension using comprehension questions was not deemed viable for the stimuli texts, since they were very short. In comparison, comprehension was evaluated using five multiple-choice questions after each text in Chapter 3, which would be very difficult with the much shorter texts used in Chapter 6. Better working memory capacity, however, should allow for a stronger memory trace of the context, since it has been linked to better comprehension (see Daneman & Merikle, 1996, for an overview, but also see Van Dyke et al., 2014). According to the good enough approach, however, the comprehension system only delivers a representation that is good enough, using less information than previously assumed (Ferreira & Patson,

2007). In light of this, it is not surprising that higher OSPAN absolute scores do not significantly affect the number of explicitly detected incongruities in the current study. This could signify that working memory capacity does not affect the integration of conflicting processing routes during reading. No interactions with working memory were found for first fixation durations, total reading times, or revisits to the critical word in either condition. It is therefore possible that expectation-based output selection is not strongly linked to comprehension. An alternative explanation is that comprehension is better predicted by other measures such as IQ, as pointed out by Van Dyke et al. (2014). This explanation, however, would imply that the integration of conflicting interpretations is dependent on comprehension to a high degree. Importantly, the results regarding working memory capacity in this study are insufficient to base any firm conclusions upon, since they were non-significant.

6.6 General discussion

Two eye-tracking experiments investigated how the explicit detection of antonym incongruities in texts affects reading, and how the costs of implicit detection are manifested in the reading process.

When an antonymic incongruity was not explicitly detected, no cost was seen in early measures in either of the two experiments, indicating that participants read the word as if it corresponded correctly to the context. Effects were found in the late measures in both experiments, however, with the effects in Experiment 2 substantially greater than in Experiment 1. The results from the current study thus strongly suggest that an incongruity is analyzed during explicit detection, but—crucially—that it is partially analyzed during implicit detection as well. The output similarity evaluator uses expectation-based output selection to create the meaning representation of the text with the heuristics analysis. Since the incongruity in this study is substantial, however, the partial output from the word-meaning retrieval created lingering effects similar to explicit detection.

In spoken language processing, events such as speech disfluencies and filled pauses frequently occur, and are disregarded by listeners to a high degree (Pickering & Garrod, 2013). In fact, pauses and filled pauses actually help processing to a large extent (Bosker et al., 2014; Collard et al., 2008; Corley & Hartsuiker, 2011; MacGregor et al., 2010). It can therefore be said that based on previous experiences, listeners are not very sensitive to disfluencies and incongruities. It seems that the implicit strategy is to either disregard them, or to perform relevant tasks to actually benefit from them. Ferreira and Patson (2007) submit that similar processes are present for both written and spoken language, suggesting that heuristics are responsible for these effects in both cases. Indeed, tolerance for incoherent text has

been shown to be high, and textual incongruities have been shown to often be missed by people (Bohan & Sanford, 2008; Daneman et al., 2007; Ferreira & Engelhardt, 2009; Roberts et al., 2016; Sanford & Sturt, 2002). Importantly, speech disfluencies that are disregarded or “repaired” are not completely erased, but produce lingering effects in subsequent processing (Ferreira, Lau, & Bailey, 2004; Ferreira & Patson, 2007). Lingering memory effects from garden path sentences in text have been shown to arise from incorrect outputs that are not completely erased (Slattery et al., 2013). The digging-in effect also shows that incorrect analyses can produce longer reading times (Tabor & Hutchins, 2004). This is similar to how implicitly detected incongruities produced lingering effects in this study, further strengthening the argument for similar processes being used.

To my knowledge, experimental results for implicit detection during reading have not previously been produced. A possible explanation for the effects found in the current study is that the stimuli were not primarily constructed to avoid detection. Rather, the incongruities were designed to be immediately perceived as erroneous in isolation, while the context lowered the detection rates. This approach was successful, since detection rates in both experiments were around 50%. Antonyms were used as incongruities, since they can be considered both maximally different and maximally similar at the same time. This created a beneficial situation in which clearly erroneous and isolated incongruities avoided explicit detection, producing the lingering effects of a clearly erroneous analysis. Compared to Experiment 1, the classification of implicit detection in Experiment 2 was much more conservative, with only the cases in which participants stated there had been no errors in the text being classified as implicit detection. This made the classification of implicit detection more reliable, at the cost of explicit detection. It is therefore likely that the results did not arise from trials that were mistakenly classified as implicit detection, or trials in which participants had unrelated problems with the text. In addition, Daneman et al. (2007) suggested that the classification of trials in their experiments was possibly not sensitive enough to show effects of implicit detection, further highlighting the importance of stringent classification of trials.

When participants explicitly detected an antonymic incongruity, early effects were found in Experiment 1, but not in Experiment 2. The results in Experiment 1 indicated that antonymic incongruities behaved in the opposite way as compared to syntactic incongruities (Braze et al., 2002; Rayner et al., 2004), with shorter first fixation durations. However, the results in Experiment 2 indicated that antonymic incongruities behaved like semantic incongruities, with no effects found in early measures. In previous research, texts with Moses illusion-type incongruities were read more quickly when participants explicitly detected the incongruity, compared to when they were not detected (Oostendorp & De Mul, 1990; Reder & Kusbit, 1991). The manipulation in the Moses illusion is called a semantic illusion, but this knowledge hinges on cultural knowledge about historical or literary matters that have very little to do with the world knowledge that is needed to understand language

(Paradis, 2003). Notably, several eye-tracking studies have found no early effects for pragmatic anomalies (Braze et al., 2002) or semantic anomalies (Bohan & Sanford, 2008; Daneman et al., 2007). The divergent results in this study could be caused by the focus on implicit detection in Experiment 2, with the classification procedure of implicit detection being more conservative compared to Experiment 1. It is also possible, however, that the absence of early effects for explicit detection in Experiment 2 was more reliable due to the larger number of participants.

The results in late measures for explicit detection in this study correlate with previous research done on detected semantic anomalies and incongruities, showing large interruptions in the reading process when participants reread the text to examine it in more detail (Bohan & Sanford, 2008; Daneman et al., 2007). In the good enough approach, the reading process largely favors using an analysis made by heuristics based on context, expectations, and previous experiences, rather than a full morphosyntactic analysis of the text (Christianson et al., 2010). Both of these processes are argued to occur in parallel, and since the heuristics are assumed to be slightly faster, the word-meaning retrieval is not completed. In the case of incongruities, the two parallel processes deliver different outputs. During reading, one output is chosen by the output similarity evaluator to be included in the meaning representation, and the route taken by the reading process differs depending on how the violation is assessed. If the violation is considered an error that requires attention from the reader, the incomplete word-meaning retrieval is used, rather than expectation-based output selection, during the creation of the meaning representation of the text. Importantly, this word-meaning retrieval is assumed to not be completed fully when a word is seen for the first time. If it had been completed, this would be seen in longer first-pass reading times, which is not the case. Instead, it is noted that this completion is not present in early measures but rather in total reading times. Explicit detection and subsequent rereading is signaled by the output similarity evaluator, since the difference between the two outputs was large. Headlines showed no increases in total reading times, suggesting that participants did not return to the information stated in the headlines when trying to understand the incongruity. This is surprising, since the incongruities were constructed to directly contradict the headlines. A possible explanation is that participants remembered the text sufficiently to correctly identify the incongruity, without needing to return to the headlines.

Table 6.8 and Table 6.9 show a summary of the behavior and measured effects, and how they correlate with the hypotheses and hypothesized effects presented in Section 6.3.2.

Table 6.8. Behavior and measured effects on the experimental measures compared to control condition, Experiment 1 (critical word) and Experiment 2 (critical word, last sentence of text, headline, and entire text)

Hypotheses are given both for how the proposed model reacts and for the changes in the participant's reading behavior. Hypotheses presented in Section 6.3.2 supported by the data in **bold**. All behavior and effects in the table are supported by the data. Differences are statistically significant; non-significant results are described as *no significant difference*.

Detection type	Behavior compared to the control condition	Measured effects compared to the control condition
Implicit detection	<u>Model:</u> The output similarity evaluator judges the two outputs from the two analyses as sufficiently similar, and uses expectation-based output selection to insert the output from the heuristics analysis into the meaning representation.	<u>On critical word:</u> • Longer total reading times in Experiment 2 but not in Experiment 1, and more revisits in both experiments, but not to the extent that is found for explicit detection. • No significant difference in first fixation durations. <u>On last sentence, headline, and entire text (Experiment 2):</u> • Shorter first-pass reading times on last sentence, but not to the extent that is found for explicit detection. • Longer whole-text reading times on entire text, but not to the extent that is found for explicit detection.
	<u>Behavior:</u> The participants continue their reading without consciously reacting to the incongruity. The output with the incongruent word, however, results in more rereading of the incongruent word. The lingering effects also stop the reading of the subsequent sentence more quickly, since it is incongruent with the previous one. Also, reading of the entire text takes longer.	
Explicit detection	<u>Model:</u> The output similarity evaluator judges the two outputs from the two analyses as different enough to require reanalysis. Expectation-based output selection is not used, and heuristics usage is halted in favor of conscious rereading of the sentence.	<u>On critical word:</u> • Longer total reading times and more revisits. • Shorter first fixation durations in Experiment 1, no significant difference in first fixation durations in Experiment 2. <u>On last sentence, headline, and entire text (Experiment 2):</u> • Shorter first-pass reading times and longer total reading times on last sentence. • No significant difference in total reading times on headline. • Longer whole-text reading times on entire text.
	<u>Behavior:</u> The participants halt their reading in favor of rereading the incongruent word. The effects were observable in both early and late measures in Experiment 1, and in late measures only in Experiment 2. The participants interrupt their reading of the last sentence more quickly, since it is incongruent with the previous one, and begin rereading the previous material. The participants reread the last sentence and entire text to try to make sense of the incongruity. The participants, however, did not read the headlines more thoroughly.	

Table 6.9. Behavior and measured effects on explicit detection, Experiment 2

Working memory capacity is represented by OSPAN score. Differences are statistically significant; non-significant results are described as *no significant difference*.

	Behavior	Measured effects
Working memory capacity	Higher working memory capacity is correlated with better text comprehension in previous research. It did not, however, assist participants in explicitly identifying irregularities and errors in the text.	• No significant difference in number of explicitly detected incongruities for participants with high working memory capacity, as measured by OSPAN absolute score.

6.7 Conclusions

During reading, two parallel processes are operating: one uses heuristics based on expectations, previous knowledge, and prediction to create an output, while the other uses word-meaning retrieval to analyze the actual word on the page. The heuristics analysis is preferred during reading. The mechanism that compares these two outputs during reading is called the output similarity evaluator in this chapter.

When there is no conflict between the two outputs (as is normally the case), the output from the heuristics analysis is used in the meaning representation of the text. This process is called expectation-based output selection in this chapter. Since the heuristics-based analysis is completed slightly more quickly, the word-meaning retrieval is left only partially completed. But when an incongruity is encountered in a text, the two processes deliver conflicting outputs. During every-day reading, this conflict is usually minor, and incongruities are therefore commonly resolved using expectation-based output selection. The parsing process assumes that the encountered incongruity does not require conscious involvement from the reader, possibly in the same manner that spelling errors, grammatical errors, or unintentional mistakes can be disregarded. This process is similar to when disfluencies, filled pauses, and errors in spoken language are compensated for by listeners. When the conflict is large, however, the output similarity evaluator uses the word-meaning retrieval instead, and the incongruity is explicitly detected. This blocks heuristics usage during subsequent reading, and a reanalysis of the incongruity and the rest of the text is initiated. In Chapter 6, explicitly detected incongruities resulted in longer total reading times and more revisits, and longer total reading times were also found for other parts of the text. Importantly, the incomplete analysis is not completed, since no early eye movement effects increased for explicit detection. Therefore, it is assumed that the output similarity evaluator signals that expectation-based output selection is not used because the difference between the outputs is too great.

When an incongruity was not explicitly detected in the experiments presented in this chapter, the reading process reacted similarly to when an incongruity was explicitly detected. Longer total reading times on the incongruent word and a higher number of revisits were found, suggesting that the incongruities were in fact implicitly detected. The lingering effects are hypothesized to arise from the partial analysis that is causing disruptions to the reading process, since the incongruity is in fact substantial. Future research is needed to investigate how different the two outputs need to be for the output similarity evaluator to complete the word-meaning retrieval instead of using expectation-based output selection, thereby allowing explicit detection of the incongruity in the text.

In sum, disruptions to the reading process were found when an incongruity was not explicitly noticed. I argue that the mind retains a partial output of the incongruity when using the expected word meaning, and that the disruptions are caused by this

unused output. This provides eye movement results that support the theory of reading using two parallel processes, which has been suggested by previous research. If the conflict between the partial output and the expected word meaning is considered to be too great, however, the incongruity is explicitly detected, heuristics usage is halted, and a thorough rereading of the text is initiated. If the incongruity is not too severe, on the other hand, even though the incongruity is processed, the mind substitutes the incongruent word with the expected word, and heuristics usage is maintained. This is similar to what was seen in Chapter 4, in which people read neutral texts that they expected to be either positively or negatively biased. When expecting a positively biased text, people who did not feel strongly about the topic “read” a positively biased text, resulting in their opinions on the topic becoming more positive when measured after 20 minutes. This was also seen when expecting a negatively biased text, with opinions becoming more negative after reading, albeit not significantly so. Participants who had strong opinions on the topic, however, resisted this influence, and did not change their opinions after reading.

Substantial disruptions to the reading process were found in this chapter, both for explicit and implicit detection of an incongruity in a text. Chapter 8 investigates whether these effects are present in more limited contexts by giving people sentences with semantic incongruities and then evaluating the detection of the incongruities after reading. A manipulation that produces sentences perceived to use poor grammar will first be developed in Chapter 7, however, in order to also examine expectation-based output selection. The output similarity evaluator prefers to use the heuristics analysis during normal reading, rather than word meaning retrieval of the word on the page. This is why egregious errors—such as the word *save* instead of *lose*—are sometimes missed. Comprehension, as estimated by the working memory capacity of the participant, was used in Chapter 6, but seemed not to alter the overall usage of expectation-based output selection when reading an incongruent word. In Chapter 8, sentences that participants expected to have poorer grammar were hypothesized to decrease the preference of the heuristics analysis, thus decreasing the likelihood of the output similarity evaluator using expectation-based output selection. In Chapter 7, participants are given sentences expected to originate from either well-respected or less-respected news sources, and then judge the language usage of these sentences after reading.

Chapter 7

You perceive the language in “less-respected” news sources as being poorer than in “well-respected” ones

7.1 Introduction

As discussed in 6.2.1, semantic manipulations in previous research often used incongruities that relied on cultural knowledge rather than linguistic information. In Chapter 6, I used semantic incongruities that relied on purely linguistic information (antonyms), in paragraphs of text, and showed experimental effects in online measures. This chapter uses the same type of semantic incongruity in isolated sentences, combined with a very common technique in linguistics research: sentence judgments tasks. These tasks have been used to examine people’s perception of grammatical violations (Schütze & Sprouse, 2013) or semantic acceptability (Gibson et al., 2011), for example. I wanted to develop a manipulation that made people perceive a sentence as displaying poorer language usage, without explicitly stating, “the following sentences will display poorer language usage.” Online measures were then investigated in Chapter 8, to see if the disruptions seen during paragraph reading were still present in sentence reading. This source bias manipulation has the potential to affect how the output similarity evaluator operates, thereby increasing or decreasing the number of incongruities that are explicitly detected by participants. Furthermore, people’s ability to correctly classify these types of semantic incongruities as erroneous in an offline task can be evaluated, thus highlighting the differences between online and offline measures.

This chapter investigates how sentence judgments are affected by expectations of sentences originating from either less-respected or well-respected news sources, and successful and failed detection of semantic incongruities in sentences. When the source of information is considered trustworthy, credible, and reliable, people find it more convincing and more persuasive (Hahn, Harris, & Corner, 2009; Miller & Baseheart, 1969; Pornpitakpan, 2004). Conversely, features of the text, such as correct use of capitalization, make people consider the text more credible (Rubin &

Liddy, 2006; Su, Huang, & Chen, 2010; Weerkamp & Rijke, 2008). This suggests that people believe sources with lower standards have poorer language usage compared to better sources. In order to investigate sentence judgment scores for incongruities in sentences, and for sentences expected to originate from either less-respected or well-respected news sources, two experiments were conducted.

In Experiment 1, the source of the sentences was only stated as “various news sources.” It was hypothesized that acceptability judgments would be lower for the semantically incongruent sentences, compared to the congruent ones. In order to attempt to elicit even lower sentence judgments for the semantically incongruent sentences, Experiment 2 examined sentence judgments when changing the participants’ expectations regarding the source of the sentence. Before each group of sentences, it was stated that they originated from either “well-respected” or “less-respected” news sources. It was hypothesized that the expectations of sentences from less-respected news sources would be a “bad-source” bias. Therefore, participants were hypothesized to be more likely to judge all sentences more harshly, not just the incongruent sentences. It was hypothesized that participants would give lower ratings for all sentences in all three tasks involving the form of the sentence, and specifically the incongruent sentences. The task in which the participants stated their interest in reading the hypothetical remainder of the text, when only considering the topic and not the form of the sentence, was hypothesized to be unaffected by the source bias. Conversely, the expectation of sentences from a well-respected news source, a “good-source” bias, would make participants more likely to judge the sentences less harshly.

This study also presented an opportunity to examine whether people are able to differentiate between grammaticality and semantics when it comes to judgment tasks. Sentence judgment tasks generally use a generic statement such as “how natural does the sentence feel to you?” regardless of whether the manipulation made to the sentence is grammatical or semantic in nature (e.g., Hofmeister, Culicover, & Winkler, 2015). Previous research suggests that there is no benefit of using two separate acceptability judgments, with one being grammatical and one being semantic in nature (Ariel, 2010; Sprouse, Schütze, & Almeida, 2013). I investigated this claim using separate judgment tasks for grammaticality and semantics.

7.2 Previous research

7.2.1 Judgments of acceptability and grammaticality

Acceptability and grammaticality judgments of ill-formed sentences are common in linguistic research. Judgment tasks are important because they provide information that other tasks do not, such as the perceived level of grammaticality of utterances (Schütze & Sprouse, 2013). The semantic acceptability of sentences is also often

verified using these judgment tasks (Gibson et al., 2011). The terms *acceptability* and *grammaticality* are used interchangeably about both grammatical and semantic violations, and some studies investigate syntactic or grammatical violations using the term *acceptability judgment* (Maunder, Fromkin, & Cornell, 1993). The difference between the terms is often unclear, with linguistic intuition investigated in similar ways using tasks defined as both *grammaticality judgments* (Nagata, 1988) and *acceptability judgments* (Dąbrowska, 2010). Grammaticality judgments have been used for studies of people with aphasia (Berndt, Salasoo, Mitchum, & Blumstein, 1988; Wulfeck, Bates, & Capasso, 1991) and studies with second-language learners (Gass, 1994; Leow, 1996). In these judgment tasks, a sentence is often judged by participants in terms of how natural it is (Hofmeister et al., 2015). A wide variety of manipulations are used, making it difficult to compare the results from different studies (Sprouse, 2011; Sprouse & Almeida, 2011). Most studies, however, only investigate either grammatical or semantic acceptability in isolation. It is argued in one study that while the terms *acceptability* and *grammaticality* are distinctly different, judgments on the grammatical correctness of a sentence are very difficult, since it is generally assumed that speakers cannot access their mental grammar consciously (Sprouse et al., 2013). Ariel (2010, p. 77) has also stated that people have difficulties differentiating between grammatical and pragmatic violations, because they are not necessarily intuitively aware of grammaticality judgments.

In addition, it is argued that acceptability judgments can be affected by more than just the grammaticality of the sentence, for instance the working memory capacity of the participant (Gibson, 1998; Lewis, 1996; Lewis & Vasisht, 2005). It is known that a linguist's judgments differ from a layperson's judgment, and that stimuli that has not been appropriately normed can pose problems (Gibson & Fedorenko, 2013). Importantly, while all judgments can be said to be somewhat subjective, they are argued to be reliable and valid for speakers' assessments and intuitions about the language/s they speak (Schütze, 1996, p. 120).

When performing judgment tasks, Likert scales are commonly used, with participants indicating how much they agree or disagree with statements or questions such as "How natural did the sentence sound to you" (Burton, 2000; DeVellis, 2003, p. 79). Another type of scale is the visual analog scale (DeVellis, 2003, pp. 81-82; Mayer, 1978), which is ungraded, and therefore offers a much higher resolution of the responses. This level of sensitivity should make it possible to investigate incongruities in greater detail, possibly allowing more subtle manipulations to be used. For a thorough explanation of Likert scales and visual analog scales, see Section 2.4.4.

7.2.2 Source credibility and source reliability

An opinionated message from a highly trustworthy communicator has been found to be more persuasive than if initial trustworthiness is low (Miller & Baseheart, 1969),

and participants have been found to be more likely to discount information from “untrustworthy” sources (Hovland & Weiss, 1951). Another study gave participants information from sources that were either reliable (“taken from the journal Science”) or unreliable (“a circular email from excitingnews@wowe.com”). The study found that information from a reliable source was considered to be more convincing than information from one that was unreliable (Hahn et al., 2009). An extensive review of five decades of research of source credibility also showed that high-credibility sources were commonly considered to be more persuasive both for attitude change and for gaining behavioral compliance, compared to low-credibility sources (Pornpitakpan, 2004).

A great deal of research has been performed on trust, such as the importance of consumer trust in organizations (McEvily, Perrone, & Zaheer, 2003; K. J. Stewart, 2003). Research has also been performed with regard to trustworthiness detection in text, suggesting that the trustworthiness of text context can be identified using two dimensions: features related to the information distributors, and the text itself (Fuoli, 2017; Su et al., 2010). With regard to internet blog posts, factors such as correct spelling, correct use of capitalization, and language usage similar to the corresponding news stories are suggested to improve credibility (Rubin & Liddy, 2006; Weerkamp & Rijke, 2008). Since expectations influence our behavior in many aspects of our lives (Barber & Silver, 1968; Moerman & Jonas, 2002; Pickering & Garrod, 2013; Stewart-Williams & Podd, 2004; Van Berkum et al., 2005), it should be possible to induce bias merely by giving people different expectations regarding the source, and this bias should affect their perceptions and judgments of sentences.

7.3 The study of source bias on sentence judgments

In two experiments, I investigated how sentence judgments are affected by semantic incongruities. The incongruent sentences were created by replacing a word like *bad* with its antonymic partner *good*. Since antonyms operate within the same meaning dimension but express opposite meaning relations within this dimension, they are maximally similar and maximally different at the same time (Paradis et al., 2015; Paradis & Willners, 2011). The semantically incongruent sentences in this study are thus highly incongruent, since the opposite word is used. They are also vague and not necessarily immediately understood as incongruent, however, since they share so many features with the corresponding congruent sentence. In Experiment 1, the sentences were only specified as being from “various” news sources. In Experiment 2, the sentences were stated to be from either “well-respected” or “less-respected” news sources. This was done in order to create an expectation of worse language usage, and better language usage, respectively. Since judgment tasks are considered to be fairly subjective (Schütze, 1996, p. 120), this source bias was hypothesized to affect all four

types of sentence judgments performed by the participants. These results would be able to show which bias is the default setting when reading the sentences in this study, and more broadly, when reading news. This study also investigated how well native speakers could differentiate between grammatical and semantic judgments, and how this affected the sentence judgment.

7.3.1 Terms and classifications

Four sentence rating terms were used in this study, based on the responses to the four different questions posed to the participants after each sentence. The *grammatical acceptability rating* was the response to the question “Based on only the form of the sentence and disregarding the missing context: How correct did the sentence feel?” This measure reported the participants’ opinion on the grammatical acceptability of the sentence. The *semantic acceptability rating* was the response to the question “Based on only the form of the sentence and disregarding the missing context: How clear and coherent did the sentence feel?” This measure reported the participants’ opinion on the semantic acceptability of the sentence. The *form-based interest rating* was the response to the question “Based on only the form of the sentence: How much does it make you want to read the whole article?” This measure reported the level of interest the participants had in reading the hypothetical remainder of the text, based on the grammaticality and semantics of the sentence. The *topic-based interest rating* was the response to the question “Based on only the topic of the sentence: How much does it make you want to read the whole article?” This measure reported the level of interest the participants had in reading the hypothetical remainder of the text, based on the topic of the sentence.

7.3.2 Hypotheses

Semantic incongruities rely on context to be correctly interpreted as incongruent. In this study, the stimuli were created to be internally incongruent, and only rely on the linguistic material present in the sentence. It has been claimed that people are only able to make intuitive acceptability judgments (Ariel, 2010, p. 77) and cannot access their mental grammar consciously (Sprouse et al., 2013). The current study investigated this by separating judgments of grammatical and semantic acceptability, but the results were expected to reflect the findings by Ariel (2010) and Sprouse et al. (2013).

In both experiments, the sentences were divided into 10 groups of eight sentences each. In Experiment 1, the participants were only told that all the sentences originated from various news sources. Significantly lower grammatical and semantic acceptability ratings were hypothesized for the incongruent critical sentences, compared to the congruent critical sentences.

Table 7.1. Hypotheses and hypothesized effects on the experimental measures for incongruent sentences, compared to the control condition containing the congruent word, Experiment 1 (no source bias) and Experiment 2 (less-respected and well-respected news sources)

Separate hypotheses are given for the form of the sentence, and for the interest in reading the hypothetical remainder of the text. The hypotheses and effects supported by the experimental data are reported in **bold** in Table 7.5 in Section 7.6.

Source bias	Hypothesis compared to the control condition	Hypothesized effects compared to the control condition
No source bias	Form: The incongruent sentences are interpreted as being sufficiently incongruent to affect people's judgments. However, people are not able to differentiate between grammatical and semantic acceptability, so they will give lower scores on both scales.	• Lower grammatical acceptability rating. • Lower semantic acceptability rating.
	Interest: The incongruent sentences should not affect people's stated interest in reading the hypothetical remainder of the text.	• No difference in form-based interest rating. • No difference in topic-based interest rating.
Less-respected news source	Form: The incongruent sentences are interpreted as being sufficiently incongruent to affect people's judgments. However, people are not able to differentiate between grammatical and semantic acceptability, so they will give lower scores on both scales.	• Lower grammatical acceptability rating. • Lower semantic acceptability rating.
	Interest: The incongruent sentences should not affect people's stated interest in reading the hypothetical remainder of the text.	• No difference in form-based interest rating. • No difference in topic-based interest rating.
Well-respected news source	Form: The incongruent sentences are not interpreted as being sufficiently incongruent to affect people's judgments. It is possible that they are interpreted as a stylistic choice, or that the participant is not sufficiently certain that they are errors and decides to "let it slip."	• No difference in grammatical acceptability rating. • No difference in semantic acceptability rating.
	Interest: The incongruent sentences should not affect people's stated interest in reading the hypothetical remainder of the text.	• No difference in form-based interest rating. • No difference in topic-based interest rating.

In Experiment 2, the participants were told that sentences originated from either well-respected or less-respected news sources. Since trust and credibility have been linked to proper spelling and language usage (Rubin & Liddy, 2006; Su et al., 2010; Weerkamp & Rijke, 2008), this source bias was hypothesized to make participants expect a higher level of correctness from the well-respected sources, and a lower level of correctness from the less-respected sources. Introducing sentences as originating from less-respected sources is assumed to create an expectation of improper language usage, and make the incongruities more evident, thereby making participants more likely to judge incongruent sentences more harshly. Introducing sentences as originating from well-respected sources is assumed to create an expectation of proper language usage, and make the incongruities less evident. Significantly lower grammatical and semantic acceptability ratings were therefore hypothesized specifically for the incongruent critical sentences, compared to the congruent critical sentences, but in the less-respected source condition only. It is also assumed, however, that these expectations would affect all sentence judgment scores. Consequently, the

grammatical and semantic acceptability ratings were expected to be significantly lower for all less-respected sentences because of this bias. Furthermore, this bias was expected to affect the ratings for how interested the participants were in reading the hypothetical remainder of the text. Two separate interest ratings are given, one based solely on the form of the sentence (form-based interest rating), and one based solely on the topic of the sentence (topic-based interest rating). Significantly lower form-based interest ratings were expected, since participants would be affected by their perception of an ill-formed sentence. Since participants were instructed to ignore the language usage for topic-based interest ratings, however, they were not hypothesized to be affected by the source bias. If this were the case, it would seem that people's interest in reading news is only marginally affected by the news source.

A summary of the hypotheses and the hypothesized effects are given in Table 7.1 and Table 7.2.

Table 7.2. Hypotheses and hypothesized effects on the experimental measures for all sentences, Experiment 2

Separate hypotheses are given for the form of the sentence, and for the interest in reading the hypothetical remainder of the text. The hypotheses and effects supported by the experimental data are reported in **bold** in Table 7.6 in Section 7.6.

Source bias	Hypothesis	Hypothesized effects
Less-respected news source	<u>Form</u> : People expect sentences taken from less-respected news sources to contain worse language usage, textual errors, and incongruities.	• Lower grammatical acceptability rating compared to well-respected news sources. • Lower semantic acceptability rating compared to well-respected news sources.
	<u>Interest</u> : When only considering the language usage, which they assume is poorer, people are less interested in reading news from less-respected news sources. When disregarding the language usage, their interest in reading the hypothetical remainder of the text is not affected.	• Lower form-based interest rating compared to well-respected news sources. • No difference in topic-based interest rating compared to well-respected news sources.
Well-respected news source	<u>Form</u> : People expect sentences taken from well-respected news sources to contain better language usage, and fewer textual errors and incongruities.	• Higher grammatical acceptability rating compared to less-respected news sources. • Higher semantic acceptability rating compared to less-respected news sources.
	<u>Interest</u> : When only considering the language usage, which they assume is better, people are more interested in reading news from well-respected news sources. When disregarding the language usage, their interest in reading the hypothetical remainder of the text is not affected.	• Higher form-based interest rating compared to less-respected news sources. • No difference in topic-based interest rating compared to less-respected news sources.

7.4 Experiment 1

7.4.1 Method

7.4.1.1 Participants

Forty-six native speakers of English (28 female; 18 to 26 years of age; $M=20.3$ years, $SD=2.02$) participated in the experiment. The participants were recruited with SONA Experiment Scheduling at the Rayner Lab, University of California, San Diego, United States. The participants were naïve to the purpose of the experiment.

7.4.1.2 Apparatus

The experiment took place at the Rayner Eye Tracking Lab, Department of Psychology, University of California, San Diego, United States. Stimuli were displayed on an HP 1230 20" 4:3 CRT monitor with a resolution of $1,280 \times 1,024$ pixels. The pre-experiment questionnaire was completed in PsychoPy (1.80.00) on a separate laptop computer.

7.4.1.3 Design and stimuli

The participants first completed a pre-experiment questionnaire where they entered age, gender, and questions about how much time they spent reading or watching the news each week. The participants were also asked to state their level of trust, as well as their level of respect regarding several specific news publications: Newsweek, the National Enquirer, the New York Times, The Sun, Time Magazine, and the Washington Post. Using Newsweek as an example, the two statements were worded as "I trust the information in Newsweek" and "I consider Newsweek to be a well-respected publication." Ungraded visual analog scales using the two extremes *completely disagree* and *completely agree* were used for all responses. The visual analog scales produced results from 0 to 100 in increments of 1. The questions on trust and respect regarding news sources were used to prime the participants for news reading and to cause them to think about how news sources can differ, as well as putting a name to the two types of sources introduced during the sentence reading experiment.

For the sentence judgment portion of the experiment, a total of 80 sentences were used in four experimental lists. Before reading a group of eight sentences, written instructions informed the participants that the upcoming sentences were taken from "various news sources." Of the 10 groups of sentences, five were said to originate from well-respected sources and five were said to originate from less-respected sources. The groups were shown in an alternating fashion, with the order of the groups randomized in Experiment Builder. The stated source of each sentence was counterbalanced, so all sentences were said to be from a well-respected source to half of the participants, and from a less-respected source to the other half.

The sentences were divided into groups of eight. Before reading a group of eight sentences, written instructions informed the participants that the upcoming sentences were taken from “various sources.”

All sentences were taken from the section of the Corpus of Contemporary American English (COCA) with material from news sources. Each list contained 20 congruent and 20 incongruent sentences. These were created by using 10 one-syllable adjective antonym pairs, including, for example, *large* and *small*, *soft* and *hard*, *good* and *bad*, and *slow* and *fast*. Two different sentences containing each word of the 10 antonym pairs were taken from the corpus and adapted for use in the study, resulting in 40 different sentences. Twenty sentences were displayed using the congruent adjective, while the remaining 20 were displayed using the corresponding incongruent adjective in the pair. These were distributed so that each list contained four sentences derived from each antonym pair, with each word used in both one congruent and one incongruent sentence. In addition, each list contained 40 filler sentences displayed in the same version to all participants, using 20 different one-syllable and 20 different two-syllable non-antonym adjectives. The adjectives were placed toward the ends of the sentences, since the semantic incongruities in the critical sentences needed context in order to be evident. All stimuli sentences can be found in Appendix 7.A.

An example sentence is shown in (19), with the antonym pair indicated in italics, and the incongruity also in bold. No words were displayed in bold or italics during the experiment, and only one version was shown to each participant.

- (19) That the airline is threatening to lay off employees is making the situation for workers insecure, and is creating *bad/good* public relations as well.

The sentences were displayed using Courier New 18-point font on two lines, with the second line containing the adjective used for stimuli creation. Groups of eight sentences were displayed, two congruent critical, two incongruent critical, and four filler sentences. Participants were instructed to click the left mouse button when they had understood the sentence. The order of the groups (10 in total), was randomized in Experiment Builder.

After each sentence, the participants were asked to perform four judgments on four separate screens using visual analog scales. See Section 7.3.1 for the specific written questions used during the experiment. The range of the scales was 0 to 100 in increments of 1, but the scales shown to participants were unmarked, and the two extremes were labeled *Not at all* on the left and *Extremely* on the right. The terms used are grammatical acceptability rating, semantic acceptability rating, form-based interest rating, and topic-based interest ratings.

7.4.1.4 Procedure

Each participant was introduced to the experiment and recorded individually. The participants were welcomed to the study, and completed a pre-experiment

questionnaire in which their age and gender were noted. Oral instructions were then given to participants that explained the experimental procedure and the four different judgment tasks. Participants were specifically told that the first judgment task only concerned how correct the sentence was, and that the second scale only concerned how clear and easy to understand the sentence was, based on the form of the sentence. They were also told that the form of the sentence only signified the way the sentence was written and how good the sentence sounded to the participant, while the topic of the sentence only signified what the sentence was about. Participants were also instructed that some sentences would originate from less-respected news sources, and some would originate from well-respected sources. During the oral instructions prior to the experiment, the participants were told that the sentences originated from “various news sources.”

Participants first completed four practice trials with sentences and accompanying rating scales. The main experiment then began with participants reading the sentences and rating each one immediately after reading it. After reading all the sentences and finishing the post-experiment opinion questionnaire, the participants were thanked for their participation in the study, and they were given one course credit as compensation for participating in the study.

7.4.1.5 Data analysis

The data processing, figures, and statistical analyses were made using R (3.2.2) (R Development Core Team, 2015) and the lme4 package (Bates et al., 2015) in R Studio (0.99.473). Linear mixed effect models were used to examine the effects of semantic incongruities on four different sentence ratings: the grammatical acceptability rating, the semantic acceptability rating, the form-based interest rating, and the topic-based interest rating. A fixed effect of sentence type with three levels was used (congruent critical, incongruent critical, congruent filler), also examining the interaction of the expected source of the sentence, which was also used as a factor. Models used random intercepts and slopes of condition within group for participant and item. Intercept-slope correlation factors for participants and items were not used, since no models converged. The lmerTest package (2.0-29) was used to estimate *p*-values with Satterthwaite approximation.

7.4.2 Results

Even though no significant results were found, Figure 7.1 shows mean grammatical acceptability, semantic acceptability, form-based interest, and topic-based interest ratings for congruent critical, incongruent critical, and congruent filler sentences. No significant differences were found between the congruent and incongruent critical sentences (grammatical acceptability: Estimate=0.574, *SE*=1.29, *t*=0.45, *p*=.66; semantic acceptability: Estimate=0.644, *SE*=1.35, *t*=0.48, *p*=.63; form-based interest: Estimate=-0.641, *SE*=1.16, *t*=-0.55, *p*=.58; topic-based interest: Estimate=1.72,

$SE=1.53$, $t=1.12$, $p=.26$) or the congruent fillers (grammatical acceptability: Estimate=-0.166, $SE=1.11$, $t=-0.15$, $p=.88$; semantic acceptability: Estimate=-1.09, $SE=1.17$, $t=-0.93$, $p=.36$; form-based interest: Estimate=-1.16, $SE=1.01$, $t=-1.15$, $p=.25$; topic-based interest: Estimate=2.55, $SE=1.33$, $t=1.92$, $p=.06$).

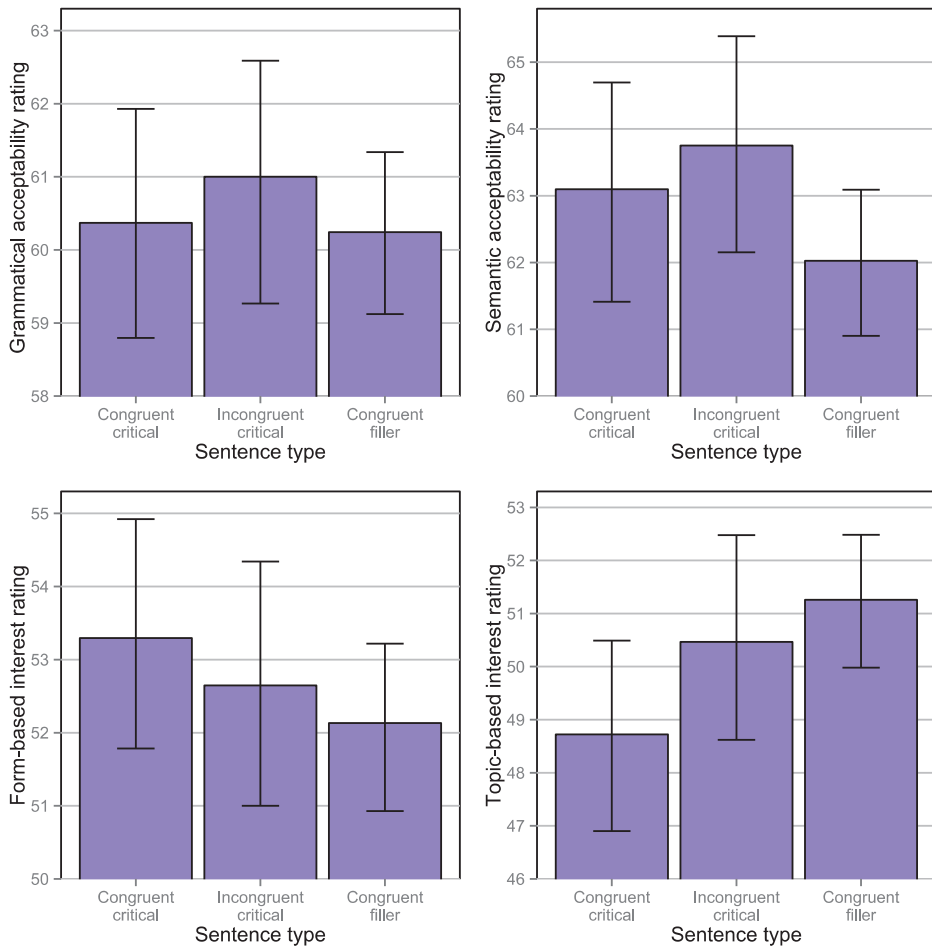


Figure 7.1. Mean grammatical acceptability, semantic acceptability, form-based interest, and topic-based interest ratings for congruent critical, incongruent critical, and congruent filler sentences, Experiment 1
Error bars denote standard errors of the mean.

7.4.3 Discussion

The definition of what constitutes a semantic anomaly, semantic incongruity, or semantic illusion is varied, with different studies using different types of manipulations (Bohan et al., 2012; Daneman et al., 2007; Kutas & Hillyard, 1980; Nieuwland & Van Berkum, 2005). Comparing results from different studies is therefore difficult, as is trying to replicate the results. In order to specify the linguistic properties of the semantic incongruity, antonyms were used as manipulations, since they are maximally similar and maximally different at the same time (see Section 6.2.3 for a more thorough explanation of the suitability of antonyms for experimental studies).

No significant differences were found for the two acceptability ratings when comparing the semantically incongruent sentences with the congruent critical sentences. The automaticity and small-but-large difference of antonymy, combined with the fact that judgment tasks are argued to be subjective to some extent (Schütze, 1996, p. 50 & p. 120), showed that the incongruity left a lot of room for interpretation. Because of this, participants were possibly finding acceptable interpretations of the incongruities and therefore did not give lower acceptability ratings to the sentences.

In order to further examine the sentence judgments of semantic incongruities, another experiment was performed. In Experiment 2, a manipulation designed to lower people's expectations for correct language was needed. Giving people specific instructions that "the language might be incorrect" might affect people's reading, as seen in Chapter 3, when instructing people to spell check the text significantly affected their reading patterns and comprehension. This was not desirable, since the research interest was to investigate natural reading. A more subtle instruction was therefore used: either stating that sentences originated from "less-respected" news sources (a "bad-source bias" intended to make participants judge the sentence more harshly) or from "well-respected" news sources (a "good-source bias" intended to make participants judge the sentence less harshly).

7.5 Experiment 2

Except for the differences explained in the Method section below, Experiment 2 was identical to Experiment 1.

7.5.1 Method

7.5.1.1 Participants

Seventy-eight native speakers of English (54 female; 18 to 26 years of age; $M=20.3$ years, $SD=1.57$) participated in the experiment.

7.5.1.2 Apparatus

Identical to Experiment 1.

7.5.1.3 Design and stimuli

The 80 stimuli sentences from Experiment 1 were divided into four experimental lists: two related to the semantic incongruence manipulation, and two related to source expectation. Before reading each group of eight sentences, participants were told that the sentences originated from news sources considered to be either “trusted and well-respected according to polling data,” or “less trusted and less-respected according to polling data.” Henceforth, these sources will be referred to as *well-respected* and *less-respected*. Five groups were said to be from well-respected sources, and five groups were said to be from less-respected sources, with the groups presented in an alternating fashion. The order of the groups was randomized in Experiment Builder, and the source condition of all sentences was counterbalanced on the lists.

7.5.1.4 Procedure

During the oral instructions prior to the experiment, the participants were told that the sentences originated from either less-respected news sources or from well-respected ones, as described in 8.5.1.3 above.

After reading all the sentences and finishing the post-experiment opinion questionnaire, the source condition manipulation was revealed.

7.5.1.5 Data analysis

Models used random intercepts and slopes of condition within group for participant and item, with intercept-slope correlation factors for participants and items. The intercept-slope correlation factors were dropped from the model for topic-based interest ratings in order for it to converge.

7.5.2 Results

Figure 7.2 shows grammatical acceptability, semantic acceptability, form-based interest, and topic-based interest ratings for congruent critical, incongruent critical, and congruent filler sentences. The ratings for the expected sentence sources are shown using separate bars.

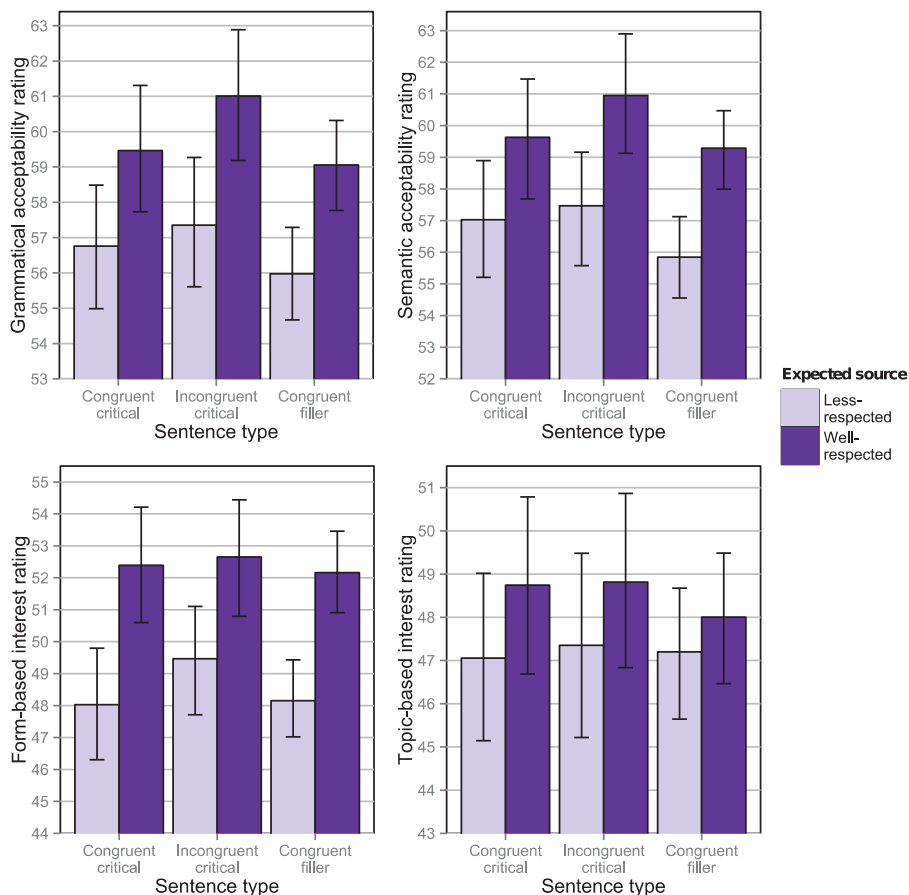


Figure 7.2. Mean grammatical acceptability, semantic acceptability, form-based interest, and topic-based interest ratings for congruent critical, incongruent critical, and congruent filler sentences, divided by the expected source of the sentence, Experiment 2
Error bars denote standard errors of the mean.

The grammatical acceptability ratings for sentences introduced as being from less-respected sources were significantly lower compared to sentences introduced as being from well-respected sources (Estimate=-2.71, $SE=1.18$, $t=-2.30$, $p<.02$). No significant differences were found between the congruent and incongruent critical sentences (Estimate=1.58, $SE=1.39$, $t=1.14$, $p=.26$) or the congruent fillers (Estimate=-1.08, $SE=1.22$, $t=-0.88$, $p=.38$). No interactions were found between source condition and the incongruent critical sentences (Estimate=-1.03, $SE=1.66$, $t=-0.62$, $p=.54$) or the congruent fillers (Estimate=0.208, $SE=1.46$, $t=0.14$, $p=.89$).

The semantic acceptability ratings for sentences introduced as being from less-respected sources were significantly lower compared to sentences introduced as being from well-respected sources (Estimate=-2.85, $SE=1.29$, $t=-2.21$, $p<.03$). No

significant differences were found between the congruent and incongruent critical sentences (Estimate=1.35, $SE=1.48$, $t=0.92$, $p=.36$) or the congruent fillers (Estimate=-0.744, $SE=1.30$, $t=-0.57$, $p=.57$). No interactions were found between source condition and the incongruent critical sentences (Estimate=-0.566, $SE=1.81$, $t=-0.31$, $p=.76$) or the congruent fillers (Estimate=-0.523, $SE=1.59$, $t=-0.33$, $p=.74$).

The form-based interest ratings for sentences introduced as being from less-respected sources were significantly lower compared to sentences introduced as being from well-respected sources (Estimate=-4.47, $SE=1.19$, $t=-3.76$, $p<.0005$). No significant differences were found between the congruent and incongruent critical sentences (Estimate=0.179, $SE=1.35$, $t=0.13$, $p=.89$) or the congruent fillers (Estimate=-0.295, $SE=1.17$, $t=-0.25$, $p=.80$). No interactions were found between source condition and the incongruent critical sentences (Estimate=1.26, $SE=1.68$, $t=0.75$, $p=.45$) or the congruent fillers (Estimate=0.439, $SE=1.45$, $t=0.30$, $p=.76$).

No significant differences were found for the source bias (Estimate=-1.75, $SE=1.39$, $t=-1.26$, $p=.21$). No significant differences were found between the congruent and incongruent critical sentences (Estimate=0.060, $SE=1.57$, $t=0.04$, $p=.97$) or the congruent fillers (Estimate=-0.690, $SE=1.37$, $t=-0.51$, $p=.61$). No interactions were found between source condition and the incongruent critical sentences (Estimate=0.200, $SE=1.96$, $t=0.10$, $p=.92$) or the congruent fillers (Estimate=0.889, $SE=1.70$, $t=0.52$, $p=.60$).

7.5.3 Discussion

Experiment 2 investigated whether acceptability judgments could be affected by the origin of the sentence being judged. Or, more specifically, it examined whether the mere expectation of a sentence from a well-respected or less-respected news source would alter the ratings. In essence, specifying that a news source is well-respected would mean that, in other people's eyes, the source is trustworthy. Specifying that a news source is less-respected would mean that, in other people's eyes, the source is less trustworthy. Importantly, the stimuli sentences were unrelated to the source specified to the participants. All three ratings involving the form of the sentence—grammatical acceptability, semantic acceptability, and form-based interest ratings—were significantly lower for the sentences said to originate from less-respected sources compared to the sentences said to originate from well-respected sources. This suggests that participants had a negative bias toward the grammar of less-respected news sources that affected their ratings of the sentences. As predicted, both grammatical and semantic acceptability judgments behaved similarly.

Surprisingly, grammatical and semantic acceptability ratings were not affected for semantically incongruent sentences, regardless of source bias. The sentence manipulations were subtle, and differed only in one respect, namely being opposites. While antonymy is argued to be a crucial part of human perception and cognition (Bianchi et al., 2011; Paradis, 2001, 2005, 2013), it seems like antonymic semantic

incongruities are not sufficiently strong to elicit results in acceptability judgment tasks. Even with the negative bias toward the less-respected sources, participants were not less forgiving of the incongruent sentences. It should be noted, however, that the main purpose of the incongruity was not to elicit results in sentence judgment tasks, but rather to elicit results in eye movement measures, which are much more sensitive. Nevertheless, these results were surprising, as the incongruent nature of the word was clear, as the stimulus example in (19), with the antonymic incongruity in bold, shows: “That the airline is threatening to lay off employees is making the situation for workers insecure, and is creating **good** public relations as well.” Note that no words were shown in bold during the experiment.

7.6 General discussion

Two experiments investigated how the explicit detection of antonym incongruities, as well as the expectations due to the origin of the sentence, affects sentence judgments. Judgment tasks are important because they provide information that other tasks do not, such as the perceived level of grammaticality of utterances (Schütze & Sprouse, 2013). It has been suggested that acceptability judgments are affected by factors other than the structure of the sentence, such as working memory capacity (Gibson, 1998; Lewis, 1996; Lewis & Vasishth, 2005). It seems, however, that very little research has investigated how bias and expectations affect acceptability judgments.

Awareness of antonymy is considered to be intuitive (Bentin, 1987; Bianchi et al., 2011; Kretchmar et al., 2009; Murphy & Jones, 2008; Paradis et al., 2009; Roehm et al., 2007; Tribushinina, 2013; van de Weijer et al., 2012, 2014). It could therefore be considered surprising that grammatical and semantic acceptability judgment scores were not significantly lower for the incongruent sentences in either experiment. I instead argue, however, that the incongruities were likely too subtle to affect these instantaneous, conscious judgments. The results demonstrate people’s inability to perform correct conscious judgments of acceptability and grammaticality when a manipulation leaves room for interpretation, such as in the case of antonymic incongruities that are maximally similar while being maximally different at the same time (Paradis et al., 2015; Paradis & Willners, 2011). Since this was the case in both Experiment 1 and Experiment 2, it suggests that people are fairly bad at performing spontaneous sentence judgments.

Research has highlighted the importance of trust and trustworthiness (McEvily et al., 2003; K. J. Stewart, 2003). Previous research has shown that trustworthy, reliable, and credible sources are considered to be more persuasive (Hovland & Weiss, 1951; Miller & Baseheart, 1969; Pornpitakpan, 2004) and more convincing (Hahn et al., 2009). It has been suggested that trustworthiness is even related to how the text is written (Fuoli, 2017; Su et al., 2010), with proper spelling and language usage found

to improve credibility in Internet blog posts (Rubin & Liddy, 2006; Weerkamp & Rijke, 2008). This idea formed the basis for creating the source bias, with the purpose of making people expect ill-formed sentences. Consequently, for all sentences said to originate from less-respected news sources in Experiment 2, the three judgments regarding the form of the sentence were significantly lower than for the sentences said to be from well-respected sources. This shows that participants are biased against the language usage of these news sources and automatically consider them less-than-perfect. This was the goal for this manipulation, since specifically stating “the following sentences have poor grammar” would be very undesirable. This confirms the argument that sentence judgments are fairly subjective measures (Schütze, 1996, p. 50 & p. 120). With an expectation of a less grammatically correct sentence, it was also hypothesized that the semantic incongruities would be judged as errors rather than seen as creative language usage, uncommon wordings, or maybe even metaphors. This was not the case, since the sentence judgments for the incongruent sentences were not lower under either bias. In Experiment 1, it was hypothesized that the semantically incongruent sentences would receive lower grammatical and semantic acceptability scores; this was not the case. In Experiment 2, it was hypothesized that the semantically incongruent sentences said to be from less-respected sources would have lower grammatical and semantic acceptability scores; this was also not the case. In light of this, the hypothesis about incongruent sentences said to be from well-respected sources should be questioned. The hypothesis was that grammatical and semantic acceptability scores would be unaffected by the semantically incongruent sentences said to be from well-respected sources. The experimental results support the hypothesis, but the behavior behind these results should be viewed in light of the unconfirmed hypotheses for the semantically incongruent sentences. The most likely interpretation is instead that the manipulation was not strong enough to elicit lower scores in sentence judgment tasks. Sentence judgment tasks involve instant decisions of the acceptability of a sentence, and many sentences are given during an experiment. In fact, the participants are sometimes even instructed to answer quickly and give spontaneous opinions, rather than think too long and analyze the sentence too thoroughly. Consequently, not much thought is given to each sentence before giving a judgment, showing the importance of using a strong manipulation in sentence judgment tasks.

I also had the opportunity to investigate if any benefit would be gained by separating grammatical and semantic acceptability judgments in sentence judgment tasks. The terms acceptability and grammaticality are used for a wide variety of sentence manipulations, with regard to judgment tasks in language studies (Berndt et al., 1988; Dąbrowska, 2010; Gass, 1994; Leow, 1996; Mauner et al., 1993; Nagata, 1988; Sprouse, 2011; Sprouse & Almeida, 2011; Wulfeck et al., 1991). The task is commonly to judge how natural a sentence feels to the participant (Hofmeister et al., 2015), a judgment that makes no distinction between grammatical or semantic acceptability. These were therefore separated, in order to also examine if people would

be able to distinguish between grammatical and semantic acceptability. By using visual analog scales instead of Likert scales in all judgment tasks, results were much more fine-grained (Burton, 2000; DeVellis, 2003; J. M. Mayer, 1978). Despite this fact, grammatical and semantic acceptability ratings were highly similar in both experiments. The operationalization of grammatical judgments and semantic judgments in this study were to ask for *correct* and *clear and coherent* judgments respectively. The results suggest that these two ratings should not be separated in sentence judgment tasks of the type performed in this study, and that a vague statement such as “how natural does the following sentence sound to you” is sufficient. These conclusions support the claims made in previous research (Ariel, 2010, p. 77; Sprouse et al., 2013), and were therefore hypothesized in this study.

Table 7.5 and Table 7.6 show a summary of the behavior and measured effects, and how they correlate with the hypotheses and hypothesized effects presented in Section 7.3.2.

Table 7.5. Behavior and measured effects on the experimental measures for incongruent sentences, compared to the control condition containing the congruent word, Experiment 1 (no source bias) and Experiment 2 (less-respected and well-respected news sources)

Separate hypotheses are given for the form of the sentence, and for the interest in reading the hypothetical remainder of the text. Hypotheses presented in Section 7.3.2 supported by the data in **bold**. All behavior and effects in the table are supported by the data. Differences are statistically significant; non-significant results are described as *no significant difference*.

Source bias	Behavior compared to the control condition	Measured effects compared to the control condition
No source bias	<u>Form</u> : The incongruent sentences are not interpreted as being sufficiently erroneous to affect people's judgments.	- No significant difference in grammatical acceptability rating. - No significant difference in semantic acceptability rating.
	<u>Interest</u> : The incongruent sentences did not affect people's stated interest in reading the hypothetical remainder of the text.	No significant difference in form-based interest rating. No significant difference in topic-based interest rating.
Less-respected news source	<u>Form</u> : The incongruent sentences are not interpreted as being sufficiently incongruent to affect people's judgments.	- No significant difference in grammatical acceptability rating. - No significant difference in semantic acceptability rating.
	<u>Interest</u> : The incongruent sentences did not affect people's stated interest in reading the hypothetical remainder of the text.	No significant difference in form-based interest rating. No significant difference in topic-based interest rating.
Well-respected news source	<u>Form</u> : The incongruent sentences are not interpreted as being sufficiently erroneous to affect people's judgments. It is unlikely, however, that they are interpreted as a stylistic choice, or that the participant decides to “let it slip.” Since no differences were found in either experiment, the more likely interpretation is that sentence judgments made immediately after reading are not affected unless stronger manipulations are used.	No significant difference in grammatical acceptability rating. No significant difference in semantic acceptability rating.
	<u>Interest</u> : The incongruent sentences did not affect people's stated interest in reading the hypothetical remainder of the text.	No significant difference in form-based interest rating. No significant difference in topic-based interest rating.

Table 7.6. Hypotheses and hypothesized effects on the experimental measures for all sentences, Experiment 2
 Separate hypotheses are given for the form of the sentence, and for the interest in reading the hypothetical remainder of the text. Hypotheses presented in Section 7.3.2 supported by the data in **bold**. All behavior and effects in the table are supported by the data. Differences are statistically significant; non-significant results are described as *no significant difference*.

Source bias	Behavior	Measured effects
Less-respected news source	Form: People expect sentences taken from less-respected news sources to contain worse language usage, textual errors, and incongruities.	• Lower grammatical acceptability rating compared to well-respected news sources. • Lower semantic acceptability rating compared to well-respected news sources.
	Interest: When only considering the language usage, which they assume is poorer, people are less interested in reading news from less-respected news sources. When disregarding the language usage, their interest in reading the hypothetical remainder of the text is not affected.	• Lower form-based interest rating compared to well-respected news sources. • No significant difference in topic-based interest rating compared to well-respected news sources.
Well-respected news source	Form: People expect sentences taken from well-respected news sources to contain better language usage, and fewer textual errors and incongruities.	• Higher grammatical acceptability rating compared to less-respected news sources. • Higher semantic acceptability rating compared to less-respected news sources.
	Interest: When only considering the language usage, which they assume is better, people are more interested in reading news from well-respected news sources. When disregarding the language usage, their interest in reading the hypothetical remainder of the text is not affected.	• Higher form-based interest rating compared to less-respected news sources. • No significant difference in topic-based interest rating compared to less-respected news sources.

7.7 Conclusions

This chapter shows that when giving a sentence judgment task to the participant immediately after reading a sentence, sentences believed to originate from less-respected news sources are perceived as using poorer grammar compared to sentences believed to originate from well-respected news sources. It can be argued that a news source that is well-respected is considered trustworthy by other people, and a less-respected source is considered to be less trustworthy. Based on previous research into trust and trustworthiness, information from less trustworthy sources is less convincing, more likely to be disregarded, and less persuasive. In fact, correct spelling was seen to improve credibility. It was thus hypothesized that sentence judgment tasks would be affected by this “bad-source bias” in this chapter, and that the language usage would be perceived as being poorer for the sentences expected to originate from less-respected news sources.

This chapter also shows that sentence judgment scores for semantically incongruent sentences are not lower compared to those for corresponding congruent sentences. I

hypothesized that semantically incongruent sentences would result in lower sentence judgment scores, since incongruities were made clearly erroneous by using incongruent antonyms such as *bad* instead of the congruent *good*. It was also hypothesized that semantically incongruent sentences would not affect acceptability judgment scores when the participants were expecting sentences from well-respected sources. It is possible that people perceived the semantic incongruities as metaphors, linguistic flair, or vague statements. It is more likely, however, that the participants were not capable of consistently making correct acceptability judgments. Performing spontaneous sentence judgments after a large number of sentences might require a more explicit incongruity than that used in this chapter. In addition, since grammatical acceptability produced similar results to semantic acceptability, not much benefit seems to be gained from separating these two judgments.

The semantic incongruities did not affect the sentence judgments in this chapter, but the source bias manipulation did. Chapter 8 investigates if effects from the semantic incongruities are present in eye movements during reading, as was the case in Chapter 6. The source bias manipulation, however, affected people in the hypothesized way, indicating it can be used to study the operation of the output similarity evaluator in Chapter 8.

Chapter 8

You sometimes read the word you expect instead of an incongruent word, but you process the incongruity nevertheless
Part 2: Sentences

8.1 Introduction

Chapter 6 used antonym pairs such as *high* and *low* to create words that would be either congruent or incongruent to the texts that were read by the participants, and measured the reading times on these words. The total reading times on the incongruent words were longer than congruent ones, both when people consciously noted the incongruent words and when they failed to detect them. If this is an important feature of the reading process, it is plausible to think that this should also be evident when using single sentences as stimuli. This chapter therefore investigates how reading is affected by successful and failed detection of semantic incongruities in sentences. This chapter also investigates how detection is affected by expectations of poor language usage in the sentences, operationalized as the expectation of sentences originating from less-respected or well-respected news sources. Chapter 7 verified that this manipulation indeed made people perceive the language usage in sentences from less-respected news sources as being poorer than the ones from well-respected ones.

Previous research shows that people can miss things they look directly at (Beanland & Pammer, 2010) and that they look at items located at incongruent locations longer than congruent locations, such as a toothbrush instead of a pen on a wooden desk by a window, but do not remember seeing them there when asked about it (Cornelissen & Võ, 2016). Previous research also shows that incorrectly performed analyses produce lingering effects in both eye movements and recall tasks, even if they are immediately corrected (Ferreira et al., 2001; Slattery et al., 2013; Tabor & Hutchins, 2004). Combined, it seems like this is evidence of processing beyond what people are aware of. As explained in Section 2.3.4, it is argued that parallel processes are used during reading, which can deliver different analyses (Clahsen & Felser, 2006;

Jackendoff, 2007). When encountering an incongruity, two different outputs are created. According to the good enough approach to language comprehension, one output is created by using the incongruent word in a syntactic analysis, and one output is created by using the word expected or predicted in the context in a heuristics analysis (Ferreira & Patson, 2007). One of these outputs needs to be selected and integrated in the meaning representation, while the other is disregarded. The proposed model for this process can be found in Section 6.3.1.1. The syntactic analysis proposed in the good enough approach is here named word-meaning retrieval for clarity. In this chapter, these results are further investigated by using sentences instead of paragraphs, again using antonyms such as *good* and *bad*. As in Chapter 6, disruptions in the reading process were expected both for explicitly detected incongruities and for incongruities that were not explicitly detected, which would indicate effects of implicit detection. The effects during implicit detection were hypothesized to arise from the discarded output containing the incongruent word, expressing similar lingering eye movement effects produced by erroneous outputs. In Chapter 6, explicit detection of incongruities did not produce reliable effects during the first visit to the word, but rather during later reading. However, because of the stronger experimental control of the sentences in this chapter, as well as a higher number of trials, disruptions in the reading process were expected as early as the first visit to the word during explicit detection. Explicit detection signifies a conscious reaction to an incongruity, which might be evident in initial stages of reading in addition to the later stages as well. Implicit detection on the other hand signifies more of an unconscious reaction, which should express effects in later stages of reading only.

In order to investigate the reading process during successful and failed detection of incongruities in sentences, two eye-tracking experiments were conducted. In Experiment 1, sentences were read by participants, with some sentences containing an incongruent word. The incongruity was created by replacing a correct word, such as *good*, with its corresponding antonym, such as *bad*. Just as in Chapter 6, detection was investigated in a post-experiment interview, asking about each incongruent text in turn. In Experiment 2, the only difference was to introduce a source bias before a group of sentences was read, by stating that the sentences originated from “well-respected” or “less-respected” news sources. Since the language usage of sentences from “less-respected” news sources was perceived as being poorer than sentences from “well-respected” news sources in Chapter 7, it is conceivable that people are much more critical of sentences from “less-respected” sources. Consequently, hearing that a sentence originated from “less-respected” sources was expected to increase the number of explicitly detected incongruities, and hearing that sentences originated from “well-respected” sources was expected to decrease the number of explicitly detected incongruities. Eye movement differences were investigated between the incongruities that had been successfully detected, the incongruities that were not explicitly detected, and the texts with the correct word. The sentences that participants were not able to

recall were classified in two ways: either they were categorized as not explicitly detected, or they were removed from the data (both analyses are shown).

8.2 Previous research

When investigating how incongruities affect the reading process, many types of incongruities have been used. A discussion of this was presented in Section 6.2.1. When incongruities are encountered and expectations are violated, the text or sentence needs to be reanalyzed. In some cases, the initial analysis leaves a memory trace. A discussion of this was presented in Section 6.2.2. Antonyms have been researched and used as experimental stimuli in many studies. A discussion of this was presented in Section 6.2.3.

A great deal of research has been carried out regarding people's trust in terms of information from different sources. Features such as correct spelling, for example, improved the credibility scores for internet blog posts (Rubin & Liddy, 2006; Weerkamp & Rijke, 2008). A discussion of this was presented in Section 7.2.2.

8.3 The study of implicit detection of incongruities during sentence reading

In two eye-tracking experiments, I investigated the effects of successful and failed detection of antonymic incongruities when participants were reading a sentence. To make sure the incongruities would be clearly incongruent in their contexts, while still having a high chance of not being detected, antonyms were used. The advantages of using antonyms as stimuli are briefly summarized here for clarity, but see Section 6.3 for a more thorough explanation. Antonyms are polar opposites within a meaning space with all other features being identical, making antonyms maximally similar and maximally different at the same time. This semantic relation is fundamental for native speakers, relies very little on world knowledge, the onset of detection is clear, and the binary meaning relation is identical in all manipulations.

Detection of incongruities was determined in a post-experiment interview, in which four fillers and each critical sentence were presented in a recall task. If a participant stated that anything had been strange with a sentence, it was classified as explicit detection; otherwise, it was classified as implicit detection. The sentences the participants said they forgot were also noted, and two different types of analyses were made. One classified forgotten sentences as implicitly detected, and one removed forgotten sentences from the data. It is important to note that the incongruities were highly conspicuous in the overall context, and clearly erroneous. All analyses investigated the critical word only.

An example sentence can be found in (20) below. All stimuli sentences can be found in Appendix 7.A. In (20), the antonym pair is indicated in italics, with the incongruent word also in bold. Only one word was given to each participant, with *hot* being the word given in the correct version of the sentence and *cold* being the word given in the incongruent version of the sentence. No words were displayed in bold or italics during the experiment, and only one version was shown to each participant.

- (20) As the high temperatures persisted, tourism in beach areas was greatly boosted by the *hot/cold* weather.

8.3.1 Terms and classifications

Terms and classifications used in this chapter are the same as those used in Chapter 6. They are briefly summarized here for clarity, but see also Section 6.3.1. Previous research suggests that two parallel processes operate during reading: one process that syntactically analyzes the text, and another that uses heuristics based on expectations, prediction, context, and previous knowledge. Since the manipulation is purely semantic in this study, the syntactic analysis in the good enough approach is referred to as *word-meaning retrieval*. Note that I do not argue for the validity of the term syntactic analysis, but merely extend the notion to also incorporate semantic details. The term word-meaning retrieval does not differentiate between semantic and syntactic details. The term word-meaning retrieval does not differentiate between semantic details and syntactic ones. It can plausibly be said that erroneous syntax is discovered by retrieving the meaning of the word and noticing that it does not fit.

8.3.1.1 Proposed model for when an incongruent word is fixated

This study introduces the two mechanisms *output similarity evaluator* and *expectation-based output selection* to explain the effects that explicit and implicit detection of incongruities have on the reading process. The proposed model is visualized in Figure 8.1a, Figure 8.1b, and Figure 8.1c. The two critical words used in the figures are taken from the incongruent version of (20) in Section 8.3 above. This means that *hot* is the word expected in the context, but *cold* is the actual word that the reader encounters. The model is briefly summarized here for clarity, but see Section 6.3.1.1 for a more thorough explanation. When an incongruent word is fixated, the two parallel processes that operate during reading create two outputs for every word that is fixated. Of these two outputs, only one is used during the creation of the cumulative meaning representation of the text. Figure 8.1a shows a schematic diagram of the output similarity evaluator when fixating on an incongruent word and comparing the two outputs. As mentioned in Section 6.3.1.1, the output similarity evaluator is biased toward using the heuristics analysis; in other words, the expected word is commonly favored. It is possible that this preference may increase or decrease depending on factors such as expectations of sentences with better or poorer language

usage. In Chapter 6, the claim that this preference may increase or decrease depending on better comprehension of the material was examined. Higher working memory capacity has been correlated with higher reading comprehension in several previous studies (see Daneman & Merikle, 1996, for an overview, but also see Van Dyke et al., 2014). In Chapter 3, comprehension increased with thorough reading, and decreased with skimming as compared to regular reading. These reading types were evaluated using total reading times, necessitating another measure to possibly evaluate comprehension in Chapter 6. The comprehension of the material was therefore estimated by measuring the working memory capacity of the participants. Working memory capacity did not affect the number of explicitly detected incongruities, however. Figure 8.1b shows a schematic diagram of expectation-based output selection under normal circumstances; that is, when the two outputs are considered sufficiently similar for regular reading to continue. Expectation-based output selection uses the heuristics output in the meaning representation of the text, and the reader remains unaware of the error. The incomplete output from the word-meaning retrieval is left incomplete, and is disregarded by the reading process. Figure 8.1c shows a schematic diagram of expectation-based output selection during explicit detection of an incongruity; that is, when the two outputs are considered sufficiently different to warrant further inspection by the reader. Expectation-based output selection is not used, and the word-meaning retrieval is added to the meaning representation. As mentioned in the introduction, a successfully detected incongruity is considered to be a case of explicit detection, and failure to notice any issues in the sentence is considered to be a case of implicit detection. Note that the model is simplified, describing only the selection process for whether or not the heuristics analysis is used during reading on a fixation-by-fixation basis.

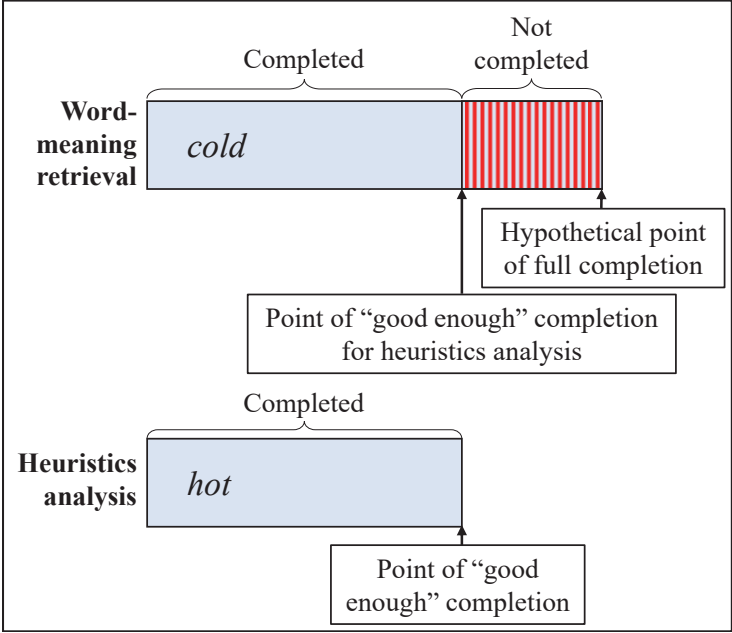


Figure 8.1a. Schematic diagram of the two different outputs created by the two parallel reading processes after fixating on an incongruent word
 The figure represents a schematic diagram of a fixation on the incongruent critical word in the example sentence (20) above.

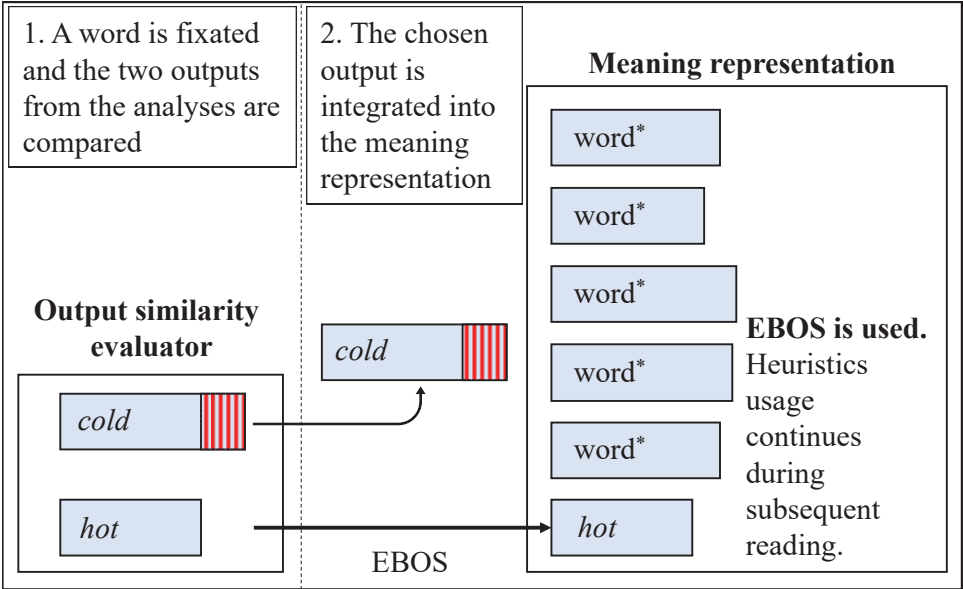


Figure 8.1b. Schematic diagram of the output similarity evaluator during normal operation
 The difference between the two outputs does not elicit a reaction, and expectation-based output selection (EBOS) is therefore used. The * denotes one small piece of the parsing process that is entered into the ongoing meaning representation. The example word in the figure is the incongruent critical word in the example sentence (20) above.

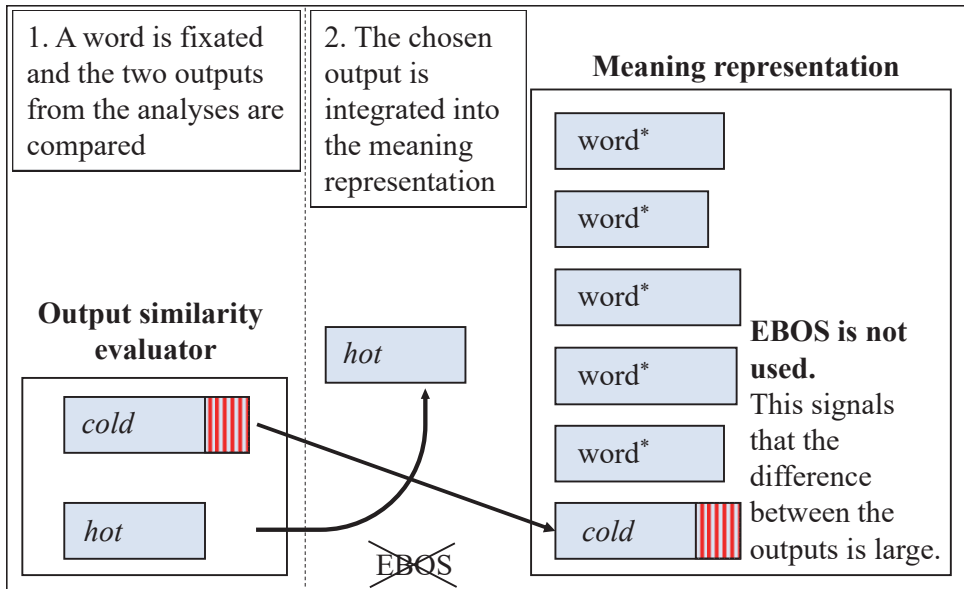


Figure 8.1c. Schematic diagram of the output similarity evaluator when the difference between the two outputs elicits a reaction

The difference between the two outputs elicits a reaction, and expectation-based output selection (EBOS) is therefore not used. The * denotes one small piece of the parsing process that is entered into the ongoing meaning representation. The example word in the figure is the incongruent critical word in the example sentence (20) above.

8.3.2 Hypotheses

As in Chapter 6, disruptions to the reading process are expected both during explicit detection of the incongruity, and also during failures to explicitly detect the incongruity, which is referred to as implicit detection. As is stated in Section 6.3.2, the disruptions during explicit detection are hypothesized to be the result of the participant noticing the incongruity, and rereading the incongruent word and the rest of the text as a result. The disruptions during implicit detection are hypothesized to be the result of the disregarded output from the parallel reading process that contains the incongruent word. The experimental manipulation can be said to be large, since antonym pairs are maximally different within the same meaning dimension. This creates a large contrast between the two parallel outputs, which should create visible effects for both detection and non-detection of the incongruity, compared to when the correct word is read. By using only one sentence, very little context is introduced. It should be more difficult to correctly identify the incongruity, since less context will highlight the incongruent nature of the word, but it should also be more likely to be explicitly detected, since there is less text to hide the error.

As in Section 6.3.2, evidence of implicit detection should be seen in eye movement effects similar to explicit detection, in the form of lingering effects arising from the incongruent output from the parallel analysis (Bottoms et al., 2010; Christianson et

al., 2001; Lau & Ferreira, 2005; Slattery et al., 2013; Tabor & Hutchins, 2004). Longer total reading times are expected for the incongruent word as compared to the congruent, control condition. First-pass reading times are not expected to be affected by detection, since previous research has shown that semantic incongruities do not affect early measures. Since the incongruities in this study are expected to be strong enough to trigger disruptions in reading immediately upon first seeing them, however, regression path durations and selective regression path durations should be longer for explicitly detected incongruities. Implicitly detected incongruities are only expected to affect late measures, since the reading process will not react initially, but only during subsequent reading. All effects of implicit detection should be smaller, but still significant when compared to the control condition.

To attempt to influence the total number of explicitly detected incongruities, a source bias was introduced in Experiment 2. Before a block of sentences was presented, a written instruction was given stating that they originated from either well-respected or less-respected news sources. Believing that a sentence originates from a well-respected news source is hypothesized to make the participants trust that the sentence is grammatically and semantically correct. This should result in a lower number of explicitly detected incongruities. Conversely, believing that a sentence originates from a less-respected news source is hypothesized to make the participants more likely to expect a sentence that is grammatically and semantically incorrect. This should result in a higher number of explicitly detected incongruities.

A summary of the hypotheses and the hypothesized effects are given in Table 8.1 and Table 8.2.

Table 8.1. Hypotheses and hypothesized effects on the experimental measures on the critical word compared to the control condition, Experiment 1 and Experiment 2

Hypotheses are given both for how the proposed model reacts and for the changes in the participant's reading behavior. The hypotheses and effects supported by the experimental data are reported in **bold** in Table 8.7 in Section 8.6.

Detection type	Hypothesis compared to the control condition	Hypothesized effects compared to the control condition
Implicit detection	<u>Model</u> : The output similarity evaluator judges the two outputs from the two analyses to be sufficiently similar, and uses expectation-based output selection to insert the output from the heuristics analysis into the meaning representation.	• Longer total reading times, but not to the extent that is found for explicit detection. • No difference in first-pass reading times. • No difference in regression path durations. • No difference in selective regression path durations. • No difference in second-pass reading times.
	<u>Behavior</u> : The participants continue their reading without consciously reacting to the incongruity. The output with the incongruent word, however, results in more rereading of the incongruent word. Effects should be observable in late measures only.	
Explicit detection	<u>Model</u> : The output similarity evaluator judges the two outputs from the two analyses to be different enough to require reanalysis. Expectation-based output selection is not used, and heuristics usage is halted in favor of conscious rereading of the sentence. The interruption is not immediate, but should occur during the first visit to the word.	• Longer total reading times. • No difference in first-pass reading times, since the interruption should only be visible after processing is completed. • Longer regression path durations. • Longer selective regression path durations. • Longer second-pass reading times.
	<u>Behavior</u> : The participants halt their reading in favor of rereading the incongruent word. Effects should be observable in all but the earliest stages of reading.	

Table 8.2. Hypothesis and hypothesized effects of source bias on explicit detection, Experiment 2

The hypotheses and effects supported by the experimental data are reported in **bold** in Table 8.8 in Section 8.6.

	Hypothesis	Hypothesized effects
Well-respected or less-respected news source	When hearing that sentences originate from less-respected news sources, people might implicitly assume that they contain more errors. Consequently, they should be better at discovering irregularities and errors in the sentences.	• A higher number of explicitly detected incongruities in sentences presented as originating from less-respected news sources, compared to the sentences from well-respected news sources.

8.4 Experiment 1

8.4.1 Method

8.4.1.1 Participants

Forty-six native speakers of English (28 female; 18 to 26 years of age) participated in the experiment. The participants were recruited with SONA Experiment Scheduling at the Rayner Lab, University of California, San Diego, United States. All participants had normal or corrected-to-normal vision. One participant was excluded from the results due to technical problems with the eye-tracking equipment. This resulted in 45 participants (27 female; $M=20.4$ years, $SD=2.04$). The participants were naïve to the purpose of the study.

8.4.1.2 Apparatus

Eye movement data were recorded with an EyeLink 1000 plus at 500 Hz (SR Research, Ottawa, Canada). Recordings were monocular using the right eye, which is the dominant eye for the majority of people. The recordings were conducted at the Rayner Lab, Department of Psychology, University of California, San Diego, United States. The head of the participant was stabilized using a chin and forehead rest during recording. The distance between the computer monitor and the participant's eyes was approximately 60 cm. Stimuli were displayed on an HP 1230 20" 4:3 CRT monitor at 1,280 × 1,024 pixels (40.5 × 30.5 cm; approximately 37.4 × 28.5 visual degrees at a distance of 60 cm). The eye-tracking system was controlled on a data recording computer using the host software for EyeLink 1000. Stimuli presentation, a 13-point calibration, and a 13-point validation of the calibration accuracy were handled using Experiment Builder (1.10.165) running on a separate stimulus computer.

The pre-experiment questionnaire was given in PsychoPy (1.80.00) on a separate laptop computer and completed by the participant. The post-experiment interview used PsychoPy (1.80.00) on a separate laptop computer to present the sentences for the participant, and the responses were recorded by the experimenter.

8.4.1.3 Design and stimuli

A total of 80 sentences were used in two experimental lists, related to the semantic incongruence manipulation. The sentences were taken from the Corpus of Contemporary American English (COCA), using the portion of the corpus with material taken from news sources. One-syllable adjective antonym pairs, such as *large/small*, *soft/hard*, *cold/hot*, *good/bad*, and *slow/fast*, were chosen for the stimuli; 10 antonym pairs in total were used. In addition, 40 additional one- and two-syllable adjectives were chosen for usage as fillers in the experiment. Two sentences with each word in the antonym pairs, and one sentence with each filler word were chosen from the corpus, and adapted for the experiment. The selected word needed to be used at the end of the sentence, and the first part of the sentence was used to introduce some context in order to make the semantic incongruity as evident as possible. The corresponding antonym in each pair was substituted in each of the 40 sentences to create 40 sentences containing semantic incongruities. A total of 20 congruent critical sentences and a total of 20 sentences with semantic incongruities were used for each experimental list, meaning that each word was used both as a semantic incongruity and a regular adjective in two different sentences. The remaining 40 sentences with the additional adjectives were shown to all participants. All stimuli sentences can be found in Appendix 7.A.

The sentences were displayed using a monospaced font, Courier New 18-point (approximately 4.5 × 7.0 mm, approximately 0.30 × 0.45 degrees of visual angle at a distance of 70 cm). All sentences were displayed on two lines, with the critical word

always shown on the second line. An example stimulus sentence can be found in (20) in Section 7.3.

The sentences were presented in 10 groups of eight sentences. Each group contained two congruent critical sentences, two sentences containing semantic incongruities, and four filler sentences. The order of the sentences was randomized using Experiment Builder, and the two experimental lists were counterbalanced. Written instructions before each block directed the participants to click the left mouse button when they were satisfied that they had understood the sentence, and they would then be taken to the next item. A task unrelated to the current study—a sentence judgment task—was performed by the participants to ensure they were reading the sentences.

8.4.1.4 Procedure

Each participant was introduced to the experiment and recorded individually. The participants were welcomed to the study, and completed a pre-experiment questionnaire in which their age and gender were noted. The experimental procedure was explained, with the participants told that a sentence would be presented in isolation on screen. The participants were to read the sentence until they understood it, and then press the space bar. After each sentence, the participants answered questions about the sentence. The questions were present in the experiment partly to make sure the participant was paying attention, and these data were not used in this study (the reader has probably noticed that these data were analyzed in Chapter 7). The participants were then seated in the eye tracker, and the experiment began with four practice trials. If the participants had any questions, they could ask them after the practice trials. If not, the main experiment started, beginning with a calibration and validation procedure. The calibration was repeated until the deviation was under 0.7 degrees of visual angle in both the horizontal and the vertical directions. The 80 sentences were then presented, one per screen. The eye-tracking recording was monitored, and a calibration and validation procedure was initiated and completed during the experiment if needed.

A post-experiment interview was carried out immediately after the eye-tracking portion of the experiment. The first half to two-thirds of one sentence was shown to participants, who were asked to “recall the next word in the sentence” in order to decrease suspicion of the experimental manipulation. They were then asked if they remembered if there was anything wrong with the sentence during the experiment. If a participant stated that they saw something wrong with the sentence, regardless of what this was, the trial was classified as explicit detection. If a participant said there had been nothing wrong with the sentence, the trial was classified as implicit detection. If a participant stated that they did not remember the sentence, the trial was classified as a forgotten trial. Rather than specifically asking “try to remember this sentence, and think about if it had any errors in it,” the task to recall the next word in the sentence was created to lower the likelihood of the participants understanding the

research interest, but still make a deliberate and conscious effort to recall the sentence. The task began with two congruent filler sentences, followed by all 20 incongruent sentences. The incongruent word was never shown to participants.

After the post-experiment interview, the participants were shown examples of incongruent sentences, and the research question was explained. The participants were again thanked for their participation, and given a course credit as compensation.

8.4.1.5 Data analysis

Velocity-based high-speed event detection was performed in SR Research Data Viewer using the default settings (acceleration threshold=8,000°/s², velocity threshold=30°/s), which transformed the raw data into fixations and saccades. Eye tracking data were cleaned for usage in reading research with the in-built automatic process in Data Viewer using the default settings, with the exception that fixations from 80ms to 1,000 ms were used. Only trials in which the critical word was fixated at least once were included in the models.

The data processing, plots, and statistical models were made using R (3.2.2) (R Development Core Team, 2015) and the lme4 package (Bates et al., 2015) in R Studio (0.99.473). Linear mixed effect models were used to examine the effect of explicit and implicit detection on first-pass, second-pass, and total reading times, as well as regression path durations and selective regression path durations on the critical word. Only words that were fixated at least once were used in the analyses. All data were log-transformed to better fit a Gaussian distribution. A fixed effect with three levels (congruent control condition, explicitly detected incongruity, implicitly detected incongruity) was used. Models used random intercepts and slopes of condition within group for participant and item, but no models converged with intercept-slope correlation factors for participants and items. The lmerTest package was used to estimate *p*-values with Satterthwaite approximation.

Two analyses were performed for all measures: one in which forgotten sentences were classified as implicit detection, and one in which forgotten sentences were removed from the statistical analyses. The decision to classify forgotten sentences as implicitly detected was made in cases in which nothing in the sentence made the participant react.

8.4.2 Results

All statistical models used log-transformed data, but in order to improve readability, no log-transformation was performed on the data in the figures.

8.4.2.1 Performance

Table 8.3 shows detection of incongruities for Experiment 1, excluding the trials where the critical word was not fixated. Results are shown as both the total number of trials, and the proportion of detection type (implicit detection, explicit detection, or forgotten sentence). If classifying the forgotten sentences as implicit detection, as is done in the main analyses in this chapter, around 71.8% of incongruities were not explicitly detected by participants. This is the preferred classification in this chapter. When removing the forgotten sentences, around 58.3% of incongruities were not explicitly detected by participants. Note that analyses with this classification are also presented in the results section.

Table 8.3. Explicit and implicit detection of incongruities, as well as sentences that were forgotten by the participants, Experiment 1

Only trials where the critical word was fixated at least once are reported, both the total number of trials (count) and the proportion of trials in the critical condition (percent).

	Trials	Proportion
Control condition	798	-
Implicit detection	299	39.4
Explicit detection	214	28.2
Forgotten sentence	246	32.4

8.4.2.2 Eye movement measures

Table 8.4 shows all eye movement results for Experiment 1. Total reading times were significantly longer for both explicitly detected incongruities (Estimate=0.202, $SE=0.062$, $t=3.27$, $p<.005$) and implicitly detected incongruities (Estimate=0.126, $SE=0.053$, $t=2.36$, $p<.02$), when classifying forgotten sentences as implicit detection. When removing forgotten sentences from the analysis, total reading times were significantly longer for explicitly detected incongruities (Estimate=0.212, $SE=0.061$, $t=3.47$, $p<.001$), but only trending towards significance for implicitly detected incongruities (Estimate=0.104, $SE=0.057$, $t=1.81$, $p=.08$). Figure 8.2 shows total reading times for the control condition, explicit detection, and implicit detection with forgotten sentences included. The figure showing implicit detection with forgotten sentences removed from the data can be found in Appendix 8.A.

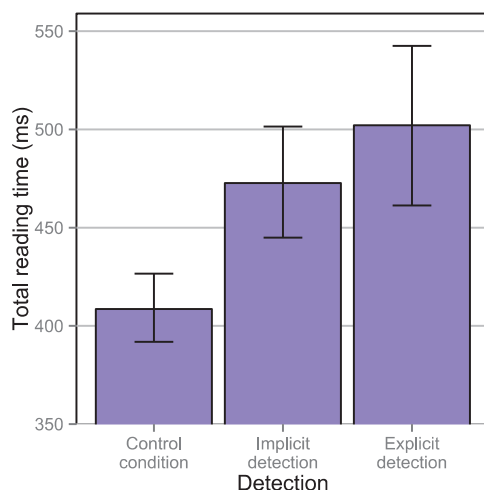


Figure 8.2. Mean total reading times (ms) on the critical word for the control condition, implicit detection, and explicit detection, Experiment 1
Error bars denote standard errors of the mean.

Table 8.4. Mean eye movement results on the critical word, Experiment 1

All measures are given in milliseconds with standard errors in parentheses. Statistically significant differences when comparing with the control condition are presented in **bold**. * indicates that the LMM reported $p < .05$, ** indicates $p < .01$, † indicates that the results were only trending towards significance ($p = .08$).

	Control condition	Implicit detection (including forgotten sentences)	Implicit detection (without forgotten sentences)	Explicit detection
<i>Total reading time</i>	409 (9.12)	473* (14.5)	451† (17.5)	502** (20.0)
<i>First-pass reading time</i>	239 (3.56)	242 (4.45)	243 (5.59)	249 (8.19)
<i>Regression path duration</i>	498 (27.4)	467 (34.9)	468 (49.7)	648 (75.8)
<i>Selective regression path duration</i>	264 (4.84)	267 (6.36)	264 (8.13)	292† (12.7)
<i>Second-pass reading time</i>	225 (5.88)	232 (6.42)	235 (9.11)	228 (9.07)

Figure 8.3 shows selective regression path durations for the control condition, explicit detection, and implicit detection with forgotten sentences included. The figure showing implicit detection with forgotten sentences removed from the data can be found in Appendix 8.A.

No significant differences in first-pass reading times were found for explicitly detected incongruities (forgotten sentences as implicit detection: Estimate=0.027, $SE=0.038$, $t=0.73$, $p=.47$; forgotten sentences removed: Estimate=0.032, $SE=0.039$, $t=0.80$, $p=.43$) or implicitly detected incongruities (forgotten sentences as implicit detection: Estimate=0.0040, $SE=0.032$, $t=0.13$, $p=.90$; forgotten sentences removed: Estimate=0.020, $SE=0.037$, $t=0.55$, $p=.59$).

No significant differences in regression path durations were found for explicitly detected incongruities (forgotten sentences as implicit detection: Estimate=0.085,

$SE=0.062$, $t=1.38$, $p=.17$; forgotten sentences removed: Estimate=0.091, $SE=0.063$, $t=1.45$, $p=.15$) or implicitly detected incongruities (forgotten sentences as implicit detection: Estimate=-0.040, $SE=0.044$, $t=-0.91$, $p=.36$; forgotten sentences removed: Estimate= -0.052, $SE=0.055$, $t=-0.95$, $p=.35$).

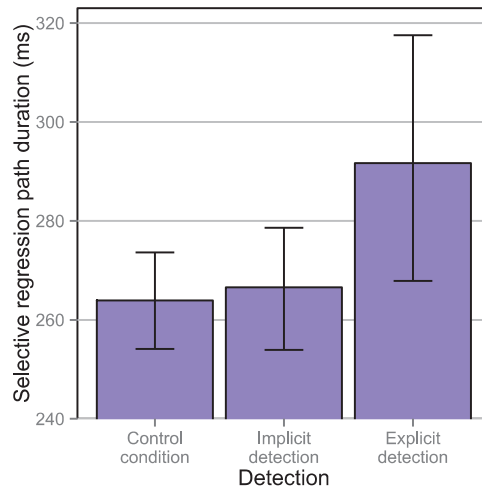


Figure 8.3. Mean selective regression path durations (ms) on the critical word for the control condition, implicitly detected incongruent words, and explicitly detected incongruent words, Experiment 1
Error bars denote standard errors of the mean.

Selective regression path durations were longer for explicitly detected incongruities, but only trending towards significance (forgotten sentences as implicit detection: Estimate=0.060, $SE=0.034$, $t=1.78$, $p=.08$; forgotten sentences removed: Estimate=0.061, $SE=0.033$, $t=1.82$, $p=.07$). No significant differences were found for implicitly detected incongruities (forgotten sentences as implicit detection: Estimate=-0.0066, $SE=0.024$, $t=-0.27$, $p=.78$; forgotten sentences removed: Estimate=-0.0079, $SE=0.029$, $t=-0.27$, $p=.79$).

No significant differences in second-pass reading times were found for explicitly detected incongruities (forgotten sentences as implicit detection: Estimate=0.016, $SE=0.041$, $t=0.38$, $p=.71$; forgotten sentences removed: Estimate=0.018, $SE=0.041$, $t=0.43$, $p=.67$) or implicitly detected incongruities (forgotten sentences as implicit detection: Estimate=0.021, $SE=0.032$, $t=0.65$, $p=.52$; forgotten sentences removed: Estimate=0.038, $SE=0.039$, $t=-0.97$, $p=.33$).

8.4.3 Discussion

It has been suggested that reading uses two parallel processes, one relying on syntactic information and one relying on heuristics based on expectations and previous

experience (Christianson et al., 2010; Ferreira, 2003; Ferreira & Engelhardt, 2009; Ferreira & Patson, 2007). The many cases in which semantic anomalies are missed is considered to be evidence of this (Barton & Sanford, 1993; Bohan & Sanford, 2008; Croft & Cruse, 2004; Hannon & Daneman, 2004; Paradis, 2015). The parallel process argued to rely on syntactic information is called word-meaning retrieval in the current study, in order to avoid confusion between syntax and semantics. When encountering an incongruity, two different outputs are produced by these parallel reading processes: one correctly identifies the incongruity, and the other uses heuristics to predict a word that is correct in the context, and therefore does not correctly correspond to the actual word. One of these outputs is integrated into the meaning representation, and one needs to be disregarded. The model is thoroughly explained in Section 6.3.1.1. A thorough explanation of how the model explains the results for both implicit and explicit detection can be found in Section 6.4.3.

When incongruities were not explicitly detected in this study, significantly longer total reading times were found for the critical word, suggesting effects of implicit detection. The difference between the output from the heuristics analysis and the output from the word-meaning retrieval were not deemed sufficiently large by the output similarity evaluator to elicit a reaction. Instead, the participant continued reading without consciously noticing that anything is wrong. It is plausible that the lingering effects in eye movement measures are disruptions to the reading process caused by this disregarded analysis. If so, the effects can be related to the findings that after an already performed analysis of text is changed, lingering effects of the first analysis have been seen in both recall (Christianson et al., 2001; Ferreira et al., 2001; Lau & Ferreira, 2005) and eye movement measures (Slattery et al., 2013; Tabor & Hutchins, 2004). These effects are also similar to how implicit detection of incongruities manifests in scene viewing, in which incongruent items not consciously detected by participants nevertheless produce increased total dwell times (Cornelissen & Võ, 2016). It has also been suggested that people may be unaware of their eye movement patterns (Võ et al., 2016), providing further strength for the argument that language processing occurs outside of awareness, but is present in eye movements. In the analysis without forgotten sentences, results only trended toward significance ($p=.08$). This is not surprising, since these trials constitute more than 30% of the data in the incongruent condition.

When incongruities were explicitly detected in this study, significantly longer total reading times were found for the critical word. The difference between the output from the heuristics analysis and the output from the word-meaning retrieval was deemed sufficiently large by the output similarity evaluator to elicit a reaction. Expectation-based output selection was not used. Instead, the word-meaning retrieval was added into the meaning representation, making the reader aware of the error. The sentence was then reread.

Earlier effects for explicit detection were also found, with increases in selective regression path durations for the critical word. Note that these results were only

trending toward significance. Another experiment was thus necessary in order to replicate and corroborate the results for selective regression path durations for the critical words for explicitly detected incongruities, since they were only trending toward significance. In addition, total reading times for implicit detection were only trending toward significance without forgotten sentences, further suggesting the need for another experiment. In Experiment 2, no changes were made to the stimuli or procedure. In order to try to influence the number of explicitly detected incongruities, a source bias was introduced by instructing the participants that the upcoming sentences were either taken from “well-respected” or “less-respected” news sources. In Chapter 7, it was found that the sentences believed to be taken from well-respected news sources were considered more accurate and well-formed compared to sentences believed to be taken from less-respected news sources. Therefore, the number of explicitly detected incongruities is hypothesized to be lower when the participants expect sentences from well-respected news sources, and higher when expecting sentences from less-respected news sources. In a way, this could be described as a type of peer influence, since it describes how other people consider the sources, much in the way that texts in Chapter 4 and Chapter 5 were said to be about controversial topics that other people felt strongly about. Calling something a well-respected news source could signify that other people consider it to be trustworthy, reliable, and fair. Calling something a less-respected news source, on the other hand, could signify that other people consider it to be untrustworthy, unreliable, and biased.

8.5 Experiment 2

Except for the differences explained in the Method section below, Experiment 2 was identical to Experiment 1.

8.5.1 Method

8.5.1.1 *Participants*

Seventy-eight native speakers of English (54 female; 18 to 26 years of age) participated in the experiment. Seven participants were excluded from the results due to problems with the eye-tracking equipment. This resulted in 71 participants in total (47 female; $M=20.2$ years, $SD=1.50$).

8.5.1.2 *Apparatus*

Identical to Experiment 1.

8.5.1.3 Design and stimuli

The 80 stimuli sentences from Experiment 1 were divided into four experimental lists: two related to the semantic incongruence manipulation, and two related to source expectation. Before each group of eight sentences, participants were told that the sentences originated from news sources considered to be either “trusted and well-respected according to polling data,” or “less trusted and less-respected according to polling data.” Henceforth, these sources will be referred to as *well-respected* and *less-respected*. Five groups were said to be from well-respected sources, and five groups were said to be from less-respected sources, with the groups presented in an alternating fashion. The order of the groups was randomized in Experiment Builder, and the source condition of all sentences was counterbalanced on the lists.

8.5.1.4 Procedure

The participants were given oral instructions that before each group of sentences, they would be informed whether the sentences were taken from well-respected or less-respected sources.

The participants were told about the true nature of the source condition after the experiment.

8.5.1.5 Data analysis

The source condition was used as a factor when examining differences in detection between sentences believed to be from well-respected and less-respected sources. The source condition was also used as a factor in all eye-tracking analyses.

8.5.2 Results

8.5.2.1 Performance

Table 8.5 shows detection of incongruities for Experiment 2, only using trials where the critical word was fixated at least once. Results are shown as both the total number of trials, and the proportion of detection type (implicit detection, explicit detection, or forgotten sentence). Source condition did not significantly affect the results (forgotten sentences as implicit detection: Estimate=-0.055, $SE=0.131$, $z=-0.42$, $p=.67$; forgotten sentences removed: Estimate=-0.095, $SE=0.144$, $z=-0.66$, $p=.51$). Figure 8.4 shows the proportion of explicit detection, implicit detection, and forgotten sentences depending on source condition. If classifying the forgotten sentences as implicit detection, as is done in the main analyses in this chapter, around 67.4% of incongruities were not explicitly detected by participants. This is the preferred classification in this chapter. When removing the forgotten sentences, around 55.2% of incongruities were not explicitly detected by participants. Note that analyses with this classification are also presented in the results section.

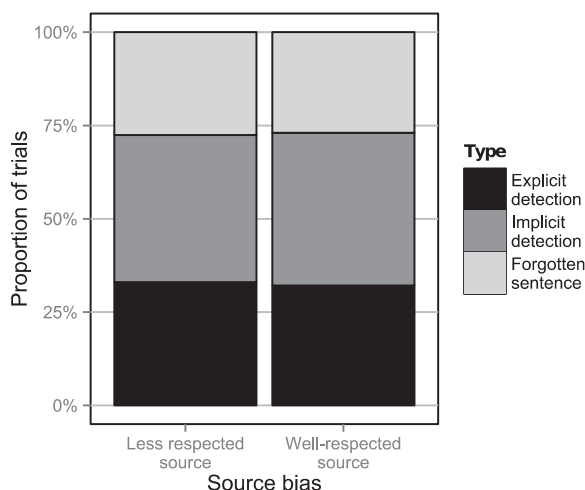


Figure 8.4. Proportion of explicit detections, implicit detections, and forgotten sentences, depending on source bias condition, Experiment 2

Only trials where the critical word was fixated at least once are reported in the figure.

Table 8.5. Explicit and implicit detection of incongruities, as well as sentences that were forgotten by the participants, Experiment 2

Only trials where the critical word was fixated at least once are reported, both the total number of trials (count) and the proportion of trials in the critical condition (percent).

	Well-respected sources		Less-respected sources	
	Trials	Proportion	Trials	Proportion
Control condition	610	-	610	-
Implicit detection	253	40.9	239	39.4
Explicit detection	199	32.2	200	33.0
Forgotten sentence	167	27.0	167	27.6

8.5.2.2 Eye movement measures

Table 8.6 shows all eye movement results for Experiment 2. Figure 8.5 shows total reading times for the control condition, explicit detection, and implicit detection with forgotten sentences included, divided into source conditions. The figure showing implicit detection with forgotten sentences removed can be found in Appendix 8.A. Total reading times were significantly longer for both explicitly detected incongruities (forgotten sentences as implicit detection: Estimate=0.288, $SE=0.047$, $t=6.08$, $p<.0001$; forgotten sentences removed: Estimate=0.285, $SE=0.046$, $t=6.22$, $p<.0001$) and implicitly detected incongruities (forgotten sentences as implicit detection: Estimate=0.121, $SE=0.043$, $t=2.84$, $p<.01$; forgotten sentences removed: Estimate=0.129, $SE=0.044$, $t=2.94$, $p<.01$). Source condition did not significantly affect the results (forgotten sentences as implicit detection: Estimate=-0.010, $SE=0.022$, $t=-0.46$, $p=.65$; forgotten sentences removed: Estimate=-0.021, $SE=0.24$, $t=-0.87$, $p=.38$).

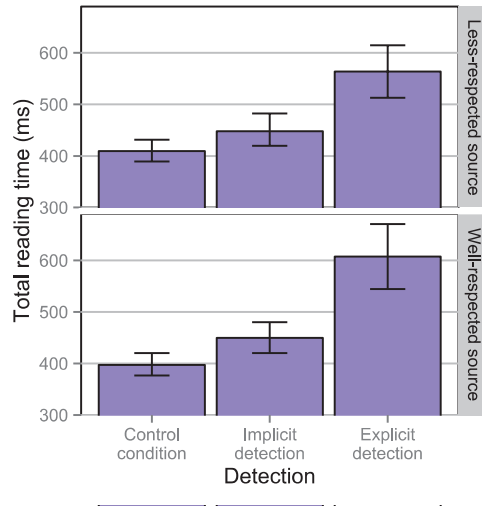


Figure 8.5. Mean total reading times (ms) on the critical word for control condition, implicitly detected incongruent words, and explicitly detected incongruent words, depending on source bias condition, Experiment 2

Error bars denote standard errors of the mean.

Table 8.6. Mean eye movement results on the critical word, Experiment 2

All measures are given in milliseconds with standard errors in parentheses. Statistically significant differences when comparing with the control condition are presented in **bold**. * indicates that the LMM reported $p < .05$, ** indicates $p < .01$, *** indicates $p < .001$. No significant effects of source condition were found.

	Control condition	Implicit detection (including forgotten sentences)	Implicit detection (without forgotten sentences)	Explicit detection
<i>Total reading time</i>				
Well-respected source	397 (10.9)	448 (15.3)	443 (19.7)	608 (31.4)
Less-respected source	410 (10.9)	450 (15.8)	464 (22.0)	564 (26.0)
Total	403 (7.73)	449** (11.0)	453** (14.7)	586*** (20.4)
<i>First-pass reading time</i>				
Well-respected source	242 (4.16)	235 (5.10)	243 (6.71)	252 (6.93)
Less-respected source	246 (4.19)	247 (5.00)	240 (6.19)	244 (7.35)
Total	244 (2.95)	241 (3.58)	242 (4.57)	248 (5.05)
<i>Regression path duration</i>				
Well-respected source	481 (28.8)	471 (35.6)	464 (49.1)	681 (84.0)
Less-respected source	464 (29.1)	462 (36.5)	441 (44.2)	766 (117.2)
Total	472 (20.5)	467 (25.5)	453 (33.1)	724 (72.1)
<i>Selective regression path duration</i>				
Well-respected source	267 (5.27)	268 (6.98)	262 (9.19)	302 (13.5)
Less-respected source	270 (5.46)	270 (6.03)	270 (7.44)	302 (13.2)
Total	269 (3.79)	269 (4.62)	266 (5.94)	302* (9.46)
<i>Second-pass reading time</i>				
Well-respected source	234 (7.71)	228 (6.72)	232 (8.85)	243 (8.85)
Less-respected source	234 (6.96)	232 (7.88)	230 (10.7)	233 (8.26)
Total	234 (5.18)	230 (5.16)	231 (6.97)	238 (6.04)

Figure 8.6 shows selective regression path durations for the control condition, explicit detection, and implicit detection with forgotten sentences included, divided into source conditions. The figure showing implicit detection with forgotten sentences removed can be found in Appendix 8.A. No significant differences in first-pass reading times were found for explicitly detected incongruities (forgotten sentences as implicit detection: Estimate=0.016, $SE=0.026$, $t=0.61$, $p=.54$; forgotten sentences removed: Estimate=0.014, $SE=0.025$, $t=0.58$, $p=.57$) or implicitly detected incongruities (forgotten sentences as implicit detection: Estimate=-0.011, $SE=0.022$, $t=-0.50$, $p=.62$; forgotten sentences removed: Estimate=-0.0083, $SE=0.024$, $t=-0.35$, $p=.73$). Source condition did not significantly affect the results (forgotten sentences as implicit detection: Estimate=-0.019, $SE=0.015$, $t=-1.29$, $p=.20$; forgotten sentences removed: Estimate=-0.0074, $SE=0.016$, $t=-0.47$, $p=.64$).

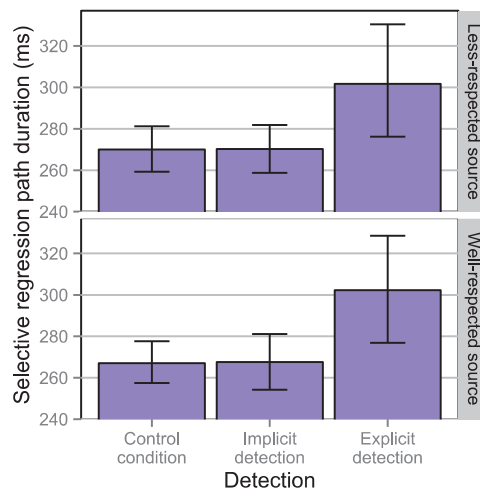


Figure 8.6. Mean selective regression path durations (ms) on the critical word for the control condition, implicitly detected incongruent words, and explicitly detected incongruent words, depending on source bias condition, Experiment 2
Error bars denote standard errors of the mean.

No significant differences in regression path durations were found for explicitly detected incongruities (forgotten sentences as implicit detection: Estimate=0.125, $SE=0.077$, $t=1.62$, $p=.11$; forgotten sentences removed: Estimate=0.114, $SE=0.084$, $t=1.35$, $p=.18$) or implicitly detected incongruities (forgotten sentences as implicit detection: Estimate=0.018, $SE=0.072$, $t=0.25$, $p=.80$; forgotten sentences removed: Estimate=-0.015, $SE=0.083$, $t=-0.19$, $p=.85$). Source condition did not significantly affect the results (forgotten sentences as implicit detection: Estimate=0.0011, $SE=0.030$, $t=0.04$, $p=.97$; forgotten sentences removed: Estimate=0.015, $SE=0.032$, $t=0.47$, $p=.64$).

Selective regression path durations were significantly longer for explicitly detected incongruities (forgotten sentences as implicit detection: Estimate=0.070, $SE=0.032$, $t=2.20$, $p<.03$; forgotten sentences removed: Estimate=0.067, $SE=0.033$, $t=2.05$, $p<.05$). No significant differences were found for implicitly detected incongruities (forgotten sentences as implicit detection: Estimate=0.0041, $SE=0.028$, $t=0.15$, $p=.88$; forgotten sentences removed: Estimate=-0.0086, $SE=0.031$, $t=-0.28$, $p=.78$). Source condition did not significantly affect the results (forgotten sentences as implicit detection: Estimate=-0.016, $SE=0.017$, $t=-0.96$, $p=.33$; forgotten sentences removed: Estimate=-0.0071, $SE=0.018$, $t=-0.39$, $p=.69$).

No significant differences in second-pass reading times were found for explicitly detected incongruities (forgotten sentences as implicit detection: Estimate=0.028, $SE=0.031$, $t=0.92$, $p=.36$; forgotten sentences removed: Estimate=0.027, $SE=0.031$, $t=0.87$, $p=.39$) or implicitly detected incongruities (forgotten sentences as implicit detection: Estimate=-0.0086, $SE=0.027$, $t=-0.32$, $p=.75$; forgotten sentences removed: Estimate=0.0002, $SE=0.032$, $t=0.01$, $p=.99$). Source condition did not significantly affect the results (forgotten sentences as implicit detection: Estimate=0.0076, $SE=0.023$, $t=0.34$, $p=.74$; forgotten sentences removed: Estimate=0.011, $SE=0.024$, $t=0.46$, $p=.65$).

The figure showing total reading times per character on the words in the filler sentences can be found in Appendix 8.B. No significant differences in total reading times per character were found for the words in the filler sentences depending on source condition (all words in the sentence: Estimate=0.005, $SE=0.012$, $t=0.40$, $p=.69$; only the adjective: Estimate=0.039, $SE=0.026$, $t=1.52$, $p=.14$).

8.5.3 Discussion

Significantly longer total reading times were found for both explicit and implicit detection compared to the control condition. During explicit detection, the participant is aware of the incongruity, and therefore rereads the sentence. During implicit detection, the disregarded word-meaning retrieval is likely producing lingering effects that are occurring outside of the participant's awareness.

The current study found significantly longer selective regression path durations for explicitly detected incongruities on the critical word when compared to the correct control condition, but not for implicitly detected incongruities, highlighting a difference between the two types of detection. When not investigating detection, Braze et al. (2002) found that pragmatic anomalies produced longer first-pass reading times on the word following the anomaly compared to the correct control condition. Bohan and Sanford (2008) used semantic incongruities to investigate the detection of incongruities, and found longer regression-path durations for explicitly detected incongruities in the region following the one containing the incongruity, but not on the incongruity itself. Daneman et al. (2007) also used semantic incongruities, and also found effects of detection in eye movements, but again only in late eye-tracking

measures. First-pass reading times showed earlier effects on the reading process than selective regression path durations, but no significant differences were found in first-pass reading times in this study. Selective regression path duration, however, can be considered a fairly early measure, since the forward saccade needs to be interrupted at a very early stage in order to not be launched. This suggests that detection might occur earlier than suggested by other research, and is visible in eye movement measures as early as the first visit to the incongruent word. In addition, it suggests that antonymic pairs are well-suited for usage as incongruities in stimuli sentences,

Expertise has not been shown to affect the explicit detection of inattentional blindness (Drew et al., 2013), and even when participants look directly at the stimulus, they often fail to explicitly detect it (Beanland & Pammer, 2010; Drew et al., 2013; Simons & Chabris, 1999). In an attempt to change the number of explicitly detected incongruities in the current study, a sentence bias manipulation was introduced that stated that sentences were either taken from well-respected or less-respected news sources. A higher level of trust or respect regarding the source was expected to increase the likelihood of explicitly detecting incongruities. No significant differences were found depending on the source bias, however, with detection being almost identical in the two conditions.

8.6 General discussion

Two eye-tracking experiments investigated how explicit detection of antonym incongruities in sentences affects reading, as well as how the costs of implicit detection are manifested in the reading process.

Previous research on semantic incongruities has shown effects on eye movements in the region after the critical word, but little evidence has been found for the critical word. Furthermore, little research has been performed on incongruities that were not detected by participants, and only paragraphs have been used in these studies. Effects of implicit detection, however, have been found in several studies investigating visual search and scene semantics, as shown, for instance, in total dwell times on the critical area (Cornelissen & Vö, 2016).

Since paragraphs were used as stimuli in Chapter 6, in Chapter 8 I intended to see whether these increases in total reading times would also be found for isolated sentences when using antonym incongruities. In Chapter 7, these incongruities produced no effects in sentence judgments. This is likely because of the fact, however, that sentence judgments are offline measures. The source bias did produce perceptions of poorer language usage in sentences believed to originate from less-respected news sources compared to the ones believed to originate from well-respected sources. The source bias was thus valuable when analyzing the output similarity evaluator in this chapter, see Section 6.6 for a thorough analysis and discussion. The

discussion is based on the suggested argument that reading uses separate processing routes (Jackendoff, 2007; Kim & Osterhout, 2005; van Herten et al., 2006), with the one based on expectations and previous experiences generally favored because it facilitates and automates processing (Bohan & Sanford, 2008; Christianson et al., 2010; Ferreira, 2003).

Longer total reading times were found for incongruities that were not explicitly detected when compared with the correct control condition in both experiments. Since the incongruity is not seen as an error requiring explicit attention from the reader, the output similarity evaluator selects the heuristics analysis and integrates it into the meaning representation. The term expectation-based output selection is coined to describe this process, which occurs in the experiments presented in both this chapter and in Chapter 6. It is very possible that the analysis kept outside of awareness by the output similarity evaluator is what produces the lingering effects in late eye-tracking measures. These effects of implicit detection are similar to the way incorrectly formed analyses can affect subsequent reading (Christianson et al., 2001; Ferreira et al., 2001; Lau & Ferreira, 2005; Slattery et al., 2013; Tabor & Hutchins, 2004). It seems that semantic information is indeed used when an incongruity is not explicitly detected (Sanford et al., 2011). The good enough approach states that language processing stops when it is considered good enough, while the shallow processing account states that language processing is shallow (Ferreira & Patson, 2007; Sanford & Graesser, 2006; Sanford et al., 2011). This would suggest somewhat deeper processing of the incongruent word in the good enough approach. It could thus be argued that the good enough approach fits the results better than the shallow processing account.

Longer total reading times were also found for explicitly detected incongruities when compared with the correct control condition in both experiments. Since the incongruity is seen as an error requiring explicit attention from the reader, the output similarity evaluator selects the word-meaning retrieval, which is only partially completed, and integrates it into the meaning representation. Thus, expectation-based output revision is not performed, which indicates that the difference between the two reading processes is large and that the text warrants more thorough reading. The heuristics usage is therefore paused as the reader is made aware of the error. Interestingly, evidence of early disruptions to the reading process was also found for explicitly detected incongruities, with longer selective regression path durations in both experiments. The increase was only trending toward significance in Experiment 1, but was significantly longer in Experiment 2. This indicates that the word-meaning retrieval was sufficiently complete to produce effects in early eye-tracking measures, since the forward saccade was interrupted in favor of a regression to previous parts of the sentence. First-pass reading times, however, showed no effects on the reading process. This is not surprising, since it is an early measure, and longer first-pass reading times are generally observed for effects such as word length and word frequency (Rayner, 1998, 2009). Longer regression-path durations have been found

for explicitly detected incongruities, but only for the region after the incongruity (Bohan & Sanford, 2008). This further suggests that antonyms are well-suited for use in language studies, and that speakers are intuitively aware of antonymy, something that has been suggested in previous research (Bianchi et al., 2011; Kretschmar et al., 2009; Murphy & Jones, 2008; Paradis et al., 2009; Tribushinina, 2013; van de Weijer et al., 2014). Note that regression path durations also increased, but that the increase was not significant in either experiment. This is likely caused by the variability of the sentences. The stimuli sentences in this study are of varying length, with only the incongruities being strictly controlled for length and frequency. This introduces high variability between trials when analyzing the entire sentences. Regression path durations are affected by the entire sentence, since they include the time during which previous parts of the sentence were read before reading continued to the right of the critical word after the first visit, while selective regression path durations only include the actual fixations on the critical word during this time. Longer sentences have more words to read, thus increasing the regression path durations.

Source bias had no effect on the number of explicitly detected incongruities. This suggests that the preference of the output similarity evaluator to use the output from the heuristics analysis, and consequently to perform expectation-based output selection, is unaffected by the bias. It is possible that a stronger bias is needed in order to produce a higher number of explicitly detected incongruities. Source bias did not significantly affect how long participants read the sentences, so the participants did not spend more time inspecting the less-respected sentences for errors. The total reading times per character for the filler sentences were actually slightly shorter for the less-respected sentences, compared to well-respected ones.

All analyses were performed twice, with forgotten sentences either removed or classified as implicit detection, but only minor differences were found in significance in Experiment 1, with total reading times for implicit detection changing from significant to marginally significant only (models shown in the general discussion classify forgotten sentences as implicit detection for brevity).

Table 8.7 and Table 8.8 show a summary of the behavior and measured effects, and how they correlate with the hypotheses and hypothesized effects presented in Section 8.3.2.

Table 8.7. Behavior and measured effects on the experimental measures on the critical word compared to the control condition, Experiment 1 and Experiment 2

Behavior and measured effects on the experimental measures on the critical word compared to the control condition, Experiment 1 and Experiment 2. Hypotheses are given both for how the proposed model reacts and for the changes in the participant's reading behavior. Hypotheses presented in Section 8.3.2 supported by the data in **bold**. All behavior and effects in the table are supported by the data. Differences are statistically significant; non-significant results are described as *no significant difference*.

Detection type	Behavior compared to the control condition	Measured effects compared to the control condition
Implicit detection	Model: The output similarity evaluator judges the two outputs from the two analyses to be sufficiently similar, and uses expectation-based output selection to insert the output from the heuristics analysis into the meaning representation.	• Longer total reading times, but not to the extent that is found for explicit detection. • No significant difference in first-pass reading times. • No significant difference in regression path durations. • No significant difference in selective regression path durations. • No significant difference in second-pass reading times.
	Behavior: The participants continue their reading without consciously reacting to the incongruity. The output with the incongruent word, however, results in more rereading of the incongruent word. Effects were observable in late measures only.	
Explicit detection	Model: The output similarity evaluator judges the two outputs from the two analyses to be different enough to require reanalysis. Expectation-based output selection is not used, and heuristics usage is halted in favor of conscious rereading of the sentence. The interruption is not immediate, but should occur during the first visit to the word.	• Longer total reading times. • No significant difference in first-pass reading times, since the interruption was only visible after processing is completed. • No significant difference in regression path durations (actual mean increases in regression path durations were 150 ms in Experiment 1 and 252 ms in Experiment 2). • Longer selective regression path durations. • No significant difference in second-pass reading times.
	Behavior: The participants halt their reading in favor of rereading the incongruent word. Effects were observable in early stages of reading, but not during the second return to the word.	

Table 8.8. Behavior and measured effects of source bias on explicit detection, Experiment 2

Differences are statistically significant; non-significant results are described as *no significant difference*.

	Behavior	Measured effects
Well-respected or less-respected news source	When hearing that sentences originate from less-respected news sources, people might implicitly assume that there are more errors in them, but this had no effect on the discovery of irregularities and errors in the sentences.	• No significant difference in explicitly detected incongruities in sentences presented as originating from less-respected news sources, compared to the sentences from well-respected news sources.

8.7 Conclusions

Semantically incongruent sentences resulted in clear disruptions to the reading process when using online measures in this chapter, in the form of longer total reading times on the incongruent word compared to the correct, control condition. These disruptions occurred even when the participants were not aware of the incongruity. These results were also found in Chapter 6, in which paragraphs of text were used as stimuli. More thorough conclusions about the cause of these disruptions

can be found in Section 6.7, but are also summarized here. Two parallel processes operate during reading, creating two separate outputs that need to be integrated into one meaning representation. One process uses heuristics that are based on expectations, previous knowledge, and prediction, while the other uses some form of word-meaning retrieval to analyze the actual word on the page. I use the term output similarity evaluator for the process that compares these two outputs. Normally, the output similarity evaluator selects the output from the heuristics analysis. I use the term expectation-based output selection for the process that integrates the output from the heuristics analysis into the meaning representation. Since the heuristics analysis is assumed to be slightly faster, the word-meaning retrieval is left only partially completed.

When an incongruity is encountered during reading, the two parallel reading processes deliver different outputs. The output similarity evaluator notes the contrasting outputs, but normally assumes that conscious involvement from the reader is not needed, as is the case with grammatical errors, spelling errors, and other mistakes. Expectation-based output selection delivers an incorrect representation of the text, and the word-meaning retrieval is disregarded. The disruptions in the reading process when an incongruity was not consciously reacted to are hypothesized to stem from this disregarded output. These disruptions can be said to be effects of implicit detection. Total reading times on the incongruent words were longer compared to the congruent versions, showing that participants spent more time on an incongruent word, even though they stated that there was nothing wrong with the sentence. Previous research has shown similar disruptions to the reading process stemming from an incorrect output that is subsequently corrected, and previous research has also shown that we are not completely aware of our own eye movements.

When the output similarity evaluator notes the contrasting outputs, this is because it is dealing with something more significant than a spelling error or a word with a slightly different meaning. Expectation-based output selection is therefore *not* used and the heuristics analysis is disregarded. Instead, the word-meaning retrieval containing the incongruity is added into the meaning representation. This makes the reader consciously aware of the incongruity, temporary halting the heuristics analysis in favor of rereading and reanalyzing the sentence. After explicit detection, total reading times on the incongruent words were longer compared to the congruent versions, indicating that rereading took place. Early effects of explicit detection on the reading process were also observed, with participants taking longer to continue past the incongruent word when explicitly detecting it due to rereading of the previous material. This suggests that detected antonymic incongruities are processed more quickly than what is assumed to be the case for semantic incongruities, but only for explicitly detected incongruities.

In Chapter 9, additional analyses are performed based on the conclusions from all chapters. One of these analyses investigates whether sentence judgments are affected by classifying explicit and implicit detection.

Chapter 9

Further analyses

Having completed all the studies in this book, it became evident that more questions could be answered by combining them and thinking of ways in which one study could inform another. Therefore, additional analyses were performed on data from these studies, based on what was learned in other chapters.

9.1 Your comprehension increases based on the time spent on the words, but the type of reading is more important

9.1.1 Introduction

In Chapter 3 of this book, eye movement characteristics for different types of reading were investigated by giving different instructions to the participants immediately prior to reading a text. There were three instructions for “normal” reading: to read regularly, to read the text thoroughly, and to skim the text. A fourth instruction asked the participants to spell check the text, which can be considered a different task than just “normal” reading. After reading each text, the participants were tested on their comprehension of the material. Total reading times were used extensively, but no specific investigation was made regarding the correlation between total reading times and comprehension. In a previous study, my colleagues and I investigated the disfluency hypothesis, which posits that a slight perceptual manipulation of a text will cause people to engage with it more text, resulting in better learning outcomes (Strukelj, Scheiter, et al., 2016). A slight blur was applied to the texts as the disfluency manipulation. When compared with non-blurred text, the text was read less thoroughly until half of the text had been read, after which the text was read more thoroughly. The task was to read the text thoroughly in order to learn it, and the participants were tested on the content 30 minutes later. No improvements to learning outcomes were found based on disfluency manipulation, and it seems likely that the blurring interferes with reading. To investigate the connection between reading time and comprehension, the data from Chapter 3 were reanalyzed to show how comprehension is affected by looking longer at the words of a text (regardless of

what task is performed by the participant), without any changes being made to the text itself.

9.1.2 Method

The data from the study in Chapter 3 were reanalyzed, using total reading time as a predictor for comprehension scores. In Chapter 3, after reading each text, five comprehension questions were given to the participants, for a maximum of five points per text. The data processing, figures, and statistical analyses were made using R (3.2.2) (R Development Core Team, 2015) and the lme4 package (Bates et al., 2015) in R Studio (0.99.473). The recordings from the right eye were used, since the right eye is dominant for the majority of people. Skipped words were not removed from the analysis, since skipped words are important to consider for skimming. The lmerTest package (2.0-29) was used to estimate p -values with Satterthwaite approximation. The models used random intercepts and slopes of condition within group for participant and item, with intercept-slope correlation factors for participants and items.

9.1.3 Results

Figure 9.1 shows mean comprehension scores as well as mean total reading times/character for regular reading, thorough reading, skimming, and spell checking. Figure 9.2 shows total reading time/character (ms) for instructions of regular reading, thorough reading, skimming, and spell checking, depending on comprehension scores. When comparing with regular reading, a significant increase was found for the comprehension scores in the condition with thorough reading (Estimate=0.349, $SE=0.135$, $t=2.59$, $p<.02$), and a significant decrease was found for skimming (Estimate=-0.687, $SE=0.165$, $t=-4.16$, $p<.001$) and spell checking (Estimate=-0.719, $SE=0.152$, $t=-4.72$, $p<.001$). Higher total reading times signified higher comprehension scores (Estimate=2.28, $SE=0.674$, $t=3.38$, $p<.001$). Overall, total reading times were significantly shorter for skimming (Estimate=-12.2, $SE=4.68$, $t=-2.60$, $p<.02$), and significantly longer for thorough reading (Estimate=15.8, $SE=4.88$, $t=3.24$, $p<.005$) and spell checking (Estimate=18.6, $SE=4.85$, $t=3.85$, $p<.005$), compared with regular reading. A significant interaction was also found between comprehension scores and skimming, showing that total reading times did not improve comprehension scores as much for skimming (Estimate=-3.31, $SE=0.884$, $t=-3.75$, $p<.0005$), compared with regular reading. The other two interactions were not significant (thorough reading: Estimate=-0.769, $SE=1.01$, $t=-0.77$, $p=.44$; spell checking: Estimate=0.450, $SE=0.846$, $t=0.53$, $p=.59$). Full statistical model output can be found in Appendix 9.A.

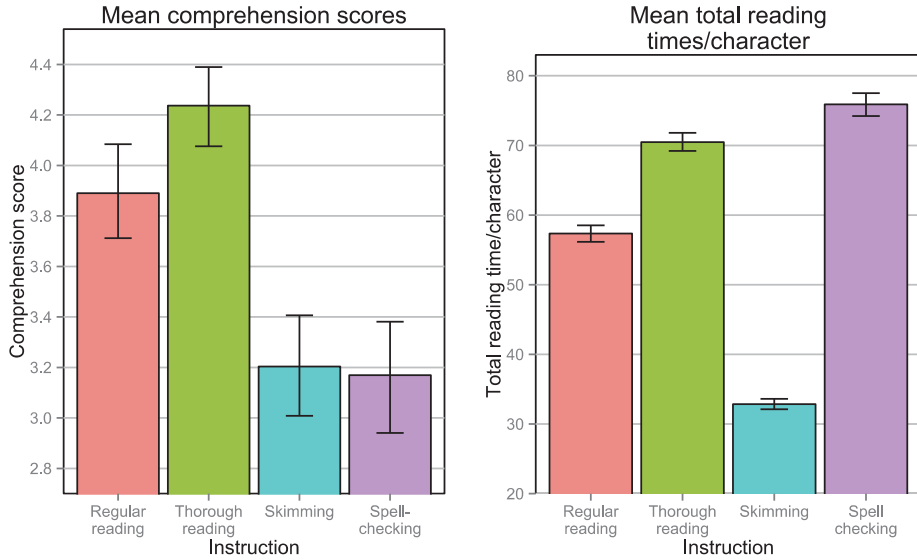


Figure 9.1. Mean comprehension scores and mean total reading times/character for regular reading, thorough reading, skimming, and spell checking, Chapter 3
Error bars denote standard errors of the mean.

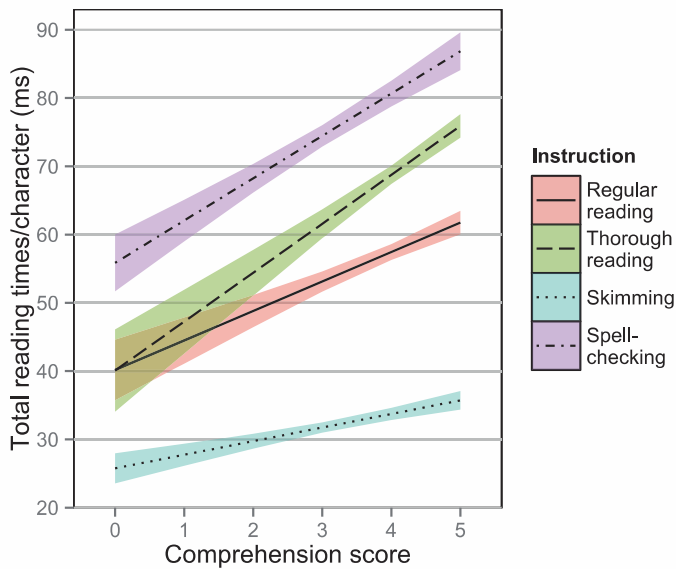


Figure 9.2. Total reading time/character (ms) for regular reading, thorough reading, skimming, and spell checking, depending on comprehension scores, Chapter 3
Regression lines include 95% confidence intervals.

9.1.4 Discussion

In Chapter 3, participants read short texts in four different ways, and were given five multiple-choice comprehension questions immediately after each text. Better or worse comprehension was demonstrated by significantly lower or higher comprehension scores compared to those for regular reading. Comprehension scores and total reading times/character increased significantly for thorough reading, and decreased significantly for skimming. In a previous study, my colleagues and I investigated the disfluency hypothesis (Strukelj, Scheiter, et al., 2016). This hypothesis posits that people engage more and learn better from text that has been made slightly more difficult to read by applying a “disfluency manipulation.” In our case, we applied a slight blurring to the text in the disfluent condition. Arguably, higher engagement would entail more thorough reading, as evidenced in longer total reading times. We found that the time spent on the text indeed changed, but that this depended on the duration of reading of the disfluent text, and that learning did not improve (Strukelj, Scheiter, et al., 2016).

With the relation between comprehension and reading in mind, the data in Chapter 3 were reanalyzed. Since spell checking was found to increase total reading times/character, but decrease comprehension, specific attention was paid to these results. It was found that total reading times/character were correlated with higher comprehension scores; in other words, spending more time reading the text resulted in higher comprehension scores. An interaction between increasing comprehension scores and longer total reading times/character showed that this increase was smaller for skimming compared to regular reading. In other words, when skimming a text, although spending more time on the text was correlated with a smaller benefit in comprehension compared to regular reading, a benefit still existed, as seen in Figure 9.2. It was expected that a significantly smaller benefit would be found for spell checking too, or that there would be entirely no benefit to spending more time on the text. However, this was not the case. No significant interaction between comprehension scores and total reading times was found for spell checking compared to regular reading. In other words, compared to how longer total reading times/character led to higher comprehension scores for regular reading, no significant deviation in this pattern was found for spell checking. Figure 9.2 indeed shows highly similar relations between comprehension scores and total reading times/character for spell checking, regular reading, thorough reading, and to a smaller extent (which was indeed *significantly* smaller), skimming. This suggests a relationship between comprehension and looking longer at the text, even when the text is being checked for spelling errors. Importantly, searching for spelling errors made comprehension scores significantly worse overall, showing a strong task effect. If people spent more time on the text when performing this task, however, their comprehension did actually improve (albeit from a severely decreased point compared to regular reading).

9.2 Your openness does not affect your opinions when expecting biased texts but actually reading neutral ones

9.2.1 Introduction

In Chapter 4, I investigated how expectations of biased texts affected opinions, and in Chapter 5, I investigated how these expectations affected eye movements during reading. I based this on common and well-researched cognitive biases: people generally gravitate toward information that confirms their opinions and away from information that disagrees with their opinions (this is known as confirmation bias, myside bias, and cognitive dissonance). In both chapters, only the headlines were changed to either agree or disagree with the participants' opinions, with no changes made to the texts themselves. In Chapter 4, in participants who said that their opinions on the topic were weak, opinions did not shift in line with cognitive biases, but rather in line with the bias of the headline. In Chapter 5, myside bias did not operate as strongly on eye movements as hypothesized. Instead, the openness of the participant combined with having a weak or strong opinion on the topic was found to strongly affect how the end-sections of the texts were read. End-sections of texts commonly contain conclusions and summaries, and are often very important to the argumentation in an essay or opinion piece in a newspaper. For the most part, agreement or disagreement did not affect reading. When we read a text we agree with, however, our current opinion can be amended by new arguments and new conclusions. Therefore, openness will arguably operate both when we agree and when we disagree with a certain opinion. In most cases in the second experiment of Chapter 5, end-sections were read less thoroughly when openness was low, and more thoroughly when openness was high. For a thorough discussion, see Section 5.7.2. This suggests that participants did not care about the conclusions when they were not open to other opinions, since they were satisfied with their own opinions and knowledge of the topic, and were not interested in hearing other opinions. Conversely, participants cared about the conclusions when they were open to other opinions, since they were open to amending, changing, and extending their current knowledge on the topic. When participants had weak opinions on the topic and were expecting a text that agreed with their opinions, the reverse was found. End-sections were read more thoroughly when openness was low, and less thoroughly when openness was high. This was interpreted as the participants wanting only to confirm their opinions, and not spending unnecessary effort on the topic, respectively. In other words, myside bias in combination with openness affected reading when opinions were weak and the text was expected to agree with currently-held opinions. It is thus plausible that higher degrees of openness would result in larger opinion shifts. This section therefore analyzes how opinions are affected by the openness of the participant.

9.2.2 Method

The opinion data from the study in Chapter 4 were reanalyzed, using the results for openness from the study in Chapter 5 as a predictor. The agreement score for each topic before reading were subtracted from agreement scores for each topic after reading. The absolute value was used, meaning that all negative scores were turned into positive scores. This resulted in the measure *absolute difference in agreement scores before and after reading*. This measure showed opinion shifts when disregarding whether the opinions became more positive or more negative, which was the goal. In other words, the measure shows only the amount of opinion shift, and not the type.

The data processing, figures, and statistical analyses were made using R (3.2.2) (R Development Core Team, 2015) and the lme4 package (Bates et al., 2015) in R Studio (0.99.473). The lmerTest package (2.0-29) was used to estimate p -values with Satterthwaite approximation. The model used random intercepts and slopes of condition within group for participants.

9.2.3 Results

No significant differences in agreement scores before and after reading were depending on openness were found (main effects: all $ps > .46$; interactions: all $ps > .09$). Figure of agreement scores after reading the texts depending on the openness, and full statistical model output, can be found in Appendix 9.A.

9.2.4 Discussion

The results were surprising, since higher openness did not result in bigger opinion shifts, represented by the absolute difference in agreement scores before and after reading. A likely explanation is that openness does not operate on the type of opinion shift operationalized in Chapter 4. Participants self-reported their opinions on the topics before reading, read neutral texts with positively or negatively biased headlines, had a 20-minute filler task, and then self-reported their opinions again. In Chapter 4, it was found that the expectations induced by the evaluative bias of the headlines resulted in opinion shifts for people with weak opinions, but not for people with strong opinions. Expectations alone are thus sufficient to cause opinion shifts in some cases. It seems, however, that expectations are not sufficient to cause opinion shifts with regard to the openness of the participants. As discussed in Section 5.2.3, the openness personality trait has been found to positively correlate with self-direction (autonomy of thought, openness to new ideas), universalism (understanding and tolerance for all people and ideas) (Roccas et al., 2002), and being broadminded, and being imaginative (Dollinger et al., 1996). It is possible that the effects of openness are only evident after texts with actual bias are read, after longer texts are read, after a

longer time has passed, or all of the above. Further research is needed to answer these questions.

9.3 You cannot consciously identify an incongruent sentence quickly, but your reading is instantly disrupted nevertheless

9.3.1 Introduction

In Chapter 8, I expected that participants would explicitly detect more incongruities when expecting sentences with poorer grammar or poorer language usage. My hypothesis was that people would expect poorer language usage in sentences they believed to originate from less-respected sources, and better language usage in sentences they believed to originate from well-respected sources. In Chapter 7, I indeed found this when analyzing sentence judgment scores relating to grammar and language usage, which were lower for sentences said to be from less-respected news sources compared to those said to be from well-respected sources. The number of explicitly detected incongruities in Chapter 8 was not affected by this manipulation, however. An alternative explanation for the results would be that people read sentences from less-respected news sources more critically, and therefore gave them lower sentence judgment scores. In Chapter 8, however, analyses of the filler sentences showed no significant difference in how thoroughly they were read, neither for total reading times per character on all words in the sentences, or on the adjective where the manipulation was located in the critical sentences. This explanation is thus unlikely, since effects should be evident in total reading times. The figure showing the total reading times per character for the less-respected and the well-respected sentences can be found in Appendix 8.B.

In Chapter 7, the semantically incongruent sentences did not affect the sentence judgment scores, when comparing the scores for congruent and incongruent sentences. I did not, however, use the classifications of implicit and explicit detection for the sentence judgments. It is possible that explicitly detected incongruities would produce significantly lower sentence judgment scores than those that were implicitly detected or those for correct sentences. This section therefore analyzes how sentence judgments are affected by successful and failed detection of semantic incongruities in text.

9.3.2 Method

The sentence judgment data from the study in Chapter 7 were reanalyzed, using the classifications of implicit or explicit detection in Chapter 8 as predictors. The data processing, figures, and statistical analyses were made using R (3.2.2) (R Development Core Team, 2015) and the lme4 package (Bates et al., 2015) in R Studio (0.99.473). All statistical models used log-transformed data to better fit a Gaussian distribution, but in order to improve readability, no log-transformation was performed on the data in the figures. The lmerTest package (2.0-29) was used to estimate p -values with Satterthwaite approximation. Models used random intercepts and slopes for participants and items.

9.3.3 Results

Models used random intercepts and slopes of condition within group for participants and items. Figure 9.3 shows grammatical and semantic acceptability ratings for the sentences in Experiment 1 from Chapter 7, and Figure 9.4 shows grammatical and semantic acceptability ratings for the sentences in Experiment 2 from Chapter 7. Four levels were used, namely congruent critical, implicitly detected, explicitly detected, and congruent filler sentences.

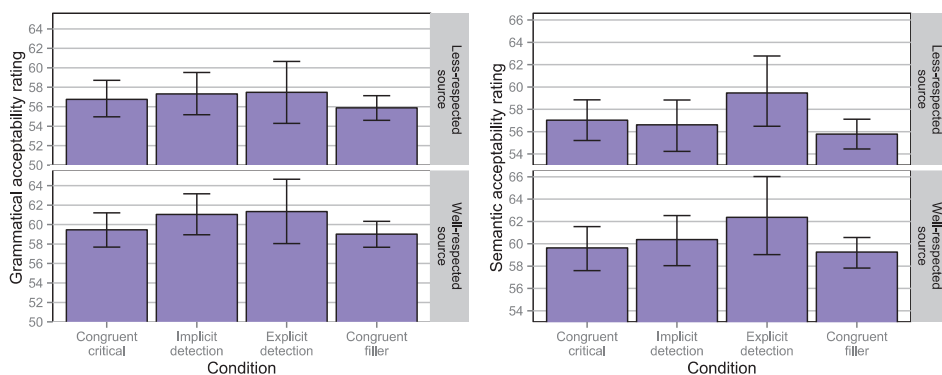


Figure 9.3. Grammatical and semantic acceptability ratings for congruent sentences, implicitly detected incongruent sentences, explicitly detected incongruent sentences, and filler sentences, Chapter 7, Experiment 1

Error bars denote standard errors of the mean.

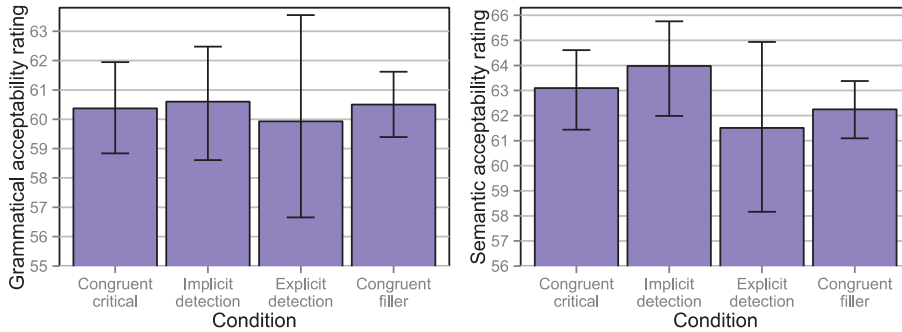


Figure 9.4. Grammatical and semantic acceptability ratings for congruent sentences, implicitly detected incongruent sentences, explicitly detected incongruent sentences, and filler sentences, divided into source bias condition, Chapter 7, Experiment 2
Error bars denote standard errors of the mean.

In Experiment 1, no differences were found between the congruent critical sentences and the other sentences (grammatical acceptability: all $ps > .96$; semantic acceptability: all $ps > .39$). In Experiment 2, all sentences said to be from well-respected sources had significantly higher grammatical and semantic acceptability ratings compared to sentences said to be from less-respected sources (grammatical acceptability: Estimate=2.70, $SE=1.19$, $t=2.28$, $p < .03$; semantic acceptability: Estimate=2.67, $SE=1.27$, $t=2.11$, $p < .05$). When comparing with the congruent critical sentences, no sentence judgments were significantly different (grammatical acceptability: all $ps > .25$; semantic acceptability: all $ps > .09$), with no interaction from source condition (grammatical acceptability: all $ps > .48$; semantic acceptability: all $ps > .55$). Full statistical model outputs can be found in Appendix 9.A.

9.3.4 Discussion

In Chapter 7, people did not give significantly lower judgment scores to sentences with semantic incongruities in either experiment. In Chapter 8, eye movement measures were disrupted by the semantic incongruities, regardless of whether the incongruities were explicitly or implicitly detected. When using the classification in Chapter 8 of explicit and implicit detection of incongruities on the judgment scores in Chapter 7, no significant effects were found. The hypothesized result was that explicitly detected incongruities would produce significantly lower sentence judgments than those that were implicitly detected. When re-performing the analysis, however, this was shown to not be the case. The sentence judgment scores given immediately after reading a sentence were not significantly affected by the semantic incongruities that were explicitly detected by participants in the follow-up interview.

In Chapter 8, where the reading of incongruities in sentences was investigated, total reading times on the incongruent words were longer both when participants explicitly detected the incongruity and when participants said there was nothing

wrong with the text or sentence, thus indicating the existence of implicit detection. This was also the case in Chapter 6, where the reading of incongruities in texts was investigated, further indicating the existence of both explicit and implicit detection. In addition, selective regression path durations in Chapter 8 were longer only when during explicit detection; that is, when participants stated that they noticed something was wrong with the sentence. This shows that the incongruent words were strong enough to give rise to experimental effects, and that the classification was reliable. The classification of explicit or implicit detection in Chapter 8 (and, indeed, Chapter 6) was performed during a post-experiment interview. The beginning of each incongruent sentence was shown to the participant (and a few congruent fillers in order to minimize the risk of the participant guessing the research interest), and they were asked to recall the sentence and state if there was anything wrong with it. This thus involved a fairly thorough evaluation of each sentence in question. In Chapter 7, explicit detection did not affect the sentence judgment results, but in Chapter 8, explicit detection was highly visible in eye movement results. This suggests that people need to extensively analyze even an obvious error to become sufficiently aware of it, in order to affect their conscious responses during sentence judgment tasks. It demonstrates that people are bad at making correct spontaneous evaluations. It also highlights the difference between online methods such as eye tracking, and offline methods such as sentence judgment tasks, showing that a lot is going on outside of our awareness during reading.

Chapter 10

Summary and final conclusions

This book is situated in the relatively new field of experimental pragmatics, and the aim was to investigate how reader expectations about texts and sentences affect eye movements during reading, and opinions and sentence judgments after reading. Since predictive processing is essential during reading, changes to the expectations of what will be read should change how the predictions operate during reading. The expectation of a context activates the predictions related to this context, and the behavior can then be investigated. Since prediction is a crucial characteristic of our minds and our cognitive processes, it is essential that experimental pragmatics takes this characteristic into account (Van Berkum, 2010). The aim of this book was to attempt to change people's interpretations during reading—people's predictive processing—by changing their expectations of what they would read.

An initial eye-tracking study investigated the eye movement characteristics of four different reading types during paragraph reading: “regular” reading, thorough reading, skimming, and spell checking. Five studies then investigated the effects of different kinds of expectations on reading, opinions, and sentence judgments. The expectations of controversial topics or less-respected news sources, compared to the expectations of neutral topics and well-respected news sources, revealed people's perception of these concepts. The expectations of a text with a bias that agrees with the reader, compared to one that disagrees with the reader, revealed how people change their reading and opinions as a result of agreeing or disagreeing with a text. The expectations of texts and sentences without errors showed that when the reader misses egregious errors, the reading process reacts to the error without the reader being consciously aware of it, revealing experimental evidence for parallel reading processes. Finally, the expectations of texts or sentences without errors, which should be considered the default setting when reading, revealed how people's predictive processing operates during normal operation.

To evaluate how expectations affect reading, the eye movement characteristics during thorough reading and skimming of paragraphs of text first needed to be established. This led to the first question of this book:

1. What are the eye movement characteristics during regular reading, thorough reading, skimming, and spell checking, and how do they differ?

In an eye-tracking experiment, participants were instructed to read a text in specific ways; to read it as they normally would; to read it thoroughly; to only skim, and to search for spelling errors. All texts were balanced in terms of length (16 lines of text, 684 characters in 146 words), with identical readability indexes (LIX 35). Among the findings, total reading times increased significantly for texts that were read thoroughly, and decreased significantly for texts that were skimmed, when comparing with texts that were read regularly (that is, the way the participant would normally read an informative text for comprehension). This showed that total reading time was a reliable measure for how thoroughly a text was read, and thus an indicator for overall processing.

The next two chapters were based on research on cognitive biases and research on media psychology. Confirmation bias and the more recent theory of myside bias (which is favored in this book), show that we are biased toward information that we agree with and against information we disagree with. Cognitive dissonance shows that when we are faced with something that contradicts our opinions, we need to consolidate these two contrasting pieces of information, a task that is fairly difficult. With regard to how opinions are affected in light of whether a text is presented as being either positive or negative, affective priming shows that we are indeed positively influenced by a positive text, and negatively influenced by a negative text. In the first experiment, participants read neutral texts with neutral headlines and topics described as being either “highly controversial” topics that people “feel very strongly about,” or merely “various topics.” In the second experiment, participants read neutral texts with positively and negatively biased headlines, such as “The good effects of divorce” or “The bad effects of divorce.” This allowed me to investigate how identical texts were read, and draw conclusions about the experimental results. Importantly, it also investigated if these cognitive biases could be introduced by expectations alone, and if so, to what extent. Since these cognitive biases are mainly concerned with how opinions change, the second question was:

2. How are opinions affected by reading a neutral text expected to be about a topic other people feel strongly about, or expected to have positive or negative evaluative bias?

Data on the opinions of the participants both before and after reading the texts were gathered in self-report tasks. It was hypothesized that merely saying that texts would be on highly controversial topics that others felt strongly about would make people feel more strongly about their own opinions. This was indeed the case. Participants were divided into having either positive or negative prior opinions on each topic and the results were transformed to remove regressions to the mean. This showed that people’s opinions resisted opinion shifts significantly for controversial topics compared to the neutral topics. Previous research suggests that controversial issues divide society, with groups having conflicting opinions based on their alternative

values (Oulton et al., 2004, p. 2; Stradling, 1985, p. 9). The findings in this book show that the mere expectation of a “controversial” topic makes people more guarded and hesitant to change their opinions. This adds to our understanding of how the opinions regarding controversial issues operate, and why controversial issues are described the way they are, by helping to explain the irrational stubbornness that can be found when controversial topics are discussed.

Previous research has suggested that people are affected by the evaluative bias of news presentation, called affective bias. Accordingly, opinions became more positive after reading texts with positively biased headlines and more negative after reading texts with negatively biased headlines, but only if the participants did not have a strong opinion to begin with on the topics. This suggests that affective priming can be induced by expectations alone. Previous research has also shown that people are biased toward texts agreeing with currently-held opinions and against texts disagreeing with currently-held opinions. I did not find that opinions held firm, however, when dividing people according to having either positive or negative opinions regarding a topic prior to reading. This suggests that expectations of biased texts are not sufficient to produce these bias effects. It is important to understand how people’s opinions change as well as stay firm in the face of contradicting evidence, and a great deal of research has been done in this field. We know that agreeing and disagreeing with the topic indeed affects how people read a text, so it is plausible to hypothesize that the effect was too weak to show up in the measures of opinion shift. It therefore became even more important to investigate how the participants were reading the texts. The third question was therefore:

3. How are eye movements during reading affected by reading a neutral text expected to be about a topic other people feel strongly about, or expected to have a positive or negative evaluative bias?

Eye-tracking data during reading were gathered during both of the experiments. When participants read texts expected to be on controversial topics that others had strong opinions about, the participants with weak opinions on the topic read the text more thoroughly compared to when the texts were introduced as being on various topics. When participants read texts with biased headlines, the participants with high openness read the ends of texts more thoroughly, and participants with low openness read the ends of texts less thoroughly. The opposite trend was found for participants with weak opinions on the topic, however, who were expecting to read a text that they would agree with. The results suggest that knowledge of other people’s opinions affects reading. With regard to openness, the results show that interest in the concluding lines of a text generally increases with a higher degree of openness, both when the participant agrees and disagrees with the bias. This contradicts the claims in myside bias, namely that people are biased toward information that agrees with their opinions and tend to avoid information that disagrees with their opinions (Mercier,

2017). Instead, it seems that openness overrides myside bias when only expectations are used as manipulations, except for the case in which the opposite trend was found. People with strong opinions on a topic should be more likely to stick to their opinions, and this is also what was found in Chapter 4. While this is seen in opinion shifts, it is shown that openness affects reading to a larger extent. In the case in which the opposite was found, the participants expected to agree with a text about a topic they did not feel strongly about. In Chapter 4, opinion shifts in line with the expected bias were found for the participants with weak opinions, suggesting a disinterest in maintaining their opinions. A possible explanation for the deviating results is an interaction between myside bias and openness in this case, namely a higher interest in the information *only* when the risk of changing opinions is low. The combination of low openness—a low chance of changing one’s opinions—and weak opinions on a topic—a disinterest in maintaining one’s opinions—results in a low likelihood of changing one’s opinions. This made it “safe” to read information more thoroughly. Because the texts were designed to be neutral, no biased information had been read when reaching the end of the text, which resulted in a higher interest in the conclusion. Conversely, the combination of high openness—a high chance of changing one’s opinions—and weak opinions on a topic—a disinterest in maintaining one’s opinions—results in a high likelihood of changing one’s opinions. This made it “unsafe” to read information more thoroughly. Because the texts were designed to be neutral, no biased information had been read when reaching the end of the text, which resulted in a lower interest in, or possibly even an avoidance of, the conclusion.

Delving deeper into the reading process, the next two chapters investigated in detail what happens when an error or incongruity in a text or sentence elicits no reaction from a reader. Paragraphs were investigated in Chapter 6, and sentences were investigated in Chapter 7. It is reasonable to assume that people’s previous experience of text is that it is congruent and without errors. When a reader encounters an incongruity without being aware of it, the expected word is used in its place and reading continues unhindered. Previous research has shown that this indeed frequently happens, but how much of the incongruent word is processed? The studies that have investigated failed detection during reading have generally found no experimental evidence of differences between failed detection and congruent text. I believed that this was because the stimuli were not sensitive enough to catch these very small experimental effects, with an over-reliance on a very small number of highly controlled stimuli rather than using a larger number of stimuli without focusing too much on controlling the texts or sentences for every possible factor. After all, the effect, however small, should behave very similarly across all stimuli. The fourth question was:

4. How are eye movements during reading affected by failed and successful detection of semantic incongruities in *texts* that are expected to contain no errors?

In Chapter 6, two eye-tracking experiments investigated processing during both failed and successful detection of incongruities during paragraph reading. Frequent antonym pairs designed to be unambiguously erroneous in their contexts, such as *high* and *low*, were used to create semantically congruent and incongruent sentences, with detection evaluated during a post-experiment interview. Reading behavior changed during both successful and failed detection, as indicated by factors including a higher numbers of revisits (significant in both experiments) and longer total reading times (not significant in the first experiment, but significant in the second experiment) for the critical word in the erroneous conditions compared to the correct condition.

Previous research suggests that parallel processes operate during reading: one uses a syntactic analysis, and the other uses heuristics based on expectations, context, and previous experiences. The syntactic analysis is called word-meaning retrieval in this book. It is argued that the heuristics analysis is normally favored in order to minimize effort and speed up processing. During failed detection, the heuristics analysis makes an incorrect (but expected) interpretation, resulting in a meaning representation with the expected word rather than the incongruent word. The reader is not aware of the incongruity, but the results suggest that the disregarded word-meaning retrieval produces lingering effects during subsequent reading. During successful detection, the word-meaning retrieval takes precedence and the incongruity is correctly identified, resulting in a meaning representation with the incongruent word. The reader becomes aware of the incongruity and rereads the sentence. The term implicit detection was introduced for cases of failed detection, since the eye movement measures clearly showed that the incongruity was at least partly processed, and the term explicit detection was introduced for successful detection. A model for the process was proposed as follows. Initially, what is termed the output similarity evaluator compares the two different outputs. If the two outputs are judged to be sufficiently similar, expectation-based output selection inserts the output from the heuristics analysis into the meaning representation. Reading can continue without a conscious reaction to the incongruity. The incongruent output is not completely erased, however, and results in more rereading of the incongruent word, which is seen in eye movement measures. If the two outputs are judged as being different enough to require reanalysis, expectation-based output selection is not used, and the incongruent word is used in the meaning representation instead. A conscious reaction to the incongruity can therefore be elicited, making the reader suspend the reliance on the heuristics analysis during subsequent reading.

So, effects of both explicit and implicit detection were found for paragraphs of text, but would these effects be found for single sentences? With single sentences, the output similarity evaluator would need to operate with much less context. A larger number of trials and a higher level of experimental control would also give better eye movement data, investigating if differences in earlier eye-movement measures could be found between explicit and implicit detection. In addition, the strength of the proposed expectation-based output selection would be examined, and the results

would be replicated in a much stricter setting. The output similarity evaluator was also examined with regard to the preference of using the output of the heuristics analysis during reading. A manipulation was created that was supposed to lower expectation-based output selection, and thus raise the number of explicitly detected incongruities. Telling the participants to look for incongruities would likely raise the number, but this would be a very undesirable way of achieving this. The presence of poorer or better language usage needed to be suggested without explicitly stating that language usage was poorer or better.

Self-rated judgment tests are often used to understand how people perceive sentences as more or less acceptable or grammatically correct. It is difficult to correctly analyze a sentence as more or less semantically acceptable intuitively, and research also suggests that differentiating between grammaticality (grammatical acceptability) and semantic correctness (semantic acceptability) is not possible during spontaneous tasks such as performing sentence judgments. I therefore used both grammatical and semantic acceptability to investigate this claim. Furthermore, it has been suggested that sentence judgment tasks are subjective, but very little research has investigated how people's expectations regarding the upcoming sentences can affect their responses in acceptability judgment tasks. I therefore introduced an expectation that the source of the sentences would either be well-respected or less-respected news sources, in order to investigate how people regarded different news sources. In cases in which less-respected news sources were indicated, this was hypothesized to implicitly lower people's expectation of semantic (and, for that matter, grammatical) acceptability, so that sentences would be spontaneously judged as having poorer grammar because of people's preconceived notions associated with them. I also wanted to see if participants were able to spontaneously judge semantically incongruent sentences as less acceptable than the corresponding congruent sentences in sentence judgment tasks. The fifth question was:

5. How are sentence judgments affected by semantic incongruities and by expectations of well-respected or less-respected news sources?

Participants read semantically congruent and incongruent sentences said to originate from various news sources in the first experiment, and said to originate from either well-respected or less-respected news sources in the second experiment. Semantic incongruities were created by replacing a word with its antonymic partner, such as *soft* instead of *hard*. Separate grammatical and semantic acceptability judgments were used, and two additional judgments investigated the participants' interest in reading the hypothetical remainder of the text, resulting in form-based and topic-based interest ratings. Grammatical acceptability, semantic acceptability, and form-based interest ratings decreased significantly for sentences said to be from less-respected news sources compared to well-respected sources. This shows that people consider the language usage of well-respected news sources to be superior to the less-respected

sources, and that this affected their perceptions of the sentences. Grammatical and semantic acceptability ratings, however, were not significantly lower for the semantic incongruities in either experiment, as had been hypothesized. This suggests that the incongruent sentences were not sufficiently incongruent to affect the conscious, spontaneous judgments of how well-formed the sentences were. This shows the difficulty of providing accurate spontaneous judgments of sentences. Sentence judgment tasks can be a good experimental method, but the results show that care needs to be taken when using them in conjunction with incongruent sentences.

Importantly, the manipulation was shown to affect people's overall perception of the sentences: sentences believed to originate from less-respected news sources were perceived as having poorer language usage compared to sentences believed to originate from well-respected sources. The manipulation was thus deemed appropriate to investigate whether poorer language usage affects how the output similarity evaluator uses the heuristics output during reading. This investigation was only a small part of the chapter, however. As mentioned above, the main focus of the chapter was to continue the investigation of the usage of expectation-based output selection of the output similarity evaluator, using stimuli that contained much less context and a larger number of experimental trials. The sixth and final question was:

6. How are eye movements during reading affected by failed and successful detection of semantic incongruities in *sentences* that are expected to contain no errors?

In Chapter 8, two eye-tracking experiments investigated processing during both failed and successful detection of incongruities during sentence reading. Again, common antonym pairs designed to be unambiguously erroneous in their contexts, such as *bad* and *good*, were used to create semantically congruent and incongruent sentences, with detection evaluated during a post-experiment interview. Reading behavior again changed during both successful and failed detection, indicated by longer total reading times (significant in both experiments) on the critical word in the erroneous conditions compared to the correct condition. In addition, early effects were observed during successful detection only, with longer selective regression path durations (trending toward significance in the first experiment, significant in the second experiment) for the critical word in the erroneous conditions compared to the correct condition.

As in Chapter 6, eye movements showed clear effects for both implicit and explicit detection. This further suggests that the output similarity evaluator compares the outputs from the parallel processes, and either uses expectation-based output selection—which leads only to implicit detection of the incongruity and continues the usage of heuristics during subsequent reading—or avoids using expectation-based output selection, which leads to explicit detection of the incongruity and conscious rereading of the sentence. Additionally, the results suggest that the rereading during explicit detection is initiated when the incongruent word is seen for the first time.

To better understand how the output similarity evaluator overrides the word-meaning retrieval during reading, I investigated whether the expectation of a text from a less-respected source would result in more explicit detections. Even though I saw that sentences said to be from less-respected sources were considered to have poorer grammar than ones said to be from well-respected sources in Chapter 7, this did not increase the number of explicit detections in Chapter 8. For the texts in Chapter 6, I investigated whether a higher working memory capacity would result in more explicit detections, since a higher working memory capacity has been correlated with higher text comprehension and should therefore enable participants to remember more of the text during reading. Even though working memory capacity has been connected to comprehension and learning in previous research (see Daneman & Merikle, 1996, for an overview, but also see Van Dyke et al., 2014), higher working memory capacity did not increase the number of explicit detections in Chapter 6. Both of these results suggest that our reading process is capable of strongly filtering the input text. It appears that the heuristics usage is strong and robust.

The semantic incongruities used in this book elicited strong online effects during reading in Chapter 6 and Chapter 8, but sentence judgments in Chapter 7 did not reflect these effects, showing that a strong manipulation is needed to affect the sentence judgments. This suggests that while people might be bad at performing sentence judgment tasks, and might misread text, our cognitive processes notice much more than we are aware of, and help us much more than we know. Since the source bias manipulation affected people's judgments of sentences, it showed that the mere expectation of a "less than perfect" sentence can lower sentence judgment scores, further highlighting the effect that preconceived notions and opinions can have on self-report judgment tasks. Furthermore, no benefit was gained by separating grammatical and semantic acceptability judgments, since both seemed to behave similarly.

Based on what I learned from these studies, several other questions materialized. Additional analyses were therefore performed.

In Chapter 3, participants read short texts and were given instructions on how to read each one, with five comprehension questions given after each text. The instructions were to perform regular reading, thorough reading, skimming, and spell checking. In a previous study, my colleagues and I investigated a claim that more time spent on text would improve learning outcomes, something that we found to be inaccurate (Strukelj, Scheiter, et al., 2016). The hypothesis in question, the disfluency hypothesis, claims that by making text slightly more difficult to read, people engage more in the text, and therefore learn better. With this in mind, the data in Chapter 3 were reanalyzed, investigating the effect of eye movements on comprehension, with specific interest in the spell-checking task. For spell checking, the comprehension scores were significantly worse, and the total reading times/character were significantly longer. The relation between comprehension scores and total reading times showed, however, that comprehension scores improved when the participants

spent longer on the text. In other words, there seems to be a relationship between looking longer and comprehending more, even with the drastic deterioration in comprehension resulting from searching for spelling errors rather than actually reading the text.

In Chapter 4, opinions were more resistant to opinion shifts for participants with strong opinions on a topic, and participants with weak opinions changed their opinions based solely on a positively or negatively biased headline. When accounting for openness in Chapter 9, increasing levels of openness were seen to result in significantly larger differences in agreement scores before and after reading for the participants with weak opinions only. When a person has a weak opinion on a topic, a higher degree of openness increases opinion shifts. When a person has a strong opinion on a topic, openness is not sufficient to cause opinion shifts (in fact, it correlated with *smaller* opinion shifts). In Chapter 5, the participants with low opinion strengths read the ends of texts differently depending on their openness when the headlines agreed with their opinions. They read more thoroughly when openness was low, and less thoroughly when openness was high. These results are interpreted as the participants being more (or less) likely to look for the information stated in the headline, which the neutral text naturally has not yet provided. With high openness, the participants are more willing to change their opinions, but with low opinion strength, the participants consider their opinions on the topic as being weak. In order to minimize effort, they trust the headline and their own opinion, skip the end of the text, and move on to the next one in order to preserve their currently held opinion. Unbeknownst to the participants, the texts were in fact natural, so this strategy was in vain. In Chapter 9, however, it was shown that openness did not affect opinion shifts as operationalized in this book. It is possible that openness only affects opinions over longer periods of time, or that merely inducing the expectation of a biased text is not sufficient to affect how openness operates on people's opinions.

In Chapter 7, it was found that sentences described as originating from less-respected news sources were given significantly lower sentence judgment scores relating to the form of the sentence, compared to sentences described as originating from well-respected news sources. People seemed to implicitly consider the language usage in these news sources to be poor, which also affected their judgments of the sentences. It was assumed that this would increase explicit detections of semantic incongruities in Chapter 8, but this did not occur. The online measures (eye movements) showed clear effects of both implicit and explicit detection in Chapter 6 and Chapter 8, but no significant effects were found in offline measures (sentence judgments), when using implicit and explicit detection of incongruities as classifications in Chapter 9. This shows that, in addition to being strongly affected by preconceived notions and implicit opinions when performing sentence judgments, people are very bad at judging sentences that are made semantically incongruent with antonyms.

10.1 Final conclusions

In this book, I first showed the distinct eye movement characteristics of four different types of reading using paragraphs as stimuli. I then showed the large effects that different expectations have on reading, on opinions, and on sentence judgments. People are much less likely to change their opinions when expecting texts that deal with controversial topics, compared to neutral topics. People with weak opinions on the topic change their opinions in line with the bias when expecting a text with positive or negative evaluative bias, but this is not the case for people with strong opinions. People think the language usage is poorer when expecting sentences that originate from less-respected news sources, as compared to well-respected news sources. In addition, the reading process can effortlessly replace an incongruent word with the word people expect without conscious attention being drawn to the incongruity, but effects of this replacement is visible in subsequent reading.

This book shows that when we read, it can be said that we are also reading our expectations. These reading expectations are shortcuts that help us when dealing with the large amounts of text we constantly encounter in our daily lives, but can also make us misinterpret information, misjudge grammatical correctness, and fail to notice errors in text. Readers should be aware that their expectations affect such disparate things as processing information, judging grammatical correctness, and reading incongruent text. Newspapers should be conscious that the expectations from headlines affect how people read the text. Researchers should be aware that paragraphs are well-suited for the analysis of reading using eye tracking, and should be vigilant about the significant effect that expectations can have on reading. In this book, I believe that I have contributed to providing a better understanding of how reading expectations affect many aspects of how texts and sentence are read.

References

- Abbott, M. J., Angele, B., Ahn, Y. D., & Rayner, K. (2015). Skipping syntactically illegal the previews: The role of predictability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(6), 1703–1714.
- Abelson, R. (1963). Computer simulation of “hot” cognition. In S. Tomkins & D. Messick (Eds.), *Computer simulation of personality* (pp. 277–298). New York: Wiley.
- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, 73(3), 247–64.
- Altmann, G. T. M., van Nice, K. Y., Garnham, A., & Henstra, J.-A. (1998). Late closure in context. *Journal of Memory and Language*, 38(4), 459–484.
- Andersson, R., Ferreira, F., & Henderson, J. M. (2011). I see what you’re saying: The integration of complex speech and scenes during language comprehension. *Acta Psychologica*, 137(2), 208–216.
- Ariel, M. (2010). *Defining pragmatics*. Cambridge: Cambridge University Press.
- Arnold, J. E., Wasow, T., Asudeh, A., & Alrenga, P. (2004). Avoiding attachment ambiguities: The role of constituent ordering. *Journal of Memory and Language*, 51(1), 55–70.
- Baayen, R. H., Davidson, D. J., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412.
- Baker, L., & Beall, L. C. (2009). Metacognitive processes and reading comprehension. In S. E. Israel & G. G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 373–388). New York: Routledge.
- Baldwin, T. (2006). Philosophy of language in the twentieth century. In E. Lepore & B. C. Smith (Eds.), *The Oxford handbook of philosophy of language* (pp. 60–102). Oxford: Oxford University Press.
- Barber, T. X., & Silver, M. J. (1968). Fact, fiction, and the experimenter bias effect. *Psychological Bulletin*, 70(6 Pt. 2), 1–29.
- Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, efficiency, intention, and control in social cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of Social Cognition* (2nd ed., Vol. 1, pp. 1–40). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology*, 34(1), 215–220.
- Barton, S. B., & Sanford, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition*, 21(4), 477–487.

- Bastardi, A., Uhlmann, E. L., & Ross, L. (2011). Wishful thinking: Belief, desire, and the motivated evaluation of scientific evidence. *Psychological Science: A Journal of the American Psychological Society / APS*, 22(6), 731–732.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1).
- Beanland, V., & Pammer, K. (2010). Looking without seeing or seeing without looking? Eye movements in sustained inattention blindness. *Vision Research*, 50(10), 977–988.
- Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, 74(3), 183–200.
- Benford, R. D., & Snow, D. A. (2000). Framing processes and social movements: An overview and assessment. *Annual Review of Sociology*, 26, 611–639.
- Bentin, S. (1987). Event-related potentials, semantic processes, and expectancy factors in word recognition. *Vision Research*, 31(2), 308–327.
- Bergen, L., & Goodman, N. D. (2015). The strategic use of noise in pragmatic reasoning. *Topics in Cognitive Science*, 7(2), 336–350.
- Berndt, R. S., Salasoo, A., Mitchum, C. C., & Blumstein, S. E. (1988). The role of intonation cues in aphasic patients' performance of the grammaticality judgment task. *Brain and Language*, 34(1), 65–97.
- Bettinghaus, E. P. (1986). Health promotion and the knowledge-attitude-behavior continuum. *Preventive medicine*, 15, 475–491.
- Bettinghaus, E. P., & Baseheart, J. R. (1969). Some specific factors affecting attitude change. *Journal of Communication*, 19, 227–238.
- Bianchi, I., Savardi, U., & Kubovy, M. (2011). Dimensions and their poles: A metric and topological approach to opposites. *Language and Cognitive Processes*, 26(8), 1232–1265.
- Bicknell, K., & Levy, R. (2012). Word predictability and frequency effects in a rational model of reading. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Conference* (Vol. 1, pp. 126–131).
- Biedert, R., Hees, J., Dengel, A., & Buscher, G. (2012). A robust realtime reading-skimming classifier. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Vol. 1, pp. 123–130). New York, NY, USA.
- Blackburn, P., Ghidini, C., Turner, R., & Giunchiglia, F. (Eds.). (2003). Modeling and using context. In *4th International and Interdisciplinary Conference CONTEXT 2003*. Springer.
- Bland, J. M., & Altman, D. G. (1994). Statistics notes: Regression towards the mean. *British Medical Journal*, 309(6957), 780.
- Blom, J. N., & Hansen, K. R. (2015). Click bait: Forward-reference as lure in online news headlines. *Journal of Pragmatics*, 76, 87–100.
- Bohan, J., Leuthold, H., Hijikata, Y., & Sanford, A. J. (2012). The processing of good-fit semantic anomalies: An ERP investigation. *Neuropsychologia*, 50(14), 3174–3184.
- Bohan, J., & Sanford, A. J. (2008). Semantic anomalies at the borderline of consciousness: An eye-tracking investigation. *Quarterly Journal of Experimental Psychology*, 61(2), 232–239.
- Boland, J. E. (2005). Visual arguments. *Cognition*, 95(3), 237–274.

- Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes*, 101(2), 127–151.
- Bonhage, C. E., Mueller, J. L., Friederici, A. D., & Fiebach, C. J. (2015). Combined eye tracking and fMRI reveals neural basis of linguistic predictions during sentence comprehension. *Cortex*, 68, 33–47.
- Borovsky, A., Elman, J. L., & Kutas, M. (2012). Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. *Language Learning and Development*, 8(3), 278–302.
- Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2014). Native “um”s elicit prediction of low-frequency referents, but non-native “um”s do not. *Journal of Memory and Language*, 75, 104–116.
- Boston, M. F., Hale, J. T., Kliegl, R., & Vasishth, S. (2008). Surprising parser actions and reading difficulty. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies Short Papers - HLT '08* (pp. 5–8).
- Boston, M. F., Hale, J. T., Vasishth, S., & Kliegl, R. (2011). Parallel processing and sentence comprehension difficulty. *Language and Cognitive Processes*, 26(3), 301–349.
- Bott, O., & Solstad, T. (2014). From verbs to discourse: A novel account of implicit causality. In *Psycholinguistic approaches to meaning and understanding across languages* (pp. 213–251). Springer International Publishing.
- Bottoms, H. C., Eslick, A. N., & Marsh, E. J. (2010). Memory and the Moses illusion: Failures to detect contradictions with stored knowledge yield negative memorial consequences. *Memory*, 18(6), 670–678.
- Boucheix, J. M., & Lowe, R. K. (2010). An eye tracking comparison of external pointing cues and internal continuous cues in learning with complex animations. *Learning and Instruction*, 20(2), 123–135.
- Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 717–726.
- Braze, D., Shankweiler, D., Ni, W., & Palumbo, L. C. (2002). Readers’ eye movements distinguish anomalies of form and content. *Journal of Psycholinguistic Research*, 31(1), 25–44.
- Bredart, S., & Modolo, K. (1988). Moses strikes again: Focalization effect on a semantic illusion. *Acta Psychologica*, 67(2), 135–144.
- Britt, M. A., Perfetti, C. A., Garrod, S., & Rayner, K. (1992). Parsing in discourse: Context effects and their limits. *Journal of Memory and Language*, 31(3), 293–314.
- Bryant, J., & Miron, D. (2004). Theory and research in mass communication. *Journal of Communication*, 54(4), 662–704.
- Brysbaert, M., Drieghe, D., & Vitu, F. (2005). Word skipping: Implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Cognitive processes in eye guidance* (pp. 53–77). Oxford: Oxford University Press.

- Bucaria, C. (2004). Lexical and syntactic ambiguity as a source of humor: The case of newspaper headlines. *Humor - International Journal of Humor Research*, 17(3), 279–309.
- Burgess, C., & Lund, K. (1997). Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes*, 12(January 2015), 177–210.
- Burton, D. (2000). Questionnaire design. In *Research Training for Social Scientists* (pp. 335–347). London: SAGE Publications.
- Büttner, A. C. (2002). *Moses and the Ark: Exploring semantic illusions*. Oxford Brookes University.
- Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction*, 20(2), 155–166.
- Carragee, K. M., & Roefs, W. (2004). The neglect of power in recent framing research. *Journal of Communication*, 54(2), 214–233.
- Chen, Y., Conroy, N. J., & Rubin, V. L. (2015). Misleading online content: Recognizing clickbait as “false news.” In *Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection* (pp. 15–19). New York, NY, USA: ACM.
- Chong, D., & Druckman, J. N. (2007). Framing theory. *Annual Review of Political Science*, 10, 103–126.
- Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, 42(4), 368–407.
- Christianson, K., Luke, S. G., & Ferreira, F. (2010). Effects of plausibility on structural priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(2), 538–544.
- Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, 55, 591–621.
- Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, 27(1), 3–42.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36, 181–204.
- Clark, J. K., Wegener, D. T., Habashi, M. M., & Evans, A. T. (2012). Source expertise and persuasion: The effects of perceived opposition or support on message scrutiny. *Personality & Social Psychology Bulletin*, 38(1), 90–100.
- Clifton, C., Ferreira, F., Henderson, J. M., Inhoff, A. W., Liversedge, S. P., Reichle, E. D., & Schotter, E. R. (2016). Eye movements in reading and information processing: Keith Rayner’s 40 year legacy. *Journal of Memory and Language*, 86, 1–19.
- Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. P. G. Van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 341–371). Elsevier Ltd.
- Collard, P., Corley, M., & MacGregor, L. J. (2008). Attention orienting effects of hesitations in speech: Evidence from ERPs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(3), 1–25.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6), 407–428.

- Converse, P. E. (1962). *The nature of belief systems in mass publics*. Survey Research Center, University of Michigan.
- Converse, P. E. (2006). The nature of belief systems in mass publics (1964). *Critical Review: A Journal of Politics and Society*, 18(1–3), 1–74.
- Corley, M., & Hartsuiker, R. J. (2011). Why um helps auditory word recognition: The temporal delay hypothesis. *PLoS ONE*, 6(5).
- Cornelissen, T. H. W., & Võ, M. L.-H. (2016). Stuck on semantics: Automatic processing of object-scene inconsistencies modulates ongoing gaze behavior. *Attention, Perception, & Psychophysics*, 1–35.
- Costa, P. T. (1991). Clinical use of the five-factor model: An introduction. *Journal of Personality Assessment*, 57(3), 393–398.
- Croft, W., & Cruse, A. D. (2004). *Cognitive linguistics*. Cambridge: Cambridge University Press.
- Cummins, C., & Katsos, N. (2010). Comparative and superlative quantifiers: Pragmatic effects of comparison type. *Journal of Semantics*, 27(3), 271–305.
- Dąbrowska, E. (2010). Naive v. expert intuitions: An empirical study of acceptability judgments. *The Linguistic Review*, 27, 1–23.
- Daneman, M., Lennertz, T., & Hannon, B. (2007). Shallow semantic processing of text: Evidence from eye movements. *Language and Cognitive Processes*, 22(1), 83–105.
- Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, 3(4), 422–433.
- Davies, M. (2008). The corpus of contemporary American English: 450 million words, 1990–present.
- Davis, W. (2015). Implicature. In *Oxford handbooks online*. Oxford University Press.
- de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010). Attention guidance in learning from a complex animation: Seeing is understanding? *Learning and Instruction*, 20(2), 111–122.
- 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.
- Denham, B. (2012). Book review: Thinking, fast and slow. *Journal of Communication*, 62(5), E6–E11.
- DeVellis, R. F. (2003). *Scale development: Theory and applications (Vol. 26)* (2nd ed.). London: SAGE Publications.
- DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, 93(5), 880–896.
- Dikker, S., & Pylkkänen, L. (2013). Predicting language: MEG evidence for lexical preactivation. *Brain and Language*, 127(1), 55–64.
- Dollinger, S. J., Leong, F. T. L., & Ulicni, S. K. (1996). On traits and values: With special reference to openness to experience. *Journal of Research in Personality*, 30, 23–41.
- Dor, D. (2003). On newspaper headlines as relevance optimizers. *Journal of Pragmatics*, 35(5), 695–721.

- Doyen, S., Klein, O., Pichon, C.-L., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PLoS ONE*, 7(1), 1–7.
- Drenhaus, H., Demberg, V., Köhne, J., & Delogu, F. (2014). Incremental and predictive discourse processing based on causal and concessive discourse markers: ERP studies on German and English. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, 36 (pp. 403–408).
- Drew, T., Vó, M. L.-H., & Wolfe, J. M. (2013). The invisible gorilla strikes again: Sustained inattention blindness in expert observers. *Psychological Science*, 24(9), 1848–1853.
- Drieghe, D., Desmet, T., & Brysbaert, M. (2007). How important are linguistic factors in word skipping during reading? *British Journal of Psychology*, 98(1), 157–171.
- Druckman, J. N. (2001). Evaluating framing effects. *Journal of Economic Psychology*, 22, 91–101.
- Duggan, G. B., & Payne, S. J. (2009). Text skimming: The process and effectiveness of foraging through text under time pressure. *Journal of Experimental Psychology: Applied*, 15(3), 228–242.
- Duggan, G. B., & Payne, S. J. (2011). Skim reading by satisficing: Evidence from eye tracking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1141–1150). ACM.
- Dunn, R., Beaudry, J. S., & Klavas, A. (2002). Survey of research on learning styles. *California Journal of Science Education*, 2(2), 75–98.
- Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, 24(6), 409–436.
- Edwards, G., Vetter, P., Mcgruer, F., Petro, L. S., Muckli, L., Hall, J., ... Muckli, L. (2017). Predictive feedback to V1 dynamically updates with sensory input. *Scientific Reports*, 7(16538), 1–24.
- Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology*, 71(1), 5–24.
- Eitel, A., & Kühl, T. (2016). Effects of disfluency and test expectancy on learning with text. *Metacognition and Learning*, 11(1), 1–13.
- Ekman, M., Kok, P., & de Lange, F. P. (2017). Time-compressed preplay of anticipated events in human primary visual cortex. *Nature Communications*, 8.
- Etgar, M., & Goodwin, S. A. (1982). One-sided versus two-sided comparative message appeals for new brand introductions. *Journal of Consumer Research*, 8(4), 460–465.
- Expectation. (2017). In *Oxford English Dictionary*. Oxford University Press.
- Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44(4), 491–505.
- Fein, O., Beni-Noked, S., & Giora, R. (2015). Under/standing cartoons: The suppression hypothesis revisited. *Journal of Pragmatics*, 86, 86–93.
- Ferguson, H. J., Scheepers, C., & Sanford, A. J. (2010). Expectations in counterfactual and theory of mind reasoning. *Language and Cognitive Processes*, 25(3), 297–346.

- Fernández, G., Shalom, D. E., Kliegl, R., & Sigman, M. (2014). Eye movements during reading proverbs and regular sentences: The incoming word predictability effect. *Language, Cognition and Neuroscience*, 29(3), 260–273.
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, 47(2), 164–203.
- Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11(1), 11–15.
- Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: implications for models of sentence processing and reanalysis. *Journal of Psycholinguistic Research*, 30(1), 3–20.
- Ferreira, F., & Engelhardt, P. E. (2006). Syntax and production. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics* (pp. 61–91). London, UK: Academic Press.
- Ferreira, F., & Engelhardt, P. E. (2009). Good enough language processing: A satisficing approach. In N. Taatgen, H. Rijn, J. Nerbonne, & L. Schomaker (Eds.), *Proceedings of the 31st Annual conference of the Cognitive Science Society*. (pp. 413–418).
- Ferreira, F., Lau, E. F., & Bailey, K. G. D. (2004). Disfluencies, language comprehension, and tree adjoining grammars. *Cognitive Science*, 28(5), 721–749.
- Ferreira, F., & Patson, N. D. (2007). The “good enough” approach to language comprehension. *Language and Linguistics Compass*, 1(1–2), 71–83.
- Festinger, L. (1962). *A theory of cognitive dissonance*. Stanford, California: Stanford University Press.
- Fetzer, A. (2017). Context. In Y. Huang (Ed.), *The Oxford handbook of pragmatics* (pp. 259–276). Oxford: Oxford University Press.
- Fiebach, C. J., Schleewsky, 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.
- Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE*, 8(10).
- Foraker, S., & Murphy, G. L. (2012). Polysemy in sentence comprehension: Effects of meaning dominance. *Journal of memory and language*, 67(4), 407–425.
- Foucart, A., Ruiz-Tada, E., & Costa, A. (2015). How do you know I was about to say “book”? Anticipation processes affect speech processing and lexical recognition. *Language, Cognition and Neuroscience*, 30(6), 768–780.
- Foulsham, T., Farley, J., & Kingstone, A. (2013). Mind wandering in sentence reading: Decoupling the link between mind and eye. *Canadian Journal of Experimental Psychology*, 67(1), 51–59.
- Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies*. University of Connecticut.
- Frazier, L., & Clifton, C. (2015). Without his shirt off he saved the child from almost drowning: Interpreting an uncertain input. *Language, Cognition and Neuroscience*, 30(6), 635–647.

- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14(2), 178–210.
- Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language*, 29(2), 181–200.
- Frenzel, S., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2011). Conflicts in language processing: A new perspective on the N400-P600 distinction. *Neuropsychologia*, 49(3), 574–579.
- Friedrich, J. (1993). Primary error detection and minimization (PEDMIN) strategies in social cognition: A reinterpretation of confirmation bias phenomena. *Psychological review*, 100(2), 298–319.
- Frisch, S., & Schlesewsky, M. (2001). The N400 reflects problems of thematic hierarchizing. *NeuroReport*, 12(15), 3391–3394.
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), 815–836.
- Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521), 1211–1221.
- Fuoli, M. (2017). Building a trustworthy corporate identity: A corpus-based analysis of stance in annual and corporate social responsibility reports. *Applied Linguistics*.
- Gaertner, S. L., & Dovidio, J. F. (2012). *Reducing intergroup bias: The common ingroup identity model*. New York: Routledge.
- Gagnepain, P., Henson, R. N., & Davis, M. H. (2012). Temporal predictive codes for spoken words in auditory cortex. *Current Biology*, 22(7), 615–621.
- Galesic, M., Tourangeau, R., Couper, M. P., & Conrad, F. G. (2008). Eye-tracking data: New insights on response order effects and other cognitive shortcuts in survey responding. *Public Opinion Quarterly*, 72(5), 892–913.
- Gao, J., & Suzuki, H. (2005). Long distance dependency in language modeling: An empirical study. In *Natural language processing* (pp. 396–405). Springer Berlin Heidelberg.
- Garvey, C., & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry*, 5(3), 459–464.
- Gass, S. M. (1994). The reliability of second-language grammaticality judgments. In E. E. Tarone, S. M. Gass, & A. D. Cohen (Eds.), *Research methodology in second language acquisition* (pp. 303–322). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Gazdar, G. (1979). *Pragmatics: Implicature, presupposition and logical form*. London: Academic Press.
- Gerrig, R. J., Ohaeri, J. O., & Brennan, S. E. (2000). Illusory transparency revisited. *Discourse Processes*, 29(2), 137–159.
- Gersten, R., Fuchs, L. S., Williams, J. P., & Baker, S. (2001). Teaching reading comprehension strategies to students with learning disabilities: A review of research. *Review of Educational Research*, 71(2), 279–320.

- Gibbs, R. W. (2007). Why cognitive linguists should care more about empirical methods. In M. Gonzalez-Marquez, I. Mittelberg, S. Coulson, & M. J. Spivey (Eds.), *Methods in cognitive linguistics* (pp. 2–18). Amsterdam: John Benjamins Publishing.
- Gibbs, R. W. (2017). Experimental pragmatics. In Y. Huang (Ed.), *The Oxford handbook of pragmatics* (pp. 310–325). Oxford: Oxford University Press.
- Gibbs, R. W., & Colston, H. L. (2012). *Interpreting figurative meaning*. New York, NY, USA: Cambridge University Press.
- Gibbs, R. W., Kushner, J. M., & Mills, W. R. (1991). Authorial intentions and metaphor comprehension. *Journal of Psycholinguistic Research*, 20(1), 11–30.
- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68(68), 1–76.
- Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 110, pp. 8051–8056).
- Gibson, E., & Fedorenko, E. (2013). The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes*, 28(1–2), 88–124.
- Gibson, E., Piantadosi, S., & Fedorenko, K. (2011). Using Mechanical Turk to obtain and analyze English acceptability judgments. *Linguistics and Language Compass*, 5(8), 509–524.
- Gideon, L. (2012). The art of question phrasing. In L. Gideon (Ed.), *Handbook of survey methodology for the social sciences* (pp. 91–107). New York: Springer.
- Gidlöf, K., Anikin, A., Lingonblad, M., & Wallin, A. (2017). Looking is buying. How visual attention and choice are affected by consumer preferences and properties of the supermarket shelf. *Appetite*, 116, 29–38.
- Giora, R., Balaban, N., Fein, O., & Algabets, I. (2004). Negation as positivity in disguise. In H. L. Colston & A. Katz (Eds.), *Figurative language comprehension: Social and cultural influences* (pp. 233–258). Hillsdale, NJ: Erlbaum.
- Giora, R., Raphaely, M., Fein, O., & Livnat, E. (2014). Resonating with contextually inappropriate interpretations in production: The case of irony. *Cognitive Linguistics*, 25(3), 443–455.
- Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9(3), 558–565.
- Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language*, 43(3), 379–401.
- Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, 48, 163–189.
- Grice, H. P. (1957). Meaning. *The Philosophical Review*, 66(3), 377–388.
- Grice, H. P. (1981). Presupposition and conversational implicature. *Radical Pragmatics*, 183.
- Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Guazzini, A., Yoneki, E., & Gronchi, G. (2014). Cognitive dissonance and social influence effects on preference judgements: An eye tracking based system for their automatic assessment. *International Journal of Human-Computer Studies*, 73, 12–18.

- Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Cambridge, MA: The MIT Press.
- Gärdenfors, P. (2014). *The geometry of meaning: Semantics based on conceptual spaces*. MIT Press.
- Haberlandt, K. (1982). Reader expectations in text comprehension. *Advances in Psychology*, 9, 239–249.
- Hahn, U., Harris, A. J. L., & Corner, A. (2009). Argument content and argument source: An exploration. *Informal Logic*, 29(4), 337–367.
- Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis. Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, 11(2), 194–205.
- Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *NAACL '01: Second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies 2001* (pp. 1–8).
- Hannon, B., & Daneman, M. (2004). Shallow semantic processing of text: An individual-differences account. *Discourse Processes*, 37, 187–204.
- Harmon-Jones, E., & Mills, J. (Eds.). (1999). *Cognitive dissonance: Progress on a pivotal theory in social psychology*. Washington DC: American Psychological Association.
- Hartshorne, J. K. (2013). What is implicit causality? *Language, Cognition and Neuroscience*, 29(7), 804–824.
- Hartshorne, J. K., & Snedeker, J. (2013). Verb argument structure predicts implicit causality: The advantages of finer-grained semantics. *Language and Cognitive Processes*, 28(10), 1474–1508.
- Heim, I. (1983). On the projection problem for presuppositions. In *Proceedings of the West Coast Conference on Formal Linguistics 2* (pp. 114–125).
- Hergovich, A., Schott, R., & Burger, C. (2010). Biased evaluation of abstracts depending on topic and conclusion: Further evidence of a confirmation bias within scientific psychology. *Current Psychology*, 29(3), 188–209.
- Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, 53(1), 575–604.
- Hofmeister, P., Culicover, P., & Winkler, S. (2015). Effects of processing on the acceptability of “frozen” extraposed constituents. *Syntax*, 18(4), 464–483.
- Hohwy, J. (2013). *The predictive mind*. Oxford: Oxford University Press.
- Hohwy, J. (2015). The neural organ explains the mind. In T. Metzinger & J. M. Windt (Eds.), *Open MIND* (Vol. 19, pp. 1–22). Frankfurt am Main: MIND Group.
- Holleman, B. (2000). *The forbid/allow asymmetry: On the cognitive mechanisms underlying wording effects in surveys* (Vol. 16.). Amsterdam, The Netherlands: Rodopi.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking - A comprehensive guide to methods and measures*. Oxford: Oxford University Press.
- Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly*, 15(4), 635–650.
- Huang, Y. (2000). *Anaphora: A cross-linguistic study*. Oxford: Oxford University Press.

- Huang, Y. (2007). *Pragmatics*. Oxford: Oxford University Press.
- Huang, Y. (2017). Introduction: What is pragmatics? In Y. Huang (Ed.), *The Oxford handbook of pragmatics* (pp. 1–20). Oxford: Oxford University Press.
- Huey, E. B. (1908). *The psychology and pedagogy of reading*. New York: The Macmillan Company.
- Hyönä, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction*, 20(2), 172–176.
- Hyönä, J. (2011). Foveal and parafoveal processing during reading. In S. P. Livversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 819–838). Oxford: Oxford University Press.
- Hyönä, J., Lorch, R. F., & Kaakinen, J. K. (2002). Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology*, 94(1), 44–55.
- Hyönä, J., Lorch, R. F., & Rinck, M. (2003). Eye movement measures to study global text processing. In R. Radach, J. Hyönä, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 313–334). Amsterdam: Elsevier Science.
- Hyönä, J., & Niemi, P. (1990). Eye movements during repeated reading of a text. *Acta Psychologica*, 73, 259–280.
- Hyönä, J., & Nurminen, A.-M. (2006). Do adult readers know how they read? Evidence from eye movement patterns and verbal reports. *British Journal of Psychology*, 97(1), 31–50.
- Ifantidou, E. (2009). Newspaper headlines and relevance: Ad hoc concepts in ad hoc contexts. *Journal of Pragmatics*, 41(4), 699–720.
- Inhoff, A. W. (1984). Two stages of word processing during eye fixations in the reading of prose. *Journal of Verbal Learning and Verbal Behavior*, 23(5), 612–624.
- Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language*, 86, 157–171.
- Jackendoff, R. (2007). A parallel architecture perspective on language processing. *Brain Research*, 1146, 2–22.
- Jakuszeit, M., Kotz, S. A., & Hasting, A. S. (2013). Generating predictions: Lesion evidence on the role of left inferior frontal cortex in rapid syntactic analysis. *Cortex*, 49(10), 2861–2874.
- Jarodzka, H., & Brand-Gruwel, S. (2017). Tracking the reading eye: Towards a model of real-world reading. *Journal of Computer Assisted Learning*, 33(3), 193–201.
- Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, 20(2), 146–154.
- Johansson, R., Holmqvist, K., Mossberg, F., & Lindgren, M. (2011). Eye movements and reading comprehension while listening to preferred and non-preferred study music. *Psychology of Music*.
- John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–138). New York: The Guilford Press.

- Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology*, 80(4), 557–571.
- Jones, S., Murphy, M. L., Paradis, C., & Willners, C. (2012). *Antonyms in English*. Cambridge: Cambridge University Press.
- Jullien, D. (2016). *A methodological perspective on behavioral economics and the role of language in economic rationality*. Université Nice Sophia Antipolis.
- Kaakinen, J. K., & Hyönä, J. (2007). Perspective effects in repeated reading: An eye movement study. *Memory & Cognition*, 35(6), 1323–1336.
- Kaakinen, J. K., & Hyönä, J. (2010). Task effects on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(6), 1561–1566.
- Kaakinen, J. K., Hyönä, J., & Keenan, J. M. (2003). How prior knowledge, WMC, and relevance of information affect eye fixations in expository text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(3), 447–457.
- Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.
- Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, 94(2), 113–147.
- Kamas, E. N., Reder, L. M., & Ayers, M. S. (1996). Partial matching in the Moses illusion: Response bias not sensitivity. *Memory & Cognition*, 24(6), 687–699.
- Kamide, Y. (2008). Anticipatory Processes in Sentence Processing. *Linguistics and Language Compass*, 2(4), 647–670.
- Kamoen, N., Holleman, B., Mak, P., Sanders, T., & van den Bergh, H. (2011). Agree or disagree? Cognitive processes in answering contrastive survey questions. *Discourse Processes*, 48(5), 355–385.
- Kamoen, N., Holleman, B., & van den Bergh, H. (2013). Positive, negative, and bipolar questions: The effect of question polarity on ratings of text readability. *Survey Research Methods*, 7(3), 181–189.
- Kardash, C. M., & Scholes, R. J. (1996). Effects of preexisting beliefs, epistemological beliefs, and need for cognition on interpretation of controversial issues. *Journal of Educational Psychology*, 88(2), 260–271.
- Karttunen, L. (1973). Presuppositions of compound sentences. *Linguistic Inquiry*, 4, 169–93.
- Keasey, C. B. (1973). Experimentally induced changes in moral opinions and reasoning. *Journal of Personality and Social Psychology*, 26(1), 30–38.
- Kennedy, A., Pynte, J., Murray, W. S., & Paul, S.-A. (2013). Frequency and predictability effects in the Dundee corpus: An eye movement analysis. *Quarterly Journal of Experimental Psychology*, 66(3), 601–618.
- Keysar, B., & Henly, A. S. (2002). Speakers' overestimation of their effectiveness. *Psychological Science*, 13(3), 207–212.
- Kikuchi, Y., Attaheri, A., Wilson, B., Rhone, A. E., Nourski, K. V., Gander, P. E., ... Petkov, C. I. (2017). Sequence learning modulates neural responses and oscillatory coupling in human and monkey auditory cortex. *PLoS Biology*, 15(4), e2000219.

- 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.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.
- Kiousis, S. (2005). Compelling arguments and attitude strength: Exploring the impact of second-level agenda setting on public opinion of presidential candidate images. *The Harvard International Journal of Press/Politics*, 10(2), 3–27.
- Klayman, J. (1995). Varieties of confirmation bias. In J. Busemeyer, R. Hastie, & D. L. Medin (Eds.), *Psychology of learning and motivation - Advances in research and theory* (Vol. 32, pp. 385–418). San Diego: Academic Press.
- Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, 16(1–2), 262–284.
- Koornneef, A. W., & Van Berkum, J. J. A. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language*, 54(4), 445–465.
- Kotz, S. A., Holcomb, P. J., & Osterhout, L. (2008). ERPs reveal comparable syntactic sentence processing in native and non-native readers of English. *Acta Psychologica*, 128(3), 514–527.
- Kretchmar, F., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). Parafoveal and foveal N400s dissociate spreading activation from contextual fit. *NeuroReport*, 20, 1613–1618.
- Kronrod, A., & Engel, O. (2001). Accessibility theory and referring expressions in newspaper headlines. *Journal of Pragmatics*, 33(5), 683–699.
- Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1), 32–59.
- Kuperberg, G. R., Paczynski, M., & Ditman, T. (2011). Establishing causal coherence across sentences: An ERP study. *Journal of Cognitive Neuroscience*, 23(5), 1230–1246.
- Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190–207). New York, NY, USA: Oxford University Press.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427), 203–205.
- Kuypers, J. A. (2002). *Press bias and politics: How the media frame controversial issues*. Westport, US: Greenwood Publishing Group.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2), 211–240.
- Langacker, R. (1999). *Grammar and conceptualization*. Berlin: Mouton de Gruyter.
- Lau, E. F., & Ferreira, F. (2005). Lingering effects of disfluent material on comprehension of garden path sentences. *Language and Cognitive Processes*, 20(5), 633–666.

- 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.
- Leow, R. P. (1996). Grammaticality judgment tasks and second-language development. In J. E. Alatis (Ed.), *Georgetown university round table on languages and linguistics* (pp. 126–39). Georgetown University Press.
- Levin, I. P., & Gaeth, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research*, 15(3), 374–378.
- Levin, I. P., Gaeth, G. J., Schreiber, J., & Lauriola, M. (2002). A new look at framing effects: Distribution of effect sizes, individual differences, and independence of types of effects. *Organizational Behavior and Human Decision Processes*, 88(1), 411–429.
- Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A topology and critical analysis of framing effect. *Organizational Behavior and Human Decision Processes*, 76(2), 149–188.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177.
- Levy, R. (2011a). Probabilistic linguistic expectations, uncertain input, and implications for eye movements in reading. *Studies of Psychology and Behavior*, 9(1), 52–63.
- Levy, R. (2011b). Integrating surprisal and uncertain-input models in online sentence comprehension: Formal techniques and empirical results. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies* (pp. 1055–1065).
- Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems. Proceedings of the 2006 Conference* (pp. 841–848). MIT Press.
- Lewis, R. L. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research*, 25(1), 93–115.
- Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3), 375–419.
- Liversedge, S. P., Pickering, M. J., Branigan, H. P., & van Gompel, R. P. G. (1998). Processing arguments and adjuncts in isolation and context: The case of by-phrase ambiguities in passives. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 461–475.
- Lodge, M., & Stroh, P. (1993). Inside the mental voting booth: An impression-driven model. In S. Tyengar & W. McGuire (Eds.), *Explorations in political psychology* (pp. 225–263). Durham: Duke University Press.
- Lodge, M., & Taber, C. S. (2005). The automaticity of affect for political candidates, parties and issues: An experimental test of the hot cognition hypothesis. *Political Psychology*, 26(3), 455–482.
- Long, D. L., Prat, C., Johns, C., Morris, P., & Jonathan, E. (2008). The importance of knowledge in vivid text memory: An individual-differences investigation of recollection and familiarity. *Psychonomic Bulletin & Review*, 15(3), 604–609.

- Long, D. L., & Prat, C. S. (2002). Memory for Star Trek: The role of prior knowledge in recognition revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(6), 1073–1082.
- Long, D. L., Wilson, J., Hurley, R., & Prat, C. S. (2006). Assessing text representations with recognition: The interaction of domain knowledge and text coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(4), 816–827.
- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37(11), 2098–2109.
- Lou, Y., Liu, Y., Kaakinen, J. K., & Li, X. (2017). Using support vector machines to identify literacy skills: Evidence from eye movements. *Behavior Research Methods*, 49(3), 887–895.
- MacGregor, L. J., Corley, M., & Donaldson, D. I. (2010). Listening to the sound of silence: Disfluent silent pauses in speech have consequences for listeners. *Neuropsychologia*, 48(14), 3982–3992.
- Magliano, J. P., Graesser, A. C., Eymard, L. A., Haberlandt, K., & Gholson, B. (1993). Locus of interpretive and inference processes during text comprehension: A comparison of gaze durations and word reading times. *Cognition*, 19(3), 704–709.
- Mak, W. M., & Sanders, T. J. M. (2012). The role of causality in discourse processing: Effects of expectation and coherence relations. *Language and Cognitive Processes*, 28(9), 1414–1437.
- Marquez, F. T. (1980). How accurate are the headlines? *Journal of Communication*, 30(3), 30–36.
- Marsh, D. M., & Hanlon, T. J. (2007). Seeing what we want to see: Confirmation bias in animal behavior research. *Ethology*, 113(11), 1089–1098.
- Mason, L., Tornatora, M. C., & Pluchino, P. (2013). Do fourth graders integrate text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Computers & Education*, 60, 95–109.
- Mata, A., Ferreira, M. B., & Reis, J. (2013). A process-dissociation analysis of semantic illusions. *Acta Psychologica*, 144(2), 433–443.
- Mauner, G., Fromkin, V. A., & Cornell, T. L. (1993). Comprehension and acceptability judgments in agrammatism: Disruptions in the syntax of referential dependency. *Brain and Language*, 45, 340–370.
- Mayer, J. M. (1978). Assessment of depression. In M. P. McReynolds (Ed.), *Advances in psychological assessment, Vol. 4* (pp. 358–425). San Francisco: Jossey-Bass.
- Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction*, 20(2), 167–171.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Readings in Cognitive Science: A Perspective from Psychology and Artificial Intelligence*, 88(5), 375–407.
- McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly*, 36(2), 176–187.

- McCrudden, M. T., & Sparks, P. C. (2014). Exploring the effect of task instructions on topic beliefs and topic belief justifications: A mixed methods study. *Contemporary Educational Psychology*, 39(1), 1–11.
- McEvily, B., Perrone, V., & Zaheer, A. (2003). Trust as an organizing principle. *Organization Science*, 14(1), 91–103.
- McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38, 283–312.
- Meibauer, J. (2012). What is a context? Theoretical and empirical evidence. In R. Finkbeiner, J. Meibauer, & P. B. Schumacher (Eds.), *What is a context? Linguistic approaches and challenges*. (pp. 9–32). Amsterdam: John Benjamins Publishing Company.
- Meibauer, J., & Steinbach, M. (2011). Introduction: Experimental research at the pragmatics/semantics interface. In J. Meibauer & M. Steinbach (Eds.), *Experimental pragmatics/semantics* (pp. 1–18). Amsterdam/Philadelphia: John Benjamins Publishing.
- Mercier, H. (2017). Confirmation bias - Myside bias. In R. F. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory* (2nd ed., pp. 99–114). New York: Routledge.
- Miller, G. R., & Baseheart, J. (1969). Source trustworthiness, opinionated statements, and response to persuasive communication. *Speech Monographs*, 36(1), 1–7.
- Moerman, D. E., & Jonas, W. B. (2002). Deconstructing the placebo effect and finding the meaning response. *Annals of internal medicine*, 136(12), 471–476.
- Molek-Kozakowska, K. (2013). Towards a pragma-linguistic framework for the study of sensationalism in news headlines. *Discourse & Communication*, 7(2), 173–197.
- Molek-Kozakowska, K. (2014). Coercive metaphors in news headlines: A cognitive-pragmatic approach. *Brno Studies in English*, 40(1), 149–173.
- Moravcsik, J. E., & Kintsch, W. (1993). Writing quality, reading skills, and domain knowledge as factors in text comprehension. *Canadian Journal of Experimental Psychology*, 47(2), 360–374.
- Morris, J. P., Squires, N. K., Taber, C. S., & Lodge, M. (2003). Activation of political attitudes: A psychophysiological examination of the hot cognition hypothesis. *Political Psychology*, 24(4), 727–746.
- Morris, R. K. (2006). Lexical processing and sentence context effects. In M. A. Gernsbacher & M. J. Traxler (Eds.), *Handbook of psycholinguistics* (pp. 377–401). Burlington: Academic Press.
- Moss, H. E., Ostrin, R. K., Tyler, L. K., & Marslen-Wilson, W. D. (1995). Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 863–883.
- Murphy, M. L., & Jones, S. (2008). Antonyms in children's and child-directed speech. *First Language*, 28(4), 403–430.
- Nagata, H. (1988). The relativity of linguistic intuition: The effect of repetition on grammaticality judgments. *Journal of Psycholinguistic Research*, 17(1), 1–17.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220.

- Nieuwland, M. S., & Van Berkum, J. J. A. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research*, 24(3), 691–701.
- Noveck, I., & Sperber, D. (Eds.). (2004). Experimental pragmatics. In *Experimental pragmatics*. Palgrave Macmillan.
- Noveck, I., & Sperber, D. (2012). The why and how of experimental pragmatics: The case of “scalar inferences.” In D. Wilson & D. Sperber (Eds.), *Meaning and relevance* (pp. 307–330). Cambridge: Cambridge University Press.
- Oostendorp, H. Van, & De Mul, S. (1990). Moses beats Adam: A semantic relatedness effect on a semantic illusion. *Acta Psychologica*, 74, 35–46.
- Oswald, M. E., & Grosjean, S. (2004). Confirmation bias. In R. F. Pohl (Ed.), *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory* (pp. 79–96). Hove, UK: Psychology Press.
- Otten, M., & Van Berkum, J. J. A. (2008). Discourse-based word anticipation during language processing: Prediction or priming? *Discourse Processes*, 45(6), 464–496.
- Oulton, C., Dillon, J., & Grace, M. (2004). Reconceptualizing the teaching of controversial issues. *International Journal of Science Education*, 26(4), 411–423.
- Pan, Z., & Kosicki, G. (1993). Framing analysis: An approach to news discourse. *Political Communication*, 10, 55–75.
- Paradis, C. (2001). Adjectives and boundedness. *Cognitive Linguistics*, 12(1), 47–65.
- Paradis, C. (2003). Is the notion of linguistic competence relevant in Cognitive Linguistics? *Annual Review of Cognitive Linguistics*, 1, 207–231.
- Paradis, C. (2005). Ontologies and construals in lexical semantics. *Axiomathes*, 15(4), 541–573.
- Paradis, C. (2013). Lexical semantics. In C. A. Chapeller (Ed.), *The encyclopedia of applied linguistics*. Oxford, UK: Wiley-Blackwell.
- Paradis, C. (2015). Meanings of words: Theory and application. In U. Hass & P. Storjohann (Eds.), *Handbuch Wort und Wortschatz*. Berlin: Mouton de Gruyter.
- Paradis, C., Löhdorf, S., van de Weijer, J., & Willners, C. (2015). Semantic profiles of antonymic adjectives in discourse. *Linguistics*, 53(1), 153–191.
- Paradis, C., & Willners, C. (2011). Antonymy: From convention to meaning-making. *Review of Cognitive Linguistics*, 9(2), 367–391.
- Paradis, C., Willners, C., & Jones, S. (2009). Good and bad opposites: Using textual and experimental techniques to measure antonym canonicity. *The Mental Lexicon*, 4(3), 380–429.
- Park, H., & Reder, L. M. (2004). Moses illusion: Implication for human cognition. *Cognitive Illusions*, 275–291.
- Peleg, O., Giora, R., & Fein, O. (2004). Contextual strength: The whens and hows of context effects. In I. Noveck & D. Sperber (Eds.), *Experimental pragmatics* (pp. 172–186). Palgrave Macmillan.
- Perfetti, C. A., Faux, R., Beverly, S., Bell, L., & Rodgers, K. (1987). Comprehending newspaper headlines. *Journal of Memory and Language*, 26(6), 692–713.

- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *The Behavioral and Brain Sciences*, 36(4), 329–347.
- Pickering, M. J., & van Gompel, R. P. G. (2006). Syntactic parsing. In M. A. Gernsbacher & M. J. Traxler (Eds.), *Handbook of psycholinguistics* (pp. 455–503). Burlington: Academic Press.
- Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology*, 34(2), 243–281.
- Prediction. (2017). In *Oxford English Dictionary*. Oxford University Press.
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6(7), 576–582.
- Pyykkönen, P., & Järvikivi, J. (2009). Activation and persistence of implicit causality information in spoken language comprehension. *Experimental Psychology*, 57(1), 5–16.
- R Development Core Team. (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422.
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62(8), 1457–1506.
- Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14(3), 191–201.
- Rayner, K., Garrod, S., & Perfetti, C. A. (1992). Discourse influences during parsing are delayed. *Cognition*, 45(2), 109–139.
- Rayner, K., Schotter, E. R., Masson, M. E., Potter, M. C., & Treiman, R. (2016). So much to read, so little time: How do we read, and can speed reading help? *Psychological Science in the Public Interest*, 17(1), 4–34.
- Rayner, K., Slattery, T. J., & Belanger, N. N. (2010). Eye movements, the perceptual span, and reading speed. *Psychonomic Bulletin & Review*, 17(6), 834–839.
- Rayner, K., Smith, T. J., Malcolm, G. L., & Henderson, J. M. (2009). Eye movements and visual encoding during scene perception. *Psychological Science*, 20(1), 6–10.
- Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(6), 1290–1301.
- Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, 3(4), 504–509.
- Reder, L. M., & Kusbit, G. W. (1991). Locus of the Moses illusion: Imperfect encoding, retrieval, or match? *Journal of Memory and Language*, 30, 385–406.
- Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, 26(4), 445–476–526.
- Reingold, E. M., & Rayner, K. (2006). Examining the word identification stages hypothesized by the E-Z Reader model. *Psychological Science*, 17(9), 742–746.

- Rensink, R. A. (2002). Change detection. *Annual Review of Psychology*, 53(1), 245–277.
- Roberts, G., Langstein, B., & Galantucci, B. (2016). (In)sensitivity to incoherence in human communication. *Language and Communication*, 47, 15–22.
- Roccas, S., Sagiv, L., Schwartz, S. H., & Knafo, A. (2002). The Big Five personality factors and personal values. *Personality and Social Psychology Bulletin*, 28(6), 789–801.
- Roehm, D., Bornkessel-Schlesewsky, I., Rösler, F., & Schlewsky, M. (2007). To predict or not to predict: Influences of task and strategy on the processing of semantic relations. *Journal of Cognitive Neuroscience*, 19(8), 1259–1274.
- Roland, D., Mauner, G., O'Meara, C., & Yun, H. (2012). Discourse expectations and relative clause processing. *Journal of Memory and Language*, 66(3), 479–508.
- Rubin, V. L., & Liddy, E. D. (2006). Assessing credibility of weblogs. In *Proceedings of the AAAI spring symposium: Computational approaches to analyzing weblogs* (pp. 187–190).
- Sadeghi, N., Kasim, Z. M., Tan, B. H., & Abdullah, F. S. (2012). Learning styles, personality types and reading comprehension performance. *English Language Teaching*, 5(4), 116–123.
- Sadoski, M. (1999). Comprehending comprehension. *Reading Research Quarterly*, 34(4), 493–500.
- Sanders, T. J. M., & Gernsbacher, M. A. (2004). Accessibility in text and discourse processing. *Discourse Processes*, 37(2), 79–89.
- Sanford, A. J., & Graesser, A. C. (2006). Shallow processing and underspecification. *Discourse Processes*, 42(2), 99–108.
- Sanford, A. J., Leuthold, H., Bohan, J., & Sanford, A. J. S. (2011). Anomalies at the borderline of awareness: An ERP study. *Journal of Cognitive Neuroscience*, 23(3), 514–523.
- Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, 6(9), 382.
- Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of Personality Assessment*, 63(3), 506–516.
- Schaffer, D. (1995). Shocking secrets revealed! The language of tabloid headlines. *ETC: A Review of General Semantics*, 52(1), 27–46.
- Scheufele, D. A. (1999). Framing as a theory of media effects. *Journal of Communication*, 49(1), 103–122.
- Scheufele, D. A. (2000). Agenda-setting, priming, and framing revisited: Another look at cognitive effects of political communication. *Mass Communication & Society*, 3(2–3), 297–316.
- Scheufele, D. A., & Tewksbury, D. (2007). Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of Communication*, 57(1), 9–20.
- Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010). A closer look at split visual attention in system- and self-paced instruction in multimedia learning. *Learning and Instruction*, 20(2), 100–110.
- Schotter, E. R., Bicknell, K., Howard, I., Levy, R., & Rayner, K. (2014). Task effects reveal cognitive flexibility responding to frequency and predictability: Evidence from eye movements in reading and proofreading. *Cognition*, 131(1), 1–27.

- Schotter, E. R., Tran, R., & Rayner, K. (2014). Don't believe what you read (only once): Comprehension is supported by regressions during reading. *Psychological Science*, 25(6), 1218–1226.
- Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, 29(1), 86–102.
- Schutte, N. S., & Malouff, J. M. (2004). University Student Reading Preferences in Relation To the Big Five Personality Dimensions. *Reading Psychology*, 25(4), 273–295.
- Schütze, C. T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Berlin: Language Science Press.
- Schütze, C. T., & Sprouse, J. (2013). Judgment tasks. In R. J. Podesva & D. Sharma (Eds.), *Research methods in linguistics* (pp. 27–50). New York: Cambridge University Press.
- Sheafer, T. (2007). How to evaluate it: The role of story-evaluative tone in agenda setting and priming. *Journal of Communication*, 57(1), 21–39.
- Shepard, D. S., & Finison, L. J. (1983). Blood-pressure reductions: Correcting for regression to the mean. *Preventive Medicine*, 12(2), 304–317.
- Silverman, M. E., & Mack, A. (2006). Change blindness and priming: When it does and does not occur. *Consciousness and Cognition*, 15(2), 409–422.
- Simola, J., Salojärvi, J., & Kojo, I. (2008). Using hidden Markov model to uncover processing states from eye movements in information search tasks. *Cognitive Systems Research*, 9(4), 237–251.
- Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattention blindness for dynamic events. *Perception*, 28(9), 1059–1074.
- Slattery, T. J., & Rayner, K. (2013). Effects of intraword and interword spacing on eye movements during reading: Exploring the optimal use of space in a line of text. *Attention, Perception & Psychophysics*, 75(6), 1275–1292.
- Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, 69(2), 104–120.
- Slattery, T. J., Yates, M., & Angele, B. (2016). Interword and interletter spacing effects during reading revisited: Interactions with word and font characteristics. *Journal of Experimental Psychology: Applied*, 22(4), 406–422.
- Slioussar, N. (2010). Processing of a free word order language: The role of syntax and context. *Journal of Psycholinguistic Research*, 40(4), 291–306.
- Smith, E. E., & Swinney, D. A. (1992). The role of schemas in reading text: A real-time examination. *Discourse Processes*, 15(3), 303–316.
- Smith, T. J., Lamont, P., & Henderson, J. M. (2012). The penny drops: Change blindness at fixation. *Perception*, 41(4), 489–492.
- Soames, S. (1979). A projection problem for speaker presuppositions. *Linguistic Inquiry*, 10, 623–666.
- Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *Journal of Neuroscience*, 32(25), 8443–8453.

- Solstad, T., & Bott, O. (2013). Towards a formal theory of explanatory biases in discourse. In M. Aloni, M. Franke, & F. Roelofsen (Eds.), *Proceedings of the 19th Amsterdam Colloquium* (pp. 203–210).
- Spector, B. (2013). Maxims of conversation. In *Oxford bibliographies in linguistics*.
- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind and Language*, 25(4), 359–393.
- Sperber, D., & Noveck, I. (2004). Introduction. In I. Noveck & D. Sperber (Eds.), *Experimental pragmatics*. Palgrave Macmillan.
- Sperber, D., & Wilson, D. (1995a). *Relevance: Communication and cognition* (2nd ed.). Oxford: Blackwell.
- Sperber, D., & Wilson, D. (1995b). *Relevance: Communication and cognition* (2nd ed.). Oxford: Blackwell.
- Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(6), 1521–1543.
- Sprouse, J. (2011). A test of cognitive assumptions of magnitude estimation: Commutativity does not hold for acceptability judgments. *Language*, 87(2), 274–288.
- Sprouse, J., & Almeida, D. (2011). *Power in acceptability judgment experiments and the reliability of data in syntax*. University of California, Irvine & Michigan State University.
- Sprouse, J., Schütze, C. T., & Almeida, D. (2013). A comparison of informal and formal acceptability judgments using a random sample from Linguistic Inquiry 2001–2010. *Lingua*, 134, 219–248.
- Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: A fresh view of inhibition in the retina. In *Proceedings of the Royal Society of London B: Biological Sciences* (Vol. 216, pp. 427–459).
- Stalnacker, R. C. (1973). Presuppositions. *Journal of Philosophical Logic*, 2, 447–457.
- Stanovich, K. E., West, R. F., & Toplak, M. E. (2013). Myside bias, rational thinking, and intelligence. *Current Directions in Psychological Science*, 22(4), 259–264.
- Stewart-Williams, S., & Podd, J. (2004). The placebo effect: Dissolving the expectancy versus conditioning debate. *Psychological Bulletin*, 130(2), 324–340.
- Stewart, A. J., Pickering, M. J., & Sanford, A. J. (2000). The time course of the influence of implicit causality information: Focusing versus integration accounts. *Journal of Memory and Language*, 42(3), 423–443.
- Stewart, K. J. (2003). Trust transfer on the World Wide Web. *Organization Science*, 14(1), 5–17.
- Stone, J., & Cooper, J. (2001). A self-standards model of cognitive dissonance. *Journal of Experimental Social Psychology*, 37(3), 228–243.
- Strack, F., Martin, L. L., & Schwarz, N. (1988). Priming and communication: Social determinants of information use in judgments of life satisfaction. *European Journal of Social Psychology*, 18(5), 429–442.
- Stradling, B. (1985). Controversial issues in the curriculum. *Bulletin of Environmental Education*, 170, 9–13.

- Strukelj, A., Brännström, K. J., Holmberg, N., Mossberg, F., & Holmqvist, K. (2016). The impact of sound presentations on executive control: Evidence from eye movements. *Psychology of Music*, 44(5), 996–1011.
- Strukelj, A., Foulsham, T., & Nyström, M. (2016). Social context modulates basic properties of oculomotor control. *Journal of Eye Movement Research*, 9(2), 1–9.
- Strukelj, A., & Niehorster, D. C. (2018). One page of text: Eye movements during regular and thorough reading, skimming, and spell checking. *Journal of Eye Movement Research*, 11(1), 1–22.
- Strukelj, A., Scheiter, K., Nyström, M., & Holmqvist, K. (2016). Exploring the lack of a disfluency effect: Evidence from eye movements. *Metacognition and Learning*, 11(1), 71–88.
- Sturt, P., Sanford, A. J., Stewart, A., & Dawydiak, E. (2004). Linguistic focus and good-enough representations: An application of the change-detection paradigm. *Psychonomic Bulletin & Review*, 11(5), 882–888.
- Su, Q., Huang, C.-R., & Chen, H. K. (2010). Evidentiality for text trustworthiness detection. *Proceedings of the 2010 Workshop on NLP and Linguistics: Finding the Common Ground*, (July), 10–17.
- Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3), 755–769.
- Tabor, W., & Hutchins, S. (2004). Evidence for self-organized sentence processing: Digging-in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2), 431–450.
- Thompson, E. R. (2008). Development and validation of an international English Big-Five mini-markers. *Personality and Individual Differences*, 45(6), 542–548.
- Tinker, M. A. (1927). Eye movement duration, pause duration, and reading time. *Psychological Review*, 35(5), 385–397.
- Tomasello, M. (2000). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics*, 11(1–2), 61–82.
- Trauzettel-Klosinski, S., & Dietz, K. (2012). Standardized assessment of reading performance: The new international reading speed texts IReST. *Investigative Ophthalmology and Visual Science*, 53(9), 5452–5461.
- Tribushinina, E. (2013). The acquisition of spatial adjectives: Semantic features, prefabs, and prototypes. In C. Paradis, J. Hudson, & U. Magnusson (Eds.), *The construal of spatial meaning: Windows into conceptual space*. Oxford: Oxford University Press.
- Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(3), 528–553.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453–458.
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37(3), 498–505.

- Van Berkum, J. J. A. (2010). The brain is a prediction machine that cares about good and bad - Any implications for neuropsychology? *Italian Journal of Linguistics*, 22(1), 181–208.
- Van Berkum, J. J. A., Brown, C. M., Zwitterlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3), 443–467.
- van de Weijer, J., Paradis, C., Willners, C., & Lindgren, M. (2012). As lexical as it gets: The role of co-occurrence of antonyms in a visual lexical decision experiment. In D. Divjak & S. T. Gries (Eds.), *Frequency effects in language: Linguistic representations*. Berlin: Mouton de Gruyter.
- van de Weijer, J., Paradis, C., Willners, C., & Lindgren, M. (2014). Antonym canonicity: Temporal and contextual manipulations. *Brain and Language*, 128(1), 1–8.
- 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.
- Van Gorp, B. (2007). The constructionist approach to framing: Bringing culture back in. *Journal of Communication*, 57, 60–78.
- 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.
- van Herten, M., Kolk, H. H. J., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22(2), 241–255.
- Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, 83(2), 176–190.
- van Strien, J. L. H., Kammerer, Y., Brand-Gruwel, S., & Boshuizen, H. P. A. (2016). How attitude strength biases information processing and evaluation on the web. *Computers in Human Behavior*, 60, 245–252.
- Vasishth, S., Suckow, K., Lewis, R. L., & Kern, S. (2010). Short-term forgetting in sentence comprehension: Crosslinguistic evidence from verb-final structures. *Language and Cognitive Processes*, 25(4), 533–567.
- Vitu, F. (2011). On the role of visual and oculomotor processes in reading. In S. P. Livensedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 731–749). Oxford: Oxford University Press.
- Võ, M. L.-H., Aizenman, A. M., & Wolfe, J. M. (2016). You think you know where you looked? You better look again. *Journal of Experimental Psychology: Human Perception and Performance*, 42(10), 1477–1481.
- Võ, M. L.-H., Smith, T. J., Mital, P. K., & Henderson, J. M. (2012). Do the eyes really have it? Dynamic allocation of attention when viewing moving faces. *Journal of Vision*, 12(13), 3.
- Wallin, A., Paradis, C., & Katsikopoulos, K. V. (2016). Evaluative polarity words in risky choice framing. *Journal of Pragmatics*, 106, 20–38.
- Weerkamp, W., & Rijke, M. (2008). Credibility improves topical blog post retrieval. In *Proceedings of ACL-08: HLT* (pp. 923–931). Association for Computational Linguistics.

- White, S. J., Warrington, K. L., McGowan, V. A., & Paterson, K. B. (2015). Eye movements during reading and topic scanning: Effects of word frequency. *Journal of Experimental Psychology: Human Perception and Performance*, 41(1), 233–248.
- Whitney, C. W., & Von Korff, M. (1992). Regression to the mean in treated versus untreated chronic pain. *Pain*, 50(3), 281–285.
- Wiley, J., & Rayner, K. (2000). Effects of titles on the processing of text and lexically ambiguous words: Evidence from eye movements. *Memory & Cognition*, 28(6), 1011–1021.
- Wilkinson, S. C., Reader, W., & Payne, S. J. (2012). Adaptive browsing: Sensitivity to time pressure and task difficulty. *International Journal of Human Computer Studies*, 70(1), 14–25.
- Wilson, D., & Sperber, D. (2012). *Meaning and relevance*. Cambridge: Cambridge University Press.
- Wolfe, M. B., Tanner, S. M., & Taylor, A. R. (2013). Processing and representation of arguments in one-sided texts about disputed topics. *Discourse Processes*, 50(7), 457–497.
- Wotschack, C. (2009). *Eye movements in reading strategies: How reading strategies modulate effects of distributed processing and oculomotor control*. Universitätsverlag Potsdam, Potsdam, Germany.
- Wulfeck, B., Bates, E., & Capasso, R. (1991). A crosslinguistic study of grammaticality judgments in Broca's aphasia. *Brain and Language*, 41(2), 311–336.
- Xiang, M., & Kuperberg, G. R. (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience*, 30(6), 648–672.
- Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes*, 93, 1–13.
- Yaniv, I., & Milyavsky, M. (2007). Using advice from multiple sources to revise and improve judgments. *Organizational Behavior and Human Decision Processes*, 103(1), 104–120.
- Yudkin, P. L., & Stratton, I. M. (1996). How to deal with regression to the mean in intervention studies. *Lancet*, 347(8996), 241–243.
- Zaller, J. (1992). *The nature and origins of mass opinion*. New York: Cambridge University Press.
- Zwaan, R. A., & Rapp, D. N. (2006). Discourse Comprehension. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 725–764). London, UK: Elsevier.

Appendices

Appendix 3.A Example stimuli text in English (including line breaks)

One of the many things that threaten animals and plants alike is the danger of being eaten. Some animals take recourse in camouflaging themselves, while others hide away. A number of them can fly away and others are so fast that they can flee from their enemies. Plants, which as is well known cannot run away, have developed other means to protect themselves. These include sharp spines and thorns or a tough skin. Other plants and also a number of animals protect themselves by poisons that they contain. These poisons need not necessarily be deadly. It is sufficient if they prevent other animals from eating the poisonous beings. Some animals even protect themselves by looking like others that contain poisonous substances. In the animal kingdom, a particularly striking color generally signals that the animal is inedible.

Appendix 3.B Remaining full statistical model outputs

All models except comprehension scores and total number of visits on words used log-transformed data to better fit a Gaussian distribution. In all models, p -values were generated using Satterthwaite approximation.

Table 3.B1. Comprehension scores depending on instruction

Statistically significant results in **bold**; * signifies $p < .05$, *** signifies $p < .001$.

	Estimate	SE	df	t-value	p-value	
Thorough reading	0.3488	0.1348	14.031	2.587	.021485	*
Skimming	-0.6873	0.1653	13.780	-4.157	.000999	***
Spell checking	-0.7191	0.1524	13.820	-4.720	.000340	***

Table 3.B2. Average fixation durations and saccade amplitudes depending on instruction

Statistically significant results in **bold**; * signifies $p < .05$, ** signifies $p < .01$, *** signifies $p < .001$.

	Estimate	SE	df	t-value	p-value	
<i>Average fixation durations</i>						
Thorough reading	-0.00513	0.010100	14.66	-0.508	.6188	
Skimming	-0.03125	0.011409	19.85	-2.740	.0127	*
Spell checking	0.08360	0.014926	51.62	5.602	.0000	***
<i>Saccade amplitudes</i>						
Thorough reading	0.05897	0.01503	87500	3.923	.0000	***
Skimming	0.08987	0.01825	86420	4.918	.0000	***
Spell checking	-0.16730	0.01535	84640	-10.902	.0000	***

Table 3.B3. Proportions of leftward and vertical saccades depending on instruction

Both models are binomial generalized linear mixed models. Statistically significant results in **bold**; * signifies $p < .05$, *** signifies $p < .001$.

	Estimate	SE	z-value	Pr(> z)	
<i>Leftward saccades</i>					
Thorough reading	0.02522	0.02117	1.19	.2335	
Skimming	-0.12283	0.02627	-4.68	.0000	***
Spell checking	0.04935	0.02147	2.30	.0215	*
<i>Vertical saccades</i>					
Thorough reading	0.06275	0.02569	2.44	.0146	*
Skimming	-0.17570	0.03264	-5.38	.0000	***
Spell checking	0.14861	0.02574	5.77	.0000	***

Table 3.B4. First fixation durations on words depending on instruction as a function of word frequency, when excluding skipped words

Statistically significant results in **bold**; *** signifies $p < .001$. A colon denotes an interaction.

	Estimate	SE	df	t-value	p-value	
Word frequency	0.01425	0.003052	36260	4.667	.0000	***
Thorough reading	0.02194	0.01538	36260	1.427	.1537	
Skimming	-0.02336	0.01687	36260	-1.385	.1661	
Spell checking	0.07480	0.01545	36260	4.842	.0000	***
Word frequency: Thorough reading	-0.00804	0.004239	36250	-1.897	.0578	.
Word frequency: Skimming	-0.00136	0.004575	36250	-0.298	.7658	
Word frequency: Spell checking	0.00337	0.004243	36250	0.795	.4265	

Table 3.B5. First-pass reading times on words depending on instruction as a function of word frequency, when excluding skipped words

Statistically significant results in **bold**; ** signifies $p < .01$, *** signifies $p < .001$. A colon denotes an interaction.

	Estimate	SE	df	t-value	p-value	
Word frequency	0.03998	0.003423	36250	11.681	.00000	***
Thorough reading	0.02791	0.01724	36260	1.619	.10551	
Skimming	-0.03343	0.01892	36250	-1.768	.07714	.
Spell checking	0.05643	0.01732	36260	3.258	.00112	**
Word frequency: Thorough reading	-0.00849	0.004753	36250	-1.787	.07399	.
Word frequency: Skimming	-0.00490	0.005131	36250	-0.955	.33939	
Word frequency: Spell checking	0.02386	0.004758	36250	5.016	.00000	***

Table 3.B6. Total reading times on words depending on instruction as a function of word frequency, when excluding skipped words

Statistically significant results in **bold**; * signifies $p < .05$, ** signifies $p < .01$, *** signifies $p < .001$. A colon denotes an interaction.

	Estimate	SE	df	t-value	p-value	
Word frequency	0.10090	0.004657	3.626e+04	21.662	.00000	***
Thorough reading	0.06797	0.02346	3.626e+04	2.897	.00376	**
Skimming	-0.18050	0.02574	3.626e+04	-7.013	.00000	***
Spell checking	0.05510	0.02357	3.626e+04	2.338	.01940	*
Word frequency: Thorough reading	0.01859	0.006466	3.625e+04	2.875	.00404	**
Word frequency: Skimming	-0.02970	0.006979	3.625e+04	-4.256	.00002	***
Word frequency: Spell checking	0.04359	0.006472	3.625e+04	6.734	.00000	***

Table 3.B7. First fixation durations on words depending on instruction as a function of word length, when excluding skipped words

Statistically significant results in **bold**; *** signifies $p < .001$. A colon denotes an interaction.

	Estimate	SE	df	t-value	p-value	
Word length	0.00850	0.001980	36260	4.294	.0000	***
Thorough reading	0.01840	0.01552	36260	1.185	.2359	
Skimming	-0.01489	0.01684	36260	-0.885	.3764	
Spell checking	0.06101	0.01556	36260	3.920	.0000	***
Word length: Thorough reading	-0.00447	0.002755	36250	-1.621	.1050	
Word length: Skimming	-0.00250	0.002920	36250	-0.856	.3920	
Word length: Spell checking	0.00490	0.002754	36250	1.779	.0752	.

Table 3.B8. First-pass reading times on words depending on instruction as a function of word length, when excluding skipped words

Statistically significant results in **bold**; *** signifies $p < .001$. A colon denotes an interaction.

	Estimate	SE	df	t-value	p-value	
Word length	0.02996	0.002209	36260	13.560	.0000	***
Thorough reading	0.02747	0.01732	36260	1.585	.1129	
Skimming	-0.02999	0.01879	36260	-1.596	.1105	
Spell checking	0.01890	0.01737	36260	1.088	.2765	
Word length: Thorough reading	-0.00527	0.003074	36250	-1.713	.0866	.
Word length: Skimming	-0.00411	0.003259	36250	-1.260	.2078	
Word length: Spell checking	0.02285	0.003074	36250	7.435	.0000	***

Table 3.B9. Total reading times on words depending on instruction as a function of word length, when excluding skipped words

Statistically significant results in **bold**; *** signifies $p < .001$. A colon denotes an interaction.

	Estimate	SE	df	t-value	p-value	
Word length	0.07810	0.002953	36040	26.442	.000000	***
Thorough reading	0.03592	0.04160	22	0.863	.397316	
Skimming	-0.14440	0.03975	58	-3.632	.000595	***
Spell checking	-0.01553	0.04884	46	-0.318	.751875	
Word length: Thorough reading	0.01655	0.004104	36050	4.033	.000055	***
Word length: Skimming	-0.02413	0.004356	35920	-5.540	.000000	***
Word length: Spell checking	0.03965	0.004111	36030	9.646	.000000	***

Table 3.B10. Number of visits on words depending on instruction as a function of word length

Statistically significant results in **bold**; * signifies $p < .05$, ** signifies $p < .01$, *** signifies $p < .001$. A colon denotes an interaction.

	Estimate	SE	df	t-value	p-value	
Word length	0.17110	0.004014	67720	42.624	.000000	***
Thorough reading	0.01668	0.07927	18	0.210	.835740	
Skimming	-0.16550	0.07017	30	-2.358	.025059	*
Spell checking	0.02988	0.09278	34	0.322	.749416	
Word length: Thorough reading	0.05282	0.005690	67720	9.283	.000000	***
Word length: Skimming	-0.05734	0.005670	67720	-10.113	.000000	***
Word length: Spell checking	0.03095	0.005683	67720	5.446	.000000	***

Appendix 4.A Results from the joint analysis using the alternative transformation method

In this section, the term pre-reading is used for values before reading, and the term post-reading is used for values after reading. The alternative transformation method is first described, results are then shown, and the full statistical model is then presented. In the statistical model, *p*-values were generated using Satterthwaite approximation.

$$\text{Agreement score} - \frac{\text{Mean prereading agreement score}}{\left(\frac{\text{Mean prereading agreement score}}{\text{Mean postreading agreement score for noncontroversial texts}} \right)}$$

Figure 4.A1. Mathematical formula for alternative transformation of agreement scores after reading (post-reading agreement scores), to remove regressions to the mean

Agreement score in **bold** signifies every individual agreement score, all other values are constant. Pre-reading and post-reading are referred to as prereading and postreading, in order to avoid confusion with minus signs.

$$\text{Agreement score} - \frac{\text{Mean postreading agreement score for noncontroversial texts}}{\text{Mean postreading agreement score for noncontroversial texts}}$$

Figure 4.A2. The mathematically reduced transformation described in Figure 4.A1

Agreement score in **bold** signifies every individual agreement score, all other values are constant. Post-reading is referred to as postreading, in order to avoid confusion with a minus sign.

Table 4.A1. Mean retention of agreement scores after reading depending on topic and headline, when grouping trials depending on positive or negative agreement score before reading, using alternative transformation for agreement scores, Experiment 1 and 2 combined

Standard deviations in parentheses. The alternative transformation of agreement scores can be found in Figure 4.A1 and Figure 4.A2.

Topic, Headline	Group with positive agreement scores before reading	Group with negative agreement scores before reading
Controversial topic	0.0444 (0.385)	-0.0257 (0.398)
Positive headline	0.0247 (0.409)	-0.0044 (0.337)
Negative headline	0.0133 (0.369)	0.0042 (0.386)

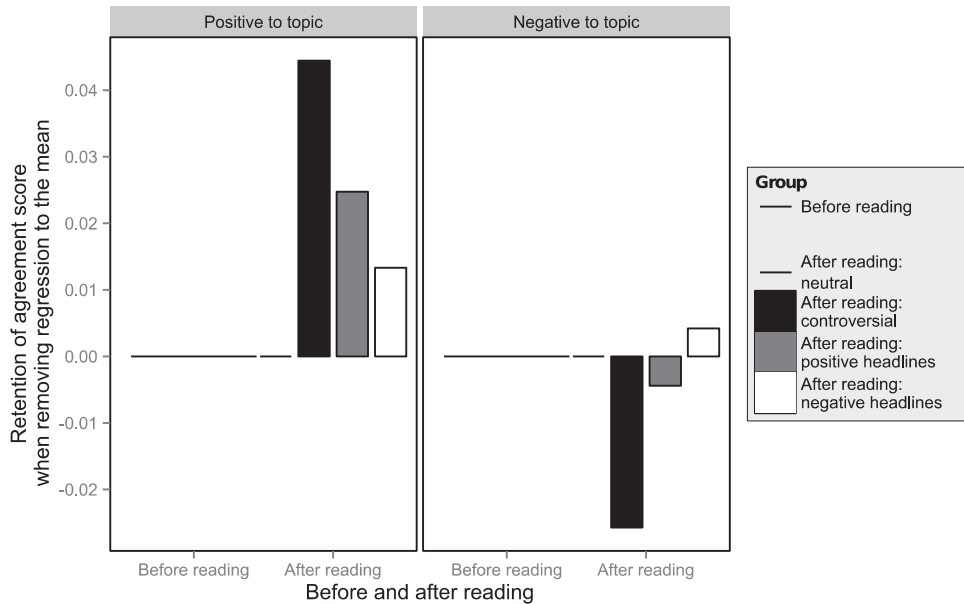


Figure 4.A3. Mean retention of agreement scores after reading depending on topic and headline, depending on trials with positive or negative agreement score before reading, using alternative transformation for agreement scores, Experiment 1 and 2 combined

The agreement scores after reading the neutral texts in Experiment 1 were used as the baseline, using the alternative transformation method in Figure 4.A2, which results in the zero values in the figure. A larger value in the figure signify a higher retention of agreement scores.

Table 4.A2. Full statistical model output: Retention of agreement scores after reading depending on topic and headline, when grouping trials depending on positive or negative agreement score before reading, using alternative transformation for agreement scores, Experiment 1 and 2 combined

Statistically significant results in **bold**; * signifies $p < .05$. A colon denotes an interaction. The alternative transformation of agreement scores can be found in Figure 4.A1 and Figure 4.A2.

	Estimate	SE	df	t-value	p-value	
<i>Experiment 1</i>						
Noncontroversial topic	-0.00139	0.02465	2986	-0.056	.9551	
Controversial topic	0.04553	0.02390	2981	1.905	.0569	.
<i>Experiment 2</i>						
Positive headline	0.02512	0.02676	2940	0.939	.3478	
Negative headline	0.01316	0.02848	2967	0.462	.6440	
<i>Experiments 1 and 2</i>						
Negative to topic	-0.00176	0.01750	3004	-0.101	.9199	
<i>Experiments 1 and 2</i>						
Noncontroversial topic: Negative to topic	0.00046	0.03695	2992	0.012	.9900	
Controversial topic: Negative to topic	-0.07402	0.03690	2993	-2.006	.0449	*
Positive headline: Negative to topic	-0.02732	0.04247	2971	-0.643	.5201	
Negative headline: Negative to topic	-0.00618	0.04154	2969	-0.149	.8818	

Appendix 4.B Full statistical model output

In the statistical model, p -values were generated using Satterthwaite approximation.

Table 4.B1. Retention of agreement scores after reading depending on topic and headline, when grouping trials depending on positive or negative agreement score before reading, Experiment 1 and 2 combined
Statistically significant results in **bold**; * signifies $p < .05$. A colon denotes an interaction.

	Estimate	SE	df	t-value	p-value	
<i>Experiment 1</i>						
Noncontroversial topic	-0.00137	0.03564	2981	-0.038	.9694	
Controversial topic	0.07408	0.03456	2974	2.143	.0322	*
<i>Experiment 2</i>						
Positive headline	0.04012	0.03868	2933	1.037	.2997	
Negative headline	0.02108	0.04118	2962	0.512	.6087	
<i>Experiments 1 and 2</i>						
Negative to topic	-0.00221	0.02533	3007	-0.087	.9305	
<i>Experiments 1 and 2</i>						
Noncontroversial topic: Negative to topic	0.00047	0.05350	2997	0.009	.9929	
Controversial topic: Negative to topic	-0.11920	0.05343	2997	-2.230	.0258	*
Positive headline: Negative to topic	-0.04500	0.06152	2974	-0.732	.4645	
Negative headline: Negative to topic	-0.01137	0.06017	2973	-0.189	.8502	

Appendix 6.A Stimuli texts used in Experiment 1

Table 6.A1 shows the stimuli texts used in Experiment 1. In the critical texts, the antonym pair is indicated in *italics*, with the incongruent word also in **bold**. No words were displayed in bold or italics during the experiment, and only one version was shown to each participant. The filler texts were given to all participants. Correct answers to the comprehension questions in *italics*.

Table 6.A1. Stimuli texts, Experiment 1

Total number of critical texts: 12. Total number of filler texts: 4. Number of words: $M=53.9$ ($SD=9.62$).

Text	Word count	Comprehension question
Pharmaceutical reforms save billions of crowns for Swedes The Swedish Pharmaceutical Benefits Council claims that Swedish consumers have saved close to seven billion Swedish crowns since pharmacies began consistently selling the <i>cheap/expensive</i> equivalent of brand name medications.	37	The Swedish consumers have saved 7 billion by not buying branded medicine. (<i>True</i>)
Lars Frölander – World Champion swimmer Lars Frölander recently won two gold medals in two days, breaking three world records in the process. Lars might thus be considered Sweden's <i>best/worst</i> athlete. He is highly satisfied with his accomplishments during the world championships in short-course swimming.	44	Lars Frölander broke two world records in swimming during the world championships. (<i>False</i>)
Large, mature elk are on the decline in Östergötland The County Board in Östergötland is concerned about a decline of mature animals in the local elk population. The board has instructed hunters to target moose calves instead of <i>adult/young</i> animals to help remedy the situation.	45	The elk population in Östergötland is currently very good. (<i>False</i>)
U.S. interest rates are on the rise The U.S. Federal Reserve recently increased the interest rate by a quarter of a percentage point to 4.75 percent. It is the <i>highest/lowest</i> level in five years. The new central bank chief said that further increases are possible in the future.	48	The U.S. interest rate is very stable at the moment. (<i>False</i>)
Petrol stations are getting rid of cash payment machines Many gas stations are removing the option to pay with cash at petrol kiosks. Cash requires additional handling and it makes the stations more vulnerable to robberies. Thus the petrol companies will <i>save/lose</i> money by removing the cash machines entirely.	49	It is getting harder to pay with cash at petrol stations nowadays. (<i>True</i>)
More and more banks joining the insurance industry The competition on the insurance market has grown recently. The reason for this is that banks have recently entered the insurance market, and have <i>started/stopped</i> selling insurance. Nordea is one recent example, who have joined forces with the Danish insurance company TrygVesta.	50	Insurance is now sold by both insurance companies and banks. (<i>True</i>)
Many Dutch people are moving to Karlshamn The Swedish city of Karlshamn recently participated in an emigration fair in Holland. So far 20 Dutch families have expressed interest in moving to the beautiful and affordable city. The families liked the city's <i>low/high</i> housing prices as well as its closeness to nature.	51	Being close to nature is one of the factors why Dutch families move to Karlshamn. (<i>True</i>)

Consumer confidence hits year high Swedish consumer confidence hit its highest level in a year last February. The GFK index of household confidence showed an increase to 6.0 points, from 5.1 points in October 2011. This <i>rise/fall</i> in February is thought to have been driven by rising income expectations, despite the euro-zone debt crisis.	54	The index for Swedish consumption is unusually low at the moment. (False)
Flooding hits Southern Sweden Many basements and roads in the south of Sweden have been affected by flooding. The cause of the large quantities of water is that the snow has melted rapidly on account of the recent <i>warm/cold</i> weather. Because of the situation on the roads, the traffic information center for the Swedish Road Administration has been very busy lately.	61	The large amounts of water in Southern Sweden is caused by rain. (False)
Many African Men Fail to Get HIV Treatment A new study states that men in sub-Saharan Africa are not accessing HIV/AIDS treatment nearly as often as women. In other words, many men are dying prematurely. Researchers are calling for a more <i>balanced/imbalanced</i> approach to gender in fighting the epidemic, so that men and women would have the same opportunities to receive treatment.	62	Many men are dying prematurely in Sub-Saharan Africa. (True)
The number of professors has doubled in Sweden According to new study from Statistics Sweden, the number of professors in Sweden almost doubled during the last ten years. Converted to full-time positions, this means that in October last year there were over 3900 professors. Statistics Sweden believes that the <i>increase/decrease</i> stems from recent reforms that enabled academic lecturers to be promoted to professors.	63	The number of professors in Sweden has been constant over the past ten years. (False)
Prosecutor wants to deport convicted shoplifter The Supreme Court will consider whether a man jailed for theft should be deported as well. The prosecutor firmly believes that the man should be deported, as the man has shown indifference when it comes to Swedish law and because there is a <i>high/low</i> risk of repeat offense. The lower courts have refused to deport him because he is unpunished in the past.	69	The prosecutor claims that the man has shown indifference to Swedish law. (True)
Södra postpones storm compensation The forest group Södra's payment to the storm-affected forest owners will not occur this June as promised. Several owners have received letters where Södra states that payment is postponed. The storm Gudrun affected most of southern Sweden.	41 Filler text	Forest owners affected by the storm will be completely without compensation. (False)
Medicines to treat severe MS approved A new drug for difficult cases of MS has been approved by the European Medicines Agency. The drug, Tysabri, is for patients who are not helped by their current medication, and whose MS is rapidly worsening. The drug is given intravenously in the hospital and will be available in Swedish hospitals in late July.	60 Filler text	The drug Tysabri is meant for people with severe cases of MS. (True)
The moon causes the solar eclipse On Wednesday, a solar eclipse will occur. Shortly after 13.00, the eclipse will be at its most extreme point. This will be a partial solar eclipse, which will not be more noticeable than the falling of twilight. About one third of the solar disc will be obscured when the moon passes between Earth and the Sun.	62 Filler text	The eclipse will occur late in the afternoon. (False)
The Pentagon is investigating the media image of the Iraq War The U.S. Department of Defense has set up a media group to counter news reports that the Pentagon considers faulty. Secretary of Defense Donald Rumsfeld has previously said that the administration is in a propaganda war with the media. Vice President Dick Cheney admitted that there has been a recent increase in bombings in Iraq.	66 Filler text	The U.S. Department of Defense will work to make news coverage of the Iraq war more accurate. (True)

Appendix 6.B Stimuli texts used in Experiment 2

Table 6.B1 shows the stimuli texts used in Experiment 2. In the critical texts, the antonym pair is indicated in *italics*, with the incongruent word also in **bold**. No words were displayed in bold or italics during the experiment, and only one version was shown to each participant. The filler texts were given to all participants. Correct answers to the comprehension questions in *italics*.

Table 6.B1. Stimuli texts, Experiment 2

Total number of critical texts: 12. Total number of filler texts: 4. Number of words: $M=62.7$ ($SD=5.39$).

Text	Word count	Comprehension question
Pharmaceutical reform saves money for Swedish consumers The Swedish Pharmaceutical Benefits Council claims that Swedish consumers have saved close to seven billion Swedish crowns since the introduction of a new pharmaceutical reform. The reform instructed pharmacies to consistently sell the <i>cheap/expensive</i> equivalent of brand name medications. The Council expects no negative effects from the reform.	55	The Swedish consumers have saved 7 billion by not buying branded medicine. (<i>True</i>)
Lars Frölander – World Champion swimmer Lars Frölander recently won two gold medals in two days, breaking three world records in the process. He is highly satisfied with his accomplishments during the world championships in short-course swimming. Lars might therefore be considered one of the <i>best/worst</i> athletes in Sweden. He is currently in training to continue improving his performance.	58	Lars Frölander broke two world records in swimming during the world championships. (<i>False</i>)
Large, mature elk are on the decline The County Board in the north of Sweden is concerned about a decline of mature animals in the local elk population. The board has therefore instructed hunters to exclusively target calves. This is a change from previous policies, where <i>old/young</i> animals were targeted. In this way, the situation is expected to be greatly alleviated.	60	The elk population in the north of Sweden is currently very good. (<i>False</i>)
U.S. interest rates are on the rise The U.S. Federal Reserve recently elevated the interest rate by a quarter of a percentage point to 4.75 percent. The new central bank chief said that further increases are possible in the future. It is the <i>highest/lowest</i> level in five years. The interest rates are expected to affect the stock market negatively in the coming months.	63	The U.S. interest rate is very stable at the moment. (<i>False</i>)
Petrol stations profit by getting rid of cash payment machines Many gas stations are removing the option to pay with cash at petrol kiosks. Cash requires additional handling and it makes the stations more vulnerable to robberies. The petrol companies are expected to <i>save/lose</i> money by removing the cash machines entirely. The companies plan to implement this during the next few years.	62	It is getting harder to pay with cash at petrol stations nowadays. (<i>True</i>)
More and more banks joining the insurance industry The competition on the insurance market has grown recently, as a result of banks entering the insurance market. Among several changes in banking during recent years, the fact that banks have <i>started/stopped</i> selling insurance could be seen as the biggest one. Nordea recently joined forces with the Danish insurance company TrygVesta.	59	Insurance is now sold by both insurance companies and banks. (<i>True</i>)
Dutch people moving to Karlshamn in order to save money The Swedish city of Karlshamn recently participated in an emigration fair in Holland. So far 20 Dutch families have expressed an interest in moving to the beautiful city. The families appreciated the city's <i>low/high</i> overall housing prices. However, many saw the closeness to nature as the most important reason to move to Sweden.	63	Being close to nature is one of the factors why Dutch families move to Karlshamn. (<i>True</i>)

Consumer confidence hits year high Swedish consumer confidence hit its highest level in a year last February, according to a survey performed by Statistics Sweden. The GFK index of household confidence showed an increase to 6.0 points, and despite the euro-zone debt crisis, the income expectations are positive. The <i>rise/fall</i> in February is thought to have been driven by these expectations. The level of consumer confidence is expected to stay constant during the coming year.	75	The index for Swedish consumption is unusually low at the moment. (False)
The flooding in Southern Sweden causes problems Many basements and roads in the south of Sweden have been affected by flooding. The cause of the large quantities of water is that the snow has melted rapidly due to the recent mild weather. This has created a <i>bad/good</i> situation on the roads. As a result, the traffic information center for the Swedish Road Administration has been very busy lately.	68	The large amounts of water in Southern Sweden is caused by rain. (False)
Many African Men Fail to Get HIV Treatment A new study states that men in sub-Saharan Africa do not receive HIV/AIDS treatment nearly as often as women. Many men die prematurely, and researchers therefore call for increased male HIV testing. This would deal with the epidemic <i>more/less</i> effectively. Also, aid given in the form of HIV/AIDS treatment would further alleviate the situation.	62	Many men are dying prematurely in Sub-Saharan Africa. (True)
The number of professors has doubled in Sweden According to new study from Statistics Sweden, the number of professors in Sweden almost doubled during the past ten years. Converted to full-time positions, this means that in October last year there were over 3900 professors. Statistics Sweden believes that this <i>increase/decrease</i> stems from recent reforms. The change enables academic lecturers to be promoted to professors without applying for a position.	69	The number of professors in Sweden has been constant over the past ten years. (False)
Prosecutor wants to deport convicted shoplifter The Supreme Court will consider whether a man jailed for theft should be deported. The prosecutor firmly believes that the man should be deported as the man has shown complete indifference to Swedish law. The prosecutor also believes there is a <i>high/low</i> risk of repeat offense. The lower courts have refused to deport him because he is unpunished in the past.	67	The prosecutor claims that the man has shown indifference to Swedish law. (True)
SCA postpones storm compensation The forest group SCA's payment to the storm-affected forest owners will not occur this June as promised. Several owners have received letters where SCA states that payment is postponed. The storm Gudrun affected most of southern Sweden. As a result, lumber prices are expected to rise in the coming years.	54 Filler text	Forest owners affected by the storm will be completely without compensation. (False)
Medicines to treat severe MS approved A new drug for difficult cases of MS has been approved by the European Medicines Agency. The drug, Tysabri, is for patients who are not helped by their current medication, and whose MS is rapidly worsening. The drug is given intravenously in the hospital and will be available in Swedish hospitals in late July.	60 Filler text	The drug Tysabri is meant for people with severe cases of MS. (True)
The moon causes the solar eclipse On Wednesday, a solar eclipse will occur. Shortly after 13.00, the eclipse will be at its most extreme point. This will be a partial solar eclipse, which will not be more noticeable than the falling of twilight. About one third of the solar disc will be obscured when the moon passes between Earth and the Sun.	62 Filler text	The eclipse will occur late in the afternoon. (False)
The Pentagon is investigating the media image of the Iraq War The U.S. Department of Defense has set up a media group to counter news reports that the Pentagon considers faulty. Secretary of Defense Donald Rumsfeld has previously said that the administration is in a propaganda war with the media. Vice President Dick Cheney admitted that there has been a recent increase in bombings in Iraq.	66 Filler text	The U.S. Department of Defense will work to make news coverage of the Iraq war more accurate. (True)

Appendix 6.C Example stimuli screens

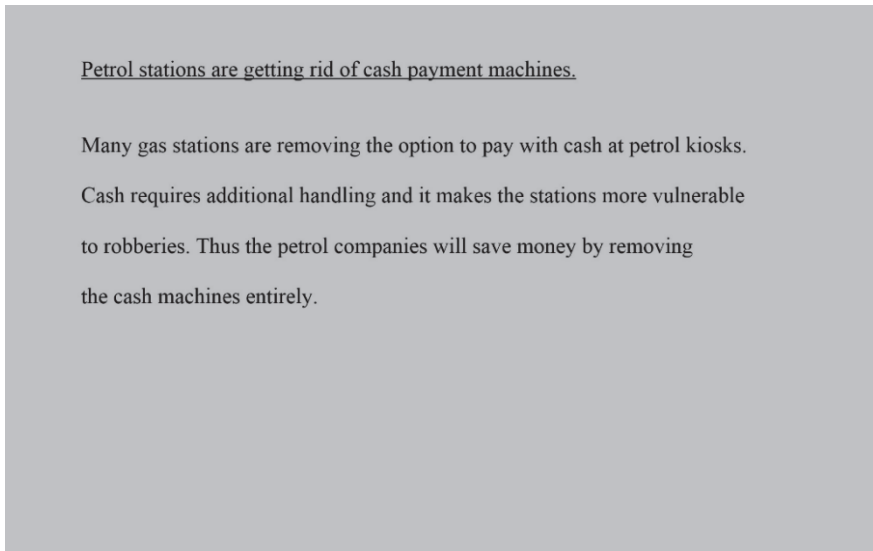


Figure 6.C1. Example stimuli (control condition), Experiment 1

The example stimulus is shown using the identical proportions of the experimental setup. The word *save* is changed into *lose* in the incongruent condition.

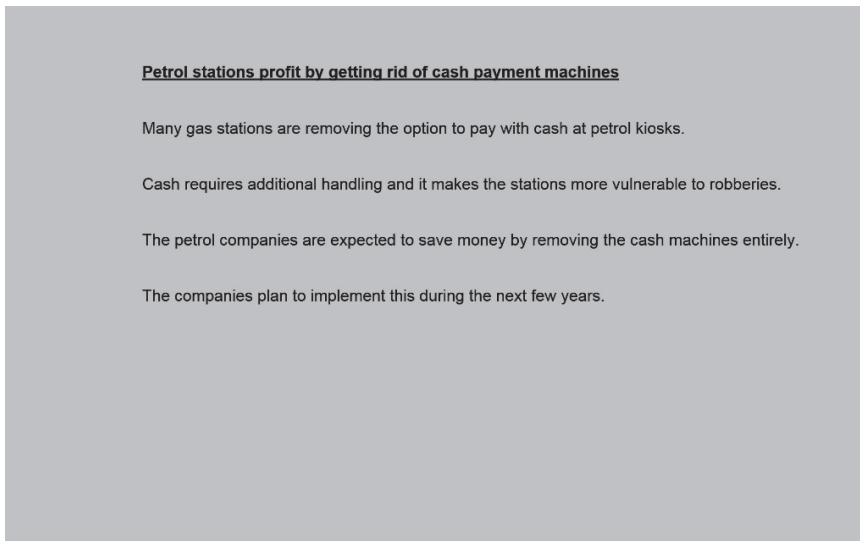


Figure 6.C2. Example stimuli (control condition), Experiment 2

The example stimulus is shown using the identical proportions of the experimental setup. The word *save* is changed into *lose* in the incongruent condition.

Appendix 7.A Stimuli sentences

In the critical sentences, the antonym pair is indicated in *italics*, with the incongruent word also in **bold**. No words were displayed in bold or italics during the experiment, and only one version was shown to each participant. The filler sentences were given to all participants. The filler sentences have the adjective used for the analyses in Chapter 8 in *italics*.

Critical sentences

One syllable:

1. As the high temperatures persisted, tourism in beach areas was greatly boosted by the *hot/cold* weather.
2. Rising population trends show that outside the cozy warmth of their homes, residents face risky and expensive winters in a *cold/hot* economic climate.
3. It has become a major consumer rights issue and is about to become a *hot/cold* political issue, too.
4. The glow from the bulb is incandescent, with a nice warm light instead of *cold/hot* fluorescent glare.
5. The candidate excelled at fund raising, and had many supporters because he was young, energetic and a *good/bad* public figure.
6. That the airline is threatening to lay off employees is making the situation for workers insecure, and is creating *bad/good* public relations as well.
7. The applicants compete against other applicants and are ultimately selected if they look like a *good/bad* fit for the position.
8. The financial advisor predicts that about 20 banks in the area will be closed next year because of *bad/good* real estate loans.
9. Immigrants have dramatically upped the pace of startups, despite the deep recession and *slow/fast* economic recovery.
10. The organization is an advocate for large and *fast/slow* growing school systems.
11. Due to dwindling animal populations, declining interest and cold weather, hunting has been *slow/fast* this year.
12. The computer maker would like more manufacturers to use its blazingly *fast/slow* chip, the RS/6000.
13. Viewership for its programs suffered greatly when the network lost eight important affiliate stations and their *large/small* local audiences.
14. The problem is that the specific airport was never meant for such traffic, as it was merely a *small/large* private air strip in the 1930s.
15. The festival had set up a loudspeaker in lower Main Street, to project the sound from the two *large/small* outdoor TV screens.
16. The band says that they gladly sleep in a van if it helps to perpetuate the stereotype of a band on a *small/large* independent label.
17. The governor expresses beliefs in free markets, low taxes and limited government, and a commitment to a *strong/weak* national defense.
18. Even with the prospect of profitability, the company is still in a relatively *weak/strong* financial position.

19. They were granted extraordinary access to financial records, in part because of a *strong/weak* public demand for greater accountability.
20. The bug is widespread in nursing homes because the elderly and those with *weak/strong* immune systems are especially vulnerable.
21. In the suburbs, first-time buyers tend to be *young/old* professional couples in their mid-30s.
22. The areas have grown sharply even as construction, traditionally considered one of the *old/young* economic mainstays, cuts back.
23. The company favors films by women or third-world film-makers, and also passes on established directors in favor of *young/old* directors.
24. With regards to this type of policy, everybody needs to be on the same page, like *a(n) old/young* married couple.
25. Observers credited the victory to the extensive preparations and the *sharp/dull* political instincts of the campaign.
26. The magazine is plagued by the slumping sales caused by the inclusion of badly written, *dull/sharp* serialized novels.
27. The incompetent handling of the case has drawn *sharp/dull* public criticism from firefighters and their supporters.
28. Not one for gimmicks, he is, by his own admission, an unemotional speaker and a *dull/sharp* public figure.
29. The critical situation makes it likely that taxpayers are getting a *hard/soft* kick in the rear end.
30. She is a smart and tough negotiator, but nevertheless, she is still *soft/hard* spoken and laid back.
31. The economic and social problems throw the administration into a *hard/soft* new reality that must be faced immediately.
32. In the film the actress displays a jaded quality, even though she is young and beautiful with an almost babyish *soft/hard* face.
33. Fearing strangers is useful sometimes, but can also build *thick/thin* walls around us.
34. The bureau re-evaluated all materials, which bolstered a previously *thin/thick* political evidence report.
35. While walking in the barren area, she protected her feet from the jagged rocks by wearing shoes with *thick/thin* rubber soles.
36. Determined to hold onto office at any cost to the nation, he was a man who consciously skated along a *thin/thick* legal edge.
37. Businesses have had a hard time obtaining commercial credit, and that has worsened in *rough/smooth* economic times, several brokers said.
38. It is an unseemly spectacle for a company that prides itself with having a *smooth/rough* corporate culture.
39. His sister died of cancer a year after his dad, which explains the *rough/smooth* emotional terrain he had to travel.
40. Many investors are interested in the substantial financial returns and *smooth/rough* working relationships.

Filler sentences

One syllable:

1. Despite the initial success, the inevitable high point was soon reached, with local market saturation of *light* industrial goods.
2. The site provides extensive and detailed event listings, plus profiles of dozens of *cool* local bands.
3. The government enjoys massive public support, in spite of *strict* monetary policies.
4. Some argue that it would be advisable to extend the concept of *free* public education to four years of college or vocational school.
5. Health care experts say the hospitals facing the biggest strain are those in urban and *poor* rural settings.
6. Those living near power lines complain about the unsightly appearance of the towers and the *loud* buzzing sound they emit.
7. When producing their art, the agency combines digital prints, plaids and patterns in *fresh* ways.
8. Research shows that it does not spell economic ruin for communities to adopt comprehensive *clean* indoor air laws.
9. What is turning out to be a veritable goldmine, online auctions do brisk business in *used* electronic equipment.
10. It is unlikely that the most ambitious arguments put forward by policy makers will gain *broad* international support.
11. The market includes only two hospital networks that can provide *wide* geographic coverage.
12. Not going into the specifics may actually have been the *smart* political move for both parties.
13. A larger-than-life governor attracts *bright* young people to his presidential campaign.
14. His large office features a working fireplace and a *long* wooden conference table with deep-burgundy leather chairs.
15. If we are to design a better finance system, getting history right for this *dark* economic period is critical.
16. By witnessing gruesome deaths and the physical injuries of war, ex-combatants still bear *deep* emotional wounds that need time to heal.
17. Uniformed guards patrol corridors under *harsh* lights that never dim, day or night.
18. There is a concern around the country about what legislators in this very *tough* economic time are going to say.
19. The new museum will hopefully bring in many tourists, and a renewed interest in the state's *rich* musical heritage.
20. The area suffers from harsh temperatures, extremely *short* growing season, and a lack of fresh water.

Two syllables:

21. The airport was placed in a flawed location to satisfy *narrow* political and economic ambitions.
22. Aviation experts say the location is the logical spot if the government decides that a *shiny* new airport should be built.
23. Muscle-flexing by national governments has also made it more complicated to tackle *global* environmental issues.
24. The manager is waiting for it, ever fearful that it will make an *awful* economic situation even worse.

25. We need to move toward solutions to seemingly intractable problems that don't easily fit the *tired* old paradigm.
26. Just days after the football season ended, owners complained about *heavy* financial losses and began looking for more sponsors.
27. The governor's concern was expressed in *recent* public speeches and private conversations.
28. This reluctance certainly reflects the *normal* human reluctance to admit problems or mistakes.
29. The prime minister stated he will try to further strengthen the relationship based on the *common* cultural bonds between his country and the United States.
30. The company scrambled to prevent a relatively *minor* technical defect from becoming a huge public relations blunder.
31. The new directors want to prevent the development of a *useful* public transit system in the city.
32. Bankers signaled that they were prepared to take action to shore up *nervous* financial markets after Sunday's vote.
33. Within the profession, graduates of *foreign* medical schools say they are facing new levels of bias.
34. Fortunately, the last thing anyone really wants to see at a fashion show are some *boring* old clothes.
35. There is open drug dealing on the streets, especially near the *ugly* concrete stadium on the south edge of downtown.
36. The proposal is likely rife with loopholes, which *clever* political operatives and election law experts can find ways around.
37. Doctors are no more or no less immune to this, and to the *basic* human drives of empathy and pity, than anyone else.
38. Few hopes for survival can be had for the new diseases, most of which were previously unheard of among *healthy* young men.
39. Aggressive competition may force the company to reduce prices, as well as *rapid* technological changes in the telecommunications industry.
40. Existing sales reached a low for the year, and prices have been sliding in most *major* metropolitan areas.

Appendix 8.A Figures for analyses without forgotten sentences

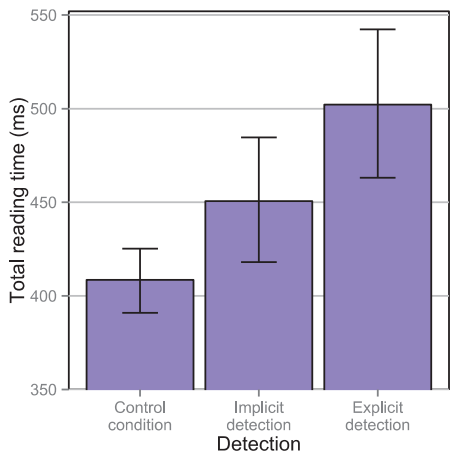


Figure 8.A1. Mean total reading times (ms) for control condition, implicitly detected incongruent words, and explicitly detected incongruent words, when removing forgotten sentences from the analysis, Experiment 1
Error bars denote standard errors of the mean.

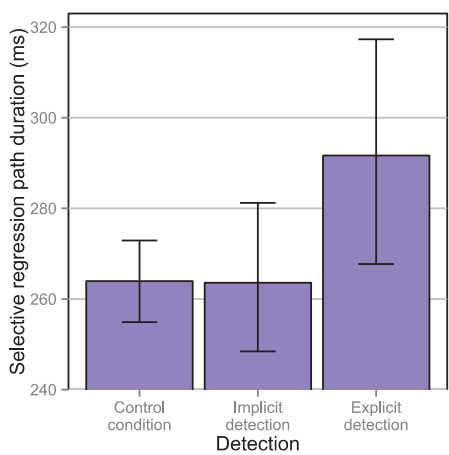


Figure 8.A2. Mean selective regression path durations (ms) on the critical word for the control condition, implicit detection, and explicit detection, when removing forgotten sentences from the analysis, Experiment 1
Error bars denote standard errors of the mean.

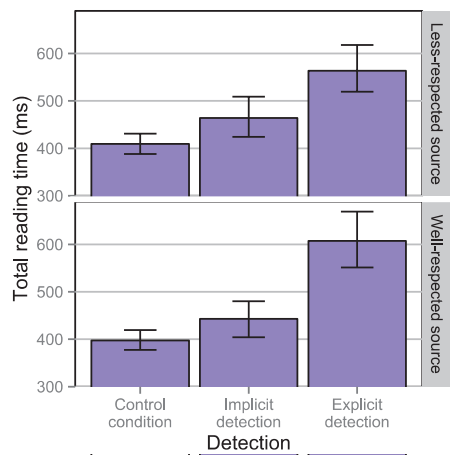


Figure 8.A3. Mean total reading times (ms) for control condition, on the critical word for the control condition, implicit detection, and explicit detection, when removing forgotten sentences from the analysis, Experiment 2
Error bars denote standard errors of the mean.

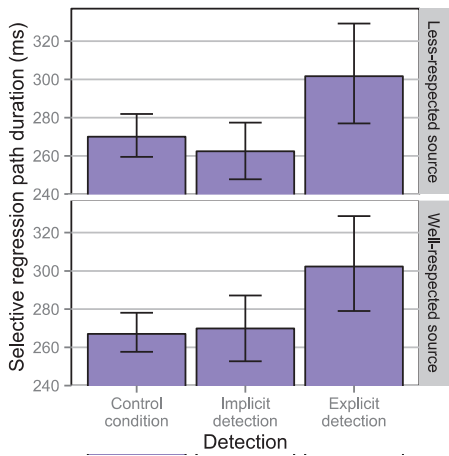


Figure 8.A4. Mean selective regression path durations (ms) on the critical word for the control condition, implicit detection, and explicit detection, when removing forgotten sentences from the analysis, Experiment 2
Error bars denote standard errors of the mean.

Appendix 8.B Figure for analyses of filler sentences in Experiment 2 (non-significant)



Figure 8.B1. Mean total reading times per character (ms) for words in filler sentences, depending on source bias condition, Experiment 2

Figures for all words in the sentences, as well as only the adjective used as the critical word in the manipulated sentences. Error bars denote standard errors of the mean.

Appendix 9.A Full statistical model outputs from Sections 9.1, 9.2, and 9.3, and remaining figure from Section 9.2

All models used log-transformed data to better fit a Gaussian distribution. In all models, p -values were generated using Satterthwaite approximation.

New analysis of data from Chapter 3

Table 9.A1. Total reading time/character depending on comprehension scores for different reading types

Statistically significant results in **bold**; * signifies $p < .05$, ** signifies $p < .01$, *** signifies $p < .001$. A colon denotes an interaction.

	Estimate	SE	df	t-value	p-value	
Comprehension	2.2811	0.6742	13890	3.383	.000718	***
Thorough reading	15.8176	4.8784	55	3.242	.002018	**
Skimming	-12.1502	4.6762	21	-2.598	.016600	*
Spell checking	18.6378	4.8469	16	3.845	.001393	**
Comprehension: Thorough reading	-0.7688	1.0052	3702	-0.765	.444379	
Comprehension: Skimming	-3.3129	0.8836	21413	-3.749	.000178	***
Comprehension: Spell checking	0.4501	0.8462	23257	0.532	.594814	

New analysis of data from Chapter 4 and Chapter 5

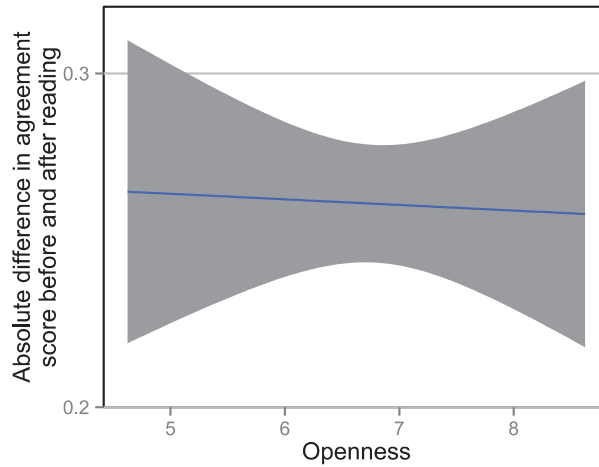


Figure 9.A1. Absolute difference in agreement scores before and after reading depending on openness, Chapter 4, Experiment 2, and Chapter 5, Experiment 2

Regression lines include 95% confidence intervals. Note that no results are significant (see Table 9.A2).

Table 9.A2. Agreement scores after reading depending on headline (agreeing or disagreeing with opinions), when grouping trials depending on low or high opinion strength (weak or strong opinions), accounting for openness

No statistically significant results were found. A colon denotes an interaction.

	Estimate	SE	df	t-value	p-value	
Openness	-0.01386	0.018834	277.7	-0.736	.4625	
Headline agrees with opinion	0.02462	0.175634	681.3	0.140	.8885	
Weak opinion	0.04615	0.199771	683.1	0.231	.8174	
Openness: Headline agrees with opinion	0.00374	0.025844	677.0	0.145	.8849	
Openness: Weak opinion	-0.00163	0.029321	681.8	-0.056	.9558	
Headline agrees with opinion: Weak opinion	-0.46637	0.276155	680.3	-1.689	.0917	.
Openness: Headline agrees with opinion: Weak opinion	0.05652	0.040324	678.8	1.402	.1615	

New analyses of data from Chapter 8

Table 9.A3. Grammatical acceptability rating depending on detection, Experiment 1

No statistically significant results were found.

	Estimate	SE	df	t-value	p-value	
Implicit detection	0.03318	1.39464	103.17	0.024	.981	
Explicit detection	0.09724	1.86262	290.86	0.052	.958	
Congruent filler	0.05829	1.11248	77.53	0.052	.958	

Table 9.A4. Semantic acceptability rating depending on detection, Experiment 1

No statistically significant results were found.

	Estimate	SE	df	t-value	p-value	
Implicit detection	0.9856	1.4593	101.99	0.675	.501	
Explicit detection	-1.6601	1.9082	269.65	-0.870	.385	
Congruent filler	-0.9056	1.1706	77.55	-0.774	.442	

Table 9.A5. Grammatical acceptability rating depending on detection and expected news source, Experiment 2

Statistically significant results in **bold**; * signifies $p < .05$. A colon denotes an interaction.

	Estimate	SE	df	t-value	p-value	
Implicit detection	-0.2142	1.5274	262	-0.140	.8886	
Explicit detection	2.1458	1.8840	573	1.139	.2552	
Congruent filler	-0.8463	1.2088	186	-0.700	.4848	
Well-respected news source	2.6997	1.1852	6019	2.278	.0228	*
Implicit detection: Well-respected news source	1.3244	1.8793	6024	0.705	.4810	
Explicit detection: Well-respected news source	0.8219	2.4391	6044	0.337	.7362	
Congruent filler: Well-respected news source	0.4073	1.4461	6013	0.282	.7782	

Table 9.A6. Semantic acceptability rating depending on detection and expected news source, Experiment 2

Statistically significant results in **bold**; * signifies $p < .05$. A colon denotes an interaction.

	Estimate	SE	df	t-value	p-value	
Implicit detection	-0.9427	1.6338	262	-0.577	.5644	
Explicit detection	3.4467	2.0156	573	1.710	.0878	.
Congruent filler	-1.2617	1.2911	185	-0.977	.3297	
Well-respected news source	2.6777	1.2702	6041	2.108	.0351	*
Implicit detection: Well-respected news source	1.1973	2.0158	6047	0.594	.5526	
Explicit detection: Well-respected news source	0.2341	2.6159	6068	0.090	.9287	
Congruent filler: Well-respected news source	0.8500	1.5497	6035	0.549	.5834	

