

## ABSTRACT

Title of Dissertation:

MEMORY RETRIEVAL IN PARSING AND  
INTERPRETATION

Ananda Lila Zoe Schlueter, Doctor of  
Philosophy, 2017

Dissertation directed by:

Professor Ellen Lau, Linguistics

This dissertation explores the relationship between the parser and the grammar in error-driven retrieval by examining the mechanism underlying the illusory licensing of subject-verb agreement violations ('agreement attraction'). Previous work motivates a two-stage model of agreement attraction in which the parser predicts the verb's number and engages in retrieval of the agreement controller only when it detects a mismatch between the prediction and the bottom-up input (Wagers, Lau & Phillips, 2009; Lago, Shalom, Sigman, Lau & Phillips, 2015). It is the second stage of retrieval and feature-checking that is thought to be error-prone, resulting in agreement attraction. Here we investigate two central questions about the processing system that underlies this profile. First, to what extent does error-driven retrieval end up altering the structural representation of the sentence, as compared to an independent feature-checking process that can result in global inconsistencies? Using a novel dual-task paradigm combining self-paced reading and a speeded forced choice task, we show that comprehenders do not misinterpret the attractor as the subject in agreement

attraction. This indicates that the illusory licensing reflects a low-level number rechecking process that does not lead to restructuring. Second, what is the relationship between the information guiding the retrieval process and the terms that define agreement in the grammar? In a series of speeded acceptability judgment and self-paced reading experiments, we demonstrate that the number cue in error-driven retrieval is as abstract as the terms in which agreement is stated in the grammar, and that semantic features not relevant to the dependency in the grammar are not used to guide retrieval of the agreement controller. However, data from advanced Chinese learners of English suggests that it is not the case that all features relevant to the grammatical dependency will necessarily be used as retrieval cues. Taken together, these results suggest that the feature-checking repair mechanism follows grammatical principles but can result in a final structural representation of the sentence that is inconsistent with the grammar.

# MEMORY RETRIEVAL IN PARSING AND INTERPRETATION

by

Ananda Lila Zoe Schlueter

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park, in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
2017

Advisory Committee:  
Professor Ellen Lau, Chair  
Professor Jan Edwards  
Professor Norbert Hornstein  
Professor Colin Phillips  
Professor L. Robert Slevc  
Professor Alexander Williams

© Copyright by  
Ananda Lila Zoe Schlueter  
2017

## Acknowledgements

First and foremost, I would like to thank Ellen Lau, my advisor and committee chair. I am extremely grateful to Ellen for all the time and energy she invested in me over these past five years. Ellen has been the best advisor I could have hoped for and has shaped me as a researcher (and I hope also as a person). My weekly meetings with Ellen were always encouraging and I frequently walked away with some insight that she had somehow managed to extract from my jumbled thoughts. Ellen is not only an incredible scientist, she is also one of the kindest people I have ever met and, as Norbert put it, “the closest thing our department has to a saint”. She took it remarkably well when I told her in my first year that I didn’t want to do “anything with brains”. Although there are no brains in this dissertation, her Cognitive Neuroscience of Linguistics class completely changed my mind and I am very glad to have had the opportunity to learn about EEG from her.

I am also extremely grateful to my co-advisor and committee member Alexander Williams. Alexander’s hands-off approach to advising meant that our meetings were less frequent but that he was always available when I needed help to think something through. I think I never managed to fully take in all of the extremely insightful things he said during our meetings, but I always left having learned something important. I continue to be amazed by the breadth and depth of his knowledge, not just about linguistics but about life, the universe and everything.

Thank you to the members of my dissertation committee. To Colin Phillips, whose knowledge, experience and dedication to the students I benefitted from immensely. I am particularly grateful for his support during my job search. Although I was extremely lucky to get a postdoc at Edinburgh, his advice to “steel myself for rejection” and not to take it personally was something I frequently reminded myself of when the rejection letters kept coming. To Bob Slevc, who agreed to work with me as part of my IGERT lab rotation and has facilitated my foray into language production. To Norbert Hornstein, who was the first person from UMD I ever spoke to. I admire his clarity of thought and his outspokenness, and he also gives surprisingly good pep talks. To Jan Edwards, who was kind enough to agree to serve on my committee in the middle of the summer even though she had no idea who I was.

In addition to my committee, I also want to thank the entire faculty of the Linguistics Department. Not only are they a group of truly amazing scientists, but they are also extremely generous with their time and knowledge. I am grateful that all faculty members were always approachable and willing to talk to me about my work. I appreciate that they somehow managed to be supportive while not holding back any constructive criticism. Although it took me almost the whole five years to stop being nervous when giving a lab meeting, it was always clear that their criticism came from a place of actually caring about my research and I know that I benefitted immensely from their feedback.

I am grateful to Alix Kowalski for being a great friend and housemate. I met Alix the first day I arrived in the US and I don't know how I would have made it through these five years without her. I am so glad that we lived together the whole time and were able to support each other during all of the challenges we both faced. I can't imagine how different my experience would have been without Alix (and the various cats we shared our house with over the years), but I'm sure that my time here would have been a lot less fun - and I definitely would not have learned how to wrangle a hairless cat into a sweater (#sweaterbruceisbetterbruce). Moving away will be a very tough adjustment. Among many other things, I will miss watching murder mysteries together while analyzing data in R, and I hope that we can resume that as a regular activity at some point in the future (data analysis optional).

I am grateful to Peter Enns, who decided that 5 years in grad school was not the right choice for him and then had to endure it by proxy anyway. I like to think that his time in the linguistics program wasn't completely wasted since that's how we met. Although Peter had to move away for a real job, he has been extremely supportive and I am very lucky to have had him during this difficult time and beyond.

When I arrived at Maryland, I remember being completely awe-struck by how knowledgeable, confident and accomplished the older students were. In hindsight, they might not always have felt as confident as they seemed, but my research builds on their achievements and I continue to be impressed and inspired by their work. I am especially thankful to Sol Lago, Dan Parker, Shota Momma, Dave Kush, Wing Yee Chow, and Aaron Steven White. I have learned so much from them and I really appreciate their willingness - both while at UMD and after they had already left - to help me develop my ideas and come to my rescue with technical issues (even when it involved looking at an R script during a conference dinner).

I also want to thank the rest of my cohort and all the other students I overlapped with during my time here. Interacting with them in classes, at lab meetings and outside the department has had a hugely positive impact on my life. In particular: Mike Fetters, who more than once proved that he is the kind of friend who will drop everything to help out in an emergency; Chris Heffner, who I greatly admire for his boundless energy and infectious enthusiasm not only for his research but for life in general; Alia Lancaster, who is a great yoga/shouty-pilates buddy and an even better friend; Eric Pelzl and Rachel Adler, who introduced me to a plethora of board games and were a lot of fun to hang out with; and Lara Ehrenhofer, who has been a great officemate, always willing to listen to me rant and offering cookies and tea when needed.

One of the things I appreciate most about being a student at UMD is that it automatically made me a part of a bigger community. This has given me the opportunity to interact not only with current members of the department but also with a range of other people, including alumni and visitors. All of these people have been incredibly generous with their time and expertise and I am truly grateful to them. There are too many to mention, but I particularly want to thank Brian Dillon, Diogo

Almeida, Darren Tanner, Matt Tucker, Masaya Yoshida, Laurel Brehm, Shayne Sloggett, Caroline Andrews, Iria de Dios Flores, and Zuzanna Fuchs. Whether I temporarily shared an office with them or we had a long conversation at a conference, they all have in common that they have made me feel excited about my research when things were particularly difficult.

Of course, this list would not be complete without the lab managers and research assistants who provided invaluable practical support: Anna Namyst, Julia Buffinton, Hanna Muller, Jon Burnsky, Jamie Lebovics, and Rebekah Senderling. I also especially want to thank Kim Kwok, who helped me so many times and without whom the entire department would descend into chaos.

Finally, I would like to acknowledge that some of the research in this dissertation was generously funded by an NSF dissertation improvement grant (NSF-DDRI: 1651058).

It goes without saying that I am beyond grateful to my family for the unconditional love and support that they have given me, not just over these past five years, but my entire life.

# Table of Contents

<b>Acknowledgements .....</b>	ii
<b>Table of Contents .....</b>	v
<b>List of Tables .....</b>	ix
<b>List of Figures.....</b>	xii
<b>Chapter 1 .....</b>	1
1.1 Overview.....	1
1.2 Illusions of grammaticality .....	2
1.3 Agreement attraction.....	5
<i>1.3.1 Production versus comprehension .....</i>	6
<i>1.3.2 A two-stage model of agreement attraction .....</i>	8
1.4 Structural ramifications of error-driven retrieval.....	10
1.5 The use of grammatically (ir)relevant information in error-driven retrieval....	12
1.6 Outline of the dissertation.....	14
<b>Chapter 2 .....</b>	18
2.1 Outline.....	18
2.2 Error-driven retrieval in subject-verb agreement processing.....	19
2.3 Structural reanalysis in comprehension .....	22
2.2 Structure and interpretation.....	24
2.5 Agreement and interpretation .....	28
2.6 Experiment 1: adjective-choice task .....	31
<i>2.6.1 Participants.....</i>	33
<i>2.6.2 Materials .....</i>	33
<i>2.6.3 Plausibility Norming .....</i>	34
<i>2.6.4 Agreement Attraction Norming .....</i>	36
<i>2.6.5 Procedure.....</i>	38
<i>2.6.6 Analysis .....</i>	38
<i>2.6.7 Results .....</i>	39
<i>2.6.8 Discussion .....</i>	47
2.7 General Discussion .....	51
<i>2.7.1 The final representation of agreement attraction sentences .....</i>	52
<i>2.7.2 A third possibility: Revising the subject's number feature .....</i>	55
2.8 Conclusion .....	58
<b>Chapter 3 .....</b>	60
3.1 Outline.....	60
3.2 Cue-based retrieval in sentence processing .....	61
3.3 What matters in agreement processing? .....	62
<i>3.3.1 Production.....</i>	63
<i>3.3.2 Comprehension .....</i>	65
3.4 The present study .....	67
3.5 Experiment 2: conjoined NPs (speeded acceptability) .....	68
<i>3.5.1 Participants .....</i>	70

3.5.2 <i>Materials and Design</i> .....	70
3.5.3 <i>Procedure</i> .....	71
3.5.4 <i>Analysis</i> .....	72
3.5.5 <i>Results</i> .....	72
3.5.6 <i>Discussion</i> .....	74
3.6 Experiment 3: conjoined NPs (SPR).....	76
3.6.1 <i>Participants</i> .....	77
3.6.2 <i>Materials and Design</i> .....	77
3.6.3 <i>Procedure</i> .....	78
3.6.4 <i>Analysis</i> .....	79
3.6.5 <i>Results</i> .....	80
3.6.6 <i>Discussion</i> .....	83
3.7 Experiment 4: conjoined NPs without second determiner (speeded acceptability).....	84
3.7.1 <i>Participants</i> .....	85
3.7.2 <i>Materials and Design</i> .....	85
3.7.3 <i>Procedure and Analysis</i> .....	86
3.7.4 <i>Results</i> .....	87
3.7.5 <i>Discussion</i> .....	87
3.8 Experiment 5: conjoined adjectives (speeded acceptability) .....	89
3.8.1 <i>Participants</i> .....	90
3.8.2 <i>Materials and Design</i> .....	90
3.8.3 <i>Procedure and Analysis</i> .....	91
3.8.5 <i>Results</i> .....	92
3.8.6 <i>Discussion</i> .....	93
3.9 Experiment 6: conjoined adjectives in self-paced reading .....	94
3.9.1 <i>Participants</i> .....	95
3.9.2 <i>Materials and Design</i> .....	95
3.9.3 <i>Procedure and Analysis</i> .....	95
3.9.4 <i>Results</i> .....	95
3.9.5 <i>Discussion</i> .....	98
3.10 Experiment 7: conjoined vs. stacked adjectives (speeded acceptability) ....	100
3.10.1 <i>Participants</i> .....	100
3.10.2 <i>Materials and Design</i> .....	101
3.10.3 <i>Procedure and Analysis</i> .....	101
3.10.4 <i>Results</i> .....	102
3.10.5 <i>Discussion</i> .....	104
3.11 General Discussion .....	104
3.11.1 <i>The role of morphological form in memory retrieval</i> .....	105
3.11.2 <i>Notional plurality in agreement computation in comprehension</i> .....	107
3.11.3 <i>Associative Cues</i> .....	108
3.12 Conclusion .....	111
<b>Chapter 4</b> .....	113
4.1 Outline.....	113
4.2 Notional number in agreement processing .....	114
4.3 Experiment 8: Notional number.....	120

<i>4.3.1 Participants</i> .....	121
<i>4.3.2 Materials and Design</i> .....	121
<i>4.3.3 Procedure</i> .....	123
<i>4.3.4 Analysis</i> .....	123
<i>4.3.5 Results</i> .....	124
<i>4.3.6 Discussion</i> .....	127
<i>4.4 Semantic interference in comprehension</i> .....	129
<i>4.5 Experiment 9a: Plausibility</i> .....	134
<i>4.5.1 Participants</i> .....	134
<i>4.5.2 Materials and Design</i> .....	135
<i>4.5.3 Procedure</i> .....	136
<i>4.5.4 Analysis</i> .....	137
<i>4.5.5 Results</i> .....	139
<i>4.5.6 Discussion</i> .....	142
<i>4.6 Experiment 9b: Plausibility (replication)</i> .....	143
<i>4.6.1 Participants</i> .....	144
<i>4.6.2 Materials, Procedure and Analysis</i> .....	144
<i>4.6.3 Results</i> .....	144
<i>4.6.4 Discussion</i> .....	147
<i>4.7 General Discussion</i> .....	148
<i>4.7.1 Availability of number information vs semantic information</i> .....	149
<i>4.7.2 Error-driven retrieval relies on grammatically relevant features</i> .....	150
<i>4.8 Conclusion</i> .....	151
<b>Chapter 5</b> .....	154
<i>5.1 Outline</i> .....	154
<i>5.2 L2 Grammatical knowledge and online processing</i> .....	155
<i>5.2.1 Subject-verb agreement processing in advanced Chinese learners of English</i> .....	157
<i>5.2.2 Agreement attraction in L2 processing</i> .....	160
<i>5.3 The plural markedness effect</i> .....	163
<i>5.4 The present study</i> .....	165
<i>5.5 Experiment 10: L2 speeded acceptability</i> .....	166
<i>5.5.1 Participants</i> .....	168
<i>5.5.2 Materials and Design</i> .....	168
<i>5.5.3 Procedure</i> .....	170
<i>5.5.4 Analysis</i> .....	170
<i>5.5.5 Results</i> .....	171
<i>5.5.6 Discussion</i> .....	176
<i>5.6 Experiment 11: L2 SPR</i> .....	179
<i>5.6.1 Participants</i> .....	180
<i>5.6.2 Materials and Design</i> .....	181
<i>5.6.3 Procedure</i> .....	181
<i>5.6.4 Analysis</i> .....	182
<i>5.6.5 Results</i> .....	183
<i>5.6.6 Discussion</i> .....	189
<i>5.7 General Discussion</i> .....	195

<i>5.7.1 L2 Morphological Sensitivity</i> .....	196
<i>5.7.2 Processing routines</i> .....	197
<i>5.7.3 Asymmetrical number representation</i> .....	198
<i>5.7.4 Offline strategies</i> .....	199
<i>5.7.5 Automatic sensitivity</i> .....	199
<b>5.8 Conclusion</b> .....	200
<b>Chapter 6</b> .....	202
6.1 Summary .....	202
6.2 Error-driven retrieval as a rechecking operation .....	202
6.3 Grammatically (ir)relevant cues in error-driven retrieval.....	204
6.3.1 Associative cues.....	207
6.4 Implications for our understanding of cue-based retrieval .....	208
<b>Bibliography</b> .....	211

## List of Tables

Table 1: Results from mixed logit model in speeded acceptability judgment task in Experiment 1 .....	38
Table 2: Results from mixed logit model for adjective choice in Experiment 1 .....	41
Table 3: Results of linear mixed effects model of response time on the adjective- choice task in Experiment 1 (using log transformed RTs) .....	42
Table 4: Results of linear mixed effects model in the verb region in Experiment 1 (sum coded; using log transformed RTs).....	43
Table 5: Results of linear mixed effects model in the spillover region in Experiment 1 (sum coded; using log transformed RTs).....	44
Table 6: Results of linear mixed effects model in verb region in Experiment 1 for trials on which the head-matching adjective was chosen (sum coded; using log transformed RTs) .....	45
Table 7: Results of linear mixed effects model in spillover region in Experiment 1 for trials on which the head-matching adjective was chosen (sum coded; using log transformed RTs) .....	46
Table 8: Results of linear mixed effects model in verb region in Experiment 1 for trials on which the attractor-matching adjective was chosen (sum coded; using log transformed RTs) .....	47
Table 9: Results of linear mixed effects model in spillover region in Experiment 1 for trials on which the attractor-matching adjective was chosen (sum coded; using log transformed RTs) .....	47
Table 10: Results of linear mixed logit model on acceptance rates in Experiment 2. ....	74
Table 11: Mean raw reading times per condition for regions of interest in Experiment 3 (standard error of the mean in parentheses).....	81
Table 12: Results of linear mixed effects model in verb region in Experiment 3 (using log transformed RTs). ....	81
Table 13: Results of linear mixed effects model in first spillover region in Experiment 3 (using log transformed RTs). ....	81
Table 14: Results of linear mixed effects model in second spillover region in Experiment 3 ( using log transformed RTs). ....	81
Table 15: Results of mixed logit model on acceptance rates in Experiment 4. ....	86
Table 16: Results of mixed logit model on acceptance rates in Experiment 5. ....	93
Table 17: Mean raw reading times per condition for regions of interest in Experiment 6 (standard error of the mean in parantheses).....	96
Table 18: Results of linear mixed effects model in verb region in Experiment 6 (using log transformed RTs). ....	97
Table 20: Results of linear mixed effects model in second spillover region in Experiment 6 (using log transformed RTs). ....	97
Table 21: Results of linear logit model on acceptance rates in Experiment 7.....	103
Table 22: Results of the linear mixed effects model for the experimental items in the verb region in Experiment 8 (using log transformed RTs).....	125
Table 23: Results of the linear mixed effects model for the experimental items in the first spillover region in Experiment 8 (using log transformed RTs).....	125

Table 24: Results of the linear mixed effects model for the experimental items in the second spillover region in Experiment 8 (using log transferred RTs) .....	125
Table 25: Results of the linear mixed effects model in the verb region for the control manipulation in Experiment 8 (using log transformed RTs) .....	126
Table 27: Results of the linear mixed effects model in the second spillover region for the control manipulation in Experiment 8 (using log transformed RTs) .....	127
Table 28: Results of the linear mixed effects model in the verb region in Experiment 9 (using log transformed RTs) .....	140
Table 30: Results of the linear mixed effects model in the verb region when the attractor is a plausible subject for the verb, Experiment 9 (using log transformed RTs).....	142
Table 31: Results of the linear mixed effects model in the verb region when the attractor is not a plausible subject for the verb, Experiment 9 (using log transformed RTs) .....	142
Table 32: Results of linear mixed effects model in the verb region (replication), Experiment 9b (using log transformed RTs) .....	145
Table 34: Results of the linear mixed effects model in the verb region for conditions in which the attractor was a plausible subject for the verb (replication), Experiment 9b (using log transformed RTs) .....	147
Table 36: Results of mixed logit model on acceptance rates of sentences with singular head nouns in Experiment 10, including both language group (Note: Due to convergence issues model had only by-subject random intercepts.) .....	173
Table 37: Results of mixed logit model on acceptance rates of sentences with singular head nouns in Experiment 10 for the L1 group. ....	174
Table 38: Results of mixed logit model on acceptance rates of sentences with singular head nouns in Experiment 10 for the L2 group. ....	175
Table 39: Results of mixed logit model on acceptance rates for sentences with plural head nouns in Experiment 10, including both language groups (Note: Due to convergence issues model had only by-subject random intercepts.) .....	175
Table 40: Results of mixed logit model on acceptance rates for sentences with plural head nouns in Experiment 10 for the L1 group. ....	176
Table 41: Results of mixed logit model on acceptance rates for sentences with plural head nouns in Experiment 10 for the L2 group. ....	176
Table 42: Results of linear mixed effects model in the verb region of sentences with singular head nouns for both language groups, Experiment 11 (using log transformed RTs) .....	184
Table 44: Results of the linear mixed effects model in the verb region of sentences with singular head nouns for the L1 group, Experiment 11 (using log transformed RTs). ....	185
Table 46: Results of the linear mixed effects model in the verb region of sentences with singular head nouns for the L2 group, Experiment 11 (using log transformed RTs). ....	186
Table 47: Results of the linear mixed effects model in the spillover region of sentences with singular head nouns for the L2 group, Experiment 11 (using log transformed RTs). ....	187

Table 49: Results of the linear mixed effects model in the spillover region of sentences with plural head nouns for both language groups, Experiment 11 (using log transformed RTs) .....	188
Table 51: Results of the linear mixed effects model in the spillover region of sentences with plural head nouns for the L1 group, Experiment 11 (using log transformed RTs) .....	188
Table 52: Results of the linear mixed effects model in the verb region of sentences with plural head nouns for the L2 group, Experiment 11 ( using log transformed RTs).....	189
Table 53: Results of the linear mixed effects model in the spillover region of sentences with plural head nouns for the L2 group, Experiment 11 (using log transformed RTs). ....	189
Table 54: Results of mixed logit model on acceptance rates in offline acceptability judgment task for both L1 and L2 groups in Experiment 11.....	194
Table 56: Results of mixed logit model on acceptance rates in offline acceptability judgment task for the L2 group in Experiment 11.....	195

## List of Figures

Figure 1: Acceptance rates across conditions in speeded acceptability judgment task in Experiment 1 .....	37
Figure 2: Percentage of trials with a head-matching adjective choice across conditions in Experiment 1.....	40
Figure 3: mean RTs split by adjective choice (attractor-matching response in blue: 0; head-matching response in red: 1) in each experimental condition in Experiment 1. Proportion of head-noun compatible responses beneath condition labels. ....	42
Figure 4: Region-by-region mean raw reading times in Experiment 1. Error bars indicate standard error of the mean.....	43
Figure 5: Region-by-region mean reading times for trials on which the (correct) head-matching adjective was chosen (Experiment 1). Error bars indicate standard error of the mean.....	45
Figure 6: Region-by-region mean reading times for trials on which the attractor-matching adjective was chosen (Experiment 1). Error bars indicate standard error of the mean.....	47
Figure 7: Acceptance rates across conditions in Experiment 2. Error bars indicate standard error of the mean. ....	73
Figure 8: Region-by-region mean log reading times in Experiment 3. Error bars indicate standard error of the mean.....	80
Figure 9: Acceptance rates across conditions in Experiment 4. Error bars indicate standard error of the mean. ....	86
Figure 10: Acceptance rates across conditions in Experiment 5. Error bars indicate standard error of the mean. ....	92
Figure 11: Region-by-region mean log reading times in Experiment 6. Error bars indicate standard error of the mean.....	96
Figure 12: Acceptance rates across conditions in Experiment 7. Error bars indicate standard error of the mean. ....	103
Figure 13: Region-by-region mean log reading times for the experimental items in Experiment 8. Error bars indicate standard error of the mean.....	125
Figure 14: Region-by-region mean log reading times for the control items in Experiment 8. Error bars indicate standard error of the mean.....	126
Figure 15: Region-by-region mean residualized reading times across conditions in which not only the head noun but also the attractor were plausible subjects of the main verb in Experiment 9. Error bars indicate standard error of the mean....	140
Figure 16:Region-by-region mean residualized reading times across conditions in which only the head noun but not the attractor were plausible subjects of the main verb in Experiment 9. Error bars indicate standard error of the mean....	141
Figure 17: Region-by-region mean residualized reading times across conditions in which not only the head noun but also the attractor were plausible subjects of the main verb (replication), Experiment 9b. Error bars indicate standard error of the mean.....	146
Figure 19: Acceptance rates across conditions with singular head nouns for L1 group (upper panel) and L2 group (lower panel) for Experiment 10 (note different	

scales on y-axis for L1 and L2 group). Error bars indicate standard error of the mean.....	172
Figure 20: Region-by-region mean raw reading times for the L1 group of sentences with singular head nouns, Experiment 1. Error bars indicate standard error of the mean.....	185
Figure 21: Region-by-region mean raw reading times for the L2 group for sentences with singular head nouns, Experiment 11 (error bars indicate standard error of the mean).....	186
Figure 22: Acceptance rates in offline acceptability judgment task for L1 group (upper panel) and L2 group (lower panel) in Experiment 11. Error bars indicate standard error of the mean .....	194

# Chapter 1

## 1.1 Overview

The aim of this dissertation is to explore the nature of error-driven retrieval in sentence processing by examining the mechanism underlying the illusory licensing of subject-verb agreement violations. This includes investigating both the impact of error-driven retrieval on the structural representation of the sentence and the relationship between the information guiding the retrieval process and the terms that define agreement in the grammar.

Much recent work has asked whether the interpretation comprehenders arrive at always tracks the syntax. In this dissertation, I pursue this issue by investigating whether the illusory licensing of an agreement violation reflects a restructuring or a low-level rechecking operation. In agreement attraction, a subject-verb agreement violation causes a problem for integrating the verb into the structure of the sentence. Previous work has shown that when comprehenders receive input that cannot be integrated into the current parse, they often engage in successful structural reanalysis of the previous input. This illustrates that an error signal can cause large scale restructuring, but does a grammatical illusion like agreement attraction also reflect structural reanalysis? If the error signal from an agreement violation triggers a similar restructuring operation, the attractor would be misinterpreted as the subject. Although the interpretation would be very different from the input, it would track the structural representation of the sentence and the structural representation would be consistent with the grammar. However, if agreement attraction is

the result of a simple rechecking operation that does not alter the structural representation of the sentence, the final representation in some sense contains an agreement violation.

In this dissertation, I show that the illusory licensing of an agreement violation is not the result of a restructuring process that obligatorily occurs when the attractor is misretrieved in the search for the agreement controller in memory. Instead, I propose that error-driven retrieval of the agreement controller generally involves a low-level number rechecking operation., This suggests that even if the mental representation of the sentence and the derived interpretation have to be consistent, some features (such as number agreement) can be inconsistent with the grammatical constraints in the final representation.

This dissertation also demonstrates that the repair mechanism in error-driven retrieval relies on grammatical principles. Once an agreement violation is detected, the information that is used to find the agreement controller in memory uses cues as abstract as the terms in which the agreement dependency is stated in the grammar. Moreover, the parser does not seem to use additional information, such as notional number or plausibility, that is not grammatically relevant to subject-verb agreement in the error-driven retrieval operation,

## 1.2 Illusions of grammaticality

Much prior research has demonstrated that online processing can be susceptible to so-called grammatical illusions (see Phillips, Wagers & Lau, 2011, for review). In grammatical illusions, comprehenders seem to fail to notice that the linguistic input contains a grammatical violation, typically in configurations where it appears that

structurally irrelevant material in the sentence erroneously licenses an item that is not actually licensed according to the grammar. For example, a sentence with a subject-verb number agreement violation, such as '*The key to the cabinets are rusty*', might be perceived as acceptable if it contains a structurally irrelevant noun that matches the verb's number marking (Wagers, Lau & Phillips, 2009). Grammatical illusions have been observed for a number of structures (see among others Tanner, Nicol & Brehm, 2014; Dillon, Mishler, Sloggett & Phillips, 2013; Lago, Shalom, Sigman, Lau & Phillips, 2015, Parker & Phillips, 2016), but there is variability in how robust these illusions are. In this dissertation, I use subject-verb agreement attraction, which is perhaps the most well-documented example of a grammatical illusion and has been found to be robust across a number of online measures. I return to the relationship between grammatical illusions like agreement attraction and the less robust ones like negative polarity item licensing and reflexive processing in the conclusion. Importantly, grammatical illusions represent cases in which there is a discrepancy between offline and online sensitivity to a grammatical constraint. In untimed tasks such as acceptability judgments, comprehenders are sensitive to a particular configuration that constitutes a grammatical violation, but they fail to show that sensitivity in online processing only. Consequently, their discovery has been of great interest to research on language processing, as they raise important questions about the relationship between the grammar and the parser (Lewis & Phillips, 2015).

The discrepancy has led some researchers to argue that the grammar and the parser are in fact separate cognitive systems. The phenomenon of linguistic illusions certainly raises questions for a view of the language system under which grammatical

theories and processing mechanisms describe the same system. However, grammatical illusions are not necessarily evidence that these are separate cognitive systems. In fact, it is possible to account for these illusions while maintaining a one-system view as long as we take into consideration that in online processing grammatical constraints have to be implemented using general cognitive mechanisms (Phillips & Lewis, 2013; Lewis & Phillips, 2015).

One of the cognitive mechanisms that has been shown to be crucial for language processing is memory retrieval: linguistic dependencies frequently hold between non-adjacent items and require retrieval of previous items from memory. Consequently, the architecture of the memory system is an important factor in how language is processed. Grammatical illusions can be explained by the way in which comprehenders navigate linguistic representations in memory. The memory system underlying language comprehension relies on content-addressable cue-based retrieval (retrieval (McElree, 2000; McElree, Foraker & Dyer, 2003; Lewis & Vasishth, 2005; Van Dyke & McElree, 2006; Martin & McElree, 2009). Items are encoded as bundles of features and retrieved from memory if their features match the retrieval cues. However, the architecture of the memory system is noisy and retrieval is susceptible to similarity-based interference from non-target items. The misretrieval of a structurally inaccessible linguistic item can lead to an illusion of grammaticality because the erroneously retrieved item acts as an illusory licensor. In this dissertation, I follow Lewis and Phillips (2015) in adopting the assumption that grammatical theories and sentence processing mechanisms are the same cognitive system for structure generation.

### 1.3 Agreement attraction

The phenomenon of agreement attraction was first investigated systematically in a production experiment by Bock and Miller (1991). They found that in a sentence completion task participants were more likely to produce agreement errors with singular subjects if they contained a plural noun inside a prepositional modifier ('*The key to the cabinets*'). This effect has since been replicated in a large number of production studies (see for example Bock & Eberhard, 1993; Eberhard, Cutting & Bock, 2005; Brehm & Bock, 2013; among many others).

A corresponding phenomenon has also been observed in comprehension, where it involves the illusory licensing of an ungrammatical verb form. In agreement attraction in comprehension, a subject-verb agreement violation is erroneously perceived to be acceptable in the presence of a non-subject that matches the verb in number. For example, comprehenders are much less likely to notice the agreement violation in a sentence like '*The key to the cabinets are rusty*', which contains the structurally inaccessible plural noun '*cabinets*', than in the same sentence that contains the singular noun '*cabinet*'. Agreement attraction occurs not only with prepositional modifiers, but also in relative clause constructions, such as '*The musicians who the reviewer praise so highly will probably win*', in which the attractor and the verb are not contiguous (Wagers et al., 2009; Dillon et al., 2013; Pearlmutter, Garnsey & Bock, 1999; Tanner, Nicol & Brehm, 2014; Staub, 2009, 2010). Subject-verb agreement attraction is also not limited to English and has been observed crosslinguistically (Spanish: Lago et al., 2015; Arabic: Tucker, Idrissi & Almeida, 2015), indicating that it is not an oddity of English but reflects a more general processing mechanism.

### *1.3.1 Production versus comprehension*

The early experimental research on agreement attraction was mostly conducted using production paradigms. Accounts of agreement attraction in production have largely focused on representational explanations for this phenomenon. The claim is that the number feature of a singular subject is affected by the presence of a plural attractor, either through feature percolation or spreading activation (e.g. Bock & Eberhard, 1993; Pearlmuter et al., 1999; Bock et al., 2004). The most influential representational account is the Marking and Morphing model (Eberhard, Cutting & Bock, 2005). According to this model, the number information on a noun phrase is a value that can range from unambiguously singular to unambiguously plural. The number marking on the verb assumed to be probabilistic, so a singular subject with a more ambiguous number value elicits more agreement errors in production. Although a subject with a singular head noun should be valued as unambiguously singular, the presence of a plural element inside it ('*The key to the cabinets*') will raise the subject's number value and make it more ambiguous, increasing the likelihood of agreement errors.

Representational models relying on feature percolation or spreading activation like those often assumed for agreement attraction in production have sometimes been proposed to extend to comprehension (Pearlmuter, Garnsey & Bock, 1999). Although representational models like Marking and Morphing can account for the agreement attraction data in production, they fail to capture some of the comprehension data. If agreement attraction is a result of misrepresenting the number feature of the subject, this predicts that grammatical sentences should sometimes be perceived as ungrammatical in

the presence of a plural attractor ('*The key to the cabinets is...*'). However, that does not seem to be the case (Wagers et al., 2009; Tucker et al., 2015; Lago, Shalom, Sigman, Lau & Phillips, 2015; but cf. Pearlmuter et al., 1999).

In contrast, the facilitative impact of a number-matching non-subject in comprehension can be accounted for very naturally by a cue-based retrieval model (Wagers, Lau & Phillips, 2009). Subject-verb agreement is a dependency in which the syntactic number of the verb has to match the syntactic number of the subject. In order to check this the subject has to be retrieved from memory. In the cue-based memory retrieval system assumed here, the verb provides a number cue (e.g. [plural]) as well as a structural cue (e.g. [subject]). When one of the items from memory has features that match both the cues, it is highly likely to be retrieved. Note that when there is a number-matching non-subject present, this also receives an activation boost from the number retrieval cue. In ungrammatical sentences in which the subject does not match the verb in number, a number-matching non-subject (attractor) can be erroneously retrieved in a phenomenon called facilitative similarity-based interference. In this case, the subject does not receive a boost in activation from the number cue and its activation level is only raised by the structural cue. The attractor noun in turn receives a boost in activation from the number cue. In some cases, this leads to the misretrieval of the attractor instead of the actual target, which results in an amelioration of the processing difficulty associated with agreement violations. Here, I assume a cue-based retrieval model of agreement attraction in comprehension.

It is possible that the mechanism underlying attraction effects in comprehension might be different from production (Tanner, Nicol & Brehm, 2014; Acuna-Farina, 2012).

This is an assumption we have to make if we want to adopt the currently predominant view of agreement attraction in production that states that attraction is the result of misrepresenting the subject's number information. However, cue-based retrieval models have also been proposed for production (Badecker & Kuminiak, 2007; Slevc & Martin, 2016). Nevertheless, even if attraction in production and comprehension are both reflections of a cue-based retrieval mechanism, this does not necessarily guarantee that they are susceptible to similarity-based interference from the same features. Although I assume a language system in which production and comprehension are part of the same system and operate over the same kind of representations, they are fundamentally different regarding the direction of encoding: production encodes linguistic structure starting from the message-level, comprehension encodes linguistic structure to arrive at the message level. It is therefore possible that conceptual information might have a stronger impact on retrieval in production than in comprehension.

### *1.3.2 A two-stage model of agreement attraction*

In a cue-based retrieval model of agreement attraction there are two theoretical possibilities about when retrieval of the agreement controller is triggered. In principle, it is possible that subject-verb agreement processing in comprehension always involves retrieval of the agreement controller from memory based on the retrieval cues of the verb. In a grammatical sentence, the features of the subject are a perfect match for the retrieval cues on the verb: it fulfills both the structural cue of being the subject and its number feature matches the number cue. Even if there is a structurally irrelevant noun that matches the number marking on the verb, this item only receives activation from one of

the retrieval cues. Its activation level is therefore lower than that of the subject. Consequently, the appropriate target is retrieved from memory. Retrieval in a sentence with an agreement violation would be triggered in the same way (by default), but the outcome would be different.

The other possibility, and the one I will pursue in this dissertation, is that agreement attraction is an error-driven phenomenon (Wagers et al., 2009; Lago et al., 2015). Recent research has shown that predictive mechanisms are an important component of comprehenders' abilities to maintain robust language comprehension under time-pressure with noisy input. There is overwhelming evidence that language comprehension is not exclusively driven by bottom-up input and that comprehenders deploy top-down mechanisms to make use of existing information to predict upcoming input (see Kutas et al., 2010, for review). In the case of subject-verb agreement, this motivates a view in which comprehenders predict the number of the upcoming verb based on the number feature of the subject. If the bottom-up input matches their prediction, the verb's number marking is licensed and there is no need to retrieve the agreement controller. However, when the prediction is violated, this triggers an error-driven process to check whether the verb's number marking was licensed by the subject. Under this model, grammatical sentences without an agreement violation do not involve cue-based retrieval of the agreement controller. Instead, agreement checking is a two-stage process and retrieval of the agreement controller is the second step that is limited to instances where an agreement violation has been detected.

An important type of evidence in favor of this two-stage model are data indicating that comprehenders initially show sensitivity to the agreement violation even in the

presence of a number-matching attractor. Recent research has shown that attraction effects occur in the right tail of the reading time distribution, compared to the effect of grammaticality which also exerts an influence on faster reading times (Staub, 2009, 2010; Lago et al., 2015). Moreover, in eye-tracking studies, agreement violations have been observed in early reading time measures, while attraction effects were only found in late reading time measures (Dillon, Mishler, Sloggett & Phillips, 2013). This suggests that during the initial processing of the verb comprehenders are sensitive to the agreement violation even in the presence of a plural attractor. The amelioration of the processing disruption associated with this violation does not appear to occur until a later stage of processing. In this dissertation, I am assuming a two-stage model of agreement attraction in which comprehenders predict the number marking of the verb and only engage in retrieval of the agreement controller when the input mismatches their prediction. In Chapter 2, I will also briefly discuss the data from an experiment exploring misinterpretation of the attractor as the subject can be more easily accounted for if we assume that the retrieval of the agreement controller is error-driven, providing further support for this view.

#### 1.4 Structural ramifications of error-driven retrieval

The first question I tackle in this dissertation is whether or not there are structural ramifications of the error-driven retrieval process indexed by agreement attraction. If the agreement controller is retrieved following the detection of a mismatch between the verb's predicted number and the bottom-up input, the attraction effect could be indexing the building of a new structure, in which the verb's number marking is in fact licensed.

Alternatively, attraction could be the result of a low-level feature checking operation that does not trigger reanalysis of the previously assigned structure.

Under the first hypothesis, accidentally retrieving the attractor instead of the actual target from memory has far reaching consequences for the structural representation of the sentence. The parser inserts the output of the error-driven retrieval process into the subject position, drastically changing the mental representation of the sentence. Under this view, comprehenders experience an illusion of grammaticality because there is in fact no subject-verb agreement violation in the mental representation of the sentence they arrive at. This is in a sense similar to the structural reanalysis comprehenders engage in when they reach the point of disambiguation of a garden-path sentences that had initially been assigned the wrong structure (Bever, 1970; Frazier & Fodor, 1978; Frazier & Rayner, 1982). However, reanalyzing the structure of a garden-path sentence results in a structure that can accommodate the entire sentence in a way that is consistent with the grammar. In contrast, because agreement attraction sentences are in fact ungrammatical on any analysis, restructuring in the agreement attraction case would result in a part of the sentence (i.e. the subject's head noun) not being included in the final representation.

The alternative hypothesis is that misretrieval of the attractor does not automatically trigger structural reanalysis. This view assumes that error-driven checking of grammatical features can occur without reanalysis of the structure. The attraction effect arises as a result of the parser locating an item in memory that licenses the verb's number marking. This does not necessarily mean that misretrieval of the attractor is never a contributing factor for engaging in structural reanalysis, but it does imply that in cases in which the comprehender experiences attraction effect the ultimate mental

representation of the sentence is not consistent with the grammar: there is still a subject-verb agreement violation.

Based on the experiment reported in Chapter 2, I conclude that the mechanism responsible for agreement attraction does not necessarily involve restructuring the previously built representation of the sentence. Instead, it appears that agreement attraction is mostly the result of a low-level feature-checking mechanism that does not usually have a structural impact. However, the data does suggest error-driven retrieval in agreement processing and the likelihood of misrepresenting the attractor as the subject are not completely independent from each other.

## 1.5 The use of grammatically (ir)relevant information in error-driven retrieval

The second question I take up in this dissertation is what kind of information is used to guide retrieval. Subject-verb agreement is a dependency between two syntactically defined categories: the inflected verb's number marking has to match the number feature of the noun phrase in subject position. A natural assumption would be that this process uses all and only the information that defines this dependency in the grammar. Here I assume that the initial top-down process that generates the prediction of the verb's number relies strictly on the terms that define this dependency in the grammar. However, the retrieval step in agreement attraction is a type of repair process triggered by a mismatch of the input with the predicted number marking on the verb. In that way, it is a different type of mechanism from the structure generation in the prediction step. It is

conceivable that such a repair process makes use of extra-grammatical principles instead of or in addition to the grammatically relevant features.

I address different sub-questions related to this issue in the remaining chapters of this dissertation. In Chapter 3, I demonstrate that the cues that the parser uses can be as abstract as those by which agreement is defined in the grammar. I test this by showing that conjoined singular noun phrases like '*the husband and (the) wife*' cause agreement attraction when they occur in a structurally irrelevant position where they cannot license agreement. Conjoined singular noun phrases are syntactically plural (they take plural agreement when they occur in subject position), but they do not contain an unequivocal morphological correlate of syntactic plurality. Consequently, the fact that comprehenders experience attraction effects with them indicates that the number cue targets a more abstract plural feature than just an unequivocal morphological correlate.

Even if the error-driven retrieval process makes use of the abstract cues defining the dependency in the grammar, it might also use additional information in such a repair process. In Chapter 4, I show that retrieval of the agreement controller does not appear to be guided by notional number or by the plausibility match with the verb. This suggests that the retrieval process is limited to using the information that is relevant for the dependency in the grammar.

If the error-driven retrieval operation uses only relevant grammatical information to guide retrieval, does it necessarily use *all* relevant grammatical information? My findings suggest that relevant grammatical information does not necessarily have to be implemented as retrieval cues. I show that advanced Chinese learners of English are not susceptible to agreement attraction in comprehension, even though they are sensitive to

subject-verb agreement violations in online processing. The fact that they notice the ungrammaticality but are not distracted by a structurally irrelevant number-matching noun indicates that they do not use the verb's number retrieval cue to find the agreement controller. Although they have acquired the grammatical constraint on English subject-verb agreement, they are not using it in a native-like way in online processing. Interestingly, this makes them more accurate than native speakers in the sense that they are not prone to this illusion of grammaticality.

## 1.6 Outline of the dissertation

The structure of this dissertation is as follows. In Chapter 2, I ask whether comprehenders can experience illusory licensing of an ungrammatical plural verb if their mental representation of the whole sentence remains inconsistent with the grammar. In a novel dual-task paradigm combining self-paced reading with a forced choice task I test whether comprehenders mistake the attractor for the thematic subject of the sentence. The results show that while participants clearly experience agreement attraction in the self-paced reading measures, they are still very accurate in choosing the adjective that matches the subject's head noun. If agreement attraction triggers reanalysis of the previously constructed representation, this would be reflected in participants choosing the adjective that is a plausible continuation for a sentence in which the attractor is the subject when they experience agreement attraction. This indicates that the output of the error-driven retrieval process in agreement processing does not frequently trigger structural reanalysis. Instead, it suggests that agreement attraction is at least usually the result of a low-level feature checking operation without structural impact. If misretrieval

of the attractor does not necessarily trigger restructuring, that also suggests that the ultimate mental representation of agreement attraction sentences remains inconsistent with the grammar.

Chapter 3 explores the question whether the retrieval cues used in processing subject-verb agreement in comprehension are as abstract as the information that defines the dependency in the grammar. I use a series of speeded acceptability judgment tasks and self-paced reading experiments to test whether agreement attraction is sensitive to the vehicle by which syntactic plurality is introduced. If the retrieval process targets an abstract number feature, by which agreement is defined in the grammar, the way in which this abstract plurality is introduced would not be expected to have an impact. I use conjoined singular noun phrases like '*the husband and (the) wife*' in attractor position to show that agreement attraction arises with structurally irrelevant noun phrases that do not bear an unequivocal morphological correlate of syntactic plurality. This demonstrates that error-driven retrieval uses abstract grammatical features to guide the search for the agreement controller in memory.

Chapter 4 focuses on the effect information that is not relevant to the grammatical dependency has on the retrieval process. I address this question in two self-paced reading experiments. The first one uses syntactically singular collective nouns in attractor position and manipulates their bias towards either a collected group reading (notionally singular) or a distributed group reading (notionally plural). I find that participants do not show a reduction in the slow-down associated with processing an agreement violation when the attractor is notionally plural. This suggests that even in error-driven retrieval, the parser specifically uses the type of number information that is relevant for the

dependency in the grammar (i.e. syntactic number). Chapter 4 also focuses on the role of semantic information not related to number. I test whether the plausibility match between the attractor and the verb has an impact on agreement attraction in comprehension. The results show that whether or not the attractor is a plausible subject for the verb does not affect agreement processing. This is converging evidence that the retrieval operation appeals only to the grammatical information that defines the dependency in the grammar and that retrieval of the agreement controller is not guided by other types of linguistic information.

In Chapter 5, I explore whether possessing the grammatical knowledge related to a dependency means that the relevant information is necessarily implemented as retrieval cues in online processing. I report results from two studies with advanced Chinese learners of English. In a speeded acceptability judgment experiment and a self-paced reading experiment, I investigate whether they experience agreement attraction in online processing, in spite of speaking a native language that does not have subject-verb number agreement. The data show that the advanced learners are sensitive to agreement violations, so there is evidence that they have acquired the grammatical knowledge and can use it in online processing. However, they do not show any processing facilitation in ungrammatical sentences when the attractor is plural, unlike native speakers. This finding indicates that for second language learners it is possible to have grammatical knowledge about agreement that is not used to guide retrieval of the agreement controller. This potentially has interesting implications for our understanding of the retrieval process in native speakers, since it suggests that not all grammatically relevant information necessarily has to be implemented as retrieval cues in an error-driven process. However,

these results come from second language learners of English whose native language does not encode grammatical number, so it is unclear whether this possibility is limited to constraints acquired later in life as part of a second language.

Chapter 6 concludes the dissertation by summarizing the empirical evidence and discussing the implications for cue-based models of retrieval. I also return to the relationship between grammatical illusions in different types of dependencies and discuss how the findings of this dissertation can inform our understanding of the different profiles of selective fallibility we observe.

## **Chapter 2**

### 2.1 Outline

Chapter 2 investigates the nature of the error-driven process resulting in the illusory licensing of agreement violations. I assume a two-stage model of agreement attraction, in which the parser predicts the verb's number based on the subject's number and only engages in retrieval of the agreement controller when it detects a mismatch between the prediction and the bottom-up input. The aim of this chapter is to determine whether the second stage of this process, error-driven retrieval, represents a restructuring or a rechecking operation. I use a novel dual-task design that combines self-paced reading with a speeded forced choice task in which participants complete sentence fragments by choosing one of two adjectives. The adjectives are either compatible with the subject's head noun or with the attractor, making the choice an explicit measure of whether comprehenders mistake the attractor for the subject when they experience agreement attraction. As expected, the self-paced reading results show clear evidence of facilitated processing of agreement violations in the presence of a structurally irrelevant number-matching noun. However, participants overwhelmingly chose the adjective compatible with the subject's head noun even in agreement attraction configurations. This suggests that the output of the error-driven retrieval operation is not necessarily used to reanalyze the structure assigned to earlier input. We propose that illusory licensing of an agreement violation can be the result of a rechecking process that is only concerned with number and does not necessarily have any impact on the structural representation of the sentence. Interestingly, this suggests that not all error-driven repair processes lead to a unified

structure that is consistent with the grammar: in agreement attraction, the mismatch between the features on the subject and the verb seem to persist in the final representation.

## 2.2 Error-driven retrieval in subject-verb agreement processing

Subject-verb agreement in English is a morphosyntactic dependency in which the number feature on the verb has to match the number feature of the subject. As discussed in Chapter 1, it has long been observed that this dependency is susceptible to so-called agreement attraction errors in production (Bock & Miller, 1991), in which the number marking on the verb matches a structurally inaccessible plural noun rather than the singular subject (*'The key to the cabinets are rusty'*). In comprehension, these sentences are often perceived as grammatical and do not show the processing cost normally associated with agreement violations (e.g. Pearlmutter et al., 1999; Wagers et al., 2009).

Following previous work, here we will assume that agreement attraction effects in comprehension can best be accounted for by a cue-based retrieval model (Wagers et al., 2009; Tanner et al., 2014; Dillon et al., 2013; Lago et al., 2015; Tucker et al., 2015). Sentence processing frequently requires comprehenders to establish dependencies between items that are not directly adjacent to each other, which means that retrieving items from memory is central to language comprehension. According to cue-based retrieval models (e.g. McElree, 2000; Van Dyke & Lewis, 2003; Lewis & Vasishth, 2005), items are encoded in memory as bundles of features. These items are content-addressable based on the features they contain and can be accessed in parallel rather than through a serial search mechanism (Martin & McElree, 2008). When a search is triggered

and the retrieval cues on an item are used to search for a matching target in memory, activation from each cue is transferred to each item that includes a matching feature. The item with the highest activation level is retrieved. If more than one item matches a certain retrieval cue, the activation is split between them. As long as the actual target is a perfect match for all the retrieval cues, a partial match between the cues and a non-target item will not prevent the actual target from being retrieved. The target's activation level is still higher because it matches all the cues. However, when there is a partial mismatch between the target's features and the cues, the presence of a partially matching non-target item can lead to the misretrieval of this non-target item instead of the actual target, in what is called similarity-based interference.

A memory model based on cue-based retrieval can explain the facilitation that is observed in ungrammatical sentences with plural attractors as an example of similarity-based interference. The retrieval cues on the verb include both structural and number cues. When there is a number mismatch between the subject and the verb in the presence of a plural non-subject attractor ('*The key to the cabinets are...*'), the activation from the number cue raises the level of activation of the attractor, but not the subject. In a subset of cases, this leads to the misretrieval of the number-matching attractor instead of the number-mismatching subject. This is reflected in higher acceptance rates and an amelioration of the processing difficulty associated with agreement violations in online measures.

As briefly discussed in Chapter 1, in this dissertation I will be assuming a two-stage model of agreement attraction in comprehension where attraction is the result of an error-driven process. According to this view, agreement attraction does not reflect

comprehenders' failure to notice the agreement violation on a subset of trials. Instead, it arises as a result of the architecture of the memory system once detection of an agreement violation has triggered retrieval. The two-stage model assumes a parser that predictively generates structure in a top-down fashion. Upon encountering a sentence-initial noun phrase, the parser builds a structure with a subject position to which it attaches the noun phrase. This structure also includes a predictively generated VP with an empty verb slot. The number marking of the verb is predicted based on the number information of the subject. Once the bottom-up information about the verb is received, the parser integrates it into the structure that is being incrementally built. In grammatical sentences without subject-verb agreement violations, the predicted number feature of the verb matches the information from the input and the verb is inserted without difficulty. However, when there is a mismatch between the predicted number feature and the number marking in the actual input this creates an error signal, which triggers cue-based retrieval of the agreement controller. Under this view, cue-based retrieval for agreement processing is an error-driven process that only occurs when there is a mismatch between top-down and bottom-up information. This model predicts that the temporal profile of agreement attraction in online measures shows an initial disruption from the agreement violation which is then ameliorated by the misretrieval of a number-matching attractor. Indeed, Lago et al. (2015) find a pattern of reaction time distributions in self-paced reading that is consistent with this prediction, and so does Dillon et al. (2013) in eye-tracking measures.

In this chapter, I focus on the relation between the error-driven retrieval operation and the structural representation from which the interpretation is derived. I address two separate but closely related questions about error-driven retrieval and interpretation. The

first question is whether agreement attraction necessarily reflects structural reanalysis. If the parser deals with the output of the error-driven retrieval operation by misanalysing it as the subject, this would require extensive revision of the structural representation that had previously been built. Structural reanalysis is not uncommon in sentence comprehension, but if the attractor is misanalysed as the subject, this would result in an interpretation that is unlikely to be faithful to what the speaker intended.

If the process underlying agreement attraction does not necessarily result in restructuring, we can ask whether error-driven retrieval can have an impact on the structural representation – and thus the interpretation – at all. They could be completely independent if the error-driven retrieval process involves only very low-level checking of agreement features. Under this hypothesis, if the number matching attractor is erroneously retrieved, this would not lead to reanalysis of the structural representation. Instead, its number feature would simply signal that the verb's number marking was in fact licensed. In that case, it has to be possible for the final structural representation of the sentence to not align with the way in which the number feature was checked. However, even if misretrieval does not force restructuring, it is possible that it is one of a number of contributing factors that sometimes lead to structural reanalysis. In the following sections, I review evidence that at least some error signals trigger structural reanalysis and consider the impact this has on interpretations.

## 2.3 Structural reanalysis in comprehension

Language unfolds over time and sentences are often temporarily ambiguous. There is plenty of evidence that comprehenders process sentences incrementally, rather

than waiting until they have heard the entire sentence to assign it a structure. The parser has also been shown to not only process bottom-up input incrementally, but to also engage in predictive top-down structure building. When the actual bottom-up input is inconsistent with the structure assigned to the previous input, this triggers reanalysis. For example, in the sentence '*John believes the boy to be honest.*', '*the boy*' is likely to be initially incorporated as the direct object of '*believes*'. However, the following word ('*to*') signals that that cannot be the correct analysis and that '*the boy*' instead has to be the subject of the clausal complement of '*believes*'.<sup>1</sup> Temporarily ambiguous sentences in which comprehenders initially assign the wrong parse and have to engage in reanalysis at the point of disambiguation are referred to as 'garden path' sentences (Bever, 1970; Frazier & Fodor, 1978; Frazier & Rayner, 1982). Classic examples of garden-path sentences like '*The horse raced past the barn fell*' (Bever, 1970) are much more difficult to recover from than the example provided above. However, even sentences with a mild garden-path require reanalysis, indicating that the parser has to be very skilled at changing the structural analysis assigned to previous input. Importantly, this structural reanalysis is driven by an error signal when the input is incompatible with the existing mental representation of the sentence.

Under a two-step model of agreement attraction, encountering an agreement violation is also an example of an error signal from the bottom-up input. Based on what we know about the parser's ability to restructure an existing representation in garden-path sentences, it is conceivable that receiving an error signal from an agreement violation also triggers restructuring. However, it should be noted that there is an important

---

<sup>1</sup> Assuming an analysis of exceptional case-marking constructions in which '*the boy*' is in fact located in the subject position of the embedded clause.

difference between the type of structural reanalysis in garden path sentences and the potential reanalysis that might be happening in agreement attraction. In garden path sentences, the reanalysis includes all the previous input in the new structure. In contrast, structural reanalysis with agreement attraction would require the parser to assign a structural representation only to part of the linguistic input. In a sentence like '*The key to the cabinets are old*', if the attractor ('*the cabinets*') is misanalysed as the subject due to misretrieval in agreement checking, there is no clear way for the subject's actual head noun to be incorporated into this revised structure.

## 2.2 Structure and interpretation

Another important assumption I will be making in this dissertation is that the sentence-level interpretation is a faithful mapping from the syntactic structure that the comprehender has computed. I will therefore briefly review previous work that might be thought to be in conflict with this assumption, and explain why I believe that it is not.

In the past 15 years there has been mounting evidence that the interpretations comprehenders arrive at are not always uniformly consistent with the linguistic input (for recent reviews see Karimi & Ferreira, 2015; Christianson, 2016). Renewed interest in this question was first sparked by the work conducted by Ferreira and colleagues (Christianson et al., 2001; Ferreira et al., 2001) on the interpretation of garden-path sentences (Bever, 1970). In these temporarily ambiguous sentences comprehenders initially assign the wrong parse to some material and have to engage in reanalysis at the point of disambiguation (Frazier & Fodor, 1978; Frazier & Rayner, 1982). For example, reading time data indicates that in a sentence like '*While Anna dressed the baby played in*

*the crib*, the noun phrase ‘*the baby*’ is immediately integrated as the direct object of the verb ‘*dressed*’. Consequently, at the verb ‘*played*’ comprehenders have to engage in costly reanalysis to change the structural representation in memory so that ‘*the baby*’ is represented as the subject of ‘*played*’ rather than the direct object of ‘*dressed*’ (Christianson et al., 2001; Ferreira et al., 2001). Comprehension questions designed to probe participants’ final interpretation of these garden-path sentences showed that they frequently accepted interpretations not consistent with the input. For example, if asked if Anna had dressed the baby, they would answer yes. This led Ferreira and colleagues to conclude that comprehenders do not always recover completely from the initial misparse in garden-path sentences. However, more recent research (Slattery et al., 2013) suggests that the lingering misinterpretation observed with garden-path sentences is not a result of the parser’s failure to completely reanalyze the structural representation.

The interpretations comprehenders accepted in Ferreira et al.’s experiments were not licensed by the actual sentence, but they were consistent with the initial misparse. Parsing is incremental and interpretation is derived from the structure as incoming input is integrated. In garden-path sentences like the ones tested by Ferreira et al., the initial misparse that was constructed up to the point of disambiguation does not only exist on a syntactic level, but is in fact incrementally interpreted. For example, in the sentence given above, when ‘*the baby*’ is integrated as the direct object of ‘*dressed*’, this has immediate interpretive consequences. Even if the syntactic parse undergoes complete reanalysis at the point of disambiguation (‘*played*’), the initial parse of ‘*While Anna dressed the baby*’ has already been interpreted. This interpretation of the initial misparse is not licensed by the linguistic input, but it is consistent with an interpretation derived from the structure

during processing. Slattery et al. (2013) argue that this interpretation lingers in memory and can have downstream effects, even if the ultimate syntactic parse – and the ultimate sentence-level interpretation - is consistent with the input.

Further evidence that comprehenders do not always arrive at an interpretation that is faithful to the linguistic input comes from a series of experiments conducted by Ferreira (2003). These studies tested the comprehension of unambiguous but syntactically challenging sentences, such as passives, object- and subject-clefts. The non-canonical sentences were presented auditorily and participants had to name either the agent or the patient. Their responses showed that they frequently misassigned the thematic roles when the patient role had to be assigned before the agent role, i.e. in the passives and object-cleft sentences. Based on these findings, Ferreira and colleagues propose that comprehension does not always involve the construction of detailed linguistic representations via an “algorithmic route”, and that comprehenders instead frequently use what they call “shallow” or “good enough” representations and “fast and frugal” processing heuristics (Ferreira, et al., 2001; Ferreira et al., 2002; Karimi & Ferreira, 2016; but cf. Frazier, 2008; Koornneef & Reuland, 2016). However, there seems to be no clear evidence that no syntactic structure was computed in these cases, rather than a structure that did not match the input but was consistent with the interpretation. In this dissertation I will assume that interpretations can only be derived from detailed structural representations. I will discuss the reasons for this assumption further in the rest of this section.

Misinterpretations have recently also been observed for implausible but unambiguous sentences that assign thematic roles in canonical order. Gibson et al. (2013)

found that participants frequently answered comprehension questions about implausible sentences like '*The mother gave the candle the daughter*' not based on the grammatically licensed interpretation but on the plausible alternative ('*The mother gave the candle to the daughter*'). This was modulated by how many changes were required to get the plausible alternative, as well as the noise rate (modeled by the proportion of syntactic errors in the fillers). These results are consistent with a noisy-channel model of language comprehension (Levy, 2008; Levy et al., 2009) which emphasizes that linguistic input is often noisy and error prone, and consequently suggests that comprehenders engage in Bayesian rational inferencing about the meaning that a producer most likely intended given uncertain information about the linguistic input. Importantly for us, there is some evidence that comprehenders not only generate a plausible interpretation that is not licensed by the linguistic input, but that they actually build a syntactic representation of the unlicensed interpretation. Implausible sentences with a double object construction, like '*The mother gave the candle the daughter*', have been found to syntactically prime the prepositional dative construction of the plausible alternative (Momma & Slevc, 2015). This is consistent with a speech error reversal system proposed by Frazier and Clifton (2015). According to this account, comprehenders use their knowledge of the production system – specifically, what kind of speech errors frequently occur - to repair the input they receive. This does not conflict with the mechanisms suggested by the noisy channel model, but it makes the additional claim that comprehenders actually repair the input on a structural level.

Further evidence for structural repair comes from a recent study by Ivanova et al. (2017) that examines the processing of anomalous sentences. By definition, syntactically

anomalous sentences cannot be parsed in a way that would be consistent with the grammar. In their study, Ivanova et al. used sentences like '*The waitress the book to the monk*' that were missing a verb; there is no interpretation that is consistent with the anomalous linguistic input. Interestingly, they found that these anomalous sentences syntactically primed sentences with a prepositional dative construction, i.e. the structure the anomalous sentence would have had if it had not been missing a verb. This suggests that comprehenders process even anomalous sentences by constructing a structural representation that is consistent with grammatical constraints.

In summary, there is clear evidence that under certain circumstances comprehenders systematically generate an interpretation that is not faithful to the linguistic input. However, it seems possible that this involves building grammatically well-formed structural representations that are consistent with the misinterpretation, though not completely faithful to the input.

## 2.5 Agreement and interpretation

As outlined above, cue-based memory retrieval models provide a good account of the formation of morphosyntactic dependencies such as subject-verb agreement in sentence processing. However, the ultimate goal of comprehension is not to establish dependencies between items to check formal features, but to derive the intended interpretation by building a structural representation of the input. The goal of this chapter is to investigate how the output of memory retrieval operations for checking formal features impacts structural representations. If the representation is restructured and the retrieval output is integrated in the subject position, this would lead to misinterpretation

of the attractor as the thematic subject. This would result in an interpretation not consistent with the linguistic input, but not because comprehenders are engaging in shallow parsing. Instead, the misinterpretation would be a systematic result of the basic properties of the memory system subserving language comprehension. Here, we briefly review the studies that we are aware of that address the question of whether the attractor is misanalysed as the subject in agreement attraction.

Thornton and MacDonald (2003) conducted a series of experiments examining the impact of whether the attractor was a plausible or implausible subject for the verb. In two production studies, participants were presented with a preamble containing two nouns (*'The album by the classical composers'*) and a verb that had to be used to form a complete sentence. They manipulated whether the verb could have both the head noun and the attractor or only the head noun as a plausible (passive) subject and found that agreement attraction error rates were increased when the plural attractor was a plausible subject. The comprehension experiment also showed plausibility effects as reflected in an increase in reading time at the verb in the presence of a plural attractor when both the head noun and the attractor were a plausible subject, which is reminiscent of the semantic interference found by Van Dyke and McElree (2006), mentioned above. However, the comprehension experiment did not include ungrammatical sentences to test for agreement attraction effects. Therefore, the data is not directly informative about how misretrieval for formal feature checking can alter interpretations.

Pittman and Smyth (2005) replicated Thornton and MacDonald's production results. They also added a new component to the task in order to investigate whether participants misrepresent the attractor as the subject: after repeating the preamble and

completing the sentence using the given predicate, participants were presented with a choice of two predicates. They had to continue the sentence using ‘*and*’ followed by whichever of the two predicates they chose. One of the predicates was always a semantic match for the head noun and the other for the attractor. For example, for a preamble like ‘*The boy by the trees*’ the choice would be between ‘*chubby*’ and ‘*green*’. The rates of predicate selection errors show that participants were more likely to choose a predicate that was only a good semantic fit for the attractor in agreement attraction configurations with a singular head noun and a plural attractor, but only when they had made an agreement attraction error in the sentence. This indicates that participants sometimes misinterpreted the attractor as the subject and were more likely to do so when the attractor erroneously controlled agreement. However, even after an agreement error with a singular head noun and plural distractor, the rate of choosing the incorrect predicate was still only around 12%. As outlined above, if the retrieval output for agreement checking is used to change the existing parse of the sentence, a possible consequence of misretrieval in agreement attraction is that comprehenders might misinterpret the attractor as the subject of the sentence. However, as other work argues that the mechanisms responsible for agreement attraction effects may differ between production and comprehension (Acuna-Farina, 2009, 2012; Acuna-Farina et al., 2014; Tanner et al., 2014), these data are suggestive but do not allow us to draw definitive conclusions about the impact of misretrieval on the structural representation of the sentence in comprehension.

Lau et al. (2008) used inverted pseudoclefts in a self-paced reading experiment to address the question whether the attractor is misinterpreted as the subject by testing for

plausibility effects at the thematic verb. They used sentences like '*The phone by the toilets was/were what Patrick used/dialed/flushed/embarrassed*', in which they manipulated grammaticality and the plausibility of the head noun and the attractor as a thematic subject by varying the verb. If agreement attraction triggers structural reanalysis and results in the attractor occupying subject position in the mental representation, the plausibility match between the attractor and the verb should matter. However, the results only show a main effect of head noun plausibility with participants exhibiting a slow-down at the thematic verb when the head noun of the subject was not a plausible match. There was no interaction with attraction context or the plausibility of the attractor. Lau et al. conclude that the misretrieval of the attractor does not lead to thematic subject reassignment, meaning that the misretrieval is selective for formal feature satisfaction. However, this study used inverted pseudoclefts, which is not a structure used in other agreement attraction studies. It requires retrieval of the subject not just for agreement checking at the inflected auxiliary, but again at the wh-word before the main verb is encountered, which might have influenced their results. We address this question in Experiment 1 using a novel dual-task design that provides a very clear measure of which noun phrase comprehenders took to be the subject.

## 2.6 Experiment 1: adjective-choice task

In Experiment 1, we used a novel dual-task paradigm to investigate whether agreement attraction leads comprehenders to erroneously interpret the attractor as the subject of the sentence. Misinterpretation of the attractor as the thematic subject would

indicate that the retrieval output for agreement checking is integrated into the subject position of the structural representation, replacing the actual subject.

We developed a dual-task paradigm combining self-paced reading with a forced-choice task. Participants read sentence fragments and had to complete them by selecting an adjective that was either compatible with the head noun of the subject or the attractor noun. The choice of adjective on each trial is indicative of whether the attractor was misrepresented as the subject. If erroneously retrieving the attractor in the process of agreement checking is necessarily linked to structural reanalysis, all trials on which agreement attraction occurs should have a final interpretation in which the attractor is the thematic subject. We would expect to see a higher rate of participants choosing the adjective that matches only the attractor in an agreement attraction configuration, i.e. with an ungrammatical verb and a plural attractor. If, however, misretrieval of the attractor for agreement checking does not force the parser to engage in restructuring, this task also allows us to investigate whether agreement checking and the structural representation can interact at all. If the error-driven retrieval process in agreement checking is completely independent from structure building, comprehenders should not be more likely to choose the attractor-matching adjective in the agreement attraction condition.

The nature of the dual-task paradigm also makes it possible to analyze not only adjective choice and overall reading times, but also to take adjective choice on each trial into consideration when analyzing reading times. Overall, we expected to find a typical agreement attraction profile for the self-paced reading data: a slow-down in ungrammatical conditions, ameliorated by the presence of a plural attractor. If agreement

attraction causes comprehenders to mistake the attractor for the subject, this should be reflected by choosing the attractor-matching adjective. Consequently, in the reading time data we would expect a large attraction effect for trials on which the head-matching adjective was chosen. In contrast, we would not expect to see any attraction for trials that culminated in a head-matching adjective choice. However, if misretrieval of the attractor does not necessarily result in restructuring, the reading time data should show agreement attraction regardless of adjective choice. Nevertheless, if misretrieval is one of a number factors contributing to the likelihood of restructuring, the attraction effect might still be stronger for the trials on which the attractor-matching adjective was chosen.

### *2.6.1 Participants*

64 native speakers of American English, who had all passed a native speaker proficiency test, were recruited via Amazon Mechanical Turk for monetary compensation. None of the subjects participated in either of the norming studies for the experimental items.

### *2.6.2 Materials*

In Experiment 1, there were 48 items sets in 4 conditions. Each item consisted of a sentence fragment for self-paced reading and two adjectives for the sentence-final adjective-choice task. The sentence fragments all had a complex subject with a singular head noun and a prepositional modifier containing the attractor. The subject was followed by an inflected form of ‘*be*’ and two adverbs. The sentence-final adjective was displayed as a forced-choice task: one adjective was a plausible match only for the head noun of the

subject and the other only for the attractor, as illustrated in (3). We manipulated attractor number (singular/plural) and grammaticality (grammatical/ungrammatical).

(3)

- a) The boy by the tree is really very CHUBBY / GREEN
- b) The boy by the tree are really very CHUBBY / GREEN
- c) The boy by the trees is really very CHUBBY / GREEN
- d) The boy by the trees are really very CHUBBY / GREEN

The items were distributed across four lists in a Latin Square design. In addition to the 48 experimental items, each list also contained 72 filler items of similar syntactic complexity for which participants also had to choose between two possible sentence-final completions.

### *2.6.3 Plausibility Norming*

Since the premise of the dual-task paradigm is that the adjective choice is informative about whether the participant has misinterpreted the attractor as the thematic subject, it is crucial that one of the adjectives is semantically plausible only for the head noun and the other only for the attractor. We conducted a plausibility rating study of simple sentences with potential head nouns and attractor nouns in subject position, varying the predicative adjective. The aim was to select 48 item sets in which one of the adjectives was rated highly plausible only for the head noun and the other only for the attractor.

30 native speakers of English participated in an adjective norming study on Ibex in which they rated 66 items in 6 conditions for plausibility on a scale from 1 (very implausible) to 7 (very plausible). All items were grammatical and the task also included 18 plausible fillers, 16 implausible fillers and 7 control items. We constructed 66 preliminary items containing a complex subject with a prepositional modifier, followed by an inflected form of *be*, two adverbs, and a sentence-final adjective. For each item, there were 8 conditions, crossing attractor number, grammaticality, and adjective plausibility. Based on these preliminary items, we constructed 66 item sets for norming, manipulating whether the subject was the head noun or the attractor noun in the 66 preliminary items. Apart from subject type (head noun vs. attractor), we also manipulated adjective type (head-match vs. attractor-match), and subject number. Since in the materials for the dual-task paradigm the head noun of the subject is always singular, the norming study included plural versions only of the attractors. This led to a total of 6 conditions, as illustrated in (1). The ratings were used to calculate the average plausibility ratings for the plausible conditions (a, d, f) and the implausible conditions for each item (b, c, e). We then selected the 48 items with the greatest difference between plausibility ratings for the plausible and the implausible conditions.

(1)

- a) The **boy** is really very *chubby*.
- b) The **boy** is really very *green*.
- c) The **tree** is really very *chubby*.
- d) The **tree** is really very *green*.

- e) The **trees** are really very *chubby*.
- f) The **trees** are really very *green*.

#### 2.6.4 Agreement Attraction Norming

The 48 chosen items were then used in a speeded acceptability judgment task to confirm that they caused the expected agreement attraction effect. 24 native speakers of American English read sentences presented word-by-word in the center of the screen with a stimulus onset asynchrony of 400ms (inter-stimulus interval: 100ms). Following each sentence, participants had 2000ms to indicate whether the sentence had been acceptable. The instructions explicitly asked them to judge sentences based on whether they sounded like natural English rather than prescriptive rules. There were 72 fillers (half grammatical) in addition to the 48 experimental items. In order to avoid exposing participants to a large number of implausible sentences, the sentence-final adjective was always the one compatible with the head noun of the subject. In the dual-task paradigm, the attraction effect in self-paced reading is measured on the verb and its spillover regions, before participants are presented with the adjectives.

The acceptance rates across conditions were analyzed with a mixed-logit model (Jaeger, 2008), excluding trials on which no response was made within 2000ms (2.5% of all trials). The acceptance rates for each condition are plotted in Figure 1. Table 1 contains the results of the mixed-logit model with grammaticality and attractor model as fixed effects (sum-coded). The random effects structure included by-subject and by-item

random intercepts and by-subject random slopes for grammaticality.<sup>2</sup> As expected, grammatical sentences were more likely to be judged acceptable than ungrammatical sentences (89.4% vs. 16.7%). Sentences with a plural attractor were also more likely to be accepted than sentences with a singular attractor (49.5% vs. 57.1%), but this effect was driven by the higher rate of acceptance of ungrammatical sentences with plural attractors. Participants were more likely to accept an ungrammatical sentence when the number of the attractor was plural (25.1% for ungrammatical sentences with a plural attractor compared to 8.2% for those with a singular attractor). This indicates that comprehenders indeed experience attraction with this particular item set, making these materials suitable for the novel dual-task paradigm.

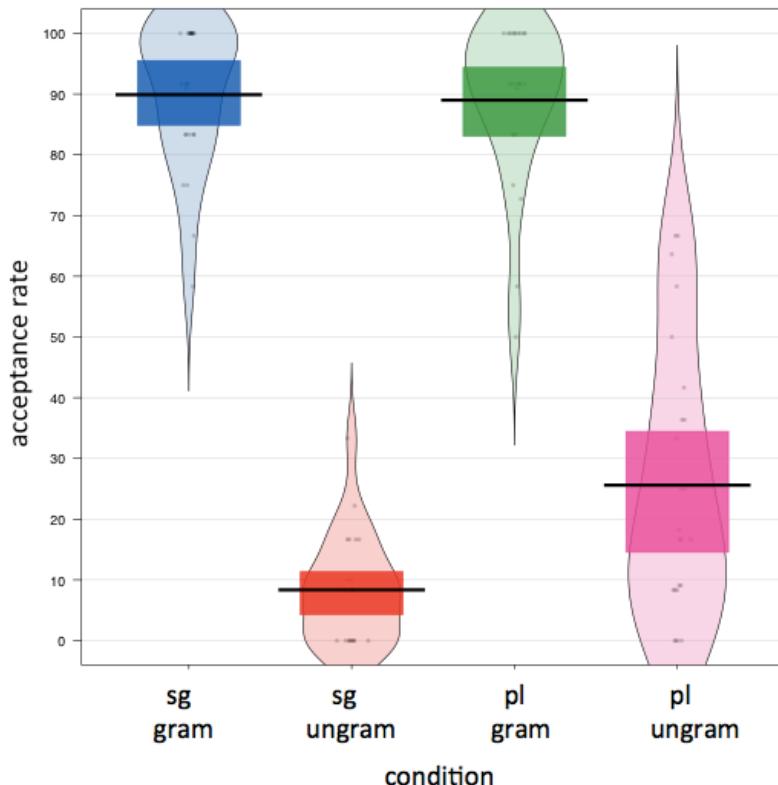


Figure 1: Acceptance rates across conditions in speeded acceptability judgment task in Experiment 1.

---

<sup>2</sup> The model also converged with by-subject and by-item random intercepts and by-subject random slopes for attractor number, but the significance of the effects does not depend on which of these models is used.

	Estimate	Std. Error	z value	p value
Intercept	0.13	0.23	0.57	0.569949
Grammaticality	2.41	0.17	13.89	< 2e-16
Attractor number	-0.39	0.10	-3.79	0.000148
Grammaticality x attractor number	0.45	0.10	4.31	1.62e-05

**Table 1:** Results from the mixed logit model in speeded acceptability judgment task in Experiment 1

### 2.6.5 Procedure

The sentences were presented in a self-paced reading paradigm with centered display using Ibex software (Drummond, 2016). Participants had to press the spacebar to see each new word and only one word at a time was visible. When they pressed the spacebar to reveal the final word of the sentence, the two adjectives for the forced-choice task appeared on the screen simultaneously, one to the left of the center and one to the right. The order in which the adjectives were displayed was randomized for each participant. Once the two adjectives appeared, participants had 3000ms to choose one of them by pressing the ‘f’-key for the one on the left or the ‘j’-key for the one on the right. If no response was made within 3000ms, the adjective-choice task timed out and the experiment moved on to the next trial.

### 2.6.6 Analysis

Trials on which there was no response within the 3000ms deadline were excluded from all analyses reported here (1.4% of experimental trials, 42 of 3072 trials). We analyzed responses to the adjective-choice task with a mixed logit model (Jaeger, 2008) using the lme4 package (Bates, Maechler, Bolker & Walker, 2015) in the R computing

environment (R Development Core Team, 2017). The model included attractor number and grammaticality as fixed effects (sum-coded) and by-subject and by-item random intercepts. The model was initially fitted with the maximal random effects structure, which was then simplified until the model converged (Barr, Levy, Scheepers & Tily, 2013).

Although the main focus of the experiment was the adjective-choice task, we also analyzed the self-paced reading data. The regions of analysis were the verb and its spillover region (first adverb). Reading times exceeding a threshold of 2000ms were not included in the analysis, leading to the exclusion of less than 0.2% of experimental trials in each region of analysis. RTs were log transformed and analyzed using linear mixed effects models with attractor number, grammaticality and adjective choice as fixed effects. The final model included random by-subject and by-item intercepts. In addition, we also split the SPR data based on adjective choice on each trial and conducted a response-contingent RT analysis.

### 2.6.7 *Results*

#### **Adjective-choice task**

The percentage of trials on which a head-noun matching adjective was chosen for each of the experimental conditions is plotted in Figure 2 and the results from the model are presented in Table 2. There was a significant main effect of grammaticality ( $p < 0.01$ ): participants were more likely to choose the adjective that matched only the subject's head noun in grammatical than in ungrammatical sentences. There was also a significant interaction between grammaticality and attractor number ( $p < 0.05$ ). In

ungrammatical sentences participants were less likely to choose the head-matching adjective when the attractor was plural. As can be seen in Figure 2, the overall accuracy rates in the forced-choice task were very high. The rate of choosing the attractor-matching adjective was only 5.6% higher in the attraction condition (ungrammatical with a plural attractor: 16.4%) than in the grammatical condition with a plural attractor (89.2%).

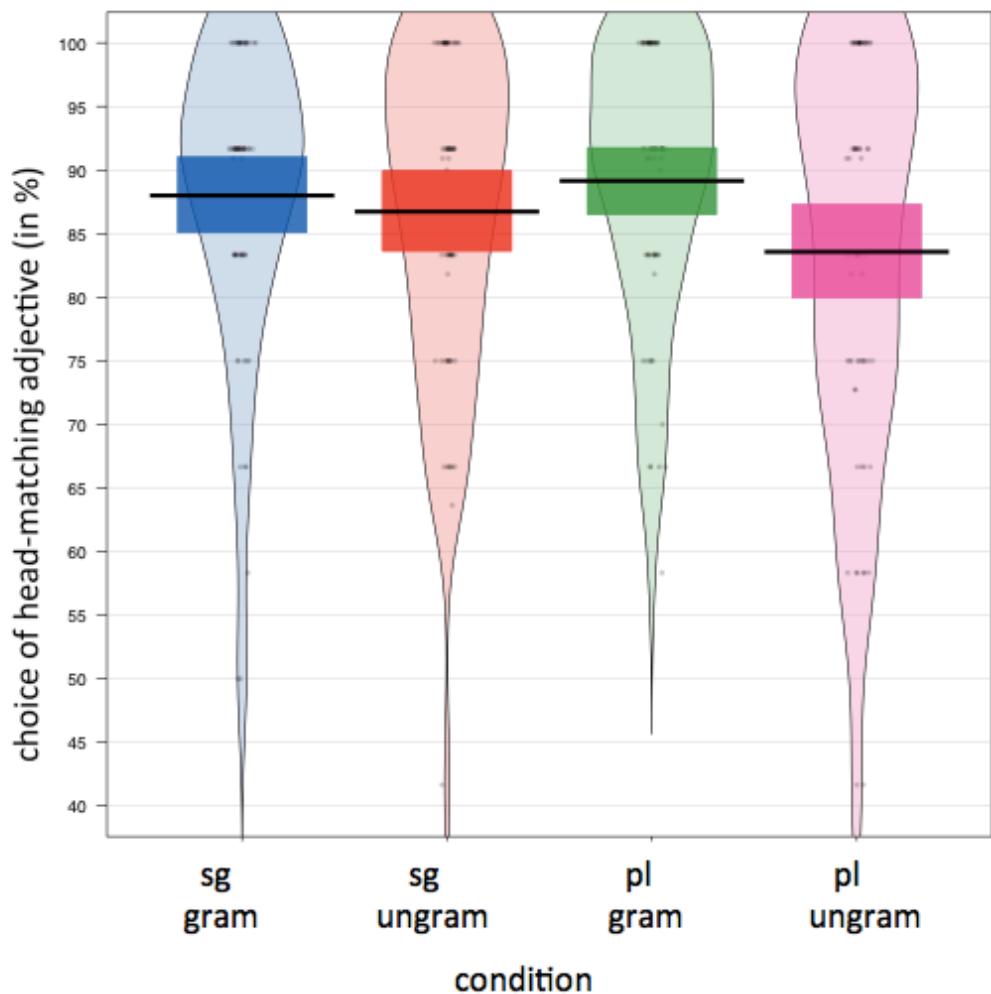


Figure 2: Percentage of trials with a head-matching adjective choice across conditions in Experiment 1

	Estimate	Std. Error	z value	p value
Intercept	2.57	0.23	11.39	< 2e-16
Grammaticality	0.18	0.06	3.00	<b>0.00271</b>
Attractor number	0.04	0.06	0.63	0.52918
Grammaticality x attractor number	-0.13	0.06	-2.16	<b>0.03064</b>

Table 2: Results from mixed logit model for adjective choice in Experiment 1

Figure 3 plots mean RTs for head-matching and attractor-matching adjective responses across conditions. For ease of readability these are raw RTs, but for the analysis RTs were log transformed. Results of the linear mixed effects model with fixed effects of grammaticality, attractor number and adjective choice are presented in Table 3. There was a significant effect of adjective choice ( $t = -3.17$ ). Visual inspection of the plot shows that trials on which the attractor-matching adjective was chosen were slower than when the head-matching adjective was chosen. The RT difference between head and attractor compatible adjective responses was larger in the grammatical than the ungrammatical conditions. However, this interaction between grammaticality and adjective choice was only marginally significant ( $t = -1.95^3$ ).

---

<sup>3</sup> The interaction is also marginally significant with raw RTs ( $t = -1.72$ ).

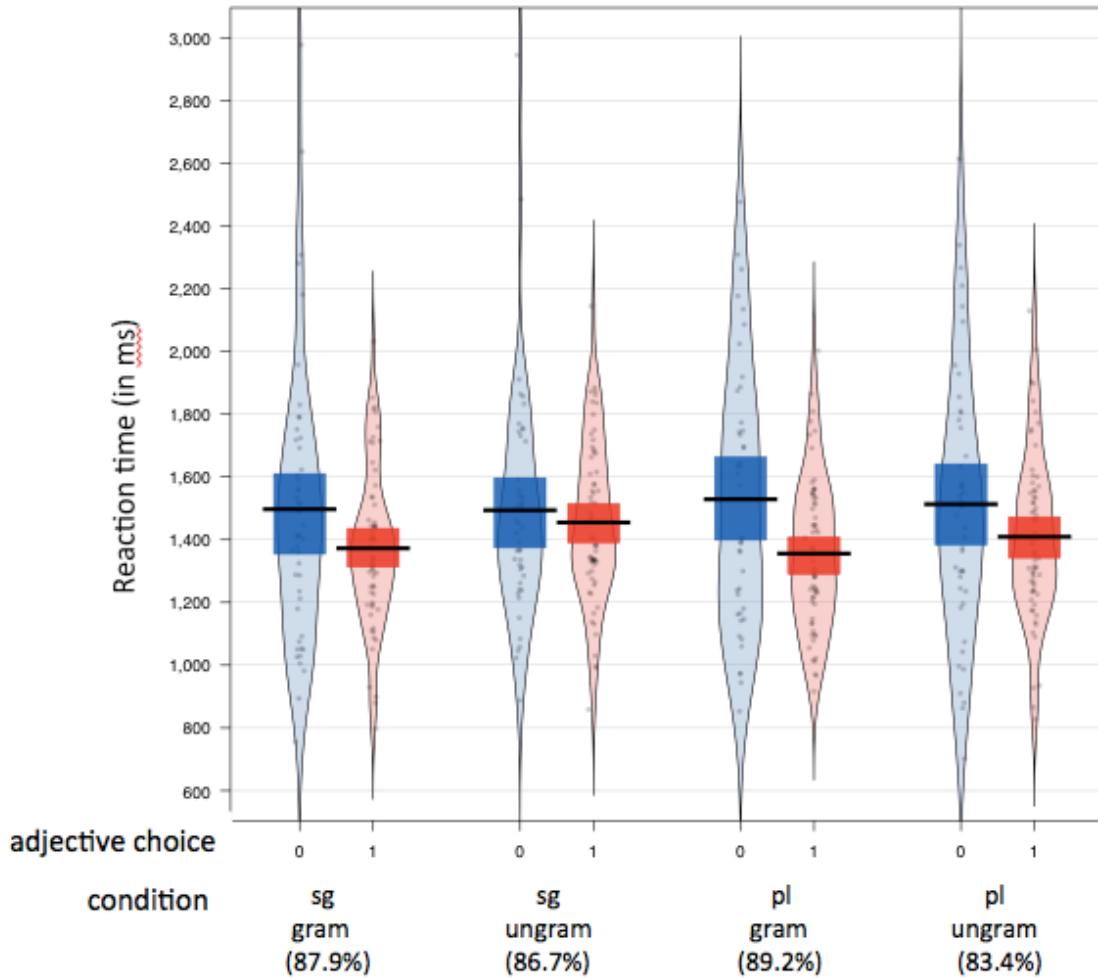


Figure 3: mean RTs split by adjective choice (attractor-matching response in blue: 0; head-matching response in red: 1) in each experimental condition in Experiment 1. Proportion of head-noun compatible responses beneath condition labels.

	Estimate	Std. Error	z value
Intercept	7.2485	0.0277	<b>262.20</b>
Grammaticality	-0.0001	0.0117	-0.01
Attractor number	-0.0016	0.0117	-0.14
Adjective choice	-0.0436	0.0138	<b>-3.17</b>
Grammaticality x Attractor number	-0.0099	0.0116	-0.85
Grammaticality x Adjective choice	-0.0245	0.0126	-1.95
Attractor number x Adjective choice	0.0130	0.0125	1.04
Grammaticality x Attractor number x Adjective choice	0.0048	0.0125	0.38

Table 3: Results of linear mixed effects model of response time on the adjective-choice task in Experiment 1 (using log transformed RTs)

## Self-paced reading

The region-by-region average reading times in Experiment 1 are plotted in Figure 4. Table 4 and 5 contain the results of the linear mixed effects models for the verb region and the spillover region. For ease of readability, Figure 4 plots raw RTs, but analyses were performed on log transformed RTs.

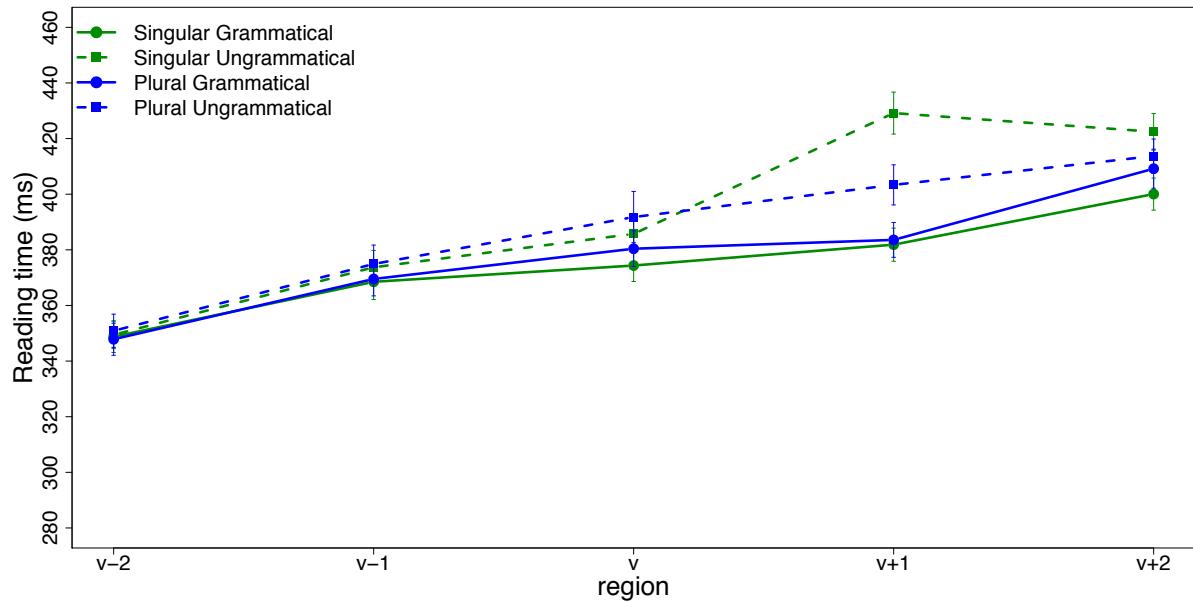


Figure 4: Region-by-region mean raw reading times in Experiment 1. Error bars indicate standard error of the mean.

	Estimate	Std. Error	z value
Intercept	5.848	0.043	134.75
Grammaticality	-0.002	0.008	-0.20
Attractor number	-0.003	0.008	-0.35
Adjective choice	-0.003	0.008	<b>-0.40</b>
Grammaticality x Attractor number	0.013	0.008	1.65
Grammaticality x Adjective choice	-0.005	0.008	-0.64
Attractor number x Adjective choice	-0.003	0.008	-0.2
Grammaticality x Attractor number x Adjective choice	0.019	0.008	<b>2.43</b>

Table 4: Results of linear mixed effects model in the verb region in Experiment 1 (sum coded; using log transformed RTs)

	Estimate	Std. Error	z value
Intercept	5.898	0.043	<b>136.06</b>

Grammaticality	-0.024	0.009	<b>-2.83</b>
Attractor number	0.009	0.009	1.08
Adjective choice	0.006	0.009	0.66
Grammaticality x Attractor number	< -0.001	0.009	-0.01
Grammaticality x Adjective choice	0.014	0.009	1.69
Attractor number x Adjective choice	0.008	0.009	-0.97
Grammaticality x Attractor number x Adjective choice	0.014	0.009	1.67

Table 5: Results of linear mixed effects model in the spillover region in Experiment 1 (sum coded; using log transformed RTs)

The only significant effect in the verb region was a three-way interaction between grammaticality, attractor number and adjective choice ( $t = 2.43$ ). In the spillover region there was a main effect of grammaticality ( $t = -2.83$ ), with increased reading times for ungrammatical sentences. The three-way interaction between grammaticality, attractor number and adjective choice remained marginally significant in the verb's spillover region ( $t = 1.67$ ).

### Response-contingent self-paced reading

The nature of the dual-task paradigm allows us to examine reaction time profiles of trials based on adjective choice. Figure 5 shows the average reading time per region for each condition for trials on which the (correct) head-matching adjective was chosen. The plot looks almost identical to the overall SPR plot. Visually, there is a very clear slow-down for the ungrammatical conditions in the verb's spillover region, which is ameliorated for ungrammatical sentences with a plural attractor. Statistical analysis confirms this: While there is no significant effect in the verb region (Table 6), in the verb's spillover region grammaticality, attractor number and their interaction all had a significant effect on reading times (Table 7). As expected, agreement violations led to a slowdown in the verb's spillover region compared to sentences with correct subject-verb

agreement, as reflected in the main effect of grammaticality ( $t = -6.61$ ). Reading times in the spillover region were longer for sentences with a singular than a plural attractor ( $t = 2.96$ ). This result was not expected and seems to be attributable to the large difference between the ungrammatical conditions with singular compared to plural attractors. Crucially, reading times show an agreement attraction pattern with the slowdown associated with a subject-verb number agreement violation being much reduced in the presence of a plural attractor (interaction between grammaticality and attractor number:  $t = -2.49$ ).

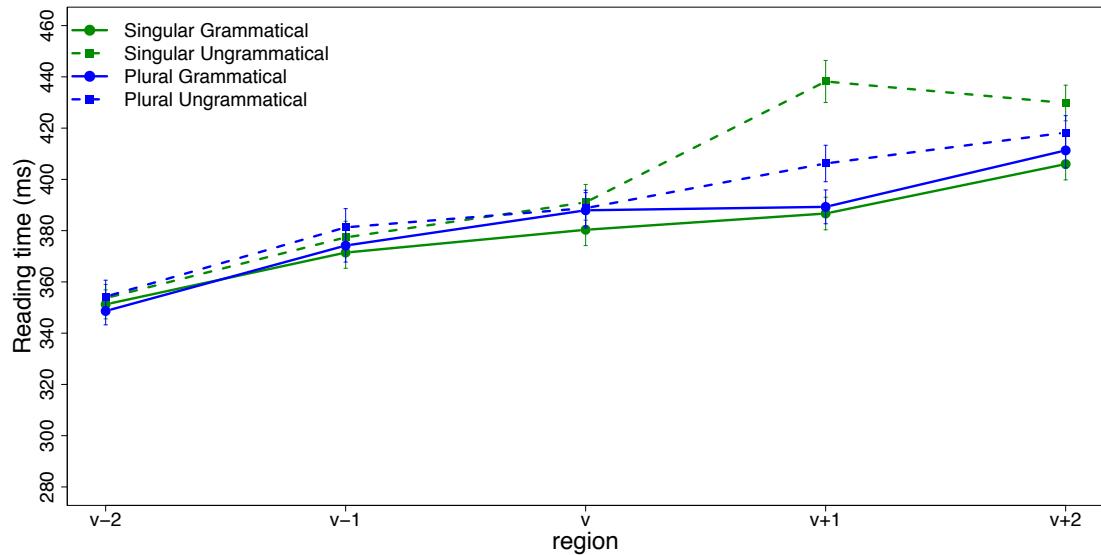


Figure 5: Region-by-region mean reading times for trials on which the (correct) head-matching adjective was chosen (Experiment 1). Error bars indicate standard error of the mean.

	Estimate	Std. Error	t value
Intercept	5.851	0.043	<b>136.17</b>
Grammaticality	0.004	0.006	0.69
Attractor number	0.001	0.006	0.11
Grammaticality x attractor number	-0.006	0.006	-1.15

Table 6: Results of linear mixed effects model in verb region in Experiment 1 for trials on which the head-matching adjective was chosen (sum coded; using log transformed RTs)

	Estimate	Std. Error	t value
Intercept	5.892	0.043	<b>137.32</b>
Grammaticality	-0.039	0.006	<b>-6.61</b>
Attractor number	0.017	0.006	<b>2.96</b>
Grammaticality x attractor number	-0.015	0.006	<b>-2.49</b>

Table 7: Results of linear mixed effects model in spillover region in Experiment 1 for trials on which the head-matching adjective was chosen (sum coded; using log transformed RTs)

Average reading times for trials on which participants chose the attractor-matching adjective are plotted in Figure 6 and the results of the statistical analyses in the verb and the verb's spillover region are provided in Table 8 and 9. It should be noted that the high accuracy on the adjective choice task meant that the sample size for this analysis was much smaller. Visual inspection of the plot reveals a very different pattern than for the head noun compatible adjective response trials. The statistical results show no significant effects in either the verb or the verb's spillover region, including no evidence for an agreement attraction effect in either the verb or the verb's spillover region. If anything, the pattern appears to be in the opposite direction. Numerically, the average reading time for ungrammatical sentences with a plural attractor is slower in the verb region than for the other three conditions.

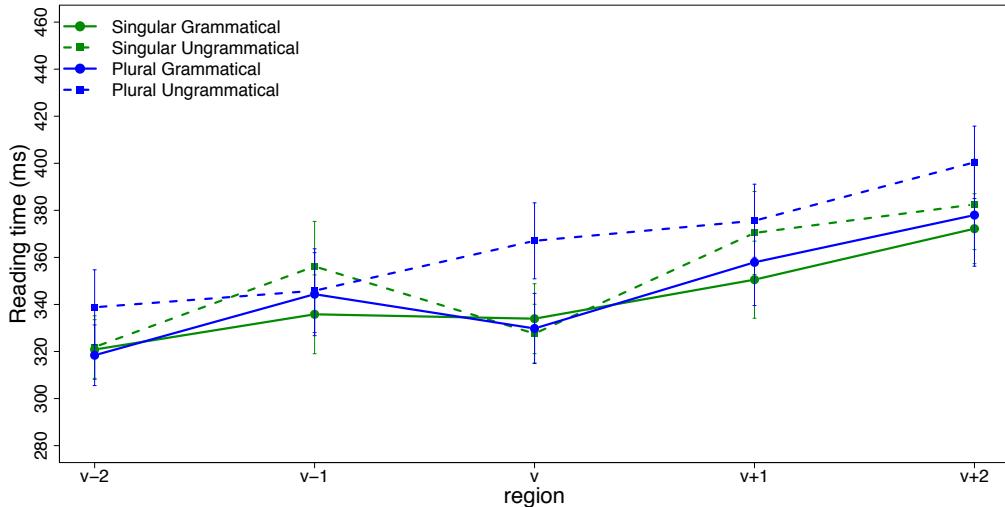


Figure 6: Region-by-region mean reading times for trials on which the attractor-matching adjective was chosen (Experiment 1). Error bars indicate standard error of the mean.

	Estimate	Std. Error	t value
Intercept	5.816	0.053	<b>110.34</b>
Grammaticality	-0.001	0.018	-0.04
Attractor number	-0.005	0.018	-0.25
Grammaticality x attractor number	0.023	0.017	1.35

Table 8: Results of linear mixed effects model in verb region in Experiment 1 for trials on which the attractor-matching adjective was chosen (sum coded; using log transformed RTs)

	Estimate	Std. Error	t value
Intercept	5.877	0.058	<b>101.58</b>
Grammaticality	-0.009	0.020	-0.46
Attractor number	-0.003	0.020	-0.16
Grammaticality x attractor number	0.003	0.019	0.14

Table 9: Results of linear mixed effects model in spillover region in Experiment 1 for trials on which the attractor-matching adjective was chosen (sum coded; using log transformed RTs)

### 2.6.8 Discussion

In Experiment 1, participants showed a clear agreement attraction effect in the overall self-paced reading data, just as expected. The self-paced reading results show that the average slowdown for ungrammatical compared to grammatical sentences with a

singular attractor was 47ms (grammaticality effect). This slowdown was reduced to 15ms when the attractor was plural, which suggests that attraction occurred on a large number of trials with an attraction configuration. If misretrieval of the attractor obligatorily leads to structural reanalysis, this would be reflected in the adjective choice: on all trials on which attraction occurred participants should have picked the attractor-matching adjective. However, the subset of trials on which participants chose the adjective that was only compatible with the attractor was extremely small across all conditions (less than 17%). Importantly, comprehenders overwhelmingly chose the adjective compatible with the subject's head noun even for ungrammatical sentences with a plural attractor. The comparison between the self-paced reading data and the adjective choice data indicates that participants experienced agreement attraction on many more trials than those on which they chose the attractor-matching adjective. This demonstrates that it cannot be the case that misretrieval of the attractor during agreement processing necessarily causes structural reanalysis.

Further evidence against the idea that agreement attraction reflects restructuring comes from the response contingent analysis of the self-paced reading data. If misretrieval of the attractor automatically triggered restructuring, we would expect to see a very strong agreement attraction effect for the trials on which the attractor-matching adjective was chosen and no attraction on trials on which the head-matching adjective was chosen. However, in fact, the self-paced reading data does not show any agreement attraction if we look only at trials with an attractor-matching response. Since there was only a very small subset of these trials, this sub-analysis is very underpowered and has to be interpreted with caution. More importantly, there is a clear pattern of agreement

attraction when we consider only the trials on which the head-matching adjective was chosen. This sub-analysis contains the majority of trials and does not suffer from being underpowered. Under a view in which misretrieval of the attractor necessarily leads the parser to reanalyze it as the subject, we would expect no attraction on these trials.

Together these response contingent analyses show that it cannot be the case that misretrieval of the attractor in error-driven retrieval necessarily triggers restructuring.

The results of Experiment 1 demonstrate that error-driven retrieval for agreement checking is not inextricably linked to restructuring, but they also suggest that misretrieval and misinterpretation do not seem to be completely independent. The advantage of the dual-task paradigm in this experiment is that we can obtain a very explicit measure of what participants interpret as the subject on each individual trial: while comprehenders very rarely chose the adjective compatible with the attractor, they did so significantly more frequently in ungrammatical sentences with plural attractors. This suggests that the attractor is at least occasionally misrepresented as the subject and that error-driven retrieval in response to the detection of an agreement violation might contribute to the likelihood of structural reanalysis. It is possible that misretrieval of the attractor triggers restructuring if the verb simultaneously contains additional semantic cues in favor of the alternative structure. Unfortunately, the nature of the task also means that the number marking always had to appear on copular ‘*be*’, which is semantically impoverished.

Another limitation of this study is that in our materials the subject’s head noun was always the first noun in the sentence, making it very salient. In fact, participants could have used a task-specific strategy in which they rely on sentence-initial position to establish subjecthood in the adjective choice task. In future research, this potential task-

specific heuristic could be prevented by including items in which subjecthood and sentence-initial position are dissociated.

While the results of Experiment 1 point towards a very interesting interaction between error-driven retrieval for agreement checking and restructuring, it should be acknowledged that a potential explanation for this pattern can be provided without assuming that it is directly linked to agreement attraction as such. The average reading times for trials with an attractor-matching response were faster than for trials on which the head-matching adjective was chosen. There were also no effects of grammaticality, attractor number or their interaction in either the verb or the verb's spillover region for these trials. Again, it needs to be noted that this was only a small subset of all the trials and this sub-analysis was very underpowered. Nevertheless, this suggests that attractor-matching responses might occur on trials on which participants were not paying attention. In that case, the mental representation of the subject might be less well encoded and more unstable than usual. On some of these trials, the attractor might have been analyzed as the subject even before the verb was encountered. Without a robust structural representation of the input prior to the verb, participants might not even have committed to a structural representation, meaning neither of the NPs is in subject position. The plural marking on the verb could then have served as a cue to pick the option with the matching number feature, explaining why attractor compatible adjectives were chosen more frequently in ungrammatical sentences with plural attractors. Although this relies on a match between the attractor's number feature and the retrieval cues of the verb, it is not identical to the mechanism we usually assume for agreement attraction. Unfortunately, we have no data on how confident participants were about their adjective choices. If attractor compatible

adjective choices really were due to inattention, participants would be expected to be less confident about their choice on these trials.

## 2.7 General Discussion

Overall, the results of Experiment 1 show that agreement attraction does not necessarily result in the misinterpretation of the attractor as the subject. The error-driven retrieval process triggered by the detection of a subject-verb agreement violation does not always lead to structural reanalysis when the attractor is misretrieved. Rather, the error-driven process appears to be mostly limited to a feature-checking step, which can apparently return an answer that is inconsistent with the overall structure and yet still result in the perception that the sentence was grammatical.

As previously discussed, the parser frequently engages in structural reanalysis when it encounters an error signal. However, it should be noted that the proposed restructuring in agreement attraction would be fundamentally different from reanalysis in garden-path sentences. In a garden-path sentence, at the point of disambiguation, it is simply impossible to integrate the new input into the existing structure without violating structural constraints. In contrast, when the parser encounters a subject-verb agreement violation, the structural configuration for integrating the verb is there. There is only a mismatch between one of the predicted features (number) and the bottom-up input. If reanalysis is costly, it might only be deployed when the error-signal is triggered by a severe violation. Moreover, in garden path sentences, the parser assigns a different analysis to the entire previous input. In agreement attraction, misrepresenting the attractor as the subject would require excluding some of the previous input from the newly built

structure. Reanalysis might only be possible if the input that has already been assigned a structure can be completely integrated into the new structure.

If the output of retrieval for agreement checking does not necessarily have a structural impact, that suggests that in most cases agreement attraction indexes a low-level feature checking operation in the following sense: Comprehenders predict the number marking of the verb based on the subject. If the verb does not match this prediction, the mismatch triggers retrieval. The aim of this error-driven retrieval process is to check whether the verb's number marking is licensed by the agreement controller. If the agreement controller is successfully retrieved, it confirms that there is a subject-verb agreement violation. However, if instead of the actual target the attractor is erroneously retrieved, its number feature can license the number marking on the verb and it is no longer perceived as an agreement violation. This relies on a low-level morphosyntactic checking mechanisms in which only the retrieved item's number feature is checked, since the misretrieved attractor does not match all of the verb's retrieval cues. The question arises what information guides retrieval for such a low-level feature checking operation. An obvious possibility would be that it is exactly the features that define the subject-verb agreement dependency in the grammar. However, this is an error-driven process, and it is possible that as a repair mechanism it also uses grammatically irrelevant information to retrieve the agreement controller from memory I turn to this question in Chapter 3 and 4.

### *2.7.1 The final representation of agreement attraction sentences*

The question whether the misretrieval of the attractor in agreement processing triggers restructuring has important implications for whether grammatical illusions can

arise with mental representations that are not actually grammatical. If misretrieval of the attractor necessarily triggers restructuring, agreement attraction would only occur when the verb's number marking is actually licensed by the final representation: with the plural attractor misrepresented in subject position, there would be no agreement violation. This would suggest that grammatical illusions arise on the basis of final representations that are not consistent with the input, but are consistent with the grammar.

In contrast, if the output of retrieval is only used to check that the number marking on the verb is consistent with the number feature of the agreement controller, misretrieval of a number matching attractor would simply signal that there is no agreement violation after all. However, the final structural representation in memory would still contain a number mismatch between the actual subject and the verb and would therefore not be consistent with the grammar.

In an ungrammatical sentence without a number-matching attractor, the error-driven search for the agreement controller returns the subject and confirm that there is an agreement violation. Thus the comprehender perceives the sentence as ungrammatical. Based on evidence from implausible and anomalous sentences discussed above (Momma & Slevc, 2015; Ivanova et al., 2017), I assume that following the rechecking failure, comprehenders revise the representation of the sentence to repair the agreement violation. According to models of sentence processing in which listeners repair speech errors by making rational inferences about noisy linguistic input, this would involve amending either the subject's or the verb's number, depending on which of these is the more reliable cue (MacWhinney, Bates & Kliegl, 1984). Therefore, the final mental representation of the sentence would be consistent with the grammar.

In contrast, if a number matching attractor is retrieved instead of the actual number-mismatching subject, that signals that there is no agreement violation after all. Due to this illusory licensing of the verb's number marking by the attractor, the comprehender does not perceive the sentence to be ungrammatical. Consequently, there is no additional repair process to revise the subject's or the verb's number and the final representation remains inconsistent with the grammar. Assuming a framework in which interpretations have to be derived from structural representations consistent with the grammar that might be considered a problem for the low-level feature checking account. However, it very much depends on when exactly we think agreement has to be licensed in online processing. If the verb's number only matters at the point at which it is integrated into the structure, illusory checking due to misretrieval of the number-matching attractor would be entirely sufficient and the discrepancy between the structure and the features that were checked does not matter in processing.

The results of Experiment 1 are not compatible with an account of agreement attraction that always involves restructuring. This means that the illusory licensing of an agreement violation must be possible without a final mental representation of the sentence in which it is actually licensed. However, the slightly higher proportion of attractor-matching adjective choices in agreement attraction configurations that we found in Experiment 1 suggests that a very small subset of trials on which the attractor is misretrieved does lead to the misrepresentation of the attractor as the subject. In this small subset, the final mental representation does actually license the verb's number marking. This suggests that what we observe as the phenomenon of agreement attraction

in measures such as speeded acceptability judgments and self-paced reading may not reflect exactly the same underlying process on all trials.

### 2.7.2 *A third possibility: Revising the subject's number feature*

The results of Experiment 1 strongly suggests that the error-driven retrieval process that results in agreement attraction is a low-level rechecking process without any structural impact. However, one could imagine a third possibility that falls in between a structural reanalysis account and a simple feature-checking model. It is possible that the representation of the sentence is altered based on the retrieval output, but much less drastically than in the complete structural reanalysis model. In particular, the parser could use the number feature of the erroneously retrieved attractor to substitute the number feature of the subject as it was originally encoded in memory. For example, in a sentence with an agreement violation and a number-matching attractor, such as '*The key to the cabinets are rusty*', the process would be the following: The subject is correctly encoded as singular and the parser predicts a singular verb. Upon encountering '*are*', there is a mismatch between the number feature of the prediction and the bottom-up input, which triggers a search for the agreement controller in memory. If the number-matching attractor is erroneously retrieved, its number feature is used to overwrite the subject's current number feature. Unlike the pure rechecking process, this renumbering account predicts interpretive consequences of misretrieval, but would result in a final representation that is consistent with the grammar as a whole and does not contain an agreement violation

. If misretrieval of the number matching attractor results in the change of the subject's number feature, this could in a sense be considered a representational account since it involves misrepresenting the number of the subject. However, it would be fundamentally different from other misrepresentation accounts: In representational accounts like feature percolation and the Marking and Morphing model, agreement attraction is a consequence of misencoding the subject's number feature prior to encountering the verb. In contrast, if the parser changes the subject's number feature based on the output of retrieval in agreement processing, misrepresenting the subject's number information would be a consequence of misretrieval, rather than the cause of it. While comprehenders would not mistake the attractor for the thematic subject, the misretrieved attractor's number feature would overwrite the number feature of the subject in memory. In that sense it is also a reanalysis account, especially compared to the low-level feature checking operation outlined above.

Patson and Husband (2016) investigated if comprehenders mistakenly interpret the number feature of the subject as plural. They used comprehension questions to explicitly probe participants' interpretation of the subject's number feature: a sentence like '*The key to the cabinets are on the table*' would be followed by the question '*Was there more than one key?*'. They found that comprehenders were more likely to agree that there were multiples of the entity denoted by the head noun when there was a plural attractor or a plural verb, and that this effect was strongest in agreement attraction configurations, in which both the attractor and the verb were plural. Patson and Husband interpret these results to show that comprehenders do indeed sometimes misrepresent the number of the complex subject noun phrase, not only in agreement attraction

configurations, but whenever there is a plural feature present on the attractor or the verb. However, a recent self-paced reading experiment by Dempsey et al. (2016) found that the processing of a plural pronoun referring back to a complex noun phrase with a singular head noun was not facilitated when it contained a plural noun inside its prepositional modifier (*'the key to the cabinets'*). This indicates that the complex NP's number information had not been misrepresented by virtue of containing a plural element. Moreover, Patson and Husband's online reading time data is not consistent with the automatic misrepresentation of complex noun phrases, as it shows no evidence of disrupted processing at the verb in grammatical sentences when the attractor was plural (*'The key to the cabinets was...'*). If comprehenders misrepresent the number feature of the subject in the presence of a plural attractor, this should be reflected in processing difficulties at the verb in grammatical sentences with plural attractors. One alternative explanation of the Patson and Husband results is that answers to explicit comprehension questions are not always an accurate reflection of the representation built during the earlier processing of the sentence. Here, having heard both a plural attractor and a plural verb might raise the confidence that there was something plural in the sentence and thus make comprehenders more likely to answer affirmatively. Consequently, while the Patson and Husband results are intriguing, they do not provide conclusive evidence that comprehenders misrepresent the subject's number feature by virtue of the presence of a plural attractor. Nevertheless, in light of the recent evidence that comprehenders sometimes carry out structural repairs on the input, the possibility that comprehenders misrepresent the subject's number information in agreement attraction in comprehension cannot be dismissed without further research.

## 2.8 Conclusion

In this chapter, we explored the relationship between the output of error-driven retrieval in agreement processing and the final structural representation of the sentence. We used a novel dual-task design to assess whether comprehenders misinterpret the attractor as the subject when they experience agreement attraction. The results show that comprehenders do not misinterpret the attractor as the subject on all trials on which agreement attraction occurs, indicating that misretrieval of the attractor does not necessarily trigger restructuring. While this is clear evidence that subject-verb agreement attraction is not a straightforward reflection of restructuring, misretrieval of the attractor does appear to increase the likelihood of misinterpreting the attractor as the subject. This suggests that the error-driven retrieval process in agreement checking generally involves low-level feature checking without integrating the output of retrieval into the agreement controller's position in the mental representation. Nevertheless, this low-level feature checking can at least sometimes contribute to the impetus for structural reanalysis.

The fact that restructuring is not automatically triggered when the attractor is misretrieved means that illusory licensing can occur even if there is no actual licensing in the final mental representation. While this is a very interesting finding, it is not clear that this would necessarily be the case for other grammatical illusions. It is conceivable that this discrepancy could be particular to agreement attraction: agreement as such does not contribute to the interpretation of a sentence and, unlike grammatical illusions involving dependencies that cannot be predicted such as reflexives or VP-ellipsis, it is an error-driven phenomenon. This potential difference between agreement attraction and non-error driven grammatical illusions certainly warrants further investigation.

So far in this dissertation I have been assuming that agreement attraction is an error-driven phenomenon based on timing evidence from self-paced reading (Lago et al., 2015) and eye-tracking (Dillon et al., 2013). However, another interesting aspect of Experiment 1 is that the results provide further independent evidence for an error-driven account of agreement attraction. Under a cue-based retrieval account, agreement attraction is a reflection of facilitative similarity-based interference, which can only occur if the cues on the verb trigger a search of the agreement controller in memory. A non-error driven account of agreement attraction, in which attraction reflects instances in which the agreement violation was not detected at all, assumes that the dependency between the subject and the verb always requires retrieval of the agreement controller. Thus, the relation between the subject and the verb is established only once the verb is encountered. Assuming that interpretation is derived from structure and that this is the point at which the structural relationship between the subject and the verb is established, misretrieval of the attractor as the agreement controller would always have to lead to misrepresentation of the attractor as the subject. However, the results of Experiment clearly demonstrate that this is not the case. This strongly suggests that the dependency between the subject and the verb is established predictively and that agreement attraction reflects misretrieval in an error-driven process rather than an obligatory retrieval process.

# Chapter 3

## 3.1 Outline

The previous chapter establishes that the error-driven retrieval operation responsible for the illusory licensing of agreement violations does not index a process of extensive restructuring. Instead, I propose that it is a low-level feature checking operation that only rarely has an impact on the structural representation of the sentence. In this chapter, I ask how faithful the retrieval cues in this error-driven process are in relation to the grammar. Evidence from a series of speeded acceptability judgments and self-paced reading experiments demonstrates that retrieval models need to include cues as abstract as the terms in which the grammatical dependencies are stated. Conjoined singular NPs, which are syntactically plural but contain only an equivocal morphological signal of plurality, caused strong attraction effects, indicating that the verb's number retrieval cue is specified in more abstract terms and does not specifically target only the unequivocal exponent of the abstract feature (plural '*-s*'). However, we also found a numerically much smaller attraction effect with attractors with conjoined adjectives, which are not syntactically plural and do not license plural agreement in the grammar. We hypothesize that this is because '*and*' frequently co-occurs with syntactic plurality and has therefore become weakly associated with the plural retrieval cue. Taken together, these findings suggest that the feature primarily targeted in memory retrieval operations linked to agreement processing are more abstract than a specific exponent of the abstract category syntactic number, but that due to the associative nature of cues and features, surface cues

that are imperfect correlates of syntactic plurality like ‘*and*’ can also interfere to a smaller extent.

### 3.2 Cue-based retrieval in sentence processing

Recent research has used a number of linguistic dependencies to investigate the architecture of the memory system underlying language comprehension. The findings suggest that it relies on cue-based retrieval of content-addressable items in memory (Wagers, Lau & Phillips, 2009; Tanner, Nicol & Brehm, 2014; Dillon, Mishler, Sloggett & Phillips, 2013; Lago, Shalom, Sigman, Lau & Phillips, 2015). Here, I will assume a cue-based retrieval system as outlined in detail by Lewis and Vasishth (2005), in which linguistic items are encoded in memory as bundles of features and are content-addressable based on the features they contain. Each item stored in memory is associated with a certain level of activation. When a comprehender encounters a retrieval cue in the input, this triggers a search for a target containing a matching feature. Due to the content-addressable nature of the system the search proceeds in a parallel rather than serial fashion (Martin & McElree, 2009). Items with a matching feature receive a boost of activation from the retrieval cue and the item with the highest activation level is retrieved from memory.

While this model gives us an outline of the process underlying memory retrieval in language comprehension, it does not specify whether the retrieval cues can be as abstract as the terms in which a dependency is stated in the grammar. In the grammar, dependencies like subject-verb agreement typically respond to very general features, such as [plural], and not more specific categories, such as *suffixal plural* or *ablauting plural*,

or even particular items, such as ‘*ducks*’ or ‘*geese*’ (Bock & Eberhard, 1993; Tucker, Idrissi & Almeida, 2015). It is possible that retrieval cues have a close relationship to the morphological exponence of a feature. The way a feature is introduced, its *vehicle*, might then have downstream effects on its encoding or retrieval in memory. However, the memory processes used to establish these dependencies might be equally abstract, displaying no sensitivity to specifically how the relevant general feature is introduced or signaled. This would necessitate the inclusion of abstract cues in our retrieval models that are only indirectly linked to morphological form.

### 3.3 What matters in agreement processing?

Subject-verb agreement is a syntactic dependency: subject and verb are syntactic categories, not phonological, morphological or semantic categories. However, the dependency involves a syntactic feature, [number], which correlates with morphological and semantic properties, if only imperfectly. For example, ‘*the tree*’ is syntactically singular in triggering singular agreement, but also morphologically singular in lacking a plural affix, and semantically singular in representing its referent as a single tree. Crucially, however, these several properties are dissociable. Noun phrases headed by a collective noun, such as ‘*fleet*’, are both syntactically and morphologically singular, at least in American English, but semantically plural: they represent their referent as a plurality of like objects. Noun phrases like ‘*the sheep*’ or ‘*the deer*’ can function as syntactically and semantically plural, despite lacking any audible morpheme to mark this. And finally, several kinds of noun phrases are plural in syntax and morphology, but not plural in semantics. These include phrases headed by *pluralia tantum*, such as ‘*the*

*scissors*'; those with the numeral '*one-point-zero*' ('*one-point-zero children*'); and those with the determiners '*no*' ('*no children*') or '*zero*' ('*zero grams*'). For agreement errors in language production, the impact of these properties has been partially teased apart as discussed below.

### 3.3.1 Production

Agreement attraction in production was first systematically investigated in a seminal study by Bock and Miller (1991). In a sentence completion task, agreement errors were more likely to be produced if a preamble with a singular subject contained a plural noun inside a prepositional modifier (*'The key to the cabinets'*). Subsequent work has used agreement attraction to try to tease apart the roles of notional, morphophonological and syntactic number in agreement production. Initially, Bock and Eberhard (1993) found no clear evidence for an impact of either morphophonological form or notional number in error elicitation tasks, as no significant increase in plural verb form errors was observed when the attractor was a syntactically singular pseudoplural ending in '*-s*' (e.g. '*course*') or a syntactically singular collective (*'fleet'*), nor did attraction rates differ for regular and irregular plurals (*'kids'* vs. *'children'*) in attractor position. However, more recent studies in Serbian, Dutch, and German do find effects of morphophonology on agreement production (Mircovic & MacDonald, 2013; Lorimor, Jackson, Spalek & van Hell, 2016). Haskell and MacDonald (2003) also observed small effects of morphological regularity on agreement production in English when there is a conflict between the subject's notional and syntactic number information.

Similarly, there is accumulating evidence that notional number impacts agreement production. Although Bock and Eberhard concluded that subject-verb agreement in

production is controlled by syntactic number, they did note a non-significant numerical trend for plural collectives in attractor position to elicit more agreement errors than plural individual nouns. There was also a correlation between how likely singular collectives were to be judged to refer to multiple entities and the frequency of agreement errors. Clearer evidence for the role of notional number in agreement production was reported by Humphreys and Bock (2005), who used collectives as the subject's head noun followed by a prepositional modifier encouraging either a collective reading ('*The gang near the motorcycles*') or a distributed reading ('*The gang on the motorcycles*'). They found that the rate at which preambles with (syntactically singular) collective head nouns elicited plural verb forms depended on whether they were construed as collective or distributed. Distributed readings more frequently led to the production of plural verbs, indicating that the notional number of the subject affects subject-verb agreement in production. Likewise, Brehm and Bock (2013) show that the likelihood of producing plural agreement with a singular subject depends on how semantically integrated its referent is: more integrated preambles ('*The drawing of the flowers*') were less likely to cause agreement errors than less integrated preambles ('*The drawing with the flowers*'). Brehm and Bock argue that this shows the effect of notional number: the less integrated a complex referent is, the more likely it is to be mentally construed as plural. The Serbian, Dutch, and German studies (Mircovic & MacDonald, 2013; Lorimor et al., 2016) also report higher rates of plural agreement for notionally plural subjects.

Agreement attraction in production is usually attributed to a misrepresentation of the subject's number information (e.g. Bock & Eberhard, 1993; Pearlmutter et al., 1999; Bock et al., 2004; Eberhard, Cutting & Bock, 2005). Under this view, it is not the case

that the number marking on the verb is unlicensed by the agreement controller, since it is consistent with the subject's faulty number information. A representational view can account for the impact of notional number on agreement errors in production by arguing that the subject's number information can be influenced from the conceptual level at the point at which it is planned. After all, the starting point in language production is the message the producer wants to convey.

### *3.3.2 Comprehension*

In comprehension, the subject's number information does not seem to be affected during the encoding stage (Wagers et al., 2009; Tucker et al., 2015; Lago, Shalom, Sigman, Lau & Phillips, 2015). Instead, it appears to be the result of an error-driven retrieval process triggered by the detection of an agreement violation, as discussed above.

If agreement errors in production are indeed the result of misrepresenting the subject's number information, there is no reason to expect that we should see the same effects in comprehension. However, while it is possible that the mechanisms underlying agreement attraction in production and comprehension are different (Tanner, Nicol & Brehm, 2014; Acuna-Farina, 2012), cue-based retrieval accounts of agreement errors in production have also been proposed (Badecker & Kumiak, 2007; Slevc & Martin, 2016). This would mean that the same mechanism underlies agreement attraction in both comprehension and production, suggesting that morphophonological and notional factors should also play a role in agreement attraction in comprehension.

However, there is an important caveat: even if agreement attraction is a result of similarity-based interference in memory retrieval in both production and comprehension,

we could still imagine a system in which they prioritise different types of information. In production, the direction of encoding starts from the message, making notional information extremely salient. It might therefore be more likely to be used in retrieval in production than in comprehension. Moreover, if we assume that agreement attraction in comprehension is an error-driven process, as I have argued in the previous chapter, this also distinguishes it from the retrieval of the agreement controller in production and could affect what type of information is used.

Regardless, investigating subject-verb agreement attraction in comprehension provides an opportunity to address the question of whether the cues used in error-driven retrieval are as abstract as the very general features in terms of which this dependency is specified in the grammar or whether this repair process uses cues that target only certain instantiations of the abstract category.

One recent study in Arabic suggests that agreement attraction effects in comprehension might depend at least partially on the way in which the syntactic plural feature is introduced, i.e. on its vehicle. Arabic has two different plural formation strategies. For suffixation plurals, a plural suffix is added to the singular, similar to the formation of the English plural by adding the suffix ‘-s’. But for ablauting plurals the plural form of the noun is formed by internal vowel change. Tucker et al. (2015) found that when the plural of the attractor was formed by suffixation, significant agreement attraction effects were observed in the reading times. However, with ablauting plurals in attractor position there was only a trend towards attraction that did not reach statistical significance.

### 3.4 The present study

In the present study, we compare agreement attraction with plurals marked by suffixation ('*the cats*') with attraction from those marked by coordination ('*the cat and the dog*'). Only the suffixal plural is an *unequivocal* sign of syntactic plurality, in this particular sense: any occurrence of the plural suffix is within a plural noun phrase, while this is not the case with '*and*'. For example, we find '*and*' within singular noun phrases with a singular referent, such as '*my wife and confidante*' or '*my cute and useful husband*'.<sup>4</sup> Here we might say that that '*and*' coincides semantically with the intersection of predicates, rather than the summing of individuals (see Heycock & Zamparelli, 2005 and Champollion, 2013, for discussion). Moreover, '*and*' also occurs between phrases of several other categories – adjective phrases, prepositional phrases, clauses – and in these cases it does not specifically mark plurality (McCloskey, 1991). Thus, while conjoined noun phrases are syntactically plural in general, the vehicle that signals this audibly, '*and*', plays this role only when it sits between noun phrases (and even then, maybe not always). Therefore it is not, in our terms, an unequivocal signal of syntactic plurality. While this distinction makes no difference in the grammar, it allows us to investigate whether the retrieval cue employed in subject-verb agreement is responsive to features as abstract as [plural] or if it targets only certain exponents of the abstract category, for instance the ones that are unequivocal correlates of syntactic plurality.

---

<sup>4</sup> Perhaps this indicates a lexical ambiguity: maybe there are two words pronounced '*and*', and only one of them occurs only within plural noun phrases (King & Dalrymple 2004). Even so, we would then still like to say that conjunction is at least *superficially equivocal*, since its homophones have similar functions, syntactically and semantically. The affixal '*-s*' might be considered ambiguous too, as between the possessive clitic and the plural affix; but it is not even superficially equivocal, since these two homophones have very different functions.

### 3.5 Experiment 2: conjoined NPs (speeded acceptability)

In Experiment 2 we used a speeded acceptability judgment task to examine whether agreement attraction in comprehension can occur even if the attractor does not contain the plural suffix ‘-s’, which is an unequivocal signal of syntactic plurality. If agreement attraction in comprehension is primarily form-driven and the number retrieval cue on the verb targets unequivocal morphological correlates of syntactic plurality in memory rather than the abstract category itself, conjoined singular noun phrases like ‘*the husband and the wife*’ should not cause agreement attraction, since they lack an unequivocal morphological correlate of syntactic plurality.

We note that conjoined singular noun phrases are certainly syntactically plural in English, since they require plural agreement on the verb when they occupy subject position (‘*The husband and the wife were/\*was next in line*’). The fact that the comprehension of such simple sentences does not appear disrupted might already seem to be evidence that the number cue used for retrieval in agreement computation is not limited to probing for plural ‘-s’, an unequivocal morphological correlate of syntactic number. However, it is important to distinguish between the process of retrieving items from memory and the process of checking agreement. In two-stage models of agreement attraction, verb number is predicted upon encountering the subject, and cue-based retrieval occurs only in mismatch cases where the prediction is violated (Wagers et al., 2009; Tanner et al. 2014). In these models, abstract syntactic number would certainly be used to generate the prediction, but might or might not be the target of the error-driven cue-based retrieval.

Several previous studies on sentence production have examined the production of agreement in sentences that contain conjoined noun phrase, but largely focusing on the different question of what factors can drive singular agreement on the verb when the true syntactic subject is a conjoined noun phrase. Brehm and Bock (2017) and Lorimor et al. (2016) showed that the semantic properties of conjoined noun phrases have an effect on whether participants choose to use singular or plural agreement: in sentence completion, singular agreement is produced more frequently when the preamble contains two abstract rather than two concrete nouns (Brehm & Bock, 2017) or two mass/deverbal nouns rather than animate/count nouns (Lorimor et al., 2016). These findings are consistent with the effects of notional number on sentence production discussed in the previous section: it is easier to separately conceptualize the referents of conjoined concrete nouns than abstract nouns, and mass/deverbal nouns are more notionally singular than animate/count nouns. Keung and Staub (2016) show that agreement with conjoined subjects is also impacted by the number of the closest conjunct (more plural verbs when the second conjunct is plural).

The focus of the current study is on agreement processing in comprehension, which may be supported by partially different mechanisms than production (Tanner et al., 2014; Acuna-Farina, 2012), and here we critically ask about the extent to which conjoined noun phrase attractors *interfere* with singular subject-verb agreement, as a means of investigating memory retrieval mechanisms.

### *3.5.1 Participants*

30 participants were recruited via the Amazon MechanicalTurk platform and received \$3 for completing the experiment. All participants were native speakers of American English and had passed a native speaker proficiency test. Data from 3 additional participants were excluded because their acceptance rate for the ungrammatical filler items was above 40%. None of the subjects participated in more than one of the acceptability judgment experiments reported here.

### *3.5.2 Materials and Design*

The materials consisted of 36 experimental item sets in a 2x3 design crosssing the factors grammatical (grammatical/ungrammatical) and attractor number (singular/plural/conjoined), resulting in six conditions per item. The subject always consisted of a singular head noun followed by a prepositional modifier containing the attractor. Since the head noun was always singular, the verb (a form of copular or auxiliary *be*) was singular in the grammatical conditions and plural in the ungrammatical conditions. Attractor type was manipulated by using either a singular noun, a suffixal plural noun, or conjoined singular noun phrases, as illustrated in (1).

(1)

- a. The slogan about *the husband was* designed to get attention.
- b. The slogan about *the husbands was* designed to get attention.
- c. The slogan about *the husband and the wife was* designed to get attention.
- d. The slogan about *the husband were* designed to get attention.

- e. The slogan about *the husbands were* designed to get attention.
- f. The slogan about *the husband and the wife were* designed to get attention.

In addition to the experimental items we included 36 grammatical and 36 ungrammatical filler items to maintain a ratio of 1:1 of grammatical to ungrammatical items. There were also 8 control items that specifically instructed participants to answer either ‘yes’ or ‘no’ in order to confirm that they were maintaining attention to the task. The experimental items were distributed across 6 lists in a Latin Square design, ensuring that each participant saw each item in only one condition. The fillers and control items were identical across lists.

### 3.5.3 Procedure

The items were displayed word by word in the center of the screen at a rate of 400ms per word using IBEX software (Drummond, 2016). The last word of each sentence was followed by a response screen asking “Was that a good sentence?”. Participants had to judge whether the sentence they had just read was acceptable or not by pressing the ‘f’-key for ‘yes’ and the ‘j’-key for ‘no’. A response had to be made within 2000ms or the display would time out and a message would be displayed telling the participant that their response was too slow. Before the start of the experiment, participants completed five practice items to familiarize them with the procedure.

### *3.5.4 Analysis*

Trials on which no response was made within the 2000ms timeout were excluded from the analysis, leading to the exclusion of 1.8% of the data in the experimental conditions. Following Jaeger (2008), we analyzed the acceptance rate for each of the six experimental conditions using a mixed logit model with the lme4 package (Version 1.1-12, Bates, Maechler, Bolker & Walker, 2015) in the R computing environment (R Development Core Team, 2016). The model had attractor type and grammaticality as fixed effects and by-subject and by-item random intercepts. This was the maximal random effects structure with which the model still converged for all acceptability judgment experiments reported here (Barr, Levy, Scheepers & Tily, 2013). We used effects coding for the effect of grammaticality (grammatical: -0.5, ungrammatical: 0.5) and a reverse Helmert scheme for attractor type. This allowed us to use one contrast to compare the singular attractor to the average of the two types of plural attractor (singular: -0.5, conjoined: 0.25, suffixal: 0.25) and one contrast to directly compare the conjoined plural attractor to the suffixal plural attractor (singular: 0, conjoined: 0.5, suffixal: -0.5).<sup>5</sup>

### *3.5.5 Results*

The proportion of ‘yes’ judgments for each of the experimental conditions in Experiment 2 is plotted in Figure 7. See Table 10 for the output of the mixed logit analysis. The results show a significant effect of grammaticality ( $p < 0.001$ ), with acceptance rates lower for ungrammatical than for grammatical sentences. There was also a significant effect of attractor type when comparing the singular attractor to the average

of the two plural attractor types ( $p < 0.001$ ). Sentences with singular attractors were accepted less frequently than sentences with a plural attractor. However, this effect was primarily driven by the low acceptance rate for ungrammatical sentences with singular attractors. The significant interaction between grammaticality and attractor type for the comparison between singular and both types of plural attractors ( $p < 0.001$ ) reflects the expected agreement attraction effect: the difference in acceptance rates for the grammatical and ungrammatical conditions (grammaticality effect) was much larger for singular attractors compared to the two types of plural attractors. Interestingly, the attraction effect for suffixal plurals was smaller than for conjoined NPs. This interaction between grammaticality and attractor type for the comparison between the conjoined and suffixal plural attractors was also significant ( $p = 0.01$ ).

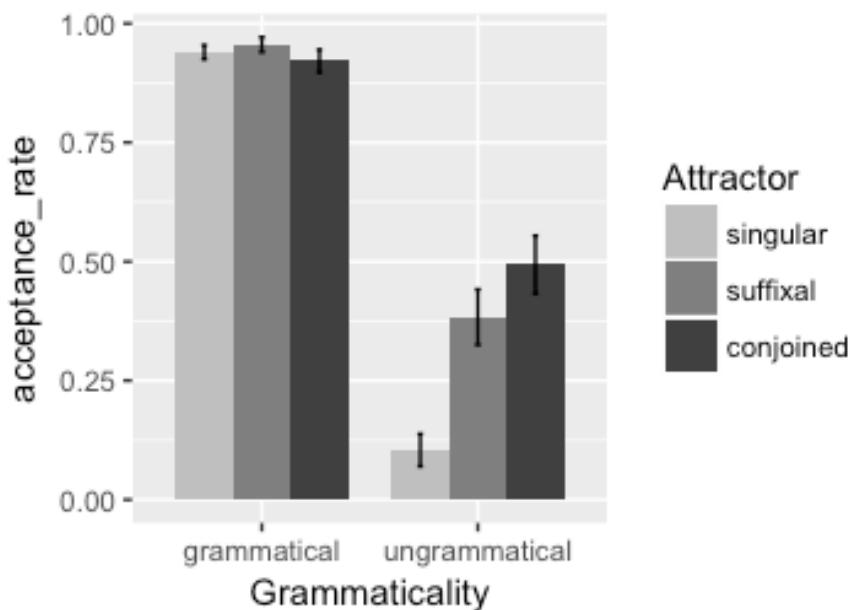


Figure 7: Acceptance rates across conditions in Experiment 2. Error bars indicate standard error of the mean.

	Estimate	Std. Error	z-value	p-value
--	----------	------------	---------	---------

Intercept	1.12	0.25	4.48	< <b>0.001</b>
Grammaticality	-4.42	0.28	-15.55	< <b>0.001</b>
Attr: Sg vs. Pl	1.35	0.34	3.94	< <b>0.001</b>
Attr: Suff vs. Conj	-0.10	0.27	-0.38	0.70
Gram x Attr Sg vs. Pl	3.26	0.70	4.68	< <b>0.001</b>
Gram x Attr Suff vs. Conj	1.36	0.55	2.50	<b>0.01</b>

Table 10: Results of linear mixed logit model on acceptance rates in Experiment 2.

### 3.5.6 Discussion

The results of Experiment 2 demonstrate that conjoined singular noun phrases of the form *determiner-noun-and-determiner-noun* cause agreement attraction effects when they occur as part of the PP-modifier of a subject with a singular head noun. As expected, we found an agreement attraction effect with plural attractors compared to the singular attractor: participants were more likely to accept ungrammatical sentences with a subject-verb agreement violation in the presence of a plural attractor, which were judged acceptable 43.8% of the time, compared to only 10% in the presence of a singular attractor. Critically, agreement attraction was observed for conjoined singular as well as suffixal plural phrases; in fact, the results show that conjoined singular noun phrases elicited even stronger attraction effects than suffixal plurals. While ungrammatical sentences were accepted 38.3% of the time in the presence of suffixal plural attractors, this rose to 49.3% for conjoined singular noun phrases. These findings indicate that it is not necessary for a potential attractor to contain an unequivocal morphological correlate of syntactic plurality to cause facilitative similarity-based interference. The number retrieval cue in subject-verb agreement processing therefore does not specifically target the plural suffix ‘-s’ in the online comprehension processes associated with agreement

attraction. Instead, the number retrieval cue seems to either target a disjunctive list of items correlating with syntactic plurality ('-s', 'and', ...), or an abstract feature shared by all exponents of syntactic plurality; we return to this question in Experiments 5-7.

These results also lend support to the claim that agreement attraction is not based on linear order (Wagers et al., 2009; Franck, Vigliocco & Nicol, 2002). In sentences with conjoined singular attractors, the linearly closest node to the verb is the second conjunct, which is singular. The syntactically plural node (the conjoined phrase) is therefore not adjacent to the verb and yet still creates attraction.

While speeded acceptability is a very powerful measure due to its binary outcome, it is not possible to draw direct conclusions about the timecourse of the observed effect. In contrast, self-paced reading data is relatively noisier but allows us to localize the effect of attraction to a particular position in the sentence. In previous work, speeded acceptability judgments and self-paced reading data for agreement attraction have frequently patterned together (Wagers et al., 2009), supporting the view in which the speed of processing tightly relates to the extent to which participants notice the ungrammaticality. However, the two measures remain complementary pieces of data. Therefore we investigate the timecourse of agreement attraction with conjoined singular attractors using self-paced reading in Experiment 3. If the increased acceptance rate of ungrammatical sentences with conjoined singular attractors in Experiment 2 reflects the same underlying process as attraction with suffixal plurals, we expect it to follow the same timecourse in self-paced reading.

We also note that while these results suggest that the number retrieval cue in agreement computation might be as abstract as the terms in which agreement is defined in

the grammar, an unintended ambiguity in our experimental materials allows an alternative explanation. We intended strings like “*the slogan about the husband and the wife*” to be parsed as singular, with ‘*and*’ embedded in the object of the preposition: “[*the slogan about [ the husband and the wife ]]*”, but participants could have parsed them differently, with ‘*and*’ unembedded, in a way that makes them plural: “[*[ the slogan about the husband ] and [ the wife ]*]”. In that case the plural form of the verb would have been grammatical. Although this parse seems intuitively unlikely given the factors of syntactic and semantic parallelism in the current materials (e.g. [*The slogan about the husband] and [the wife]*] feels quite awkward), it could account for the higher acceptance rate of ungrammatical sentences with conjoined singular NPs compared to suffixal plurals. We address this issue directly in Experiment 4, which uses conjoined singular nouns of the form *determiner-noun-and-noun*.

### 3.6 Experiment 3: conjoined NPs (SPR)

The results of Experiment 2 show that the presence of an attractor whose plurality is introduced by a vehicle that is not a perfect correlate of syntactic plurality leads to higher acceptance rates for subject-verb agreement violations. Previous research has demonstrated that an increase in the acceptance rate for ungrammatical sentences with a plural attractor in speeded acceptability judgments correlates with a reduced slowdown in those conditions in the region immediately following the verb in self-paced reading. This suggests that both of these measures provide a window into a common mechanism contributing to agreement attraction (Wagers et al., 2009). The aim of Experiment 3 was to use self-paced reading to investigate whether the attraction effect with conjoined

singular attractors in Experiment 2 follows the same timecourse during online processing that we expect with suffixal plurals.

### *3.6.1 Participants*

42 members of the University of Maryland community participated in this experiment for course credit or monetary compensation. Data from two additional participants were excluded from all analyses due to low accuracy on the comprehension questions (below 80%). All participants were native speakers of American English and provided informed consent. None of the participants took part in more than one of the experiments presented here.

### *3.6.2 Materials and Design*

To ensure that the results from Experiment 3 were comparable to those from Experiment 2, the experimental items were identical across experiments. Although in some previous self-paced reading studies a preverbal adverb was inserted to avoid spillover effects from attractor noun number on the verb (Wagers et al., 2009), in the current study the attractor and verb were directly adjacent to each other. We decided not to include preverbal adjectives here because in English they are sometimes degraded in acceptability without a very specific intonation, which might have added undesirable noise to the speeded acceptability judgment results. While spillover effects are very common in self-paced reading, the data from Wagers et al. (2009) show that the plural complexity effect lasts no more than a single region. In our study, effects in the

postverbal region (the critical verb's spillover region) should not be affected by plural spillover and can therefore be attributed to processing at the verb.

As in Experiment 2, the items were distributed across six lists in a Latin Square design, so that each participant only saw one condition per item and six items per condition. In addition to the experimental items, the materials also included 134 filler items, 102 of which belonged to four separate manipulations that are not reported here. None of these were related to agreement processing and all filler items were grammatical, meaning that 10.6% of the items were ungrammatical in total.

### *3.6.3 Procedure*

The items were presented word-by-word in a self-paced moving window paradigm (Just, Carpenter & Woolley, 1982) using Linger software (Doug Rhode, MIT) on a desktop computer. At the beginning of each trial a series of dashes appeared on the screen, masking the words of the sentence. Participants had to press the space bar to reveal each word, at which time the previous word was re-masked by a dash. Consequently, only one word at a time was visible and it was not possible for participants to re-read words that had already been re-masked. After the end of each sentence a yes/no comprehension question appeared on the screen in full. Participants had to press the 'f' key to answer 'yes' and the 'j' key to answer 'no'. The questions were simple comprehension questions and never focused on number information. Onscreen feedback was provided only when the response was incorrect. Participants were instructed to read as naturally as possible and to answer the comprehension questions as quickly and accurately as possible. Items were presented in three blocks and the order of presentation

was randomized for each participant. Before the beginning of the experiment, participants completed five practice items to familiarize themselves with the procedure.

### 3.6.4 Analysis

All trials were included in the analysis of the self-paced reading data, regardless of whether the comprehension questions were answered correctly. The regions of analysis consisted of single words and included the verb region and the two words following the critical verb (spillover regions). Reading times exceeding a threshold of 2000ms were excluded as outliers, resulting in the exclusion of less than 0.02% of all trials in the regions of analysis. RTs were log-transformed and analysed with the *lme4* package for linear mixed effects models (Bates et al., 2015) in the R computing environment (R Development Core Team, 2016). The model included grammaticality and attractor type and their interaction as fixed effects. The effects of grammaticality and attractor type were coded the same way as in Experiment 2. Following Barr et al. (2013), we initially fitted a model with the maximal random effects structure. This model failed to converge and was then progressively simplified until convergence was reached. We report results from the model with the maximal random effects structure that converged for all three regions of analysis in both of the self-paced reading experiments reported here (Experiment 3 and Experiment 6) . The final model included by-subject and by-item random intercepts and by-subject random slopes for grammaticality.

The current version of the *lme4* package (version 1.1-12) no longer implements the calculation of p-values using Markov Chain Monte Carlo (MCMC) sampling, which has previously been recommended for deriving p-values from linear mixed effects

models (Baayen, Davidson & Bates, 2008). Instead, we treat the t-statistic as a z-statistic, where a t-statistic with an absolute value larger than 2 suggests significance at the .05 level (Gelman & Hill, 2006; Kush, Lidz & Phillips, 2015).

### 3.6.5 Results

**Comprehension Accuracy.** Mean comprehension accuracy for the experimental items was 94.1%. The mean accuracy for each of the conditions ranged from 92.4% to 96.4%, indicating that participants were paying attention during the experiment.

**Self-paced reading.** The region-by-region average log-transformed RTs in Experiment 3 are plotted in Figure 8. Mean raw RTs for each condition in the verb and spillover regions are given in Table 11. The results from the mixed effects models for the verb region and the two spillover regions are presented in Table 12 to 14.

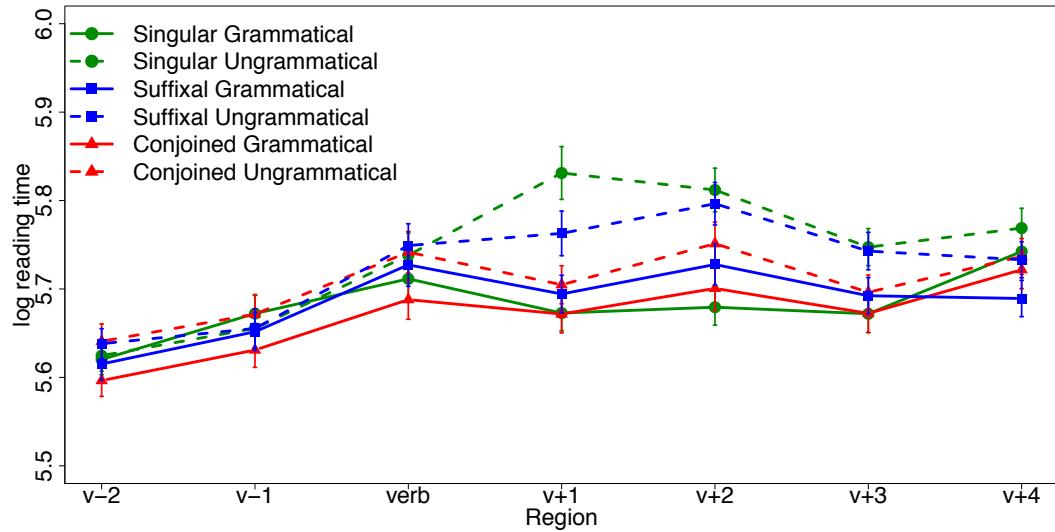


Figure 8: Region-by-region mean log reading times in Experiment 3. Error bars indicate standard error of the mean.

	Singular,	Singular,	Conjoined,	Conjoined,	Suffixal,	Suffixal,
--	-----------	-----------	------------	------------	-----------	-----------

	grammatical	ungrammatical	grammatical	ungrammatical	grammatical	ungrammatical
Verb	324.5 (9.9)	341.8 (12.4)	316.4 (8.7)	339.3 (12.0)	337.0 (12.4)	342.0 (10.8)
Verb+1	308.0 (8.1)	385.8 (14.5)	310.3 (9.2)	321.3 (9.0)	316.2 (8.0)	349.0 (11.5)
Verb+2	311.2 (8.8)	363.4 (10.6)	326.2 (10.5)	344.0 (12.0)	335.3 (12.6)	356.0 (10.0)

Table 11: Mean raw reading times per condition for regions of interest in Experiment 3 (standard error of the mean in parentheses).

	Estimate	Std. Error	t-value
Intercept	5.726	0.031	<b>185.89</b>
Grammaticality	0.034	0.017	<b>2.03</b>
Attr: Sg vs. Pl	0.003	0.023	0.15
Attr: Suff vs. Conj	-0.022	0.020	-1.11
Gram x Attr Sg vs. Pl	0.017	0.047	0.36
Gram x Attr Suff vs. Conj	0.033	0.040	0.58

Table 12: Results of linear mixed effects model in verb region in Experiment 3 (using log transformed RTs).

	Estimate	Std. Error	t-value
Intercept	5.723	0.028	<b>201.88</b>
Grammaticality	0.087	0.020	<b>4.34</b>
Attr: Sg vs. Pl	-0.059	0.024	<b>-2.48</b>
Attr: Suff vs. Conj	-0.041	0.020	-1.98
Gram x Attr Sg vs. Pl	-0.144	0.047	<b>-3.06</b>
Gram x Attr Suff vs. Conj	-0.035	0.041	-0.87

Table 13: Results of linear mixed effects model in first spillover region in Experiment 3 (using log transformed RTs).

	Estimate	Std. Error	t-value
Intercept	5.745	0.028	<b>206.87</b>
Grammaticality	0.084	0.019	<b>4.52</b>
Attr: Sg vs. Pl	-0.002	0.024	-0.10
Attr: Suff vs. Conj	-0.036	0.021	-1.77
Gram x Attr Sg vs. Pl	-0.099	0.048	<b>-2.04</b>
Gram x Attr Suff vs. Conj	-0.021	0.042	-0.50

Table 14: Results of linear mixed effects model in second spillover region in Experiment 3 ( using log transformed RTs).

In the verb region only the main effect of grammaticality was significant ( $t = 2.03$ ), with agreement violations leading to slower average reading times (grammatical = 326ms; ungrammatical = 341ms). This slowdown remained significant in the first spillover region ( $t = 4.34$ ; grammatical = 312ms; ungrammatical = 352) and the second spillover region ( $t = 4.52$ ; grammatical = 324ms; ungrammatical = 355ms). In the first spillover region there was also a significant effect of attractor type for the comparison between the singular and the two types of plural attractors ( $t = -2.48$ ). Mean reading times were slower for singular attractors than for plural attractors (singular = 347ms; plurals = 322ms). The significant interaction between grammaticality and attractor type for singular compared to plural attractors in the first spillover region ( $t = -3.06$ ; grammaticality effect singular = 78ms; grammaticality effect plurals = 22ms) and the second spillover region ( $t = -2.04$ ; grammaticality effect singular = 52ms; grammaticality effect plurals = 19ms) indicates that the slowdown associated with ungrammaticality was significantly reduced in the presence of a plural attractor. Although none of the other effects reached an absolute  $t$ -value larger than 2, the effect of attractor type for suffixal compared to conjoined plurals in the first spillover region was marginally significant ( $t = -1.98$ ). This region was read faster for conjoined attractors. Although the grammaticality effect for suffixal plurals was numerically larger than for conjoined attractors in the first spillover region, this interaction was not significant ( $t = -0.87$ ; grammaticality effect suffixal = 33ms; grammaticality effect conjoined = 11ms).

### 3.6.6 Discussion

The results of Experiment 3 are mostly consistent with the speeded acceptability judgment data from Experiment 2. As expected, subject-verb agreement violations led to slower reading times. In the critical verb's spillover regions this slowdown was reduced in the presence of plural compared to singular attractors indicating that comprehenders experienced agreement attraction from a structurally irrelevant number-matching noun. Unlike in Experiment 2, there was no evidence that conjoined singular NPs caused stronger agreement attraction than suffixal plurals. While numerically suffixal plurals showed a smaller attraction effect, this contrast was not significant.

Together, the findings from the self-paced reading task in Experiment 3 and the end-of-sentence judgment task in Experiment 2 suggest that the retrieval process that supports agreement computation in comprehension targets something more general than the plural suffix ‘-s’. However, the results are also compatible with a model in which the plural retrieval cue targets a disjunctive list of items correlating with syntactic plurality (‘-s’, ‘and’, …), rather than an abstract [plural] feature shared by all exponents of syntactic plurality. Experiment 4 was designed to rule out an alternative explanation for these results based on the coordination ambiguity.

Experiment 3 used the same experimental materials as Experiment 2, which means that there was still an unintended ambiguity in the sentences with the conjoined singular noun phrases: it is possible, if unlikely, that participants parsed them as *[subject head noun [preposition [determiner noun]]] and [determiner noun]* (*[The slogan about the husband] and [the wife]*), rather than *[subject head-noun [preposition [determiner noun and determiner noun]]]* (*The slogan about [the husband and the wife]*). In that case

the plural form of the verb would have been grammatical. We address this issue in Experiment 4, which avoids this ambiguity by using conjoined singular noun phrases of the form *determiner-noun-and-noun*.

### 3.7 Experiment 4: conjoined NPs without second determiner (speeded acceptability)

The aim of Experiment 4 was to ensure that the results we saw with conjoined singular attractors in Experiment 2 and 3 were not due to an unintended parse of this attractor type. Although conjoined attractors demonstrated a profile very similar to suffixal plural attractors, it is possible that this profile derived from a completely different source in the conjoined case. This is because the conjoined conditions had an alternative parse which is not available in the suffixal plural attractor conditions: they could be parsed as *[subject head-noun [preposition [determiner noun]]] and [determiner noun]* (*[The slogan about the husband] and [the wife]*), rather than as the intended *[subject head-noun [preposition [determiner noun and determiner noun]]]* (*The slogan about [the husband and the wife]*). Under this alternative parse, the plural form of the verb, which was intended to be a subject-verb agreement violation, would have been grammatical. This could drive increased acceptability and reduced reading times in the ‘mismatch’ condition.

Fortunately, in English it is possible to coordinate noun phrases without a second determiner, and this forces a parse in which the two local noun phrases are coordinated: *The slogan about the husband and wife*. If participants are still more likely to accept sentences with a singular subject and a plural verb when the conjoined singular attractor

does not have a second determiner, this could not be explained by parsing ambiguity and would support our original interpretation of Experiment 2 and 3.

### 3.7.1 Participants

30 native speakers of American English were recruited via the Amazon MechanicalTurk platform and received \$3 for completing the experiment. One additional participant who had an acceptance rate of 40% or above for the ungrammatical filler items was excluded from all analyses. None of the participants took part in any of the other acceptability judgment experiments reported here.

### 3.7.2 Materials and Design

The experimental items were adapted from those used in Experiment 2 and 3 by removing the determiner in front of the second noun phrase in the conjoined singular attractor (*'The slogan about the husband and wife'*). Consequently, the only possible parse for the sentences with the conjoined singular attractor was [*preposition [determiner noun and noun]]*], avoiding the unintended ambiguity in these items in Experiment 2 and 3. The items were not changed for any of the other conditions. The same 36 grammatical and 36 ungrammatical filler items plus 8 control items were included as in Experiment 2, and the experimental items were distributed across 6 lists in a Latin Square design.

### 3.7.3 Procedure and Analysis

The procedure and analysis were identical to Experiment 2. Trials on which no response was made within 2000ms were excluded, resulting in the exclusion of 0.2% of all experimental trials.

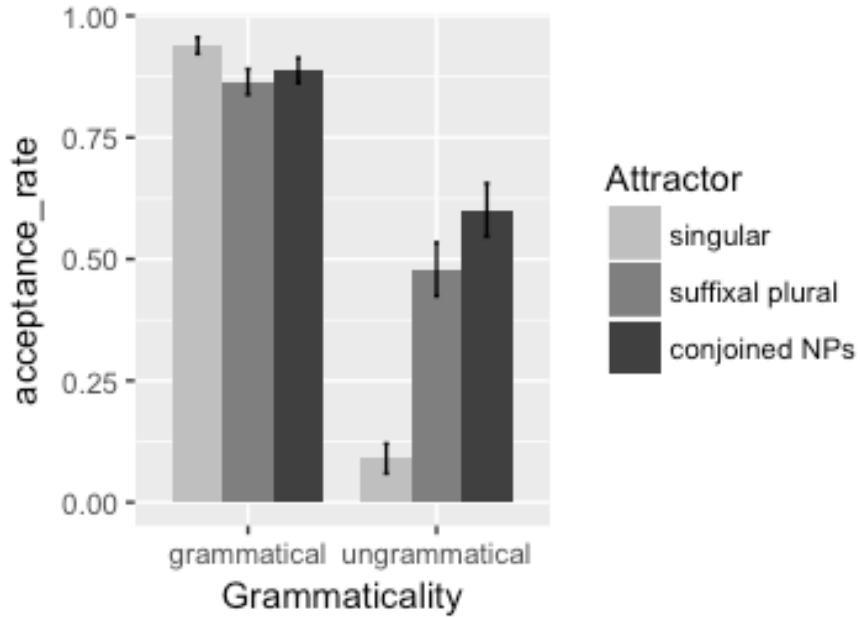


Figure 9: Acceptance rates across conditions in Experiment 4. Error bars indicate standard error of the mean.

	Estimate	Std. Error	z-value	p-value
Intercept	0.85	0.17	4.99	< 0.001
Grammaticality	-3.19	0.21	-15.03	< 0.001
Attr: Sg vs. Pl	1.30	0.31	4.15	< 0.001
Attr: Suff vs. Conj	0.38	0.20	1.88	0.06
Gram x Attr Sg vs. Pl	4.75	0.63	7.49	< 0.001
Gram x Attr Suff vs. Conj	0.34	0.40	0.85	0.40

Table 15: Results of mixed logit model on acceptance rates in Experiment 4.

### *3.7.4 Results*

Figure 9 shows the proportions of ‘yes’ judgments for each of the experimental conditions in Experiment 4. The results of the mixed logit model are presented in Table 15. There was a significant effect of grammaticality ( $p < 0.001$ ); grammatical sentences were much more likely to be judged acceptable than ungrammatical ones. The effect of attractor type was significant for the comparison between singular and the two types of plural attractors ( $p < 0.001$ ). Sentences with singular attractors were less likely to be judged acceptable than those with plural attractors. This was due to the low acceptance rate for ungrammatical sentences with a singular attractor. The interaction between grammaticality and attractor type for the comparison between singular and both types of plural attractors was highly significant ( $p < 0.01$ ) and the decrease in acceptance for ungrammatical compared to grammatical sentences was much larger for singular attractors. There was also a marginal effect of attractor type for conjoined singulars compared to suffixal plurals ( $p = 0.06$ ). Numerically sentences with conjoined singulars had a higher acceptance rate than those with suffixal plurals, but the interaction between grammaticality and attractor type was not significant ( $p = 0.4$ ) between these two attractor types.

### *3.7.5 Discussion*

The pattern in Experiment 4 is clearly consistent with the results from Experiment 2 and 3. Although no alternative parse was available for the conjoined conditions in

Experiment 4, conjoined attractors still cause agreement attraction. This rules out the alternative explanation based on an unintuitive parse of the complex subject in Experiment 2 and 3. These results further support the idea that an attractor that is syntactically plural can cause agreement attraction effects in comprehension even when its plurality is not marked by the plural ‘-s’, an unequivocal signal. The retrieval cue must be more general than just a single morpheme. While the attraction effect was numerically larger for conjoined plurals than suffixal plurals in Experiment 4, the interaction was not statistically significant (unlike in Experiment 2). It is possible that this is a very small effect that is difficult to detect. We return to this point in the General Discussion.

There are at least two options for exactly how the number retrieval cue could be general. It might be an abstract feature, [plural], shared by all exponents of syntactic plurality. Alternatively, the plural retrieval cue might target not one abstract feature, but instead a list of items that correlate with the [plural] feature (such as ‘-s’ or ‘*and*’). Under this model, the plural retrieval cue would be directly associated with the morphological exponents of syntactic plurality rather than syntactic plurality itself. In that case, the attraction seen with conjoined singular noun phrases would not be because they possess an abstract [plural] feature, but rather because the verb’s retrieval cue targets ‘*and*’. So far, we have assumed that if conjoined singular noun phrases cause agreement attraction, it must be because of the abstract [plural] feature. In Experiments 5-7, we examine the alternative possibility by considering noun phrases such as ‘*the loyal and caring husband*’ in attractor position. These include ‘*and*’ but are syntactically singular, since here the conjunction coordinates adjectives modifying a singular noun. We ask whether these too can cause errors of agreement attraction. If they do, it suggests the word ‘*and*’

has become statistically associated with syntactic plurality, to the extent that it can itself respond to the number retrieval cue triggered by the plural verb.

### 3.8 Experiment 5: conjoined adjectives (speeded acceptability)

In Experiment 5, we use singular attractors with conjoined adjectives to investigate the possibility that the number retrieval cue on the verb targets correlates of syntactic plurality rather than the abstract category itself, even in cases where the correlates do not actually introduce this category. While the results of Experiments 2-4 demonstrate that retrieval is not limited to probing for an unequivocal morphological correlate of syntactic plurality (plural ‘-s’), they do not rule out that the attraction effects we find with conjoined noun phrases is the result of retrieval targeting the word ‘*and*’, which is a correlate of syntactic plurality, although an imperfect one. Here we examine the possibility that the conjunction ‘*and*’ might be targeted in agreement computations, even though the correlation is not perfect and is not directly represented in the grammar. We can dissociate the role of abstract number and surface cues to syntactic plurality by examining the impact of singular attractors with conjoined adjectives (*the loyal and caring husband*), which contain ‘*and*’ but are not syntactically plural. If the memory processes used to establish the subject-verb agreement dependency do not just target correlates of the abstract category [plural], but are as abstract as the terms in which the dependency is stated in the grammar, this type of attractor should not cause agreement attraction effects. However, if it is morphological correlates of syntactic plurality that are targeted by the verb’s number cue in retrieval, singular attractors with conjoined adjectives should cause attraction just like suffixal plural attractors.

### *3.8.1 Participants*

We recruited 30 participants via Amazon's Mechanical Turk platform. All participants were native speakers of American English and received \$3 for participating in the experiment. Two additional participants were excluded from all analyses because they accepted the ungrammatical filler sentences more than 40% of the time.

### *3.8.2 Materials and Design*

The materials consisted of modified versions of the 36 experimental item sets from Experiment 2. The materials also included the same 36 grammatical and 36 ungrammatical fillers, as well as the 8 control items, used in the other acceptability judgment experiments reported here. The experimental items were distributed across 6 lists in a Latin Square design, with fillers and control items identical across lists. The 2x3 design crossed attractor type (singular with adjective/plural with adjective/singular with conjoined adjectives) with grammaticality (grammatical/ungrammatical), resulting in six conditions per item, see (2). As in the previous experiment, the head noun of the subject was always singular and followed by a prepositional modifier containing the attractor. The attractor took the form of the definite article *the* followed by an adjective and a singular noun (singular attractor), an adjective and a plural noun (plural attractor), or a singular noun preceded by two adjectives conjoined by *and* (conjoined adjective attractor). Participants saw each experimental item in only one condition.

(2)

- a. The slogan about *the caring husband was* designed to get attention.
- b. The slogan about *the caring husbands was* designed to get attention.
- c. The slogan about *the loyal and caring husband was* designed to get attention.
- d. The slogan about *the caring husband were* designed to get attention.
- e. The slogan about *the caring husbands were* designed to get attention.
- f. The slogan about *the loyal and caring husband were* designed to get attention.

### *3.8.3 Procedure and Analysis*

The procedure used in Experiment 5 was identical to that in Experiment 2 and 4. Trials on which no response was made within 2000ms accounted for 0.1% of all experimental trials and were excluded. Like in Experiment 2 and 4, we used effects coding for the effect of grammaticality (grammatical: -0.5, ungrammatical: 0.5). However, attractor types were coded differently, as the central question in the current experiment was whether the two singular attractors differed as a function of whether they were preceded by a single adjective or conjoined adjectives. Therefore we used one contrast to directly compare the singular attractor with a single adjective to the singular attractor with conjoined adjectives (adjective and singular noun: -0.5; adjective and plural noun: 0; conjoined adjectives and singular noun: 0.5). To keep the contrasts orthogonal, the other contrast was set to compare the attractor with a plural noun to the average of the two other attractor types (adjective and singular noun: 0.25; adjective and plural noun: -0.5; conjoined adjectives and singular noun: 0.25).

### 3.8.5 Results

The proportion of ‘yes’ judgments for all experimental conditions is illustrated in Figure 10 and Table 16 contains the output of the mixed logit model. As in the other speeded acceptability judgment experiments, grammatical sentences were more likely to be accepted than ungrammatical ones ( $p < 0.001$ ). As expected, ungrammaticality had a smaller effect on acceptance rates for sentences with attractors containing a plural noun than for sentences with attractors that contained a singular noun ( $p < 0.001$ ). Crucially, the interaction between grammaticality and attractor type was also significant for the comparison between attractors with one adjective and a singular noun and attractors with conjoined adjectives and a singular noun ( $p = 0.02$ ).

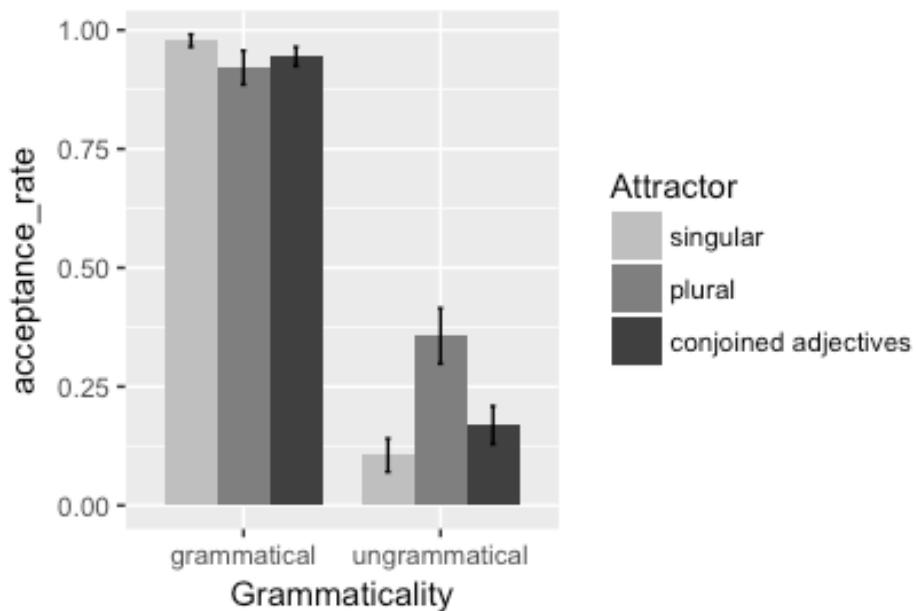


Figure 10: Acceptance rates across conditions in Experiment 5. Error bars indicate standard error of the mean.

	Estimate	Std. Error	z-value	p-value
Intercept	0.85	0.22	3.79	< 0.001

Grammaticality	-5.11	0.30	-17.05	<b>&lt; 0.001</b>
Attr: Pl. vs. Sg. & Conj. Adj.	-0.41	0.33	-1.27	0.20
Attr: Sg. vs. Conj. Adj.	-0.17	0.35	-0.50	0.62
Gram x Attr Pl. vs. Sg. & Conj. Adj.	-3.21	0.65	-4.94	<b>&lt; 0.001</b>
Gram x Attr Sg. vs. Conj. Adj.	1.61	0.70	2.30	<b>0.02</b>

Table 16: Results of mixed logit model on acceptance rates in Experiment 5.

Plural attractors had a numerically much larger attraction effect than conjoined adjective singular attractors. We performed a post-hoc test for which the model was refit with the effect of attractor type treatment coded and relevelled (Linck & Cummings, 2015). The glht function in the multcomp package (Hothorn, Bretz & Westfall, 2008) was then used to directly compare the difference between differences: effect of grammaticality for singular attractors with conjoined adjectives and for plural attractors with a single adjective. Note that this comparison is not orthogonal to the comparison between the two types of singular attractors in the main model. The post-hoc test showed that plural attractors caused a significantly larger attraction effect than singular attractors with conjoined adjectives (Estimate = -1.61; Std. Error = 0.52; z-value = -3.08; p = 0.002).

### 3.8.6 Discussion

These results provide intriguing if tentative support for the hypothesis that the word ‘*and*’ is a target for retrieval upon encountering the verb’s number cue. Singular attractors that contained the word ‘*and*’ appeared to induce a small attraction effect, leading to an increased acceptance rate for ungrammatical sentences compared to those with singular attractors with only a single adjective (grammaticality effect single adj. sg.

noun = 87.2%; conjoined adj. sg. noun= 77.5%). However, the drop in acceptance associated with subject-verb agreement violations was reduced much more by the attractor containing a single adjective and a plural noun than by the attractor with a singular noun and two conjoined adjectives (grammaticality effect single adj. plural noun = 56.4%).

In contrast to the true syntactically plural conjoined attractors examined in Experiment 2-4, the grammaticality effect was reduced less with conjoined adjective attractors than with attractors containing a (suffixal) plural. Nevertheless, the data do suggest that an attractor that does not actually signal syntactic plurality can nonetheless cause some degree of interference in agreement computation simply because it contains an imperfect correlate of syntactic plurality. In order to further investigate this possibility, Experiment 6 uses the materials from Experiment 4 in a self-paced reading experiment.

### 3.9 Experiment 6: conjoined adjectives in self-paced reading

The results from Experiment 5 suggest that the presence of the conjunction ‘*and*’ in an attractor that is not syntactically plural might cause agreement attraction. Here, we follow this up by using the same materials as in Experiment 5 in a self-paced reading task. If the presence of ‘*and*’ in the attractor is sufficient to cause agreement attraction, singular attractors with conjoined adjectives should reduce the slow-down associated with encountering an agreement violation.

### *3.9.1 Participants*

41 members of the University of Maryland community participated in this experiment for course credit or monetary compensation. The data from two additional participants were excluded from all analyses due to a low accuracy rate (below 80%) on the comprehension questions. None of the participants took part in any of the other experiments reported here.

### *3.9.2 Materials and Design*

The experimental items in Experiment 6 were identical to those used in Experiment 5, to ensure that results were easily comparable. The same set of fillers was used as in Experiment 3.

### *3.9.3 Procedure and Analysis*

The same self-paced reading procedure was used as in Experiment 3. Grammaticality and attractor type were coded the same way as in Experiment 5. Reading times exceeding a threshold of 2000ms were excluded as outliers, resulting in the exclusion of less than 0.03% of all trials in the regions of analysis.

### *3.9.4 Results*

**Comprehension Accuracy.** Mean comprehension accuracy for the experimental items was 94.5%. The mean accuracy for each of the conditions ranged between 92.6% to 96.7%.

**Self-paced reading.** Region-by-region average log-transformed RTs in Experiment 6 are plotted in Figure 11 and mean raw RTs for each experimental condition

in the regions of interest are provided in Table 17. Table 18 to 20 present the results from the linear mixed effects models for the verb region and the two spillover regions.

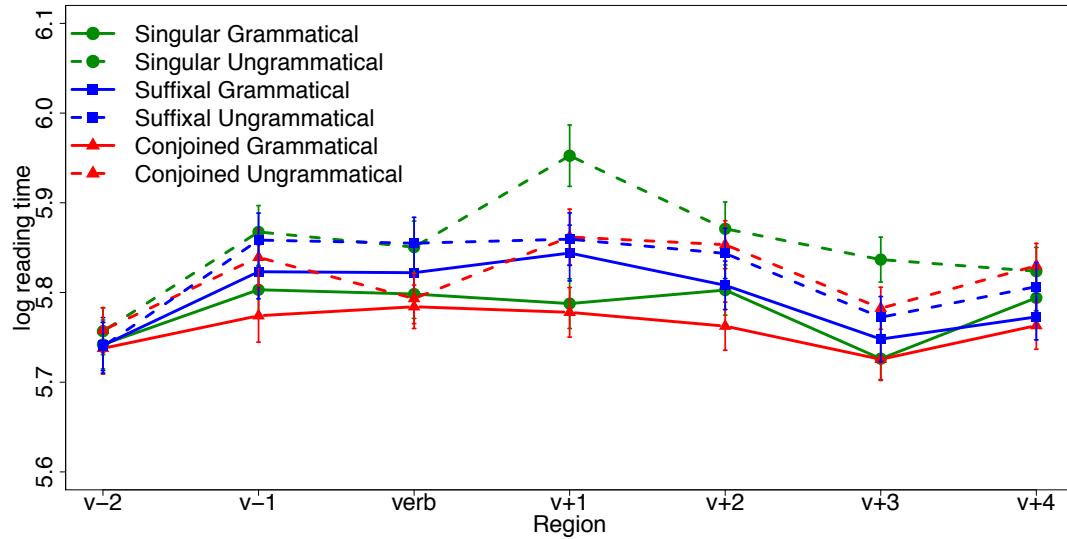


Figure 11: Region-by-region mean log reading times in Experiment 6. Error bars indicate standard error of the mean.

	Singular, grammatical	Singular, ungrammatical	Conjoined adjectives, grammatical	Conjoined adjectives, ungrammatical	Plural, grammatical	Plural, ungrammatical
Verb	363.6 (11.9)	388.9 (14.2)	349.8 (9.2)	362.7 (11.8)	376.0 (11.9)	389.9 (13.8)
Verb+1	361.1 (11.8)	448.4 (18.3)	357.7 (12.2)	399.0 (15.0)	394.7 (16.3)	391.8 (13.3)
Verb+2	367.4 (12.7)	397.7 (13.9)	350.6 (12.2)	383.7 (12.9)	365.5 (11.6)	383.3 (13.3)

Table 17: Mean raw reading times per condition for regions of interest in Experiment 6 (standard error of the mean in parentheses).

	Estimate	Std. Error	t-value
Intercept	5.818	0.046	<b>127.53</b>
Grammaticality	0.034	0.021	1.59
Attr: Pl. vs. Sg. & Conj. Adj.	-0.042	0.024	-1.76
Attr: Sg. vs. Conj. Adj.	-0.035	0.021	-1.68
Gram x Attr Pl. vs.	-0.004	0.048	-0.08

Sg. & Conj. Adj.			
Gram x Attr Sg. vs. Conj. Adj.	-0.042	0.042	-0.99

Table 18: Results of linear mixed effects model in verb region in Experiment 6 (using log transformed RTs).

	Estimate	Std. Error	t-value
Intercept	5.848	0.048	<b>120.90</b>
Grammaticality	0.088	0.022	<b>4.06</b>
Attr: Pl. vs. Sg. & Conj. Adj.	-0.011	0.027	-0.42
Attr: Sg. vs. Conj. Adj.	-0.050	0.023	<b>-2.17</b>
Gram x Attr Pl. vs. Sg. & Conj. Adj.	0.144	0.053	<b>2.70</b>
Gram x Attr Sg. vs. Conj. Adj.	-0.081	0.046	-1.75

Table 19: Results of linear mixed effects model in first spillover region in Experiment 6 (using log transformed RTs).

	Estimate	Std. Error	t-value
Intercept	5.824	0.045	<b>129.77</b>
Grammaticality	0.063	0.017	<b>3.66</b>
Attr: Pl. vs. Sg. & Conj. Adj.	-0.007	0.024	-0.26
Attr: Sg. vs. Conj. Adj.	-0.029	0.021	<b>-1.37</b>
Gram x Attr Pl. vs. Sg. & Conj. Adj.	0.059	0.049	1.21
Gram x Attr Sg. vs. Conj. Adj.	0.018	0.042	0.44

Table 20: Results of linear mixed effects model in second spillover region in Experiment 6 (using log transformed RTs).

There were no significant effects in the verb region. The effect of grammaticality became significant in the first spillover region ( $t = 4.06$ ) and remained significant in the second spillover region ( $t = 3.66$ ): reading times were slower for sentences with a subject-verb agreement violation. The effect of attractor type for the two singular

attractors was also significant in the first spillover region ( $t = -2.17$ ). Sentences with singular attractors with conjoined adjectives were read faster than those with singular attractors with only a single adjective. The interaction between grammaticality and attractor type was significant for the plural attractor compared to the average of the two types of singular attractors ( $t = 2.70$ ) in the first spillover region. The slowdown in response to an agreement violation was much reduced in the presence of a plural attractor. Unlike in Experiment 5, the interaction between grammaticality and attractor type was only marginally significant for the comparison between the attractors with a single adjectives and with conjoined adjectives ( $t = -1.75$ ). Ungrammaticality led to a numerically smaller slowdown in the presence of a singular attractor with conjoined adjectives compared to a singular attractor with a single adjective.

Numerically, the slowdown in response to an agreement violation in the first spillover region was also reduced for a plural attractor in comparison to a singular attractor with conjoined adjectives. However, a post-hoc test following the same procedure as in Experiment 5 to compare the difference between differences showed that this was not significant (Estimate = -0.068; Std. Error = 0.046; z-value = -1.47; p = 0.14).

### *3.9.5 Discussion*

The results from Experiment 6 are mostly consistent with the findings from Experiment 5. Singular attractors with conjoined adjectives increase the acceptance rate for ungrammatical sentences and reduce the associated reading time disruption. It is notable that in both experiments the attraction effect was numerically smaller for singular conjoined adjective attractors than for the plural attractors, which was not the case for the conjoined noun phrase attractors examined in Experiments 2 to 4. This suggests that the

attraction observed with conjoined noun phrases is not simply due to retrieval of the word ‘*and*’ as a correlate of syntactic plurality. Nevertheless, the fact that singular attractors with conjoined adjectives increase the acceptance rate of ungrammatical sentences and lead to a reduced slowdown in self-paced reading suggests that the presence of ‘*and*’ in the attractor causes some interference in agreement computation, even if the attractor is neither syntactically plural nor contains an unequivocal signal of syntactic plurality.

One potential explanation for the observed attraction effect with ‘*and*’ is that comprehenders are more likely to expect a plural noun following conjoined adjectives. If that were the case, their prediction of a plural noun even in the absence of one in the actual input might have caused interference in computing agreement. To rule this out, we conducted an untimed cloze task with the materials from Experiment 6. The items were cut off after the adjective/conjoined adjectives and 32 participants completed the sentences. The cloze probability of a plural noun following conjoined adjectives was only 5.6% (32 completions out of 576), This was in fact lower than after a single adjective, where it was 6.6% (38 completions out of 576). This shows that comprehenders were not more likely to expect a plural noun after conjoined adjectives. Consequently, predicting a plural cannot be the source of the attraction effect observed with conjoined adjective attractors.

However, there is a potential confound in the materials used in Experiment 5 and 6. In the conditions with singular attractors with conjoined adjectives, the head noun of the subject is separated from the verb by two additional words in comparison to the other conditions with only one adjective. In Experiment 7 we address this issue by testing singular attractors with stacked adjectives (*the loyal caring husband*), which increase the

distance between the head noun and the verb. They have a similar semantic representation to explicitly conjoined adjectives but do not include the word ‘*and*’ as a potential target for retrieval.

### 3.10 Experiment 7: conjoined vs. stacked adjectives (speeded acceptability)

The aim of Experiment was to investigate whether the apparent attraction effect observed in Experiments 5 and 6 for singular attractors with conjoined adjectives was simply due to the additional length/complexity of the attractor region rather than specifically the presence of the word ‘*and*’, which is an imperfect correlate of syntactic plurality. In Experiment 7, we adapted the materials used in Experiments 5 and 6 to include a singular attractor with stacked adjectives (‘*The slogan about the loyal caring husband*’), thereby increasing the distance between the head noun and the verb. If the higher acceptance rate for the ungrammatical sentences with conjoined adjective attractors was a result of the increased distance between the head noun and the verb, then a singular attractor with stacked adjectives should also lead to an increase in the acceptance rate. If the effect is due to the word ‘*and*’, attraction should be observed in with conjoined adjectives but not stacked adjectives.

#### 3.10.1 Participants

As with the other acceptability judgment experiments reported here, participants were recruited via Amazon’s Mechanical Turk platform. They were all native speakers of American English and received \$3 for completing the experiment. There were 30

participants, plus 4 additional participants who were excluded from all analyses because they accepted the ungrammatical filler items at a rate of 40% or above.

### 3.10.2 Materials and Design

The materials were adapted from those used in Experiment 5 and 6. Instead of using a suffixal plural attractor as one of the attractor types, we included a singular attractor preceded by two stacked adjectives (*'the loyal caring husband'*). Consequently, the three attractor types were singular attractor with single adjective (*'the caring husband'*), singular attractor with stacked adjectives (*'the loyal caring husband'*), and singular attractor with conjoined adjectives (*'the loyal and caring husband'*). For some of the items, the order of the adjectives was reversed from Experiment 5 and 6 to make the stacked adjectives sound more natural, but this was kept constant across experimental conditions. The experimental items were distributed across 6 lists in a Latin Square design, so that each participant saw each item in only one condition. Filler items (72, ratio of 1:1 grammatical vs. ungrammatical) and control items were the same as in the other acceptability judgment experiments and were identical across lists.

### 3.10.3 Procedure and Analysis

The same acceptability judgment procedure and analysis was used as in Experiment 2, 4, and 5. 0.65% of all experimental trials were excluded because no response was made within 2000ms. Grammaticality was coded the same as in all other experiments reported here (grammatical: -0.5, ungrammatical: 0.5). For attractor type, one contrast was used to directly compare the attractor with stacked adjectives to the attractor with conjoined adjectives (single adjective: 0; stacked adjectives: -0.5;

conjoined adjectives: 0.5). To keep the contrasts orthogonal, the other contrast compared the attractor with a single adjective to the average of the two other attractor types (single adjective: -0.5; stacked adjectives: 0.25; conjoined adjectives: 0.25).

### *3.10.4 Results*

Figure 12 plots the proportion of ‘yes’ judgments for each experimental condition and the results from the logit model are presented in Table 21. As expected, there was a main effect of grammaticality ( $p < 0.001$ ), with ungrammatical sentences accepted less frequently than grammatical ones. There was no significant interaction between grammaticality and attractor type for the comparison between the attractor with one adjective compared to the average of the other two attractors ( $p = 0.25$ ; grammaticality effect single adjective = 81.2%; average grammaticality effect other attractors = 76.4%). The interaction between grammaticality and attractor type was significant for the comparison between the attractor with stacked adjectives and the attractor with conjoined adjectives ( $p = 0.04$ ). The impact of ungrammaticality was smaller for attractors with conjoined adjectives compared to attractors with stacked adjectives (grammaticality effect conjoined adjectives = 71.6%; grammaticality effect stacked adjectives = 81.2%).

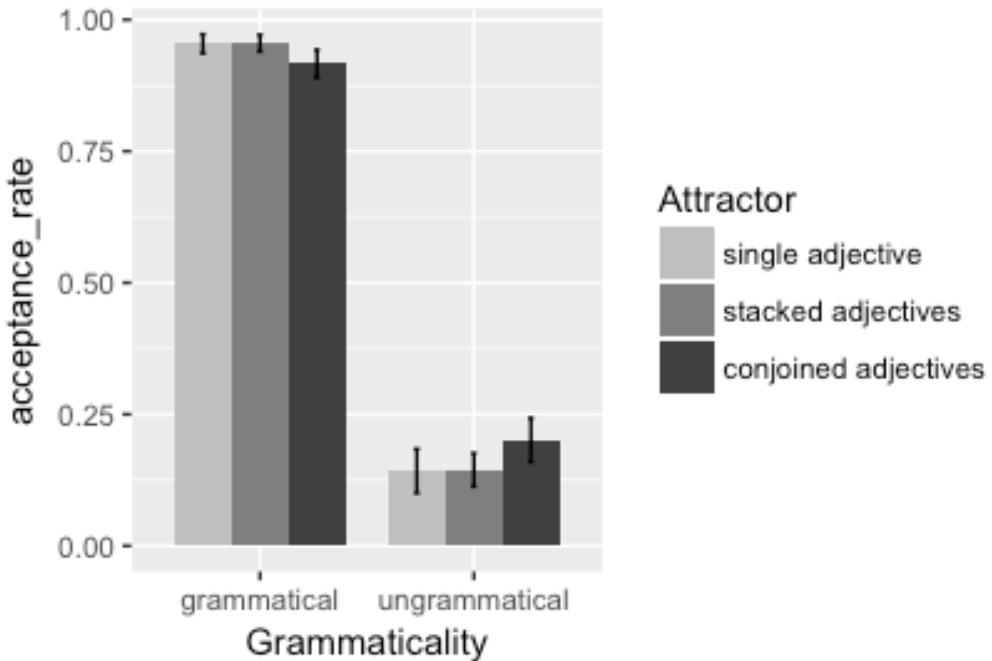


Figure 12: Acceptance rates across conditions in Experiment 7. Error bars indicate standard error of the mean.

	Estimate	Std. Error	z-value	p-value
Intercept	0.64	0.22	2.90	< <b>0.01</b>
Grammaticality	-5.23	0.31	-17.14	< <b>0.001</b>
Attr: single adj. vs. stacked & conjoined adj.	-0.05	0.35	-0.14	0.89
Attr: stacked adj. vs. conjoined adj.	-0.14	0.28	-0.49	0.63
Gram x Attr single adj. vs. stacked & conjoined adj.	0.80	0.69	1.16	0.25
Gram x Attr stacked adj. vs. conjoined adj.	1.16	0.56	2.10	<b>0.04</b>

Table 21: Results of linear logit model on acceptance rates in Experiment 7.

To test whether the attractor with conjoined adjectives caused attraction in comparison to the attractor with a single adjective, we conducted a post-hoc test

following the same procedure as in Experiment 5 and 6. The interaction between grammaticality and attractor type for attractors with a single adjective and conjoined adjectives was significant (Estimate = 1.18; Std. Error = 0.56; z-value = 2.12; p = 0.03), which is consistent with the results from Experiment 5 and 5.

### 3.10.5 Discussion

The results of Experiment 7 provide further data suggesting that syntactically singular attractors containing two conjoined adjectives can cause agreement attraction in comprehension. In contrast, the grammaticality effect for singular attractors with stacked adjectives was not at all reduced compared to the grammaticality effect for singular attractors with a single adjective. This suggests that the increase in the acceptance rate we see with the conjoined adjectives in Experiment 6 is not simply due to the increased linear distance between the subject's head noun and the verb. Instead, it seems that the presence of the conjunction '*and*' results in some degree of agreement attraction when the verb is plural, even when the noun phrase it appears in is syntactically singular as in the conjoined adjective case. We return to the question of what this means for the relationship between cues and features in the memory system in the General Discussion.

## 3.11 General Discussion

The experiments reported here investigated whether the number retrieval cue on the verb in error-driven retrieval is as abstract as the terms in which the grammatical dependency is stated, or whether it matters how the relevant property (syntactic plurality) is introduced. The experiments that used conjoined singular noun phrases as attractors

(Exp. 2-4) demonstrated that conjoined attractors that are syntactically plural but contain only an equivocal signal of syntactic plurality cause agreement attraction effects in both self-paced reading and speeded acceptability measures. In Experiment 2, attraction from conjoined singular noun phrases was significantly greater than for attractors containing plural ‘-s’, which correlates perfectly with syntactic plurality. However, this effect was not replicated in either Experiment 3 or 4. The findings suggest that syntactically plural attractors cause similarity-based interference in agreement computation regardless of whether they are marked by suffixation or conjunction. Additionally, we found an attraction effect even with syntactically singular attractors when they contained conjoined adjectives. This effect was significant in both speeded acceptability judgment tasks (Exp. 4 and 6) and marginal in self-paced reading (Exp. 5). It did not seem to be the result of a tendency of comprehenders to expect a plural noun after conjoined adjectives or of increased linear distance between the subject’s head noun and the verb (Exp. 6).

### *3.11.1 The role of morphological form in memory retrieval*

The results of Experiments 2-4 suggest that error-driven retrieval in the agreement computation process targets features more abstract than only the unequivocal exponent of syntactic plurality (plural ‘-s’) during cue-based memory retrieval. We found that comprehenders showed facilitation in ungrammatical sentences with a singular subject and a plural verb when a non-subject consisting of conjoined singular noun phrases was in attractor position, even though its syntactic plurality was not introduced by an unequivocal signal. In the self-paced reading experiment, this facilitation took the form of a reduced slow-down in the verb’s spillover region, and in the speeded acceptability

judgment tasks this was reflected in higher acceptance rates. In all three experiments, agreement attraction effects for conjoined singulars were numerically larger than for suffixal plurals, but this was only statistically significant in Experiment 2.

Of course, it is generally accepted that subject-verb agreement in the grammar of English is licensed by abstract syntactic number rather than the presence of particular morphological correlates of this abstract category. However, given previous findings that agreement attraction in comprehension is sensitive to the attractor's plural formation strategy in Arabic (Tucker et al., 2015), it could have been the case that online processing mechanisms target only certain exponents of the abstract category. Especially since agreement attraction reflects an error-driven rechecking process, it would have been possible for the parser to rely on different cues than the ones in which the dependency is stated in the grammar. Unequivocal morphological signals of syntactic plurality could have been the primary target of this memory retrieval process, but our results suggest that retrieval models need to include more abstract cues that are not directly tied to morphological form, even when retrieval is error-driven. Our results similarly indicate that syntactic number is rapidly computed for conjoined singulars, as they were able to interfere with the memory retrieval operation cued at the immediately subsequent verb.

Our findings show that retrieval for agreement checking was not sensitive to the vehicle by which syntactic plurality is introduced. While some studies on agreement production and the Arabic results in comprehension from Tucker et al. (2015) suggest that there is a tight link between morphological form and retrieval cues, the data from the studies presented here indicate that retrieval models must also include abstract cues not directly tied to the morphological exponence of a feature. As we discuss further in the

next section, some of our results in fact provide some positive evidence for the effect of surface form on the memory retrieval processes associated with agreement computations: The agreement attraction effect we observed with singular attractors containing conjoined adjectives ('*the diligent and compassionate doctor*') might be interpreted as an indication that the morpheme '*and*', which is an imperfect correlate of syntactic plurality, is targeted by the verb's number retrieval cue, even if the noun phrase in which it occurs is not syntactically plural. However, conjoined adjectives led to a markedly smaller and less reliable attraction effect than suffixal plural attractors, indicating that overt correlates of syntactic plurality are not the main target of the number retrieval cue on the verb.

### *3.11.2 Notional plurality in agreement computation in comprehension*

While our findings indicate that an unequivocal morphological correlate of syntactic plurality (plural '*-s*') is not required for a structurally inaccessible noun phrase to function as an attractor in subject-verb agreement processing, our experiments cannot clearly distinguish between the role of syntactic and notional plurality. The conjoined singular noun phrases we used as attractors in Experiments 2-4 are not only syntactically but also notionally plural. In fact, conjoined noun phrases have been argued to be even "more" plural than regular plurals in a certain sense: they introduce discourse referents that license the subsequent use of two non-coreferential pronouns, unlike plural definite descriptions (Patson, 2014). It could be argued that the numerically larger attraction effects observed with conjoined singular noun phrases hint at the impact of notional number in addition to syntactic number in agreement processing in comprehension. Production research has shown effects of notional number on agreement computation

with collective nouns (like ‘*the fleet*’) and with conjoined noun phrases (Humphreys & Bock, 2005; Brehm & Bock, 2013, 2016; Mircovic & MacDonald, 2013; Lorimor et al., 2016). However, the fact that there was no statistical difference between conjoined singular attractors and suffixal plural attractors in two out of three experiments means that we cannot conclude that notional number contributed to attraction and further research is needed. I return to this issue in Chapter 3.

The results from the experiments presented in this chapter are consistent with the idea that the factors affecting agreement processing in production and comprehension are the same. In fact, unpublished pilot data from a sentence completion task using the complex subjects in Experiment 2 as preambles suggest that conjoined singular noun phrases in attractor position also increase the likelihood of agreement errors in production. If notional number and morphophonological form have an impact on agreement attraction in comprehension, this is consistent with the idea that agreement errors in production are also a phenomenon based on similarity-based interference in memory retrieval, although of course it does not provide evidence against a representational account of agreement attraction in production.

### *3.11.3 Associative Cues*

Experiments 5-7 provided some evidence that the presence of the conjunction ‘*and*’ within the attractor caused a small interference effect even when the phrase was syntactically singular. Any conclusions should be taken as somewhat preliminary since the effect was only marginally significant in self-paced reading (Exp. 6) and the linear distance between the verb and the subject’s head noun in Experiment 7 was still greater

for the conjoined adjective construction than for the stacked adjective construction. Nevertheless, the observed agreement attraction effect with conjoined adjectives is compatible with the hypothesis that the relationship between retrieval cues and features in sentence processing is associative rather than categorical, and may not strictly follow the cue-feature relationships licensed by the grammar (Engelmann, Jaeger & Vasishth., 2016). Under this view, the relationship between cues and features is not a categorical match or mismatch; instead cues can be associated with multiple features to different extents. The association between cues and features is learned based on exposure, and while they usually reflect grammatical knowledge, over time co-occurrence patterns can lead to the association of cues with features they are not linked to in the grammar.

Engelmann et al. (2016) suggest that if two features frequently co-occur on the target item in a linguistic dependency, over time they might both become associated with the retrieval cue even if only one of them is conceptually linked with it. For example, in the case of reciprocals in English, the features *+c-command* and *+plural* always co-occur on the antecedent of the reciprocal. Consequently, in this dependency the plural retrieval cue becomes associated not only with the plural feature but also the c-command feature, and vice versa for the c-command retrieval cue. In the case of subject-verb agreement, while the actual target of the number retrieval cue is (syntactic) plurality, which controls agreement in the grammar, the presence of ‘*and*’ might have served as a kind of surface cue to plurality, even in the absence of a syntactically plural attractor.

To determine whether ‘*and*’ could potentially become associated with the plural retrieval cue through frequent co-occurrence with syntactic plurality, we conducted a small corpus-based analysis to determine its distribution. The corpus was a subset of the

Corpus of Contemporary American English (Davies, 2008) consisting of 250 sentences.

We limited the analysis to occurrences of '*and*' for which the syntactic context could be unambiguously identified, resulting in a total of 308 tokens. The distribution was as follows: 3.2% occurred between two prepositional phrases (10 tokens), 4.2% between verb phrases (13 tokens), 15.3% between adjectives (47 tokens), 23.7% between clauses (73 tokens) and 53.6% between two noun phrases (165 tokens). A native speaker of English judged whether each instance of conjoined noun phrases would take plural agreement if it occurred in subject position. Only two of the tokens of '*and*' occurred in a conjoined phrase that the native speaker considered likely to take singular agreement. This indicates that over half of all tokens of '*and*' co-occur with syntactic plurality. In summary, although singular attractors with conjoined adjectives are not actually plural, a small corpus search confirmed that there is a strong correlation between '*and*' and syntactic plurality. It is possible that its frequent co-occurrence with syntactic plurality has led '*and*' to become associated with the plural retrieval cue. This association would not be as strong as the association between the cue and its actual target feature. While a singular attractor with conjoined adjectives might receive some activation from the verb's plural retrieval cue, this would lead to the attractor being misretrieved much less frequently than a syntactically plural attractor.

An associative cue account might also provide an explanation for the effects of morphological form on attraction observed by Tucker et al. (2015) in Arabic. Interestingly, in Arabic the majority of inanimate ablauting plurals require obligatory singular agreement even in the plural (Ryding, 2005). Although the ablauting plurals used in the study by Tucker et al. referred to animates and did not trigger obligatory singular

agreement, the frequent use of ablauting plurals with singular agreement might impact to what extent this vehicle of plurality is associated with the plural retrieval cue on the verb.

Although the associative cue account provides an appealing explanation for these effects, unconstrained such an account would easily become over-powerful. Our intuition is that singular nouns ending in ‘-s’ would not drive agreement attraction effects in comprehension. But why would some features correlated with plurality act as associative cues (*‘and’*) and some not (*‘-s’*)? Similarly, what prevents associative cues from becoming weighted as strongly as the ‘true’ cues? These will be important questions to be investigated by future work.

### 3.12 Conclusion

In this chapter, we used self-paced reading and speeded acceptability judgments to demonstrate that the vehicle by which the relevant feature that licenses the subject-verb agreement dependency in the grammar, syntactic number, is introduced, does not determine whether a structurally inaccessible syntactically plural noun phrase causes similarity-based interference in error-driven retrieval. Conjoined singular NPs, which are syntactically plural but contain only an equivocal morphological signal of plurality, caused strong attraction effects. Just like suffixal plural attractors, they increased the rate at which sentences with an agreement violation were perceived as acceptable in speeded judgments and also reduced the slowdown associated with processing an agreement violation in self-paced reading. This indicates that the verb’s number retrieval cue that is used in the error-driven retrieval of the agreement controller does not specifically target only the unequivocal exponent of the abstract feature (plural ‘-s’). Instead, the results

clearly demonstrate that it is specified in more abstract terms, consistent with the way the dependency is stated in the grammar. We can conclude that although error-driven retrieval is a sort of repair mechanism that is triggered upon detecting a problem, it is still the case that the cues and features used in this process are based on the abstract linguistic knowledge about the specific dependency.

However, we also found a numerically much smaller attraction effect with attractors with conjoined adjectives, which are not syntactically plural and do not license plural agreement in the grammar. We hypothesize that this is because '*and*' frequently co-occurs with syntactic plurality and has therefore become weakly associated with the plural retrieval cue. Taken together, these findings suggest that the feature primarily targeted in error-driven memory retrieval operations linked to agreement processing are more abstract than a specific exponent of the abstract category syntactic number, but that due to the associative nature of cues and features, surface cues that are imperfect correlates of syntactic plurality like '*and*' can also interfere to a smaller extent. This pattern is reminiscent of recent production findings, which have shown that morphophonological form and notional number have an effect on agreement computation in production. If agreement attraction is affected by the same factors in production and comprehension, this might be a reason to prefer a unified account of this phenomenon across both modalities. Given that there is strong evidence that agreement attraction in comprehension is the result of similarity-based interference in memory retrieval, and that the production data appears to be consistent with either a representational or a cue-based retrieval account, to me this seems to suggest that such a unified model would be a cue-based retrieval model.

# **Chapter 4**

## **4.1 Outline**

In this chapter, I investigate whether information that is not relevant to the subject-verb agreement dependency in the grammar is used to guide retrieval of the agreement controller in error-driven retrieval. Chapter 2 suggested that agreement attraction is the result of an error-driven feature checking process, and at least in general not a reflection of wholesale restructuring of the sentence. Chapter 3 showed that it has to be possible for the morphosyntactic cues to this feature checking process to be as abstract as the terms in which the dependency is stated in the grammar. Next I investigate whether the cues to this error-driven feature checking process are limited to grammatically relevant information, or whether non-syntactic, grammatically irrelevant information is also used. Two self-paced reading experiments test the effect of notional number and the plausibility match between the attractor and the verb on agreement attraction in comprehension. The experiment on notional number uses syntactically singular collective nouns as attractors and manipulates whether they are more likely to be construed with a collected group reading (notionally singular) or a distributed group reading (notionally plural). The data show no effect of this manipulation, indicating that notional number is not targeted in error-driven retrieval for agreement checking. The second experiment varies the verb in passive sentences so that the attractor inside the complex subject either is or is not a plausible subject. Although initial results suggest that plausibility might have an effect on agreement attraction in comprehension, more fine-grained aspects of the data did not pattern as expected. A replication of the same experiment indicates that the match between the semantic content of the attractor and the verb does not impact

agreement attraction. These findings are consistent with an error-driven retrieval process that relies exclusively on the information that defines agreement in the grammar.

## 4.2 Notional number in agreement processing

Subject-verb agreement in English is a dependency in which the syntactic number of the verb has to match the syntactic number of the subject. Agreement is not determined by the notional number of the subject, as clearly demonstrated by the fact that noun phrases with the numeral '*one-point-zero*' take plural agreement ('*one-point-zero children*'), in spite of having a singular referent. However, the syntactic number of a noun phrase is not always a straightforward function from the number features of its parts. Whether a noun phrase like '*my wife and confidante*' takes singular or plural agreement depends on how it is notionally construed (Heycock & Zamparelli, 2005). If '*wife*' and '*confidante*' refer to separate entities, the entire NP is syntactically plural. If they both refer to the same entity and the entire NP has a singular referent, it is syntactically singular. This interaction between syntactic and notional information in implementing agreement has received considerable attention in the literature on sentence production and agreement attraction has been used as a tool to try to disentangle these factors.

The classic agreement attraction errors in production arise with complex subjects with a singular head noun modified by a prepositional phrase. These elicit a higher rate of agreement errors (production of ungrammatical plural verbs) when the noun inside the modifier (attractor) is plural compared to when it is singular (Bock & Miller, 1991). A follow-up study by Eberhard and Bock (1993) aimed to further explore the effects of

syntactic, notional, and morphophonological information on the production of agreement with the help of agreement attraction. A sentence completion task replicated the previous results with syntactically plural nouns in attractor position. To examine the effect of notional number, they also tested collective nouns in attractor position. In American English, collective nouns like '*fleet*' take singular agreement. However, they can be construed either with a collected group reading (notionally singular) or a distributed group reading (notionally plural). Under a collective group reading, a syntactically singular collective noun refers to a grouping of things and has a singular referent. In contrast, the distributed group reading of a collective noun refers to multiple objects and is notionally plural. There was no clear evidence for a higher agreement error rate with preambles containing syntactically singular collectives compared to individual nouns. This suggests that their notionally plural construal did not affect agreement implementation. Nevertheless, there was a (statistically non-significant) correlation between how likely a particular syntactically singular collective was judged to refer to multiple entities and the frequency with which it elicited agreement errors. Plural collectives in attractor position ('*fleets*') also led to numerically higher rates of agreement errors than plural individual nouns, further suggesting that there might be a very small effect of notional number.

The currently prevalent view is that agreement attraction errors in production can be explained by representational accounts in which the number feature of a singular subject is affected by the presence of a plural attractor, either through feature percolation or spreading activation (e.g. Bock & Eberhard, 1993; Pearlmuter et al., 1999; Bock et al., 2004). The most influential representational account is the Marking and Morphing model

(Eberhard, Cutting & Bock, 2005). According to this model, the number information on a noun phrase ranges continuously from unambiguously singular to unambiguously plural and the number marking on the verb is probabilistic. Although a subject with a singular head noun should be valued as unambiguously singular, the presence of a plural element inside it ('*The key to the cabinets*') will raise the value and make the subject number more ambiguous. The claim is that agreement errors arise because this sometimes leads to choosing the wrong verb form. Consequently, much of the production research that claims to show evidence that notional number matters in agreement attraction really focuses on the notional number of the subject as a whole, rather than of the attractor itself.

The evidence most commonly cited in support of the claim that notional number has an impact on agreement attraction in production comes from a number of studies that manipulate the notional number of the entire subject. Definite descriptions with singular individual nouns can sometimes also be construed as notionally plural. For example, '*the label on the bottles*' can denote several tokens of a label on multiple bottles.<sup>6</sup> Effects of notional number with this type of subject have been found in agreement production across a variety of languages: Italian (Vigliocco, Butterworth & Semenza, 1995), Spanish (Vigliocco, Butterworth & Garrett, 1996), French and Dutch (Vigliocco, Hartsuiker, Jarema & Kolk, 1996). However, Vigliocco et al. (1996) failed to find this effect in English. Although it is conceivable that the role of notional number differs crosslinguistically, a later study by Humphreys and Bock (2005) has been argued to show that it does matter in agreement production in English.

---

<sup>6</sup> However, '*the label on the bottles*' can also refer to a single type of label, in which case it is notionally singular.

Humphrey and Bock (2005) used complex subjects with syntactically singular collectives as head nouns. They were followed by a prepositional modifier biasing construal of the referent towards either a collected group reading ('*The gang near the motorcycles*') or a distributed group reading ('*The gang on the motorcycles*'). Humphreys and Bock found that the rate of plural verb responses depended on the construal of the collective's notional number. Preambles encouraging distributed group readings more frequently led to the production of plural verbs.

Similarly, Brehm and Bock (2013) argue that the likelihood of producing plural agreement with a singular subject depends on how semantically integrated its referent is: more integrated preambles ('*The drawing of the flowers*') were less likely to cause agreement errors than less integrated preambles ('*The drawing with the flowers*'). Brehm and Bock argue that this reflects the effect of notional number: the less integrated a complex referent is, the more likely it is to be mentally construed as plural. Studies in Serbian, Dutch, and German (Mircovic & MacDonald, 2013; Lorimor et al., 2016) also report higher rates of plural agreement for notionally plural subjects. In summary, there is accumulating evidence that notional number impacts the implementation of subject-verb agreement in production cross-linguistically. While these findings indicate that the notional number of the subject affects the implementation of subject-verb agreement in production, it should be noted that what is being tested is the role of the notional number of the referent denoted by the entire subject, not by the attractor.

Under the Marking and Morphing model, the effect of the subject's notional number on agreement production is attributable to the same mechanism that is responsible for agreement errors in the presence of a plural attractor, which is why these

studies have been taken as evidence for the impact of notional number on agreement attraction. The direction of encoding in production, from the message-level to syntactic encoding, means that notional number, which is part of the conceptual representation, can impact how the subject's number feature is valued.

It is conceivable that the impact of the subject's notional number is a result of a different process than that underlying what I have referred to as classic agreement attraction errors. It has been argued that the increased rate of agreement errors with singular subjects and plural attractors (*'the key to the cabinets'*) is not a result of misrepresenting the subject's number information but an instance of similarity-based interference (Badecker & Kuminiak, 2007; Slevc & Martin, 2016). This does not preclude the notional number of the entire subject from having an effect on agreement in production. However, the agreement errors observed with preambles like *'the gang on the motorcycles'* would reflect both the fact that the notional number of the entire subject, the actual agreement controller, can impact agreement implementation (see agreement with phrases like *'my wife and confidante'*) and that the cue-based retrieval of the agreement controller is susceptible to similarity-based interference from the attractor.

Ultimately, regardless of whether agreement errors in production are the result of misrepresentation or cue-based retrieval, we cannot draw clear conclusions about the role of notional number in agreement attraction in comprehension based on the findings from these production studies. If we assume a misrepresentation account for production, the results of the studies discussed in this section are a reflection of the same process as classic agreement attraction. However, there is convincing evidence that agreement attraction in comprehension is an example of similarity-based interference in memory

retrieval, so it would be driven by a different mechanism. A priori, we have no reason to believe that these two different mechanisms should be sensitive to exactly the same type of information. Alternatively, if agreement attraction in production is the result of cue-based retrieval in production just like in comprehension, the fact that these experiments manipulated the entire subject's notional number rather than the attractor's means that they cannot tell us anything about whether a structurally irrelevant noun phrase can be misretrieved as the agreement controller if it is notionally plural.

In prior chapters, I have argued that in comprehension attraction is the result of an error-based feature-checking process that makes use of abstract grammatical cues for retrieval. In the next experiment, I examine whether this feature-checking process might also make use of a non-grammatical cue like notional number. In a two-stage model of agreement attraction, the parser predicts the number marking on the verb based on the subject's number feature. I will assume that this predictive mechanism is faithful to grammatical constraints. If the illusory licensing of an agreement violation is an error-driven phenomenon that arises when the bottom-up input does not match the prediction, agreement attraction is the outcome of what is essentially a repair process. It is quite possible that once the parser has received an error-signal, it uses all information that could potentially be helpful in retrieving the agreement controller, even information that is irrelevant to the dependency in the grammar. While subject-verb agreement is a syntactic dependency, there is a correlation between notional number and syntactic number. This correlation is imperfect and syntactic and notional number are dissociable: several kinds of noun phrases are syntactically plural but do not have notionally plural referents. These include phrases headed by pluralia tantum ('*the scissors*') and those with

the determiners ‘*no*’ or ‘*zero*’. Nevertheless, in an error-driven retrieval rechecking operation notional number might be targeted as a proxy for syntactic number. I address this question in Experiment 8 by using syntactically singular collective nouns in attractor position and manipulating their notional number.

#### 4.3 Experiment 8: Notional number

The aim of Experiment 8 was to investigate whether notional number is targeted by the verb’s number retrieval cue in the error-driven retrieval of the agreement controller in comprehension. As outlined in the previous section, while there is evidence for the impact of notional number on the implementation of subject-verb agreement in production, the nature of the materials used in these studies means that it is not a-priori clear that attraction in comprehension would necessarily also be affected by notional number; whether we assume that the mechanism underlying (classic) agreement attraction in production is the same as in comprehension or not.

In Experiment 8, we use syntactically singular collective nouns in attractor position to test whether notional number impacts agreement attraction in comprehension. The notional number of these nouns is manipulated by a preamble sentence that exerts a bias either towards a collected group reading or a distributed group reading of the collective. While a collected group reading is notionally singular, a distributed group reading is plural on a conceptual level. If the retrieval process responsible for agreement attraction in comprehension is sensitive to notional number, we expect to see stronger

attraction effects with collectives that have a distributed group reading than collectives that have the same syntactic number but a collected group reading.

#### *4.3.1 Participants*

24 members of the University of Maryland community participated for course credit or monetary compensation. Data from one additional participant was excluded from all analyses due to low accuracy (< 80%) on the comprehension questions.

#### *4.3.2 Materials and Design*

There were 24 experimental items in a 2x2 design, crossing grammaticality and type of group reading. Each item consisted of two sentences: The first sentence (preamble) introduced a collective noun that was then repeated in attractor position in the second sentence. While all attractors were syntactically singular collectives, we manipulated whether the preamble sentence created a bias towards a collected or distributed group reading of the collective noun. Consequently, the preamble sentence in one item set varied between conditions. The second sentence contained a complex subject with a prepositional modifier, followed by an inflected form of ‘be’. The subject’s head noun was always singular, so grammatical sentences contained a singular verb and ungrammatical sentences a plural verb. The noun inside the propositional modifier was the collective introduced in the first sentence, which was either biased towards a collected or distributed group reading.

(1)

a) The **fleet** consisted of forty ships and looked very impressive.

The captain of the **fleet** *was* known for his battle skills..

b) The **fleet** consisted of forty ships and looked very impressive.

The captain of the **fleet** *were* known for his battle skills.

c) The **fleet** was powerful and looked very impressive.

The captain of the **fleet** *was* known for his battle skills.

d) The **fleet** was powerful and looked very impressive.

The captain of the **fleet** *were* known for his battle skills..

In addition to the experimental items, there were also 24 items that served as controls. There were four conditions in a 2x2 design, crossing grammaticality (grammatical/ungrammatical) and attractor type (singular/plural). The items were a subset of those used in Experiment 3 in Chapter 3, to which a preamble sentence had been added.

Experimental items and control items were distributed across 4 lists in a Latin Square Design, so that participants saw all items and only one condition per item. All of the lists also contained the same 90 filler items, all grammatical. Consequently, each participant saw 24 ungrammatical and 114 grammatical items in this experiment. The order of presentation was randomized for each participant.

#### *4.3.3 Procedure*

Items were presented on a computer screen using Linger software (Doug Rhode, MIT). The first sentence of each item was shown on the screen as a whole and participants had to press the space bar to move on to the second sentence. The second sentence was always presented word-by-word in a self-paced moving window paradigm (Just et al., 1982). On each trial, a number of dashes appeared on the screen, masking the words of the second sentence. Participants moved through the sentence by pressing the space bar. When moving to the next word, the previous word was remasked. Participants were instructed to read as naturally as possible and to make sure that they understood the sentences they were reading. All of the experimental and control items were followed by comprehension questions, as well as 30 of the filler items. Participants had to respond to comprehension questions as accurately and as fast as possible by pressing the ‘f’-key for ‘yes’ and the ‘j’-key for ‘no’. Feedback was only given when the answer was incorrect. Before the start of the experiment, there were five practice trials to familiarize participants with the procedure.

#### *4.3.4 Analysis*

The regions of interest consisted of single words and included the verb, the first word after the verb (first spillover region) and the second word after the verb (second spillover region). Reading times exceeding a threshold of 2000ms were discarded, resulting in the exclusion of less than 0.7% of the data in any of the regions of interest.

All analyses were conducted on log-transformed reading times. The data were analyzed with the *lmer* function in the *lme4* package (Version 1.1-12; Bates et al., 2015) in the statistical analysis software R (Version 3.3.2; R Core Development Team, 2017) with grammaticality and distributivity as fixed effects. We used effects coding for the fixed effects (grammatical: -0.5; ungrammatical: 0.5; collective: -0.5; distributive: 0.5). The random effects structure was initially maximally specified and then progressively simplified until the model converged for all regions of analysis (Barr et al., 2013). The final model included by-subject and by-item random intercepts, by-subject random slopes for grammaticality and attractor type, and by-item random slopes for grammaticality. The data from the control items was analyzed separately, following the same approach.

#### 4.3.5 Results

The results of the linear mixed effects models for the three regions of analysis are presented in Table 22 to 24. Table 25 to 27 contain the model's results for the control manipulation. Average log-transformed reading times are plotted in Figure 13 (experimental items) and Figure 14 (control items).

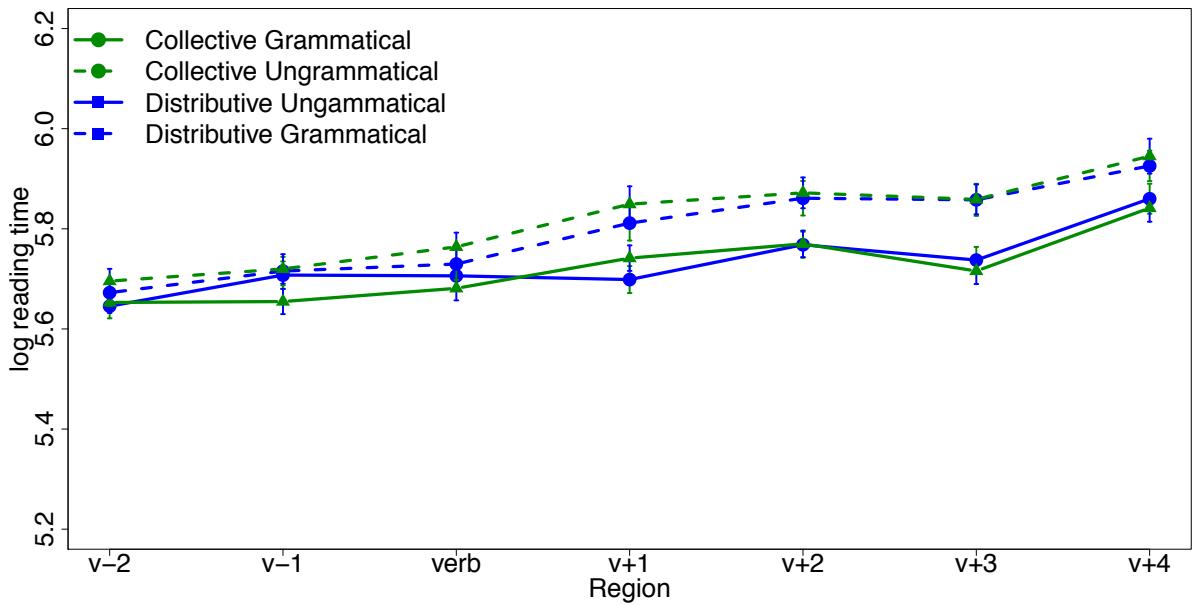


Figure 13: Region-by-region mean log reading times for the experimental items in Experiment 8. Error bars indicate standard error of the mean.

	Estimate	Std. Error	t-value
Intercept	5.72	0.05	<b>124.95</b>
Grammaticality	0.06	0.03	<b>2.07</b>
Distributivity	-0.01	0.02	-0.19
Interaction	-0.06	0.04	-1.47

Table 22: Results of the linear mixed effects model for the experimental items in the verb region in Experiment 8 (using log transformed RTs).

	Estimate	Std. Error	t-value
Intercept	5.78	0.05	<b>117.21</b>
Grammaticality	0.11	0.03	<b>3.65</b>
Distributivity	-0.04	0.02	-1.70
Interaction	0.01	0.05	0.11

Table 23: Results of the linear mixed effects model for the experimental items in the first spillover region in Experiment 8 (using log transformed RTs).

	Estimate	Std. Error	t-value
Intercept	5.82	0.04	<b>143.51</b>
Grammaticality	0.10	0.04	<b>2.53</b>
Distributivity	-0.01	0.03	-0.31
Interaction	-0.02	0.05	-0.36

Table 24: Results of the linear mixed effects model for the experimental items in the second spillover region in Experiment 8 (using log transferred RTs).

For the experimental items, there was a main effect of grammaticality in the verb region ( $t = 2.07$ ), with slower reading times for ungrammatical than grammatical items. This effect remained significant in both spillover regions (verb+1:  $t = 3.65$ ; verb+2:  $t = 2.53$ ). There were no other significant effects in any of the three regions of analysis.

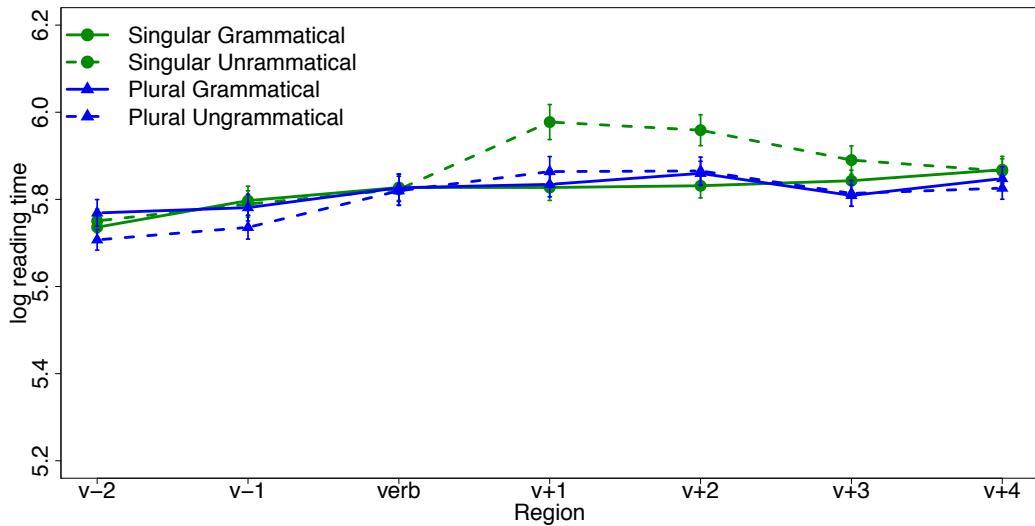


Figure 14: Region-by-region mean log reading times for the control items in Experiment 8. Error bars indicate standard error of the mean.

	Estimate	Std. Error	t-value
Intercept	5.826	0.05	<b>115.24</b>
Grammaticality	-0.002	0.03	-0.08
Number	-0.004	0.02	-0.13
Interaction	-0.009	0.05	-0.17

Table 25: Results of the linear mixed effects model in the verb region for the control manipulation in Experiment 8 (using log transformed RTs).

	Estimate	Std. Error	t-value
Intercept	5.88	0.06	<b>105.97</b>
Grammaticality	0.10	0.04	<b>2.26</b>
Number	-0.06	0.03	-1.82
Interaction	-0.13	0.05	<b>-2.63</b>

Table 26: Results of the linear mixed effects model in the first spillover region for the control manipulation in Experiment 8 (using log transformed RTs).

	Estimate	Std. Error	t-value

Intercept	5.88	0.05	<b>125.48</b>
Grammaticality	0.07	0.03	<b>2.39</b>
Number	-0.03	0.02	-0.96
Interaction	-0.12	0.05	<b>-2.55</b>

Table 27: Results of the linear mixed effects model in the second spillover region for the control manipulation in Experiment 8 (using log transformed RTs).

The results for the control items also showed a significant effect of grammaticality: subject-verb agreement violations led to slower reading times. This effect became significant in the first spillover region ( $t = 2.26$ ) and remained significant in the second spillover region ( $t = 2.39$ ). As expected, the interaction between grammaticality and attractor type was also significant in both spillover regions (verb+1:  $t = -2.63$ ; verb+2:  $t = -2.55$ ). The slowdown in response to an ungrammatical verb was reduced in the presence of a plural attractor compared to a singular attractor.

#### 4.3.6 Discussion

There was no evidence that syntactically singular collective nouns cause agreement attraction in comprehension, even when they occur in a context that encourages a (notionally plural) distributed group reading. The data for the control items are consistent with findings from previous self-paced reading studies. Subject-verb agreement violations led to slower reading times and this effect was mitigated in the presence of a plural attractor (standard agreement attraction effect). This confirms that participants in this experiment were processing the materials in such a way that agreement attraction effects could occur. For the experimental items, there was also an increase in reading times for ungrammatical sentences, as expected. This demonstrates

that participants were sensitive to subject-verb agreement violations in the experimental items. However, the slowdown was not affected by whether the collective in attractor position was biased towards a collected or distributed group reading. The lack of an interaction between grammaticality and type of group reading suggests that retrieval for agreement processing does not target notional number in comprehension.

Agreement is a syntactic dependency between the subject and the verb, so a reasonable default assumption is that the number cue on the verb specifically targets the syntactic number feature of items in memory. The data from this experiment are consistent with this assumption. This initially appears to be different from agreement processing in production, where notional number has been argued to impact agreement implementation. This claim is usually based on the results from Humphreys and Bock (2005), but there is a crucial difference between the materials in Experiment 8 and the items used by them. In their study, the collectives were the head nouns of the complex subjects ('*The gang on/near the motorcycles ...*'), whereas in the experiment reported here they were inside the prepositional modifier. Consequently, what Humphreys and Bock manipulated was whether the referent of the entire subject was construed as having a collected or distributed group reading, which is different from manipulating the notional number of a structurally irrelevant noun. The other production studies that have found effects of notional number have also used materials in which the entire subject had a notionally plural reading, not just the local noun. To our knowledge, the only production study that used collectives as attractors was the one conducted by Eberhard and Bock (1993). Interestingly, in that study they failed to find clear evidence that notional number affected agreement production.

#### 4.4 Semantic interference in comprehension

Another cue that is not relevant to subject-verb number agreement in the grammar but could be helpful in locating the agreement controller in memory is the lexical semantics of the verb; in particular the extent to which the attractor would be a plausible subject for the verb.

Similarity-based retrieval interference plays a large role in a content-addressable cue-based memory architecture (McElree 2000; Van Dyke & Lewis, 2003; Lewis & Vasishth, 2005; Martin & McElree, 2008, 2009). Under a cue-based retrieval framework, comprehenders encode items as bundles of features, which can then be directly targeted by matching retrieval cues. The features that need to be encoded in memory are usually not limited to morphosyntactic properties but also include semantic features. A word like '*boat*' obviously contains more information than just its category membership, number information and syntactic role in the sentence it appears in. It is beyond the scope of this dissertation to determine the exact nature of these semantic features. For our purposes it is sufficient to assume that in this chapter what we are referring to as semantic features is based on lexical semantics and plausibility (world knowledge).

Much recent research on retrieval interference in sentence processing has focused on morphosyntactic features in linguistic dependencies (e.g., Wagers et al., 2009; Dillon et al., 2013). However, similarity-based interference has also been observed between semantic features in retrieval operations that could have been thought to be cued only by syntactic features. A series of studies by Van Dyke and colleagues looked at semantic interference from items from within the sentence and from previously memorized lists

(Van Dyke & Johns, 2012). In a dual-task memory load and self-paced reading study, participants had to remember word lists such as ‘*table*, *sink*, *truck*’, followed by reading a sentence in a self-paced reading paradigm. The verb in the sentence was manipulated to make the items from the memorized list plausible or implausible direct objects (Van Dyke & McElree, 2006). For example, for the list above, they might see a sentence like (1).

(1) *It was the boat that the man who lived by the sea sailed/fixed after two sunny days.*

For the verb ‘*sailed*’, none of the items on the memorized list are plausible direct objects. However, all of them are plausible if the verb is changed to ‘*fixed*’. Van Dyke and McElree (2006) found that reading times at the verb were longer when the items on the list were plausible objects of the verb. This suggests that when participants encountered the verb, they used its semantic cues to retrieve the clefted object. If the items on the list also matched those semantic cues it is more difficult to retrieve the target. While the memory load task is quite different from normal sentence processing, semantic interference has also been observed from items within the same sentence. Van Dyke (2007) crossed syntactic and semantic overlap between the verb’s cues and the potential intervener in sentences like (2). Syntactic overlap was manipulated by varying whether the intervener was the subject of the relative clause. For semantic interference, Van Dyke varied whether the intervener was a plausible subject for the predicate.

(2)

a) *The frightened boy understood that the man who was swimming near the **dock/girl**  
was paranoid about dying.*

b) *The frightened boy understood that the man who said the **dock/girl** was dangerous  
was paranoid about dying.*

Syntactic interference at the predicate was stronger when the intervener shared the subject feature with the target, compared to when it was in a non-subject position.

Consistent with the findings from Van Dyke and McElree (2006), the data also showed similarity-based interference based on the plausibility match between the potential intervener and the predicate. Reading times were longer when not only the target but also the intervener were plausible subjects for the predicate. However, a follow-up study by Van Dyke and McElree (2011) failed to find semantic interference effects when the intervener did not also share a syntactic cue (here: subjecthood) with the actual target. They conclude that while semantic interference effects in comprehension clearly exist, their use might be gated by syntactic constraints or they might be weighted less heavily than syntactic cues. In the following experiment, I ask whether the parser uses plausibility information in error-driven retrieval for checking agreement, even though agreement is defined in purely formal terms in the grammar.

In production, there is evidence that the plausibility match between the attractor and the main verb affects agreement error rates. In a series of studies, Thornton and MacDonald (2003) used passives to investigate the role of the plausibility match between the attractor and the verb. In two sentence completion tasks, they tested whether

agreement error rates varied based on plausibility match. Participants were presented with verb and a preamble consisting of a complex subject with an inanimate head noun and an animate noun inside a prepositional modifier. Their task was to form a passive sentence with that verb and the preamble as the subject. One benefit of this slightly modified error elicitation task is that passive sentences require participants to produce an inflected form of ‘*be*’, which is unambiguously marked for number, unlike most verbs in English. In addition to head noun number and attractor number, Thornton and MacDonald also manipulated the plausibility match between the verb and the attractor by using different verbs. This allowed preambles to contain the same nouns across conditions. For example, a preamble like ‘*The feeling about the undergraduate student*’ was either presented with a verb like ‘*notice*’ or ‘*share*’. The subject’s head noun was always a plausible match for the verb, but plausibility as a subject varied for the attractor: Noticing a student is much more plausible than sharing one. Verb and noun pairs had been normed for plausibility in a separate plausibility rating study to ensure that changing the verb really did make a difference to how plausible the attractor was as a subject for a particular verb. Replicating previous error elicitation tasks, agreement errors were more frequent when the preamble contained a singular head noun and a plural attractor. Interestingly, this effect was modulated by plausibility. Error rates were higher when the attractor was a plausible passive subject for the verb, but only for this preamble type.

In a follow-up self-paced reading experiment, Thornton and MacDonald used the same items to test the effect of plausibility on agreement in comprehension. In order to provide participants with the number marking and the main verb’s meaning at the same time, they presented the auxiliary and the verb together while the rest of the sentence was

presented word-by-word. Similar to the production results, they found a plausibility effect only when the subject's head noun was singular and the attractor was plural. Reading times at the verb were longer when both the head noun and the attractor were plausible passive subjects for the verb. Although these results are intriguing, all of the conditions in the comprehension experiment were grammatical. If agreement attraction is the result of an error-driven checking mechanism, reading times in grammatical sentences are not informative about what type of information the error-driven retrieval process uses to find the agreement controller.

In Experiment 8, we did not find any evidence that the attractor's notional number played a role in the error-driven retrieval process triggered by subject-verb agreement violations. However, there is clear evidence that for at least some dependencies retrieval is susceptible to semantic similarity-based interference (Van Dyke & McElree, 2006). In Experiment 9, we ask whether the status of the attractor as a plausible subject for the verb impacts attraction rates. If agreement attraction is the result of an error-driven checking operation that exclusively uses the cues and features associated with the agreement dependency in the grammar, the plausibility match between the verb and the attractor should be of no consequence. However, it is possible that once an agreement violation is detected, the retrieval operation is not limited to abstract grammatical features. Instead, the parser might make use of all available information that could help retrieve the agreement controller, such as lexical semantics. In that case, an attractor that is a plausible subject for the verb would receive activation from this additional cue and would therefore be more likely to be misretrieved. This would be reflected in a stronger

attraction effect when the attractor is a plausible subject for the verb compared to when it is an implausible one.

## 4.5 Experiment 9a: Plausibility

The aim of Experiment 9 was to investigate whether comprehenders use the verb's semantic information to retrieve the agreement controller when they encounter an agreement violation. Following Thornton and MacDonald (2003), we tested this by manipulating the plausibility match between the attractor and the verb. Unlike Thornton and MacDonald, we also included agreement violations. If semantic features are used to guide retrieval, a plural attractor would be even more likely to be misretrieved when it is a plausible subject for the verb. This would be reflected in a greater reduction of the slowdown associated with agreement violations in self-paced reading. In contrast, if only the features that determine this dependency in the grammar are used, we expect to see no difference between attraction effects based on the plausibility match between the attractor and the verb.

### 4.5.1 Participants

47 participants were recruited on Amazon Mechanical Turk and received monetary compensation for their participation. They were all native speakers of American English and had passed a native speaker test. Data from one additional participant was excluded from all analyses due to low accuracy on the comprehension questions (< 80%).

#### 4.5.2 Materials and Design

The sentences in this experiment were passives with a complex subject consisting of an inanimate singular head noun and an animate attractor inside a prepositional modifier. There were 48 item sets: 40 were adapted from the items used in Experiment 3 of Thornton & MacDonald (2003) and 8 additional items following the same pattern were constructed for this experiment. Each item set had 8 conditions in a 2x2x2 design, crossing attractor number (*singular/plural*), grammaticality (*grammatical/ungrammatical*), and plausibility-match between the attractor and the main verb (*plausible/implausible*). Plausibility was manipulated by varying the main verb so that the attractor was or was not a plausible subject for it. For the Thornton and MacDonald items, the plausibility match between the noun pairs and the verbs had been normed in a separate plausibility rating study. For the additional 8 items we relied on the judgments of a native speaker. Each verb was used in two item sets and the animacy contrast between the head noun and the attractor meant that the attractor was a plausible/implausible subject for the same verb in both item sets. However, counterbalancing ensured that participants never saw the same verb twice. An example item set is provided in (3), slashes indicate how the sentences were segmented for presentation.

- (3)
- a) The feeling / about the undergraduate **student** / **was noticed** / by the dean / at the university.
  - b) The feeling / about the undergraduate **students** / **was noticed** / by the dean / at the university.
  - c) The feeling / about the undergraduate **student** / **were noticed** / by the dean / at the university.
  - d) The feeling / about the undergraduate **students** / **were noticed** / by the dean / at the university.

- e) The feeling / about the undergraduate **student** / **was shared** / by the dean / at the university.
- f) The feeling / about the undergraduate **students** / **was shared** / by the dean / at the university.
- g) The feeling / about the undergraduate **student** / **were shared** / by the dean / at the university.
- h) The feeling / about the undergraduate **student** / **were shared** / by the dean / at the university.

Since we grouped all of the words in the agentive by-phrase into one segment (*'by the dean'*), we added additional material to the end of Thornton & MacDonald's sentences to avoid potential wrap-up effects in the verb's spillover region.

Items were distributed across 8 lists so that each participant saw one condition per item. Although each verb pair was used for two items, participants never saw two items with the same verb. In addition to the experimental items, all lists also included the same 98 grammatical fillers. Since half of the experimental items each participant saw were ungrammatical, the rate of ungrammatical sentences was 16.4%. The order of presentation was randomized for each participant. Before the start of the experiment, participants completed four practice trials to familiarize themselves with the phrase-by-phrase self-paced reading paradigm.

#### 4.5.3 Procedure

Sentences were presented in a self-paced moving-window paradigm (Just et al., 1982) using Ibex software (Drummond, 2017). Each trial started with a series of dashes on the screen, masking the words of the sentence. Participants had to press the space bar to move from segment to segment. When a segment was revealed, the previous one was remasked by dashes. Participants were not able to see more than one segment on the screen at a time or to return to a segment once they had moved on. In Thornton and MacDonald's experiment, sentences were presented word-by-word except for the verb

region: the inflected form of ‘*be*’ and the main verb were presented together in one segment. We decided to present the entire sentences phrase-by-phrase to avoid drawing attention to the verb region. Experimental items were segmented in the following way: the first phrase consisted of the definite determiner and the subject’s head noun, the second segment was the complex subject’s prepositional modifier (including the attractor), followed by the inflected form of ‘*be*’ and the main verb, the agentive by-phrase and one final segment consisting of three words. The segments are marked by slashes in (3).

Participants were instructed to read as naturally as possible and to make sure that they understood the sentences they were reading. Each sentence was followed by a multiple-choice comprehension question and participants had to press the ‘*f*’-key to choose the first option or the ‘*j*’-key for the second option. The order in which the answers were displayed was randomized across participants. Participants completed four practice trials before the start of the experiment to familiarize themselves with the paradigm.

#### 4.5.4 Analysis

The regions of analysis were the verb region consisting of the inflected form of ‘*be*’ and the main verb and the spillover region made up of the agentive by-phrase. Reading times that exceeded a threshold of 3000ms were excluded as outliers<sup>7</sup>, which led to the exclusion of 0.4% of the data in the verb region and 0.6% in the spillover region.

---

<sup>7</sup> The threshold for exclusion was set higher than in the other self-paced reading experiments reported in this dissertation because of the phrase-by-phrase presentation of the materials.

The length of the verb region varied across conditions because plausibility was manipulation by varying the verb. The verbs in different conditions had different lengths, so we calculated residualized reading times based on the length of each region (Ferreira & Clifton, 1986; Trueswell et al., 1994; Thornton & MacDonald, 2003). Reading times were residualized across all experimental items and fillers. Positive residualized RTs indicate that a region was read more slowly than predicted, whereas negative residualized RTs indicate that a region was read faster than predicted.

Residualized reading times were then analyzed with linear mixed effects models using the *lmer* function in the *lme4* package ((Version 1.1-12; Bates et al., 2015) in R (Version 3.3.2; R Core Development Team, 2017). Attractor number, grammaticality and plausibility were entered into the model as fixed effects. We used effects coding for the contrasts (singular: -0.5, plural: 0.5; grammatical: -0.5, ungrammatical: 0.5; plausible: -0.5, implausible: 0.5). Following Barr et al. (2013), the random effects structure was initially maximally specified and then reduced until convergence was reached. We report results from the model that converged for both regions of analysis. The final model had by-subject and by-item random intercepts, by-subject random slopes for grammaticality and plausibility, and by-item random slopes for plausibility.

In addition to the main analysis with all three fixed effects, we also conducted additional 2x2 analyses of the verb region split by plausibility. The fixed effects were grammaticality and attractor number and the random effects included by-subject and by-item random intercepts, and by-subject random slopes for grammaticality and number.

#### 4.5.5 Results

Accuracy on the comprehension questions in the experimental conditions ranged from 90.6% to 94.8%, indicating that participants were paying attention to the task. The output of the linear mixed effects model for the verb region is presented in Table 28 and for the spillover region in Table 29. For readability, region-by-region average residualized reading times are plotted in two separate graphs split by plausibility: region-by-region average residualized RTs for conditions in which the attractor was a plausible subject for the main verb are plotted in Figure 15 and region-by-region average residualized RTs for conditions in which the attractor was not a plausible subject for the verb are plotted in Figure 16.

In the verb region, there was a main effect of grammaticality ( $t = 3.9$ ). As expected, participants had slower reading times when there was a subject-verb agreement violation. The interaction between grammaticality and attractor number was also significant ( $t = -2.44$ ). The slowdown elicited by an agreement violation was smaller in the presence of a plural attractor. Interestingly, the three-way interaction between grammaticality, attractor number and plausibility came out significant ( $t = 2.17$ ). This indicates that the interaction between grammaticality and attractor number (attraction effect) differed based on whether the attractor was a plausible subject for the main verb. We followed this up by analyzing the data split by plausibility, which we report below. In the spillover region, the only significant effect was the main effect of plausibility. The region was read faster when only the subject's head noun was a plausible match for the main verb.

	Estimate	Std. Error	t-value
--	----------	------------	---------

Intercept	90.57	13.7	<b>6.61</b>
Grammaticality	72.72	18.63	<b>3.90</b>
Number	-13.62	12.30	-1.11
Plausibility	-22.20	12.77	-1.74
Gram x Number	-60.04	24.60	<b>-2.44</b>
Gram x Plausibility	-2.85	24.59	-0.12
Number x Plausibility	-6.96	24.58	-0.28
Number x Gram x Plausibility	106.49	49.16	<b>2.17</b>

Table 28: Results of the linear mixed effects model in the verb region in Experiment 9 (using log transformed RTs).

	Estimate	Std. Error	t-value
Intercept	-10.73	16.77	-0.64
Grammaticality	-1.80	13.12	-0.14
Number	4.23	11.92	0.36
Plausibility	-35.87	13.87	<b>-2.59</b>
Gram x Number	-13.39	23.84	-0.56
Gram x Plausibility	-0.67	23.84	-0.03
Number x Plausibility	-23.73	23.84	-0.10
Number x Gram x Plausibility	-6.23	47.65	-0.13

Table 29: Results of the linear mixed effects model in the spillover region, Experiment 9 (using log transformed RTs).

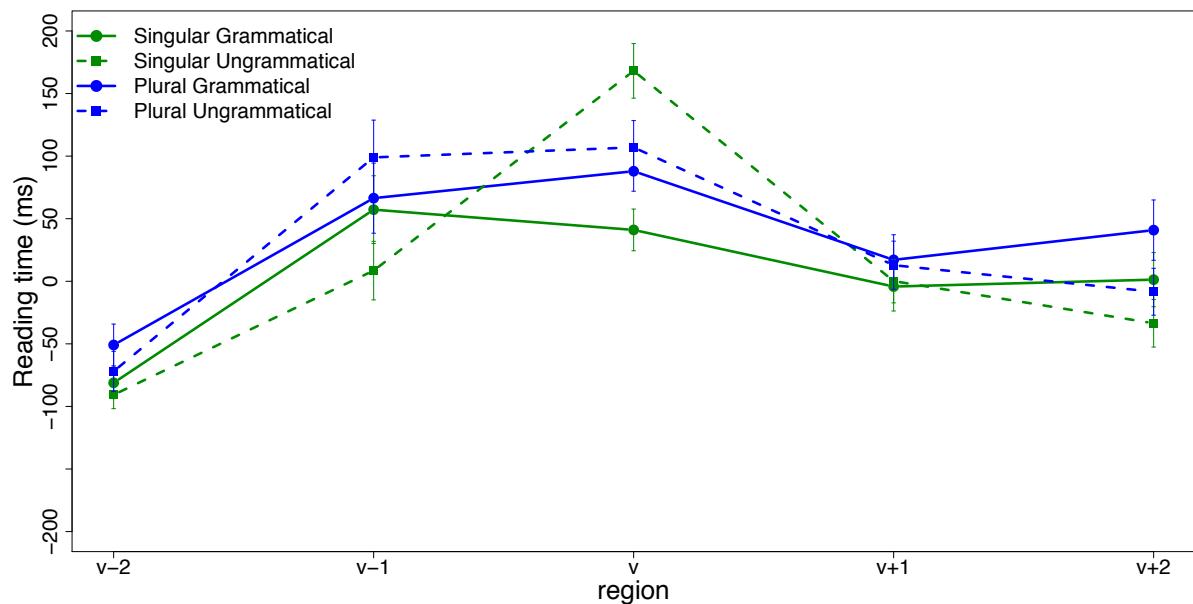


Figure 15: Region-by-region mean residualized reading times across conditions in which not only the head noun but also the attractor were plausible subjects of the main verb in Experiment 9. Error bars indicate standard error of the mean.

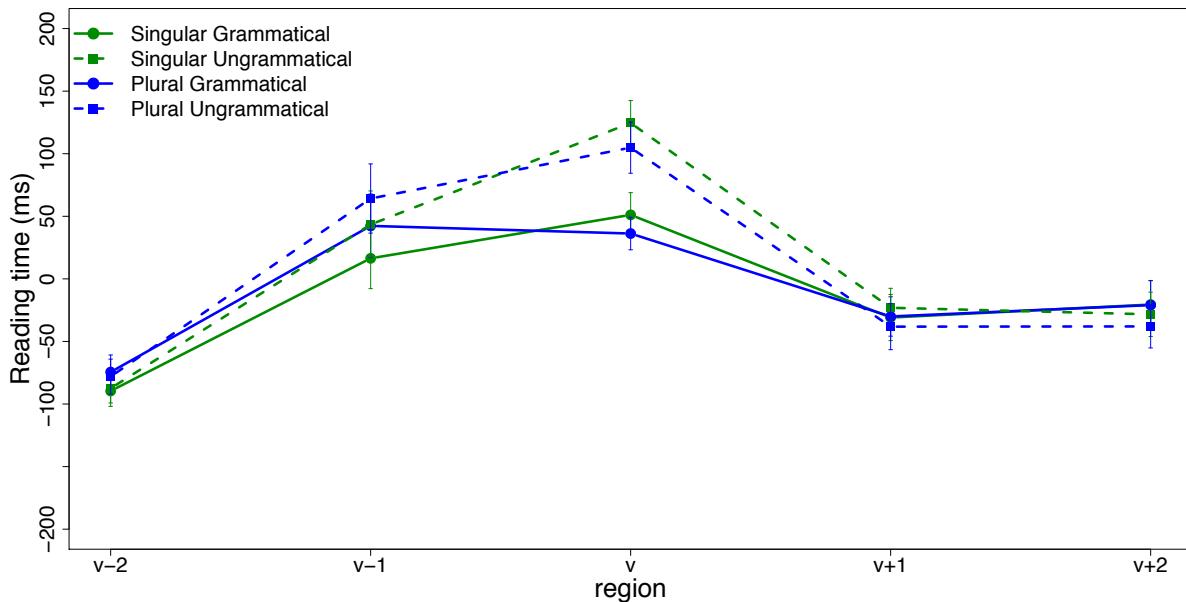


Figure 16: Region-by-region mean residualized reading times across conditions in which only the head noun but not the attractor were plausible subjects of the main verb in Experiment 9. Error bars indicate standard error of the mean.

Table 30 and 31 show the results of the linear mixed effects models for the verb region split by plausibility. Examining only the conditions in which the attractor was a plausible subject for the main verb, there was a main effect of grammaticality ( $t = 3.06$ ) and an interaction between grammaticality and attractor number ( $t = -3.13$ ). Subject-verb agreement violations resulted in a slowdown in the verb region, but the impact of ungrammaticality was reduced when the attractor was plural. No significant effects were found in the spillover region. In contrast, the model for conditions in which there was a plausibility mismatch between the main verb and the attractor only showed a main effect of grammaticality ( $t = 3.56$ ) but no interaction between grammaticality and attractor number ( $t = -0.2$ ). Agreement violations led slower reading times and this slowdown was not affected by attractor number. No significant effects were found in the spillover region.

	Estimate	Std. Error	t-value
Intercept	101.88	15.92	<b>6.40</b>
Grammaticality	74.32	24.28	<b>3.06</b>
Attractor Number	-10.13	18.40	-0.55
Interaction	-113.14	36.16	<b>-3.13</b>

Table 30: Results of the linear mixed effects model in the verb region when the attractor is a plausible subject for the verb, Experiment 9 (using log transformed RTs).

	Estimate	Std. Error	t-value
Intercept	79.26	14.12	<b>5.62</b>
Grammaticality	71.48	20.10	<b>3.56</b>
Attractor Number	-16.98	16.86	-1.01
Interaction	-6.58	33.37	-0.20

Table 31: Results of the linear mixed effects model in the verb region when the attractor is not a plausible subject for the verb, Experiment 9 (using log transformed RTs).

#### 4.5.6 Discussion

In Experiment 9a, we found the expected slowdown associated with agreement violations in the verb region. Overall, the slowdown was larger when the attractor was singular than when it was plural. However, this was only true for conditions in which the attractor was a plausible subject for the main verb. These results are consistent with a mechanism for retrieving the agreement controller from memory that uses not only the verb's structural and number information but also its semantic content. This would suggest that in error-driven retrieval for checking subject-verb agreement, the cues that are used are not limited to the terms that define the dependency in the grammar.

However, the three-way interaction between grammaticality, attractor number and plausibility did not follow the numerical pattern predicted if semantic cues were used in error-driven retrieval. Essentially, using semantic cues would predict that the agreement attraction effect should be stronger for conditions in which the attractor is a plausible

subject for the verb. Instead, the three-way interaction was driven by reading times in the sentences in which the attractor was a plausible subject for the verb: we found slower reading times in the plural grammatical condition and faster reading times in the singular ungrammatical condition compared there was no plausibility match between the attractor and the verb. This is entirely unexpected and cannot easily be accounted for by a model in which error-driven cue-based retrieval for agreement checking uses semantic information, so Experiment 9b aims to replicate the findings of Experiment 9a.

#### 4.6 Experiment 9b: Plausibility (replication)

The aim of Experiment 9b was to replicate Experiment 9a. In Experiment 9a, we found a significant three-way interaction between grammaticality, attractor number and plausibility. For sentences in which the attractor was a plausible subject for the verb, agreement violations caused a smaller slowdown in the verb region when the attractor was plural. In contrast, no such agreement attraction effect was observed when the attractor was not a plausible subject for the main verb. However, this interaction was driven by an unexpected pattern: when the attractor was a plausible subject for the verb, reading times in grammatical sentences with a plural attractor were slower and the slowdown in ungrammatical sentences with a singular attractor was larger. This is a surprising result and needs to be further investigated by replicating the experiment.

#### *4.6.1 Participants*

46 native speakers of American English participated in this Experiment. They were recruited via Amazon Mechanical Turk and received monetary compensation. Data from one additional participant was excluded from all analyses because of low accuracy on the comprehension questions (<80%).

#### *4.6.2 Materials, Procedure and Analysis*

The materials, procedure and analysis in Experiment 2b were identical to Experiment 2a. 0.3% of the data in the verb region and 0.4% in the spillover region were excluded as outliers.

#### *4.6.3 Results*

Participants answered comprehension questions accurately on between 91.5% and 95% of the trials across experimental conditions. Table 32 and 33 present the output of the linear mixed effects model with grammaticality, attractor number and plausibility as fixed effects, in the verb region and the spillover region respectively. Figure 17 shows region-by-region average residualized reading times for conditions in which the attractor was a plausible subject for the main verb and Figure 18 shows region-by-region average residualized RTs for conditions in which the attractor was not a plausible subject for the verb.

	Estimate	Std. Error	t-value
Intercept	52.46	10.31	<b>5.09</b>

Grammaticality	50.83	16.50	<b>3.08</b>
Number	-14.08	11.93	-1.18
Plausibility	11.53	13.98	0.83
Gram x Number	-42.88	23.86	-1.80
Gram x Plausibility	1.56	23.86	0.07
Number x Plausibility	-10.50	23.86	-0.44
Number x Gram x Plausibility	3.40	47.70	0.07

Table 32: Results of linear mixed effects model in the verb region (replication), Experiment 9b (using log transformed RTs).

	Estimate	Std. Error	t-value
Intercept	-32.36	12.44	<b>-2.60</b>
Grammaticality	36.33	13.99	<b>2.60</b>
Number	-6.96	11.10	-0.63
Plausibility	-0.07	14.24	-0.01
Gram x Number	10.47	22.20	0.47
Gram x Plausibility	-44.39	22.19	<b>-2.00</b>
Number x Plausibility	50.21	22.19	<b>2.26</b>
Number x Gram x Plausibility	-14.76	44.37	-0.33

Table 33: Results of the linear mixed effects model in the spillover region (replication), Experiment 9b (using log transformed RTs).

The main effect of grammaticality was significant in the verb region ( $t = 3.08$ ) and remained significant in the spillover region ( $t = 2.6$ ). As expected, agreement violations elicited slower reading times. Although the slowdown associated with ungrammaticality was already numerically smaller in the verb region when the attractor was plural compared, the interaction between grammaticality and attractor number only reached significance in the spillover region ( $t = -2$ ). There was also a significant interaction between attractor number and plausibility in the spillover region.

Importantly, the three-way interaction between grammaticality, attractor number and plausibility was not significant in either the verb region or the spillover region. For comparison with Experiment 9a, below we still provide an analysis of the data split by plausibility in Table 34 and 35.

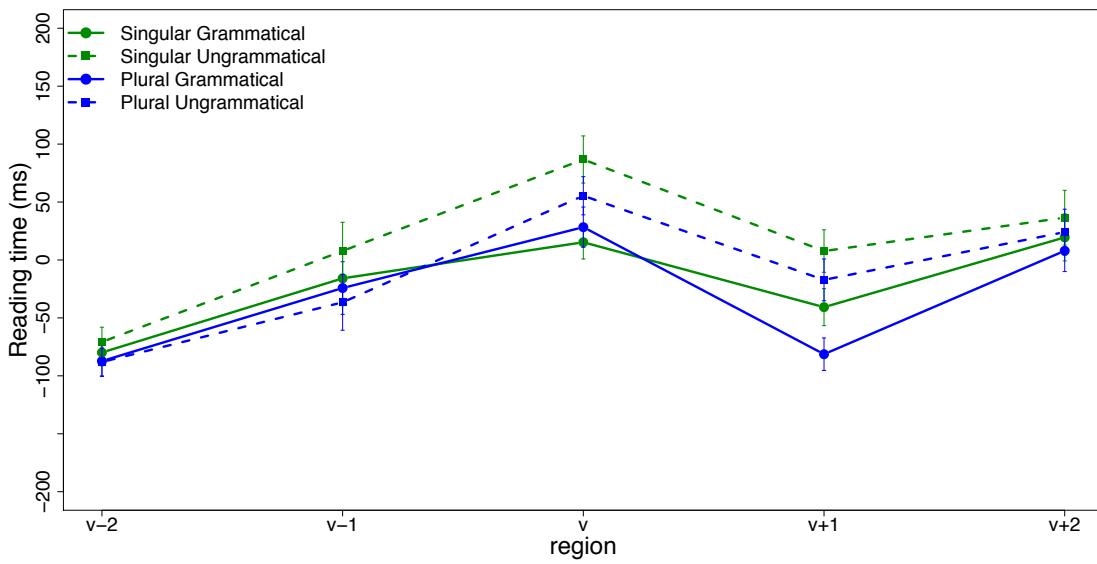


Figure 17: Region-by-region mean residualized reading times across conditions in which not only the head noun but also the attractor were plausible subjects of the main verb (replication), Experiment 9b. Error bars indicate standard error of the mean.

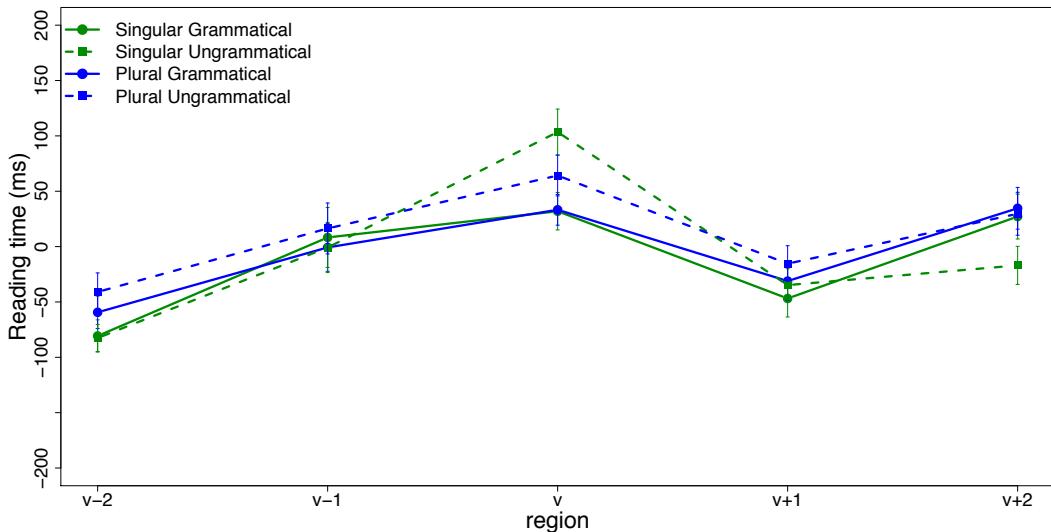


Figure 18: Region-by-region mean residualized reading times across conditions in which only the head noun but not the attractor were plausible subjects of the main verb (replication), Experiment 9b. Error bars indicate standard error of the mean.

	Estimate	Std. Error	t-value
Intercept	46.48	9.71	<b>4.79</b>
Grammaticality	49.39	20.47	<b>2.41</b>
Attractor Number	-8.49	19.25	-0.44
Interaction	-43.76	33.76	-1.30

Table 34: Results of the linear mixed effects model in the verb region for conditions in which the attractor was a plausible subject for the verb (replication), Experiment 9b (using log transformed RTs).

	Estimate	Std. Error	t-value
Intercept	58.28	14.17	<b>4.11</b>
Grammaticality	51.28	18.79	<b>-2.73</b>
Attractor Number	-19.49	17.32	-1.13
Interaction	-41.55	33.79	-1.23

Table 35: Results of the linear mixed effects model in the verb region for conditions in which the attractor was not a plausible subject for the verb (replication), Experiment 9b (using log transformed RTs).

#### 4.6.4 Discussion

The results of Experiment 9b do not replicate the findings of Experiment 9a. In both the original study and the replication, agreement violations led to a slowdown in the verb region. However, how this grammaticality effect interacted with the effects of attractor number and plausibility differed between the studies. In Experiment 9a, we found a clear agreement attraction effect: the slowdown in response to an ungrammatical verb was significantly reduced when the attractor was plural. We found the same pattern in the spillover region in Experiment 9b, but unlike in Experiment 9a, it was not significant in the verb region. However, the numerical difference between conditions was in the expected direction and the plots are clearly consistent with an agreement attraction effect on reading times in the verb region. In Experiment 9a the agreement attraction effect was modulated by plausibility, but we find no such three-way interaction in Experiment 9b. Looking at the average reading times in the different conditions, there is also no numerical trend in that direction. Whether the attractor was a plausible subject for the verb or not did not have an impact on whether it was misretrieved. These findings suggest that the verb's semantic information is not used to guide retrieval of the agreement controller in error-driven retrieval for agreement checking in comprehension.

One potential concern about this design is whether the number marking on the inflected auxiliary is processed earlier than the semantic information on the main verb. In this experiment, to make interaction between these factors possible, we presented the auxiliary and main verb simultaneously in the self-paced reading task. However, it is possible that participants still engaged in processing these items sequentially, and thus completed the error-driven agreement checking process before they had processed the semantic information of the main verb. Future work with other languages with richer verbal inflection systems could alleviate this concern to a greater extent by presenting the number marking and the semantic content of the verb on the same phonological word. This issues could also potentially be addressed in English with a relative clause construction and main verbs, like '*The rabbit/rabbits that the magician pulls/pull from the hat surprises the audience*', as is currently underway (Lago, unpublished data).

#### 4.7 General Discussion

The experiments reported in this chapter were designed to investigate whether the error-driven retrieval process triggered by an agreement violation uses information as retrieval cues that does not define agreement in the grammar. Experiment 8 found that syntactically singular collective nouns with a notionally plural distributive construal do not reduce the reading time slowdown associated with agreement violations compared to a collective attractor with a notionally singular construal. The lack of attraction effects in this experiment indicate that the search for the agreement controller is guided only by syntactic but not notional number. There is a close (though imperfect) correlation between syntactic and notional number, so it is interesting that the parser does not seem

to use the notional number information that in most cases would be helpful for finding the agreement controller. The results of Experiment 9 suggest that the parser does not use semantic/plausibility information either to search for the agreement controller. We replicated Experiment 9 due to an odd pattern in the reading times the first time we conducted it and did not find any evidence that whether the attractor is a plausible subject for the verb has an impact on agreement attraction. Together these findings suggest that error-driven repair mechanism underlying the illusory licensing of agreement violations uses only grammatically relevant retrieval cues.

#### *4.7.1 Availability of number information vs semantic information*

The results of Experiment 8 suggest that the retrieval process triggered by an agreement violation does not use lexical semantic information to guide retrieval of the agreement controller from memory. However, we need to consider the design of Experiment 8 before drawing any strong conclusions. Following Thornton and MacDonald (2003), all experimental items were passive sentences. Consequently, the number marking appeared on the inflected form of ‘be’, which does not have any lexical semantic content. In this experiment we tried to solve this issue by presenting ‘be’ together with the main verb of the sentence in the self-paced reading task. However, having both words presented on the screen simultaneously may not necessarily mean that participants process them at the same time. One possible reason why there is no evidence for the use of semantic features in agreement processing is that by the time the parser detects the agreement violation on ‘be’, the main verb has not been processed yet. This would mean that the semantic information from the main verb is simply not available to

the parser when it engages in error-driven retrieval for agreement checking. This potential timing problem could be solved by using a language with a much richer agreement system than English or using agreement attraction in English relative clauses with main verbs (Lago, unpublished data). I assume that if the number marking and the semantic content are carried by the same lexical item, the parser processes both types of information simultaneously.

Alternatively, if the number marking on the auxiliary and the lexical semantic content of the main verb are processed simultaneously, the absence of a plausibility effect could be because passive constructions are difficult to process and computing the argument roles takes extra time. In that case, we might still see a plausibility effect in active constructions. However, it is possible that computing argument roles is computationally intensive for any structure, even actives. If computing argument roles is not something that can be done rapidly upon encountering the verb, there would be no effect of plausibility in active constructions either. In either of these cases, plausibility might not have an impact on error-driven retrieval, but manipulating simple semantic association between the verb and the attractor rather than their thematic fit might still have an effect.

#### *4.7.2 Error-driven retrieval relies on grammatically relevant features*

The experiments in this chapter suggest that the parser does not use grammatically irrelevant information in the processing of subject-verb agreement in comprehension. According to the model of agreement attraction I assume in this dissertation, illusory licensing of an agreement violation consists of two steps: The first step is detecting the

mismatch between the predicted form and the bottom-up input. This requires predictive top-down processing to anticipate the verb's number marking based on the subject's number information. The second step involves error-driven retrieval of the agreement controller. In this model, retrieval of the agreement controller (subject) only occurs when the parser detects an agreement violation. We could imagine that the parser is completely faithful to the grammar when it is building new structure but that its faithfulness does not hold for error-driven processes. In that case, in top-down processing the prediction of the verb's number marking would be generated solely on the basis of the subject's number feature. However, detecting a mismatch between the verb's number marking in the bottom-up input and the predicted number is a clear signal for the parser that something has gone wrong. Information that does not bear on subject-verb agreement in the grammar, such as notional number and plausibility, might be useful in retrieving the agreement controller from memory. However, the parser does not seem to use this type of information to guide retrieval of the agreement controller. Together with the findings in the previous chapter, this result indicates that even in error-driven retrieval the parser uses only features based on which the agreement dependency is stated in the grammar. The question arises whether all features defining the dependency in the grammar necessarily have to be implemented as retrieval cues. I explore this question in the next chapter by investigating agreement processing by advanced second language learners.

## 4.8 Conclusion

In this chapter, I have presented data from two self-paced reading studies that indicate that agreement attraction in comprehension is not affected by the notional

number of the attractor or by whether the attractor is a plausible subject for the verb. This suggests that the error-driven retrieval process underlying the phenomenon of agreement attraction does not use information as retrieval cues that does not define the agreement dependency in the grammar. This is consistent with the idea that it is a low-level rechecking process, as proposed in Chapter 2.

It might seem surprising that we observed no effect of notional number on agreement attraction in our experiment, given that it is often claimed that there is overwhelming evidence for this in production. However, careful examination of the data cited in support of this reveals that it is actually still unclear whether the notional number of the attractor causes agreement attraction in production.

The failure to find an effect of the attractor's notional number on agreement attraction in Experiment 8 could potentially help us test whether agreement attraction in production should also be explained by a cue-based retrieval model, or whether a representational account is necessary. The studies examining the impact of notional number on agreement production suggest that a representational account like Marking and Morphing would predict that a syntactically singular but notionally plural attractor would interfere with the subject's number representation and therefore lead to agreement errors. However, a cue-based retrieval account for production seems to make a different prediction: if notional number is not used to guide retrieval of the agreement controller in cue-based retrieval, preambles with a syntactically singular but notionally plural attractor (like the complex subjects in Experiment 8) should not elicit a higher rate of agreement errors than when the syntactically singular attractor is notionally singular. Of course, this relies on the assumption that cue-based retrieval of the agreement controller relies on the

same information regardless of modality. It could be argued that we cannot make this assumption without further evidence, since the opposite directions of encoding in production and comprehension mean that different types of information are more salient in one than the other, which could affect what type of information is used in retrieval.

# Chapter 5

## 5.1 Outline

This chapter asks whether grammatical knowledge is necessarily implemented as retrieval cues in error-driven retrieval. I address this question by investigating the processing of subject-verb agreement in advanced Chinese learners of English. Chinese does not have subject-verb number agreement, so this constraint has to be specifically acquired as part of the second language. The experiments in this chapter ask whether Chinese learners of English not only acquire the grammatical knowledge associated with subject-verb agreement in English, but also implement it in a native-like way in online processing. In a speeded acceptability judgment task and a self-paced reading experiment, I show that advanced Chinese learners of English are automatically sensitive to agreement violations in online processing, even if they are not asked to make an acceptability judgment. However, online measures show no evidence that they experience attraction effects from a number-matching attractor; in a sense, the L2 learners actually process the input more accurately than the native speakers. This is consistent with a view in which the first ('prediction') stage of agreement processing is the same for native and non-native speakers who have acquired the L2 grammar. The difference would instead arise at the second stage of the process: When the L2 learners receive an error-signal from an agreement violation, they do not use the verb's number information to guide retrieval of the agreement controller. Instead, they seem to rely exclusively on the verb's structural cue, which also exists in their native language. These results show that even advanced second language learners might retain non-native like processing routines for dependencies they do have grammatical knowledge of. These results also suggest that not

all grammatical knowledge has to be used as retrieval cues in error-driven processing, at least by second language learners.

## 5.2 L2 Grammatical knowledge and online processing

When learning a second language (L2), learners not only have to acquire the grammar of their L2, but also have to be able to employ appropriate processing routines to understand input in real time. In this chapter, we ask two questions about the use of features and cues by second language learners. The first question is related to the use of retrieval cues in online processing. We explore whether second language learners can acquire native-like processing routines to implement knowledge of a grammatical constraint specific to their L2. The second question is representational: When the L2 grammar requires the acquisition of a feature like grammatical number that is absent in the L1, is the mental representation that L2 learners construct the same as or fundamentally different from that of native speakers? In this chapter I investigate both of these questions by looking at the processing of subject-verb number agreement in English by native speakers of Chinese.

Chinese and English are morphologically incongruent in regard to grammatical number: Unlike English, Chinese does not have subject-verb number agreement or use plural morphology on nouns (Lardiere, 2009; Jiang, 2011).<sup>8</sup> This means that this

---

<sup>8</sup> Unlike English, which has a count/mass distinction for nouns, Chinese has a classifier system. According to the Nominal Mapping Parameter (Chierchia, 1998), this means that all nouns in Chinese are plural mass nouns. Consequently, classifiers are needed for counting and there is no plural morphology. Although Mandarin Chinese has a suffix that appears to be some kind of plural marker ('-men'), its distribution is extremely restricted and it is optional even in those contexts where it can occur, so it is generally analyzed as a collective marker rather than a plural marker (Iljic, 1994; Cheng & Sybesma, 1999; but cf. Li, 1999). However, even if we take the view that '-men' is a true plural marker, Chinese does not have subject-verb number agreement.

grammatical knowledge is specific to their L2 and we can ask how Chinese learners of English mentally represent grammatical number and how they use number retrieval cues on the verb in online processing.

Native speakers of languages without grammatical number, such as Chinese, are well known to struggle with subject-verb agreement in English. This is especially apparent in production where they frequently omit the third person singular marker ‘-s’. Inflectional morphology in general is notoriously difficult for second language learners and their ultimate attainment often does not mirror that of native speakers (Lardiere, 1998). There are two types of accounts for this discrepancy: representational accounts and processing accounts. According to representational accounts, L2 learners show non-native like behavior because they have not successfully acquired the grammatical knowledge of their L2 and their mental representations are qualitatively different from those of native speakers. Under this view, L2 learners’ failure to consistently produce third person singular ‘-s’ reflects their lack of grammatical knowledge about subject-verb agreement in English (Ellis, 1988). These accounts do not claim that L2 learners have not acquired explicit knowledge about grammatical number and subject-verb agreement in English, but that they are unable to integrate this explicit knowledge into their implicit linguistic knowledge. This view was particularly emphasized in early research on second language acquisition that focused mainly on the order in which inflectional morphemes were acquired in the L2 (see for example Perkins & Larsen-Freeman, 1975; Johnson & Newport, 1989). In comparison, processing accounts argue that L2 learners’ non-native like performance can be attributed to processing difficulties rather than to a lack of grammatical knowledge (Sharwood Smith, 1986; Sorace, 2004, 2011; Sorace &

Serratrice, 2009; Hopp, 2010; Cummings, 2017). These accounts assume that the mental representations of L2 learners are qualitatively similar to those of native speakers and that the difference is quantitative in nature. According to the Missing Surface Inflection Hypothesis (Haznedar & Schwartz, 1997; Prevost & White, 2000), the lack of consistent production of subject-verb agreement by L2 learners does not indicate a lack of grammatical knowledge about subject-verb agreement. Instead, it is the result of difficulties implementing this knowledge in online processing.

### *5.2.1 Subject-verb agreement processing in advanced Chinese learners of English*

There is some evidence that native speakers of Chinese are not only able to acquire the grammatical knowledge associated with subject-verb agreement in English, but that they are also sensitive to it in online processing. In an ERP study by Chen et al. (2007), a group of Chinese learners of English and a native control group read sentences with complex subjects consisting of a singular head noun and a prepositional modifier which varied not only in grammaticality but also in the number of the structurally inaccessible noun (*'The key to the cabinet(s) was/were rusty...'* ). Each sentence was followed by a 500ms break after which participants had to make an acceptability judgment within 2000ms. The behavioral data showed that the L2 participants were generally very accurate in judging grammatical sentences acceptable and ungrammatical ones unacceptable. This was confirmed by the ERP data: ungrammatical sentences elicited an increased late frontal negativity in the L2 learners. It should be noted that this pattern is different from that in the native control group, who showed the pattern that has frequently been observed for agreement violations in native speakers of English: an early

frontal negativity (possibly a LAN) followed by a late posterior positivity (P600) in response to ungrammatical compared to grammatical sentences. Chen et al. argue that the distinct ERP responses indicate that the L2 learners are using qualitatively different neural resources for agreement processing because this is a feature specific to their L2. Nevertheless, both the behavioral and ERP results clearly show that the Chinese learners of English in this study were sensitive to subject-verb agreement violations in online processing.

Further ERP evidence for native Chinese speakers' sensitivity to subject-verb agreement violations comes from an ERP study by Armstrong, Bulkes & Tanner (2016). They investigated agreement processing in sentences with quantified subjects ('*Many/The cats meow/meows ...*'). The L2 learners showed sensitivity to agreement violations not only in the behavioral data from acceptability judgments performed after each sentence, but also in the ERP data. Like native speakers of English, the native Chinese speakers showed large and significant P600 responses to subject-verb agreement violations in both quantified and non-quantified subject conditions. In fact, in this study agreement violations elicited a P600 in the L2 learners, which is the pattern typical for native speakers, rather than the late frontal negativity observed by Chen et al. (2007). A possible explanation for this difference is that the participants in Armstrong et al.'s study were living in the US and had been immersed in an English speaking environment while the L2 group in Chen et al.'s study was not living in an immersion context and most likely did not have nearly as much experience in processing English sentences.

However, it has also been argued that even advanced Chinese learners of English remain "morphologically insensitive" to number morphology in online processing (Jiang,

2004; 2007; 2011).<sup>9</sup> In a series of self-paced reading experiments, Jiang (2004) tested how advanced Chinese learners of English process subject-verb agreement. An untimed written test confirmed that the L2 participants did have explicit knowledge about plural morphology and subject-verb agreement. For the native speakers, subject-verb agreement violations led to a significant slowdown in reading times, regardless of whether the mismatch was between a singular subject and a plural verb or a plural subject and a singular verb. No significant effect of grammaticality was found for the L2 group, leading Jiang to conclude that they are indeed not sensitive to subject-verb number agreement in online processing. However, the L2 group did show a trend in the same direction as the L1 group. Numerically they exhibited a slowdown in the ungrammatical conditions in both experiments, but this did not reach statistical significance. It is difficult to draw strong conclusions from this result, especially since the interaction between the two language groups was not tested and we do not know if there was actually a significant difference between them.

In summary, there is convincing evidence that advanced Chinese learners of English are sensitive to subject-verb agreement violations in both electrophysiological and behavioral measures. However, the studies included explicit acceptability judgment tasks, which draw attention to agreement violations in a way that is not the case for comprehension in the real world. Based on this data, it remains unclear to what extent Chinese learners of English are automatically sensitive to subject-verb agreement violations. Even for native speakers of English, it is somewhat surprising that they are

---

<sup>9</sup> Jiang actually argues that it is knowledge of the plural morpheme on nouns that is not integrated rather than agreement knowledge since the L2 group showed reliable effects of subject-verb agreement violations person mismatches between pronouns and inflected forms of 'be'. However, it is unclear how the L2 learners could have acquired grammatical knowledge about subject-verb number agreement without the acquisition of grammatical number.

sensitive to agreement violations in tasks that involve only comprehension given that they do not bear on the interpretation at all (Wagers et al., 2009). This sensitivity shows that in native English speakers agreement is processed automatically in comprehension.

Findings from other studies on agreement processing in L2 learners who had to acquire subject-verb agreement as part of their L2 suggest that they can indeed become automatically sensitive to agreement-violations in comprehension (see for example Lim & Christianson (2015) for eye-tracking with Korean learners of English).

### *5.2.2 Agreement attraction in L2 processing*

If Chinese learners of English are sensitive to subject-verb number agreement in online processing, the question arises whether they implement their grammatical knowledge in the same way as native speakers. Native speakers of English use their grammatical knowledge of number agreement to predict the verb's number marking. When they encounter input that mismatches their prediction they engage in cue-based retrieval of the agreement controller (subject) based on both the structural (subject) and morphosyntactic (number) cues of the verb. Although the structural cue would always lead to the retrieval of the correct agreement controller, we know that native speakers of English experience facilitative similarity-based interference from structurally irrelevant number-matching nouns, indicating that they also use the verb's number information to guide retrieval (e.g. Wagers et al, 2009; Tanner et al., 2014). Since Chinese does not have subject-verb agreement, Chinese learners of English have to acquire this grammatical knowledge specifically for their L2 and may or may not use it as a cue in the error-driven retrieval operation triggered by an agreement violation.

Agreement attraction offers an opportunity to examine whether processing routines and the implementation of grammatical constraints as retrieval cues are native-like in L2. Successful detection of agreement violations indicates that L2 learners have acquired the relevant grammatical knowledge. However, agreement checking in comprehension also allows us to probe whether they use all of that knowledge to guide retrieval of the agreement controller when they encounter an agreement violation. To some extent, this lets us separate out grammatical knowledge from how it is used in processing. Agreement attraction effects in L2 learners would be evidence that they implement their grammatical knowledge as retrieval cues in an error-driven process in a native-like way.

Two of the studies on subject-verb agreement processing in advanced Chinese learners of English incorporated elements of agreement attraction designs. However, experimental design issues and analysis choices prevent us from drawing strong conclusions about attraction effects from their results. The previously mentioned series of self-paced reading experiments by Jiang (2004) used complex subjects with a prepositional modifier and varied whether the number of the local noun matched or mismatched the number of the verb ('*The key to the cabinet(s) was rusty ...*'). However, the first experiment did not contain any subject-verb agreement violations, and the second and third experiment manipulated grammaticality via the number of the head noun (Experiment 2: '*The key(s) to the cabinet were rusty ...*'; Experiment 3: '*The key(s) to the cabinet was rusty ...*'). There was no systematic crossing of attractor number and grammaticality in any of these experiments.

Chen et al. (2007) also used complex subjects with prepositional modifiers in their ERP study. However, the analysis compared the grammatical condition with a singular attractor to each of the other conditions separately and did not test the interaction between grammaticality and attractor number. This means that it is not possible to draw any conclusions about whether the presence of a plural attractor has an impact on how Chinese learners of English process subject-verb agreement violations.

Native-like agreement attraction effects have been observed in other late learners of English in electrophysiological measures. In native English speakers, agreement violations elicit a large P600 response but the magnitude of this effect is attenuated by the presence of a plural attractor. The same pattern was observed in Spanish learners of English: in response to agreement violations they show a native-like (if somewhat reduced) P600 response modulated by attractor number (Tanner, 2011; Tanner et al., 2012, but cf. Tanner, Inoue & Osterhout, 2014). This indicates that the prediction and memory retrieval mechanisms underlying subject-verb agreement checking in comprehension are qualitatively similar to those of native English speakers. However, Spanish also has subject-verb number agreement and native speakers of Spanish and English show comparable attraction effects in comprehension when processing their respective L1 (Lago et al., 2015). Spanish learners of English already have processing routines to implement knowledge about subject-verb agreement from their L1 which they can transfer to the processing of their L2. Transfer effects in L2 processing are well established, for example, speakers of a morphologically rich language pay more attention to inflectional morphology in their L2 than speakers of languages with impoverished morphology (Sagarra & Ellis, 2013).

Here, we want to explore whether second language learners implement their knowledge of a grammatical constraint specific to their L2 in a native-like way by looking at agreement attraction in Chinese learners of English. They may be able to detect agreement violations, but their use of the grammatical knowledge may still be different from native speakers or L2 learners whose L1 also has subject-verb number agreement and who therefore already have processing routines to implement this knowledge.

### 5.3 The plural markedness effect

Examining agreement attraction in advanced Chinese learners of English also gives us the opportunity to investigate whether their representation of the grammatical number feature itself is native-like. According to the plural markedness hypothesis, for native speakers of English grammatical number is a privative feature, meaning that singular and plural nouns differ in whether they possess a number feature (see for example Bock & Eberhard, 1993; Eberhard, 1997). Plural is the marked form, so plural nouns possess a number feature, whereas singular is the default and singular nouns lack a number feature.

An asymmetric mental representation of the number feature predicts agreement attraction with plural but not singular attractors ('*The keys to the cabinet are rusty*'). This is consistent with findings in the literature in both production and comprehension (Kaan, 2002; Eberhard, 1997). Agreement attraction in comprehension arises because the retrieval of the subject is guided by both structural and number cues on the verb. When there is a subject-verb agreement violation between a singular subject and a plural verb,

the plural retrieval cue on the verb is one of the cues used to look for the agreement controller. A structurally irrelevant plural noun is marked with a number feature that matches the verb's number retrieval cue. Consequently, the attractor receives some activation from this partial feature-overlap and is sometimes misretrieved instead of the actual agreement controller, which does not match the verb's number cue. Compare this to the situation in which the agreement violation is between a plural subject and a singular verb. Even when there is a structurally inaccessible singular noun that matches the number marking on the verb, it is not marked for number and can therefore not receive any activation from a number retrieval cue.

The question arises whether the mental representation of number by L2 English learners whose native language does not encode grammatical number is asymmetric in the same way as that of native speakers. McCarthy (2011) argues that during the process of acquiring a feature like number, the L2 learners' mental representation is not privative. Under this view, there is no default until the feature is fully acquired; both singular and plural possess a number feature, so neither of them is more marked than the other. This would predict that L2 learners are equally likely to make mistakes with plural and singular. However, a study of L1 English learners of Spanish found no proficiency effect on number marking on verbs (McCarthy, 2011). Errors occurred more frequently for plural than for singular marking regardless of proficiency: while the speakers were very accurate in their use of singular agreement, they sometimes substituted singular agreement when plural agreement was required. This is consistent with the claim that singular is the default form and plural is marked, but not with the idea that L2 learners represent number symmetrically. It should be noted that the L2 learners in this study were

native speakers of Spanish, which has grammatical number. Assuming a model in which learners' initial knowledge of their L2 is determined by transfer from their L1 (Vainikka & Young-Scholten, 1996), these learners would have started out with an asymmetric mental representation of number.

Agreement attraction might help answer the question whether number is represented asymmetrically by L2 learners whose native language lacks this feature. If the L2 learners have a native-like privative representation of grammatical number, they should show asymmetric attraction effects like native speakers of English. If they have a non-privative representation of grammatical number, they should show symmetric attraction effects. Specifically, this would mean attraction in ungrammatical sentences with a plural subject and a singular verb when there is a number-matching singular attractor ('*The keys to the cabinet is rusty...*'). However, a caveat is that the data from these experiments can only be informative about the representational question if the L2 learners show any agreement attraction effects in online processing.

#### 5.4 The present study

The experiments reported here use speeded acceptability judgments and self-paced reading to investigate the processing of subject-verb agreement in advanced Chinese learners of English. In particular, we use the phenomenon of agreement attraction to test whether they are not only automatically sensitive to agreement violations but have also learned to use the number cue on the verb for retrieval. Based on the prior literature, we expect that advanced Chinese learners should be sensitive to agreement violations in online processing. Here we ask whether this sensitivity to agreement

violations is implemented in the same way as in native English speakers. Since grammatical number is specific to their L2, it is possible that they acquire the grammatical knowledge but do not implement it as a retrieval cue. In that case, error-driven retrieval of the agreement controller would be guided exclusively by the structural cue on the verb also available in their L1. Since retrieval would not be guided by the verb's number cue, they would show no effect of attractor number.

If Chinese learners of English have learned to use the number information on the verb as a retrieval cue, we can also ask about their mental representation of the number feature on the noun. Native speakers of English only show attraction effects in ungrammatical sentences with a plural attractor, which is thought to be a result of the asymmetric mental representation of number: only plural nouns have a number feature and singular is simply the default in the absence of a number feature. Since Chinese learners of English have to acquire grammatical number as a feature specific to their L2, their mental representation may not be asymmetrical. In online processing, this would be reflected by agreement attraction effects in ungrammatical sentences not only with plural but also singular attractors.

## 5.5 Experiment 10: L2 speeded acceptability

The aim of Experiment 10 was to investigate whether advanced Chinese learners of English can acquire native-like processing routines related to a grammatical feature specific to their L2. We manipulated attractor number, grammaticality and the number of the subject's head noun in a speeded acceptability judgment task to determine whether these learners experience agreement attraction effects. In native speakers of English,

when the subject is singular the disruption in response to an agreement violation is mitigated by the presence of a structurally irrelevant plural noun matching the number marking on the verb, leading to higher acceptance rates. Based on previous findings, we expected the L1 control group to show a clear effect of grammaticality, modulated by the presence of a plural attractor in sentences with a singular subject head noun (agreement attraction). Chinese does not have subject-verb agreement, so native speakers of Chinese have to acquire not only the grammatical knowledge associated with this constraint but also have to learn to implement it in online processing. Based on previous research reviewed above, we expected the L2 learners to show an effect of grammaticality, indexing their sensitivity to subject-agreement in online processing. However, being able to detect an agreement violation does not necessarily mean that they use native-like processing routines. The verb's number cue guiding retrieval in native English speakers is not available in their L1, so it is possible that they rely exclusively on structural cues to retrieve the agreement controller. In that case, a number matching structurally irrelevant noun would not modulate the effect of grammaticality. However, if they have not only acquired the grammatical knowledge associated with subject-verb agreement in English but have also learned to use the number information on the verb as a retrieval cue, we expect them to show an agreement attraction effect in addition to the effect of grammaticality.

In native English speakers, agreement attraction has been observed with singular subjects and plural attractors but not vice versa. However, in this experiment we also varied the number of the head noun to test whether Chinese learners of English show attraction in this configuration. For the L1 group, we expected to find only an effect of

grammaticality when the head noun was plural, and no impact of attractor number. If the L2 learners are sensitive to agreement violations and have learned to use the verb's number cue for retrieval, their behavior with plural head nouns could give us an indication of how they mentally represent grammatical number. If their mental representation of number is asymmetric in the same way as in native speakers, we expect attraction only with singular head nouns. However, if their mental representation of grammatical number is non-native like and singular is not just the absence of a number feature, we should see attraction from a singular attractor in sentences with a plural subject head noun. Of course, this representational question can only be addressed if Chinese learners of English show agreement attraction at all.

### *5.5.1 Participants*

In this experiment, the L2 group consisted of 25 native speakers of Chinese who were enrolled at the University of Maryland at the time of testing. They had all fulfilled UMD's English language proficiency requirements and can thus be considered advanced learners of English.<sup>10</sup> The L1 control group consisted of 24 native speakers of American English who were recruited via Amazon Mechanical Turk and had all passed a native speaker proficiency test for American English.

### *5.5.2 Materials and Design*

The materials were 48 item sets in a 2x2x2 design, crossing head noun number (singular/plural), attractor number (singular/plural), and grammaticality

---

<sup>10</sup> UMD requires international students who do not hold a degree from an English-speaking country to have an IELTS score of at least 7 or an IBT score of at least 100.

(grammatical/ungrammatical), as illustrated in (1). The items were adapted from Tanner (2011) and consisted of a subject in which the head noun was modified by a prepositional phrase, followed by a form of ‘*have*’ or ‘*be*’. Note that since we manipulated the number of the head noun, in this experiment the singular form of the verb is grammatical in conditions a. and c. (singular head noun), but the plural verb is grammatical in conditions e. and g. (plural head noun). This differs from most experiments on agreement attraction in comprehension, which use only singular head nouns and in which the singular verb form is thus always grammatical.

(1)

- a. The owner of the expensive **car** *has* been drinking a lot.
- b. \*The owner of the expensive **car** *have* been drinking a lot.
- c. The owner of the expensive **cars** *has* been drinking a lot.
- d. \*The owner of the expensive **cars** *have* been drinking a lot.
  
- e. The owners of the expensive **car** *have* been drinking a lot.
- f. \*The owners of the expensive **car** *has* been drinking a lot.
- g. The owners of the expensive **cars** *have* been drinking a lot.
- h. \*The owners of the expensive **cars** *has* been drinking a lot.

The 48 items were distributed across eight lists in a Latin Square design. Each participant saw one condition per item and six items per condition. Each list also

contained 48 filler items, half of which were ungrammatical, plus 8 control items which explicitly instructed participants to answer either ‘yes’ or ‘no’ on that particular trial.

### *5.5.3 Procedure*

Before the start of the experiment, both groups gave informed consent and the L2 group additionally completed a short language background questionnaire. The stimuli were displayed word by word in the center of the screen using IBEX software (Drummond, 2016). Each word was presented for 400ms, followed by a blank screen for 200ms. A response screen followed the last word of each sentence and participants had to judge whether the sentence they had just read was acceptable or not by pressing the ‘f’-key for ‘yes’ or the ‘j’-key for ‘no’. A response had to be made within 2000ms or a message would appear informing them that they had been too slow. Before the experiment, participants were familiarized with the method by completing seven practice items. The first three of those were part of a guided practice in which the participant’s response was followed by an explanation of whether they were supposed to judge the sentence as acceptable or not. This was to ensure that they understood that an acceptable sentence was one that a native speaker of English might say, regardless of plausibility.

### *5.5.4 Analysis*

Trials on which no response was made within 2000ms were excluded, resulting in the loss of 1.5% of experimental trials for the L1 group and 2.5% for the L2 group. Acceptance rates were analyzed using mixed logit models (Jaeger, 2008) with the *lme4* package for linear mixed effects models (Bates, Maechler, Bolker & Walker, 2015) in the

R computing environment (R Development Core Team, 2016). Data from trials with singular and plural head nouns were analyzed separately. For both types of head nouns, models comparing results across both groups were built with grammaticality, attractor number and language as fixed effects. Data from the L1 and the L2 group were also analyzed independently, with grammaticality and attractor number as fixed effects. For all models, the random effects structure was initially maximally specified (Barr, Levy, Scheepers & Tily, 2013) and then progressively simplified until the model converged. Unless otherwise noted, the final models all included by-subject and by-item random intercepts and all factors were deviation coded (-0.5, 0.5).

### *5.5.5 Results*

#### *Singular head nouns*

Acceptance rates across conditions for the L1 group and the L2 group are plotted in Figure 19. The results of the mixed logit models are provided in Table 36 for both groups combined, in Table 37 for the L1 group only, and in Table 38 for the L2 group only.

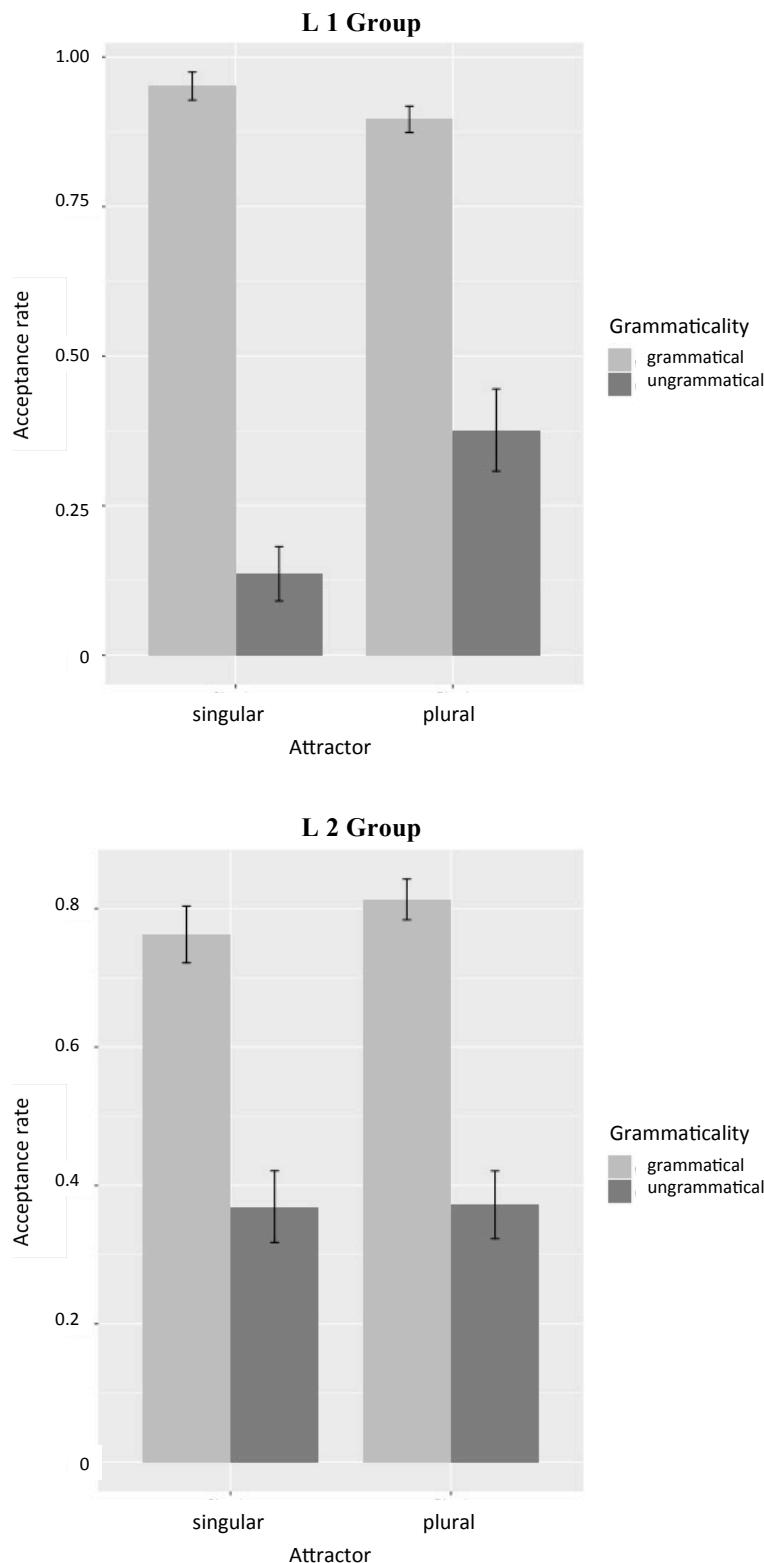


Figure 19: Acceptance rates across conditions with singular head nouns for L1 group (upper panel) and L2 group (lower panel) for Experiment 10 (note different scales on y-axis for L1 and L2 group). Error bars indicate standard error of the mean.

The model comparing the L1 and the L2 group for singular head nouns shows several statistically significant effects. Unsurprisingly, there was a significant effect of grammaticality ( $p < 0.001$ ), with ungrammatical sentences being judged acceptable less frequently than grammatical sentences. Neither the effect of attractor number nor language was significant (attractor number:  $p = 0.18$ ; language:  $p = 0.21$ ). There was a significant interaction between grammaticality and language ( $p < 0.001$ ). The acceptance rates across conditions for the two language groups plotted in Figure 19 show that compared to the L1 group the L2 group more frequently rejected grammatical sentences and accepted ungrammatical ones. There was also a significant interaction between grammaticality and attractor number ( $p < 0.003$ ), with ungrammatical sentences more likely to be accepted when the attractor was plural. Importantly, there was a significant three-way interaction between grammaticality, attractor number and language ( $p < 0.001$ ). We further investigate this interaction by looking at the data from the two language groups separately.

	Estimate	Std. Error	z-value	p-value
Intercept	0.57	0.12	4.76	< 0.001
Grammaticality	2.96	0.18	16.10	< 0.001
Attractor	-0.23	0.17	-1.34	0.18
Language	0.30	0.24	1.27	0.21
Gram x Attractor	1.02	0.35	2.94	0.003
Gram x Lang	2.02	0.36	5.67	< 0.001
Attractor x Lang	-0.13	0.35	-0.39	0.70
Gram x Attractor x Lang	2.60	0.69	3.75	< 0.001

Table 36: Results of mixed logit model on acceptance rates of sentences with singular head nouns in Experiment 10, including both language group (Note: Due to convergence issues model had only by-subject random intercepts.)

### *L1 Group*

The L1 group showed the expected main effect of grammaticality ( $p < 0.001$ ) with grammatical sentences accepted more frequently than ungrammatical sentences. They also showed the expected agreement attraction effect: There was a significant interaction between attractor number and grammaticality ( $p < 0.001$ ). Ungrammatical sentences were more likely to be accepted when the attractor was plural, indicating facilitative similarity-based interference from a structurally irrelevant noun.

	Estimate	Std. Error	z-value	p-value
Intercept	0.75	0.23	3.23	< 0.001
Grammaticality	4.17	0.33	12.67	< 0.001
Attractor	-0.34	0.29	-1.15	0.25
Gram x Attractor	2.42	0.59	4.13	< 0.001

Table 37: Results of mixed logit model on acceptance rates of sentences with singular head nouns in Experiment 10 for the L1 group.

### *L2 Group*

The L2 group also showed a significant main effect of grammaticality ( $p < 0.001$ ), with grammatical sentences accepted more frequently than ungrammatical ones. However, the interaction between grammaticality and attractor number was not significant ( $p = 0.46$ ). This pattern can be clearly seen in Figure 1: While the L2 participants had a higher acceptance rate of ungrammatical sentences than the L1 group, there was no difference in acceptability based on whether the attractor in an ungrammatical sentence was singular or plural. Although numerically the L2 participants accepted grammatical sentences with a plural attractor more frequently than with a singular attractor this difference is not statistically significant ( $p = 0.4$ ).

	Estimate	Std. Error	z-value	p-value
Intercept	0.41	0.12	3.52	0.0014
Grammaticality	1.90	0.19	9.81	< 0.001
Attractor	-0.16	0.19	-0.85	0.40

Gram x Attractor	-0.27	0.38	-0.72	0.46
------------------	-------	------	-------	------

Table 38: Results of mixed logit model on acceptance rates of sentences with singular head nouns in Experiment 10 for the L2 group.

### *Plural head nouns*

The model for plural head nouns with language, attractor and grammaticality as fixed effects is presented in Table 4. As expected, there was a significant effect of grammaticality ( $p < 0.001$ ): grammatical sentences were more likely to be accepted than ungrammatical ones. There was also a significant interaction between grammaticality and language ( $p < 0.001$ ), with the L2 group showing a smaller impact of grammaticality than the L1 group. None of the other effects were statistically significant.

	Estimate	Std. Error	z-value	p-value
Intercept	0.57	0.12	4.61	< 0.001
Grammaticality	-2.77	0.18	-15.57	< 0.001
Attractor	-0.18	0.16	-1.07	0.29
Language	-0.34	0.25	-1.39	0.16
Gram x Attractor	0.03	0.33	0.09	0.93
Gram x Lang	2.28	0.34	6.63	< 0.001
Attractor x Lang	-0.12	0.33	-0.35	0.73
Gram x Attractor x Lang	-1.12	0.66	-1.70	0.09

Table 39: Results of mixed logit model on acceptance rates for sentences with plural head nouns in Experiment 10, including both language groups (Note: Due to convergence issues model had only by-subject random intercepts.)

### *L1 group*

The model for the L1 participants for sentences with plural head nouns shows a significant effect of grammaticality ( $p < 0.001$ ), with grammatical sentences more likely to be accepted than ungrammatical ones. Neither the effect of attractor number nor the interaction between grammaticality and attractor number were significant.

	Estimate	Std. Error	z-value	p-value
Intercept	0.78	0.25	3.15	0.002
Grammaticality	-4.25	0.36	-11.95	< 0.001

Attractor	-0.12	0.28	-0.42	0.68
Gram x Attractor	0.62	0.56	1.12	0.26

Table 40: Results of mixed logit model on acceptance rates for sentences with plural head nouns in Experiment 10 for the L1 group.

### *L2 group*

For the L2 group, there was a significant effect of grammaticality ( $p < 0.001$ ), with grammatical sentences more likely to be accepted than ungrammatical sentences. The effect of attractor number and the interaction between grammaticality and attractor number were not significant.

	Estimate	Std. Error	z-value	p-value
Intercept	0.41	0.15	2.74	<b>0.006</b>
Grammaticality	-1.67	0.20	-8.37	< <b>0.001</b>
Attractor	-0.24	0.19	-1.27	0.20
Gram x Attractor	-0.55	0.38	-1.45	0.15

Table 41: Results of mixed logit model on acceptance rates for sentences with plural head nouns in Experiment 10 for the L2 group.

Table 41: Results of mixed logit model on acceptance rates for sentences with plural head nouns in Experiment 10 for the L2 group.

### *5.5.6 Discussion*

#### *Singular head nouns*

In Experiment 10, both language groups were more likely to judge grammatical sentences acceptable than ungrammatical ones when the subject had a singular head noun. This shows that both L1 and L2 participants were sensitive to subject-verb agreement violations. There was, however, an interaction between language and grammaticality, indicating that there was a difference between how grammaticality affected L1 and L2 speakers. Looking at the acceptance rates, the L2 participants were not as accurate as the native speakers in accepting grammatical sentences and rejecting

ungrammatical ones, which is hardly surprising. As expected, we find that the L2 group shows an effect of grammaticality in the same direction as the L1 group. This contradicts the claim that Chinese learners of English are insensitive to number morphology in online processing and therefore not sensitive to subject-verb agreement violations (Jiang, 2004; Jiang, 2007). However, it is consistent with findings from a recent EEG-study in which Chinese learners of English showed a smaller P600 effect than native speakers in response to agreement violations (Armstrong et al., 2016).

Although both groups clearly detected the subject-verb agreement violations, only the L1 group showed the classic attraction effect with singular head nouns. There was a three-way interaction between grammaticality, attractor number and language. While the native speakers were more likely to accept ungrammatical sentences when they had a plural attractor, no such effect was found for the L2 group. This means that while the L2 learners were sensitive to agreement violations, the number of the attractor did not have an impact on how likely they were to notice these violations. This suggests that the L2 learners do not experience facilitative similarity-based interference from a structurally irrelevant noun matching the number marking of the verb. In order for comprehenders to experience facilitative similarity-based interference from a plural attractor, they have to use not only the structural but also the number retrieval cues of the verb. The L2 learners' results indicate that, unlike native speakers, they do not use the verb's number cue to guide retrieval of the agreement controller from memory when they encounter an agreement violation. Although they have clearly acquired the grammatical constraint and have access to this knowledge in online processing, as demonstrated by their sensitivity to agreement violations, their processing routine is not entirely native-like.

### *Plural head nouns*

For plural head nouns, both the L1 and L2 group showed a significant effect of grammaticality, indicating that they were both also able to detect agreement violations when the subject was plural. As indicated by the significant interaction between grammaticality and language, there was again a difference between how grammaticality affected L1 and L2 speaker with L2 speakers showing a reduced impact of grammaticality compared to native speakers. Unlike for the singular head nouns, there was no three-way interaction between attractor number, grammaticality and language. Separate analyses of the L1 and L2 data confirmed that there were no significant interactions between grammaticality and attractor number for either group. This indicates that neither of the groups experienced agreement attraction with plural head nouns.

Unfortunately, we cannot draw any conclusion about the L2 learners' mental representation of grammatical number based on these data. If the L2 participants had experienced agreement attraction with singular but not plural head nouns, this would have indicated a native-like mental representation of grammatical number with singular as the default and plural as the marked value. However, they also failed to show attraction with singular head nouns, meaning the data are not informative about a potential representational asymmetry.

Post-hoc analyses of the L1 and L2 data with head noun number as a fixed effect in addition to attractor number and grammaticality revealed that the L1 group showed a significant three-way interaction between grammaticality, attractor number and head number ( $p < 0.5$ ). This confirms that for native speakers facilitative similarity-based

interference only arises from plural attractors, which is consistent with previous studies that have failed to find agreement attraction with plural subjects and singular attractors in native speakers (Kaan, 2002). In contrast, for the L2 group the only significant effect was grammaticality ( $p < 0.001$ ). There was no interaction with head noun number, suggesting that there is no difference in how the L2 learners process sentences with singular and plural subjects.

The results of Experiment 10 demonstrate that advanced Chinese learners of English are sensitive to subject-verb agreement violations in online processing. However, the task required an acceptability judgment which might have drawn attention to subject-verb agreement violations not otherwise typical in online processing. The data also do not tell us anything about the timecourse of subject-verb agreement processing in the L2 learners. Previous research has shown a very tight correlation between agreement attraction effects in speeded acceptability judgments and a decrease in processing difficulty in self-paced reading (Wagers et al., 2009). The aim of Experiment 11 was to explore the timecourse of subject-verb agreement processing in advanced Chinese learners of English in a self-paced reading paradigm.

## 5.6 Experiment 11: L2 SPR

Experiment 10 shows that advanced Chinese learners of English are sensitive to subject-verb agreement but do not experience agreement attraction, unlike native speakers of English. Although end-of-sentence speeded acceptability judgments are considered an online measure, they do not give us any information about the timecourse of an effect. To investigate the timing of the L2 learners' sensitivity to subject-verb

agreement in a task in which they were not asked to make any acceptability judgments, we conducted a self-paced reading experiment with the same experimental items as in Experiment 10. We predicted that both the L1 and the L2 group would show slower reading times at the verb or its spillover region in response to agreement violations. For singular head nouns, we predicted that this slowdown would be reduced with plural attractors matching the number marking of the ungrammatical verb, but only for the L1 group. For the plural head nouns, we did not expect the attractor's number to have a mitigating effect on agreement violations for either the L1 or L2 group.

### *5.6.1 Participants*

The participants in this experiment were 32 Chinese-speaking learners of English and a control group of 34 native English speakers. One additional L2 participant was excluded because they had lived in the US for one year before age 6. Data from participants with an accuracy rate below 75% on the comprehension questions for the experimental items was discarded, which led to the exclusion of data from one additional native English speaker. All participants were enrolled as students at the University of Maryland at the time of the experiments. The L2 learners had all fulfilled UMD's English language proficiency requirements and can thus be considered advanced learners of English.<sup>11</sup> None of the participants in Experiment 11 had participated in Experiment 10. Participants gave informed consent and received either course credit or monetary compensation for their participation.

---

<sup>11</sup> UMD requires international students who do not hold a degree from an English-speaking country to have an IELTS score of at least 7 or an IBT score of at least 100.

### *5.6.2 Materials and Design*

The experimental items were identical to the 48 items used in Experiment 10 and were distributed across eight lists in a Latin Square design. Each participant saw one condition per item and six items per condition. Each list also contained 100 grammatical filler items, which belonged to different manipulations not related to agreement processing and which are not reported here.

### *5.6.3 Procedure*

The items were presented word-by-word in a self-paced moving window paradigm (Just, Carpenter & Woolley, 1982) using Linger software (Doug Rhode, MIT). Each trial began with the appearance of a row of dashes, which masked the words in the sentence. Participants revealed the first word and each subsequent word by pressing the space bar. When a new word was revealed, the previous word would be re-masked, so that there was only ever one word visible at a time. Participants were instructed to read as naturally as possible, at their normal reading speed and to make sure that they understood the sentences they were reading. Half of the experimental items and 40 of the 100 fillers were followed by a comprehension question. After the end of a sentence, the entire question was displayed on the screen and participants had to respond by pressing ‘f’ for ‘yes’ or ‘j’ for ‘no’. Onscreen feedback was provided only when the answer was incorrect. Before the start of the experiment, participants completed five practice items to familiarize themselves with the procedure, followed by three experimental blocks.

After they had finished the self-paced reading task, participants completed a brief language background questionnaire and a short untimed acceptability judgment task. The materials consisted of 24 items with the same sentence structure as in the self-paced reading experiment and 24 fillers, half of which were grammatical. The experimental items in the acceptability judgment task manipulated attractor number and grammaticality, but the head noun was always singular. Each item was presented on the screen as a whole sentence and participants had as much time as they wanted to judge whether it was an acceptable sentence of English. This task was designed to test the L2 learners' processing of subject-verb agreement in English when there was no time pressure.

#### 5.6.4 Analysis

Accuracy rates on the comprehension questions were generally high, with rates ranging from 90.1% to 99% across conditions for the L1 group and from 85.1% to 94.9% for the L2 group. Data from trials with an incorrectly answered comprehension question were not excluded, since only a subset of the experimental items were followed by a comprehension question. Reading times exceeding a threshold of 3000ms were excluded as outliers. This led to the exclusion of less than 0.2% of data in any of the critical regions in both the L1 and L2 group. Analyses were carried out on log transformed RTs. The regions of analysis in this experiment were the verb region and the word following the verb (spillover region). RTs were analyzed using linear mixed effect models in the R computing environment (R Development Core Team, 2016) using the *lme4* package (Bates, Maechler, Bolker & Walker, 2015). Since the model output does not include p-

values, a t-value of magnitude 2 or above indicates significance at the 0.05 level (Gelman & Hill, 2006; Kush, Lidz & Phillips, 2015). Data for each region were analyzed with attractor number, grammaticality and language as fixed effects. Data from the L1 and the L2 groups were also analyzed separately with attractor number and grammaticality as fixed effects. The random effects structure was initially maximally specified (Barr et al., 2013) and then progressively simplified until the model converged for all regions of interest for all groups. The final model had random by-subject and by-item intercepts. All contrasts (attractor number, grammaticality, language) were sum coded.

### *5.6.5 Results*

#### *Singular head nouns*

In the model including data from both groups, there was a significant effect of grammaticality in the verb region ( $t = 2.33$ ) with agreement violations lead to increased reading times. This effect remained significant in the region immediately following the verb ( $t = 2.49$ ). The only other significant effect was language: L2 participants had slower reading times than L1 participants in both the verb region ( $t = 5.18$ ) and the verb's spillover region ( $t = 4.18$ ).

Based on the results of Experiment 10, we would have expected a significant three-way interaction between grammaticality, attractor number and language. However, unlike in Experiment 10 the interaction between grammaticality and attractor number was only marginal for the L1 group.

	Estimate	Std. Error	t-value
Intercept	5.893	0.022	<b>265.83</b>
Grammaticality	0.019	0.008	<b>2.33</b>

Attractor	-0.002	0.008	-0.22
Language	0.111	0.022	<b>5.18</b>
Gram x Attractor	0.010	0.008	1.22
Gram x Lang	0.008	0.008	0.97
Attractor x Lang	0.006	0.008	0.74
Gram x Attractor x Lang	-0.009	0.008	-1.14

Table 42: Results of linear mixed effects model in the verb region of sentences with singular head nouns for both language groups, Experiment 11 (using log transformed RTs).

	Estimate	Std. Error	t-value
Intercept	5.883	0.0273	<b>215.39</b>
Grammaticality	0.020	0.008	<b>2.49</b>
Attractor	0.003	0.008	0.34
Language	0.108	0.026	<b>4.18</b>
Gram x Attractor	-0.002	0.008	-0.30
Gram x Lang	-0.011	0.008	-1.34
Attractor x Lang	-0.007	0.008	-0.85
Gram x Attractor x Lang	-0.001	0.008	0.18

Table 43: Results of linear mixed effects model in the spillover region of sentences with singular head nouns for both language groups, Experiment 11 (using log transformed RTs).

### *L1 group*

The L1 group's region-by-region average reading times for sentences with singular head nouns are plotted in Figure 20. The results of the model for the L1 group in the regions of analysis are presented in Table 44 and 45. In the verb region, there was no significant effect of grammaticality or attractor number. The main effect of grammaticality became significant in the verb's spillover region ( $t = 3.25$ ); ungrammatical sentences led to slower reading times. There was also a marginally significant interaction between grammaticality and attractor number in the verb region ( $t = 1.84$ ), but this patterned in the opposite direction from an agreement attraction effect: ungrammatical sentences elicited longer reading times when there was a number-matching plural attractor. The numerical pattern typical of agreement attraction was

observed in the region after the spillover region (two words after the verb), but the interaction between grammaticality and attractor number did not reach significance.

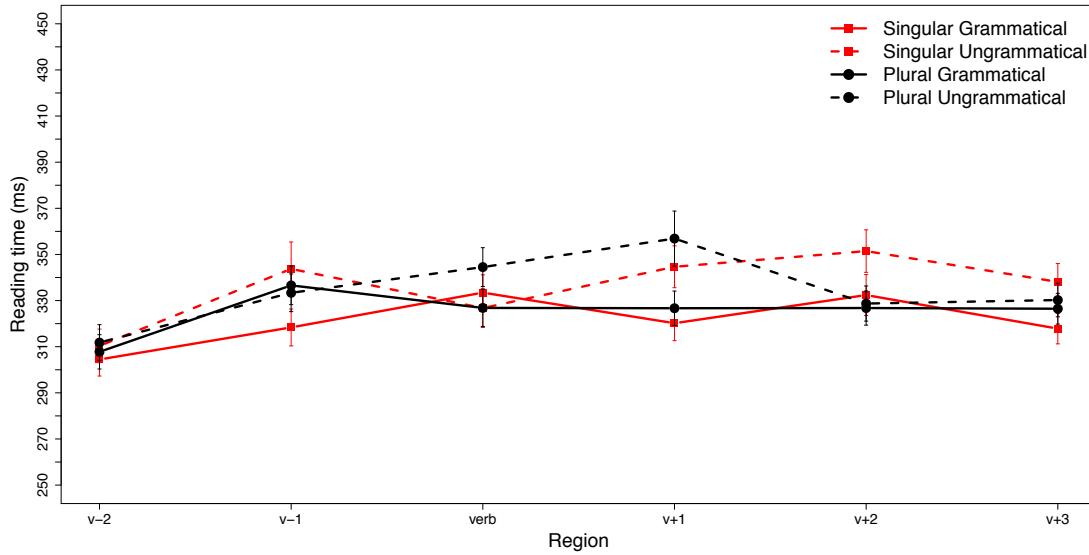


Figure 20: Region-by-region mean raw reading times for the L1 group of sentences with singular head nouns, Experiment 1. Error bars indicate standard error of the mean.

	Estimate	Std. Error	t-value
Intercept	5.78	0.04	<b>166.16</b>
Grammaticality	0.02	0.02	1.04
Attractor	-0.02	0.02	-0.73
Gram x Attractor	0.08	0.04	<b>1.84</b>

Table 44: Results of the linear mixed effects model in the verb region of sentences with singular head nouns for the L1 group, Experiment 11 (using log transformed RTs).

	Estimate	Std. Error	t-value
Intercept	5.77	0.04	<b>156.05</b>
Grammaticality	0.06	0.02	<b>3.25</b>
Attractor	0.02	0.02	1.03
Gram x Attractor	-0.02	0.04	-0.39

Table 45: Results of the linear mixed effects model in the spillover region of sentences with singular head nouns for the L1 group, Experiment 11 (using log transformed RTs).

*L2 group*

The L2 group's region-by-region average reading times for sentences with singular head nouns are plotted in Figure 21. Table 46 and 47 contain the results of the model in the two regions of analysis for the L2 group. There was a significant effect of grammaticality in the verb region ( $t = 2.17$ ): sentences with a subject-verb agreement violation were read more slowly than grammatical sentences. The interaction between grammaticality and attractor number was not significant ( $t = 0.02$ ). No significant effects were observed in the spillover region.

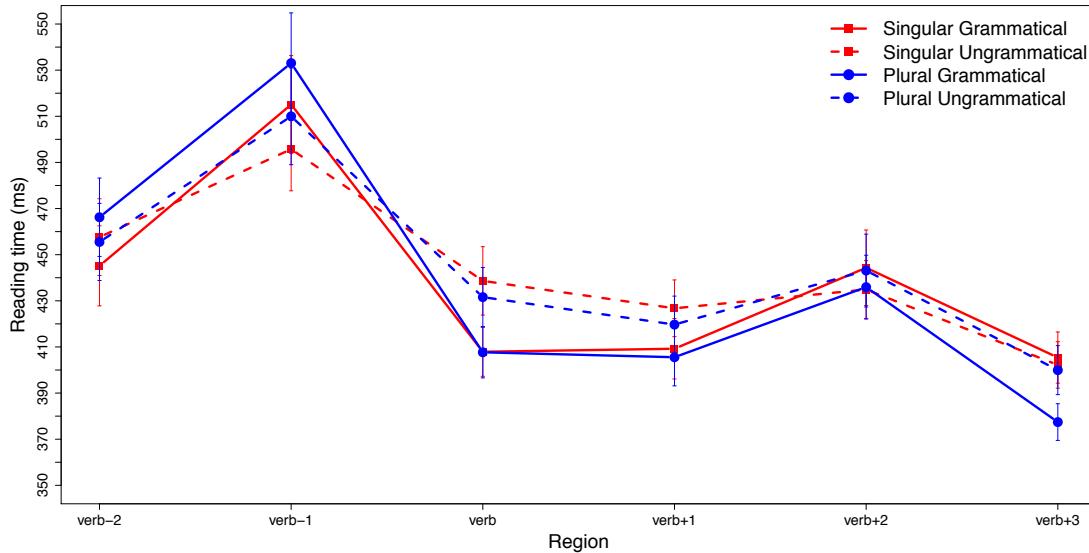


Figure 21: Region-by-region mean raw reading times for the L2 group for sentences with singular head nouns, Experiment 11 (error bars indicate standard error of the mean).

	Estimate	Std. Error	t-value
Intercept	6.004	0.026	<b>227.28</b>
Grammaticality	0.026	0.012	<b>2.17</b>
Attractor	0.004	0.012	0.31
Gram x Attractor	<0.001	0.012	0.02

Table 46: Results of the linear mixed effects model in the verb region of sentences with singular head nouns for the L2 group, Experiment 11 (using log transformed RTs).

	Estimate	Std. Error	t-value
Intercept	5.990	0.038	<b>157.06</b>
Grammaticality	0.009	0.013	0.71

Attractor	-0.004	0.013	-0.34
Gram x Attractor	-0.001	0.013	-0.06

Table 47: Results of the linear mixed effects model in the spillover region of sentences with singular head nouns for the L2 group, Experiment 11 (using log transformed RTs).

### *Plural head nouns*

Table 48 and 49 present the results of the linear mixed effects models in the verb and the spillover region for sentences with plural head nouns. In the overall model including data from both language groups, there was a significant effect of attractor number in the verb region ( $t = 2.73$ ), with plural attractors taking longer to read than singular ones. This effect remains marginally significant in the spillover region ( $t = 1.87$ ). There was also an effect of language in both the verb region ( $t = 5.00$ ) and the spillover region ( $t = 4.38$ ): the L2 group's reading times were significantly slower than the L1 group's. Surprisingly, the effect of grammaticality is not significant in either the verb region ( $t = -0.01$ ) or the spillover region ( $t = 0.02$ ).

	Estimate	Std. Error	t-value
Intercept	5.91	0.02	<b>246.81</b>
Grammaticality	< -0.01	0.02	-0.01
Attractor	0.04	0.02	<b>2.73</b>
Language	0.23	0.05	<b>5.00</b>
Gram x Attractor	0.01	0.03	0.22
Gram x Lang	-0.01	0.03	-0.35
Attractor x Lang	-0.01	0.03	-0.02
Gram x Attractor x Lang	-0.07	0.06	-1.09

Table 48: Results of the linear mixed effects model in the verb region of sentences with plural head nouns for both language groups, Experiment 11 (using log transformed RTs).

	Estimate	Std. Error	t-value
Intercept	5.88	0.03	<b>219.22</b>
Grammaticality	< 0.01	0.02	0.02
Attractor	0.03	0.02	<b>1.87</b>
Language	0.23	0.05	<b>4.38</b>

Gram x Attractor	0.01	0.03	0.20
Gram x Lang	-0.09	0.03	-2.86
Attractor x Lang	0.02	0.03	0.57
Gram x Attractor x Lang	-0.11	0.06	-1.68

Table 49: Results of the linear mixed effects model in the spillover region of sentences with plural head nouns for both language groups, Experiment 11 (using log transformed RTs).

### *L1 group*

Table 15 and 16 show the results of the linear mixed effects models for the L1 group in the verb region and the spillover region. The L1 group showed a significant slowdown with plural attractors in the verb region ( $t = 2.31$ ). The effect of grammaticality was significant only in the spillover region ( $t = 2.32$ ), with agreement violations leading to an increase in reading times.

	Estimate	Std. Error	t-value
Intercept	5.80	0.03	<b>170.66</b>
Grammaticality	0.01	0.02	0.27
Attractor	0.04	0.02	<b>2.31</b>
Gram x Attractor	0.04	0.04	1.09

Table 50: Results of the linear mixed effects model in the verb region of sentences with plural head nouns for the L1 group, Experiment 11 (using log transformed RTs).

	Estimate	Std. Error	t-value
Intercept	5.77	0.04	<b>151.52</b>
Grammaticality	0.05	0.02	<b>2.32</b>
Attractor	0.02	0.02	1.06
Gram x Attractor	0.06	0.04	1.49

Table 51: Results of the linear mixed effects model in the spillover region of sentences with plural head nouns for the L1 group, Experiment 11 (using log transformed RTs).

### *L2 group*

Table 17 and 18 show the results of the linear mixed effects models for the L1 group in the verb region and the spillover region. The L2 group did not show any significant effects in the verb region. In the spillover region, there was a marginally significant effect

of grammaticality ( $t = -1.85$ ). However, this reflected ungrammatical sentences being read faster than grammatical ones.

	Estimate	Std. Error	t-value
Intercept	5.80	0.03	<b>185.71</b>
Grammaticality	-0.01	0.03	-0.26
Attractor	0.04	0.03	1.65
Gram x Attractor	-0.03	0.05	-0.53

Table 52: Results of the linear mixed effects model in the verb region of sentences with plural head nouns for the L2 group, Experiment 11 (using log transformed RTs).

	Estimate	Std. Error	t-value
Intercept	5.99	0.04	<b>160.09</b>
Grammaticality	-0.05	0.03	<b>-1.85</b>
Attractor	0.04	0.03	1.51
Gram x Attractor	-0.05	0.05	-0.91

Table 53: Results of the linear mixed effects model in the spillover region of sentences with plural head nouns for the L2 group, Experiment 11 (using log transformed RTs).

### 5.6.6 Discussion

#### Singular head nouns

The results of the self-paced reading task in Experiment 11 are generally consistent with the findings from Experiment 10. The data indicate that L2 participants experienced processing disruption when they encountered a subject-verb agreement violation in a sentence with a singular head noun. While ungrammaticality led to the expected slowdown in the verb region, this effect was not ameliorated by the presence of a structurally inaccessible plural noun. This is consistent with the results from the speeded acceptability judgment task in Experiment 10 and provides converging evidence that advanced Chinese learners of English do not experience agreement attraction even

though they are sensitive to subject-verb agreement violations in online processing. A sensitivity to agreement violations without interference from a number-matching attractor suggests that when the L2 participants detect the violation at the verb, they engage in retrieval of the agreement controller using only the verb's structural retrieval cues and not the number information.

The data from the L1 group were not quite consistent with previous results in the literature. As expected, agreement violations increased reading times in the spillover region, but there was no indication of an attraction effect until the region after the spillover region. Even then, the reduced slowdown for ungrammatical sentences in the presence of a plural attractor was a numerical trend and not statistically significant. It is not unusual for effects in self-paced reading to occur in the spillover region, but here attraction appears to be delayed compared to previous results and it is only a numerically trend. It is unclear why this control experiment failed given all the prior replications of the attraction effect in this construction in English self-paced reading, including the SPR experiments reported in the other chapters of this dissertation. This potentially weakens the conclusions we can draw from the L2 data in this experiment.

### *Plural head nouns*

The results for the sentences with plural head nouns provide further evidence that native speakers of English do not experience agreement attraction in ungrammatical sentences when the subject is plural and the attractor is singular. In addition to a slowdown in response to an agreement violation in the verb's spillover region, the L1

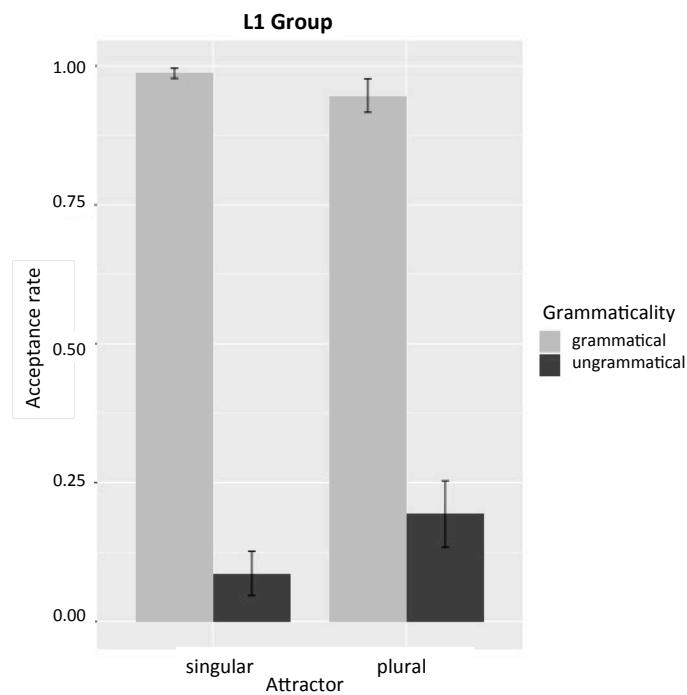
participants were also slower to read the verb region when the attractor was plural. This could be accounted for by the plural complexity effect found by Wagers et al. (2009). However, no such effect was found in this experiment when the head noun was singular. It is possible that integrating the plural attractor with the plural subject imposed additional processing demands. The L1 speakers' reading times are consistent with their behavior in Experiment 10. The data from the L2 group, however, is not consistent with the findings from Experiment 10. In the speeded acceptability judgment task, L2 learners were more likely reject ungrammatical than grammatical sentences even when the subject's head noun was plural. The self-paced reading data failed to show the corresponding slowdown for ungrammatical sentences. There was a marginally significant effect of grammaticality in the verb's spillover region, but it was in the opposite direction: grammatical sentences were read more slowly than ungrammatical ones. It is unclear why the L2 learners were able to detect agreement violations with plural subjects in the speeded acceptability task but not in the self-paced reading task. It is possible that the acceptability task itself contributed to this, as it would have made participants pay more attention to agreement violations than simply reading for comprehension. This is somewhat consistent with the results from Lim & Christianson (2015), who found that how well Korean learners of English detected subject-verb agreement violations depended on the task they had to perform. In that case, the question arises why this task-effect did not affect their sensitivity to agreement violations with singular subjects. Alternatively, plurals might be more difficult to process than singulars, leading to higher a processing load in sentences with a plural subject and affecting the L2 learners ability to process agreement.

### *Offline acceptability judgment task*

The results from the offline acceptability judgment task that participants completed after the self-paced reading experiment are presented in Table 54 – 56 and plotted in Figure 22. Both the L1 and the L2 group were more likely to accept grammatical than ungrammatical sentences ( $p < 0.001$ ). However, the L2 group had overall lower acceptance rates ( $p = 0.02$ ) and grammaticality had a smaller impact than for the L1 group ( $p < 0.001$ ). Interestingly, the attraction between attractor number and grammaticality was significant not just for the L1 but also the L2 group. Ungrammatical sentences were more likely to be judged acceptable with a plural attractor. While this is exactly what we expected for the L1 group, the L2 group's behavior contrasts with their performance in both the speeded acceptability task and the self-paced reading part of this experiment. L2 participants did not show any attraction effect in either of those two online measures. It appears that while advanced Chinese learners of English are not susceptible to the grammatical illusion of agreement attraction in online processing, they do behave like native speakers in this regard when given unlimited time.

It is possible that the apparent agreement attraction effect for the L2 learners in the offline task is driven by a different mechanism than agreement attraction in native speakers. In the offline acceptability judgment task, participants not only had unlimited time to make a judgment but they also saw the whole sentence displayed on the screen the entire time. In contrast, in the speeded acceptability judgment task and the self-paced reading task, visual information was always limited to one word at a time and there was no way to re-read previous material. Upon encountering an agreement violation at the

verb, L2 learners might have visually backtracked. Since the attractor was linearly closer to the verb than the subject's head noun it would have been the first noun they encountered. In that case, the apparent attraction would not be a result of L2 learners using the verb's number information to guide retrieval of the agreement controller, but rather an effect of proximity.



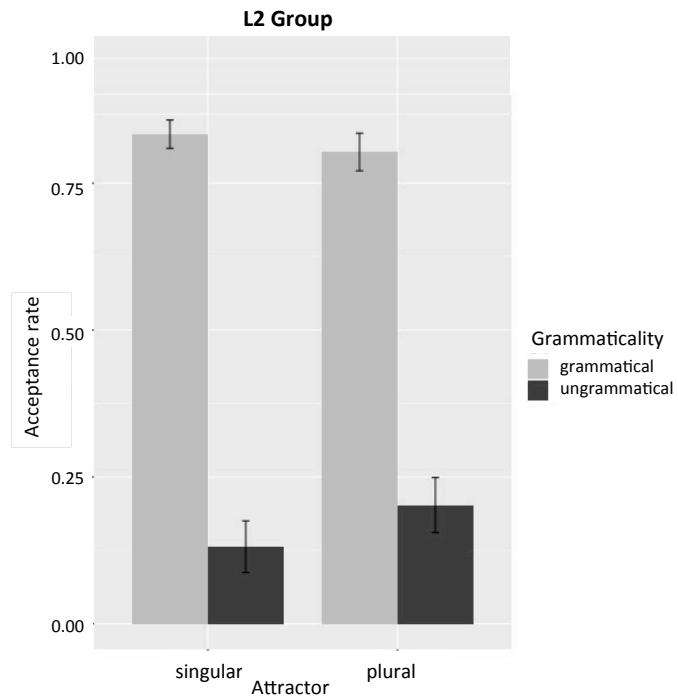


Figure 22: Acceptance rates in offline acceptability judgment task for L1 group (upper panel) and L2 group (lower panel) in Experiment 11. Error bars indicate standard error of the mean.

	Estimate	Std. Error	t-value
Intercept	0.40	0.19	<b>0.04</b>
Grammaticality	-5.13	0.29	<b>&lt; 0.001</b>
Attractor	0.08	0.25	0.74
Language	-0.92	0.38	<b>0.02</b>
Gram x Attractor	1.81	0.50	<b>&lt; 0.001</b>
Gram x Lang	2.59	0.54	<b>&lt; 0.001</b>
Attractor x Lang	0.45	0.50	0.36
Gram x Attractor x Lang	-1.63	0.99	0.10

Table 54: Results of mixed logit model on acceptance rates in offline acceptability judgment task for both L1 and L2 groups in Experiment 11.

	Estimate	Std. Error	t-value
Intercept	5.19	0.84	<b>&lt; 0.001</b>
Grammaticality	-8.36	0.95	<b>&lt; 0.001</b>
Attractor	-1.53	0.82	0.06
Gram x Attractor	2.66	0.93	<b>&lt; 0.01</b>

Table 55: Results of mixed logit model on acceptance rates in offline acceptability judgment task for the L1 group in Experiment 11.

	Estimate	Std. Error	t-value
Intercept	1.93	0.27	< <b>0.001</b>
Grammaticality	-4.28	0.34	< <b>0.001</b>
Attractor	-0.19	0.27	0.50
Gram x Attractor	0.98	0.41	<b>0.02</b>

Table 56: Results of mixed logit model on acceptance rates in offline acceptability judgment task for the L2 group in Experiment 11.

## 5.7 General Discussion

The experiments presented here investigated whether second language learners of English whose native language does not have subject-verb agreement can acquire not only the grammatical knowledge associated with this constraint but also implement it in a native-like way in online processing. Specifically, we asked whether advanced Chinese learners of English can learn to use the number information on the verb to guide the search for the agreement controller in memory when they encounter a subject-verb agreement violation, or whether they are limited to using the structural cues available in their native language. For subjects with a singular head noun, Experiment 10 showed that the L2 learners were sensitive to subject-verb agreement violations in a speeded acceptability judgment task but did not show any agreement attraction. The L1 control group showed the expected effect of grammaticality and an agreement attraction effect. The self-paced reading data from Experiment 11 confirmed that the L2 group was sensitive to agreement violations. Again there was no evidence that their processing of subject-verb agreement violations was impacted by attractor number. However, the fact that we failed to replicate reliable self-paced reading agreement attraction effects in the L1 group makes the results of Experiment 11 somewhat less informative than those of Experiment 10.

We also tested plural subjects with the aim to investigate whether the L2 learners' mental representation of grammatical number is asymmetric like in native speakers. In Experiment 10, the L2 group was sensitive to agreement violations with plural subjects and did not show an effect of attractor number. As expected, the L1 group showed the same pattern. However, in Experiment 11 the L2 learners' reading times did not reveal any sensitivity to agreement violations with plural subjects. The L1 group displayed the expected effect of grammaticality, which was not impacted by attractor number. This is consistent with the hypothesis that number is asymmetrical in the mental representation of native speakers. Surprisingly, both groups showed the same pattern of attraction in a post-experiment untimed acceptability judgment experiment, suggesting that the L2 learners used a slightly different strategy under these circumstances.

### *5.7.1 L2 Morphological Sensitivity*

The results of Experiment 10 and 11 clearly demonstrate that the advanced Chinese learners of English in our study have acquired the grammatical knowledge about subject-verb agreement in English and can implement this knowledge in online processing. This contradicts claims by Jiang (2004; 2007) about Chinese learners' inability to integrate this constraint into their implicit linguistic knowledge. However, not only are these findings consistent with later studies that found sensitivity to agreement violations across a number of measures (Chen et al., 2007; Armstrong et al., 2016), but the numerical pattern of the data on which Jiang bases the claim of morphological insensitivity appears mostly consistent with the hypothesis that the L2 learners were able

to detect agreement violations in online processing, even if the effects are not statistically significant.

It is unclear why the L2 participants in this study showed no evidence of detecting the agreement violation in the self-paced reading experiment when the subject was plural. The results from the speeded acceptability judgment task demonstrate that they are in principle able to detect such a violation in online processing.

### *5.7.2 Processing routines*

Although the L2 learners were able to detect agreement violations, the way in which they implement their grammatical knowledge during processing differs from native speakers of English. Native speakers of English are susceptible to grammatical illusions in processing subject-verb agreement violations because of the kinds of information they use to navigate linguistic representations in memory. When they encounter a subject-verb agreement violation, this triggers an error-driven process in which they use the verb's cues to retrieve the subject. This retrieval is guided by both structural and morphosyntactic cues. When there is a structurally irrelevant noun that matches the number marking of the verb, this sometimes leads to misretrieval of the attractor instead of the actual target (subject) and causes facilitative similarity-based interference. Crucially, agreement attraction effects such as increased acceptance rates and reduced slowdown in reading times for agreement violations depend on comprehenders using the verb's number cue for retrieval.

In the two experiments reported here, the L2 speakers showed no evidence of erroneously retrieving a number matching attractor instead of the subject. This indicates

that unlike native speakers, they do not actually make use of the number cue on the verb when it comes to retrieval from memory. That does not mean that they ignore the number marking on the verb in online processing, since that would result in a lack of sensitivity to agreement violations. Instead, it suggests that once they notice the mismatch between the subject's number feature and the number marking on the verb, they rely exclusively on the verb's structural cues which are also available in their native language to retrieve the agreement controller. Although the L2 learners were less accurate at judging sentences acceptable or unacceptable than the native speakers, not using information the L1 speakers were using for retrieval meant that in this situation they were not susceptible to a common processing error. Their non-native processing strategy actually meant that in a certain sense they were processing the input more accurately than the native speakers.

### *5.7.3 Asymmetrical number representation*

An additional question we hoped to address with these experiments was whether Chinese learners of English represent grammatical number in a native-like way. For L1 speakers, grammatical number is argued to be a privative feature; singular is the absence of this feature and therefore the default, while plural is the marked value. Chinese does not have grammatical number in the same way as English, so native speakers of Chinese have to acquire grammatical number as a feature when they are learning their L2. It is possible that acquiring grammatical knowledge later in life results in a very different mental representation. However, since the L2 participants did not show any attraction effects in the online tasks (and the offline task did not include plural subjects), it is

impossible to draw any conclusions about their mental representation of number from these experiments.

#### *5.7.4 Offline strategies*

One surprising finding was that L2 learners appear to show attraction effects offline but not online. We propose that this was the result of a reading strategy that is specific to situations in which the whole sentence is displayed. In that case, the L2 learners do not have to rely exclusively on their own memory representations and can instead visually backtrack to check the number information of the subject. However, since the attractor occurs linearly between the verb and the subject's head noun, regressive eye-movements are likely to land on the attractor making its number information accessible. Interestingly, this predicts that in cases of proactive interference like relative clauses the L2 learners should not show a native-like attraction pattern in an offline judgment task. It should be emphasized that the findings from this offline task cannot be extrapolated to L2 processing in real-world situations. The offline nature makes this task fundamentally different from spoken language comprehension, in which the input is fleeting and comprehenders have to rely on their own memory representations.

#### *5.7.5 Automatic sensitivity*

The L2 group's sensitivity to agreement violations (with singular subjects) in the self-paced reading task provides interesting evidence for the automatic nature of

agreement processing by these advanced learners. Although we expected them to show the observed pattern based on the results from the ERP studies (Chen et al., 2007; Armstrong et al., 2016), it is worth pointing out that in those studies participants had to make an explicit acceptability judgment after each sentence. In contrast, our self-paced reading experiment required only reading for comprehension, making the task demands more similar to real-world language processing. This suggests that the L2 learners' sensitivity to agreement violations is automatic and not more conditioned on the task than for native speakers. However, this has to remain a very tentative conclusion given that the L2 learners seemed to be sensitive to agreement violations with plural subjects only in the speeded acceptability judgment task.

## 5.8 Conclusion

We used speeded acceptability judgments and self-paced reading to show that advanced native speakers of Chinese are sensitive to subject-verb agreement violations in online processing, but do not implement their grammatical knowledge in a native-like way. The experiments reported here demonstrate that advanced Chinese learners of English do not use the verb's number information to guide the search for the agreement controller when the detection of an agreement violation triggers error-driven retrieval. Unlike native speakers of English, they do not experience an illusion of grammaticality in ungrammatical sentences with singular subjects and a structurally irrelevant noun matching the number marking on the verb.

These findings have a number of potential implications. They provide converging evidence that it is possible for L2 learners to acquire features not encoded in their L1 and

to use this knowledge in online processing. However, these constraints may not always be implemented in the same way as in native speakers. In fact, the L2 learners' sensitivity to agreement violations in conjunction with the lack of agreement attraction suggests that grammatical constraints do not necessarily have to be implemented as cues in error-driven retrieval. It is quite possible that this is only the case if the constraint is specific to the L2 and has to be acquired later in life, but it is an interesting possibility.

The results of Experiments 10 and 11 also have implications for our understanding of what is a desirable attainment regarding processing in the acquisition of a second language. The default assumption appears to be that second language learners should strive to become as native-like as possible. This would include using the verb's number cue to guide retrieval of the agreement controller when they detect a mismatch between the predicted verb form and the one in the input. However, the L2 learners' non-native like implementation of their knowledge about subject-verb agreement means that they do not seem to fall victim to a processing error that native speakers are very prone to. In this case, their failure to use a native-like strategy meant that the L2 learners were processing the linguistic input more accurately than the native English speakers. Using only the verb's structural cue to retrieve the agreement controller does not harm them in online processing but is arguably helpful, casting some doubt on the idea that native-like processing is always best.

# **Chapter 6**

## **6.1 Summary**

In this dissertation, I have argued that the illusory licensing of subject-verb agreement violations in the presence of a structurally irrelevant number-matching noun is the result of an error-driven rechecking process that does not trigger extensive structural reanalysis. I have explored the relationship between this low-level rechecking operation and the grammar, and I have argued that the retrieval cues used in the rechecking operation are not only as abstract as the terms in which agreement is stated in the grammar, but are also limited to those grammatically relevant cues. However, I also suggested that it might not be the case that all grammatical knowledge is necessarily implemented in the form of retrieval cues.

## **6.2 Error-driven retrieval as a rechecking operation**

In a study that tested whether comprehenders misinterpret the attractor as the subject when they experience agreement attraction, I explored the impact agreement attraction has on the structural representation of the sentence. The data showed that comprehenders very rarely mistake the attractor as the subject, even in the ungrammatical condition with a plural attractor. If agreement attraction occurs on a substantial number of trials with an attraction configuration, as suggested by the corresponding self-paced reading data in this study, this demonstrates that the attractor is misrepresented as the subject in only a very small subset of instances on which it was misretrieved. We can

therefore rule out the possibility that misretrieval of the attractor necessarily triggers structural reanalysis.

A conclusion that we can draw from this is that illusory licensing with subject-verb agreement violations is possible when the final mental representation of the sentence is not consistent with the grammar. If the retrieval output of the error-driven rechecking operation does not trigger restructuring of the previously encoded material, the final representation of a sentence with agreement attraction still contains a mismatch between the subject's and the verb's number. This suggests that grammatical illusions can arise without a final structural representation of the sentence that is entirely licensed by the grammar. This is an interesting discrepancy between perceived acceptability and the grammatical status of the final representation. While these results suggest that it is possible that the mental representation underlying a grammatical illusion is not consistent with the grammar, at this point we can only speculate whether that is also the case for other types of grammatical illusions. It is conceivable that being able to predictively establish the dependency between the subject and the verb plays a role in this. Error-driven retrieval of the agreement controller appears to be an operation that simply checks the verb's number marking against the retrieved item's number feature. If an item is retrieved that matches the plural number cue on the verb, this fulfills the function of licensing the plurally marked verb. The situation is very different in reflexive processing: the comprehender actually has to establish the dependency between the antecedent and the reflexive when they encounter the reflexive in the input. Unlike subject-verb agreement, the dependency between the antecedent and the reflexive cannot be predicted.

This means that retrieval of the antecedent cannot be a process that just checks if the reflexive is licensed, it must actually integrate the

While this study demonstrates that agreement attraction does clearly not involve large scale restructuring every time it occurs, we cannot dismiss the finding that the attractor-matching adjectives were chosen more frequently for ungrammatical sentences with plural attractors. Although this was a very small effect, it was statistically significant and suggests that error-driven retrieval can contribute to the likelihood of the parser engaging in structural reanalysis. This could be further explored by replicating this effect in a dual-task paradigm with materials that have been carefully edited to address any concerns possible task effects due to the structure of the sentences used in Experiment 2.

Another interesting implication of the results in Experiment 2 is that they actually support an error-driven account of agreement attraction. In a model in which retrieval of the agreement controller is obligatory, the dependency between the subject and the verb is not established predictively and has to be established upon encountering the verb. If the attractor is erroneously retrieved instead of the agreement controller, the output of that retrieval operation has to be integrated with the verb at the point at which the structure from which the interpretation is derived is generated. Consequently, misretrieval of the attractor should always result in representing it as the subject of the verb and thus misinterpretation. However, that is clearly not what we observe in Experiment 2.

### 6.3 Grammatically (ir)relevant cues in error-driven retrieval

In this dissertation, I have explored what retrieval cues are used by the repair process triggered by an agreement violation. Agreement attraction is a two-step process:

the verb's number is predicted by top-down processing based on the number of the subject; when there is a mismatch between the prediction and the bottom-up input, this triggers error-driven retrieval of the agreement controller. I adopted the assumption that the first stage, the predictive process of anticipating the verb's number based on the number of the subject, is strictly governed by how subject-verb agreement is defined in the grammar. However, even if we think that the parsing mechanism predictively generating structure is completely faithful to the grammar, it is not *a priori* clear how a repair process like error-driven retrieval might differ in its use of grammatically relevant and irrelevant information. It could be the case that once an error signal is detected in the input, the repair process uses all available information to guide retrieval of the agreement controller, even if it is not relevant to the dependency in the grammar. However, the results from a series of studies in this dissertation investigating what cues the parser uses to search for the agreement controller in memory indicate that inspite of being a sort of repair process, error-driven retrieval appears to use information that is relevant to the dependency in the grammar.

I conducted a series of experiments investigating whether the cues in the error-driven retrieval of the agreement controller are as abstract as the terms in which agreement is stated in the grammar. The results indicate that agreement attraction can arise from an attractor that is syntactically plural but does not contain an unequivocal morphological signal of syntactic plurality. This demonstrates that the error-driven retrieval operation is sensitive to the same type of plurality that determines agreement in the grammar. This of course has broader implications for our models of cue-based

retrieval in language processing in general. It is evidence that models need to include retrieval cues as abstract as the terms in which dependencies are stated in the grammar.

In contrast, there was no clear evidence that information that is not grammatically relevant to agreement was used to guide retrieval of the agreement controller. There was no indication that a notionally plural attractor can cause attraction effects in comprehension. Neither were attraction effects impacted by whether the attractor was a plausible subject for the verb. This suggests that error-driven retrieval not only uses cues as abstract as the grammar, but in fact does not use any grammatically irrelevant cues. Interestingly, the results from the study of agreement processing in advanced Chinese learners of English provide what could be interpreted as evidence against the idea that error-driven retrieval relies on using grammatically relevant cues. In this study, we found that the L2 learners had acquired the L2 specific knowledge about subject-verb number agreement which does not exist in Chinese and were sensitive to it in online processing. However, while they detected agreement violations, the disruption caused by them was not modulated by attractor number: the L2 learners did not show evidence of agreement attraction. This pattern suggests that they do not use the verb's number cue to guide retrieval of the agreement attractor, which is the reason for the facilitative similarity-based interference native speakers experience with number matching attractors. Although these learners have essentially the same grammatical knowledge about subject-verb agreement as the native speakers, it is not implemented as cues in error-driven retrieval. Of course, if error-driven retrieval uses only grammatically relevant information that does not necessarily mean that it has to use all grammatically relevant information. However, this is still an interesting discrepancy. The participants in this study were second language

learners, so it remains to be seen whether not implementing grammatical knowledge as retrieval cues is something that is specific to second language acquisition or whether this can also be the case in a native language.

### 6.3.1 Associative cues

Given the general finding that error-driven retrieval in agreement processing relies on the cues that define subject-verb agreement in the grammar, a somewhat surprising result in this dissertation was that the presence of the word '*and*' appears to cause a very small amount of agreement attraction. If only the abstract retrieval cues that define the dependency in the grammar are used, why should retrieval be sensitive to the presence of the word '*and*'? This might be explained by the frequent co-occurrence of '*and*' with syntactic plurality. However, in one of the other experiments we did not find any evidence for attraction effects with notionally plural attractors. Although they are dissociable, notional number exhibits a very strong, if imperfect, correlation with syntactic number. Based on this strong correlation between notional number and syntactic number, we would expect it to be even more associated with syntactic plurality than '*and*'. A possible way to account for this difference is that notional number and '*and*' operate on different levels of representation. Notional number relates to the conceptual representation of a referent and is not directly encoded in the surface form of the sentence. In contrast, '*and*' is very much clearly present in the input.

## 6.4 Implications for our understanding of cue-based retrieval

In this dissertation, I have presented empirical evidence suggesting that the illusory licensing of subject-verb agreement violations is the result of a low-level rechecking operation that uses only information that is relevant to subject-verb agreement in the grammar. This also suggests that the illusory licensing can occur even if the ultimate mental representation is not entirely consistent with the grammar, as it still contains an agreement error.

As mentioned in the introduction, agreement attraction is only one of the linguistic illusions comprehenders are susceptible to. Naturally, the question arises whether the findings on agreement attraction in this dissertation can be extended to other grammatical illusions. Ultimately, we are of course not only interested in grammatical illusions but in the mechanisms underlying language processing in general. In this section I summarise the implications of the experimental results in this dissertation sentence processing in general.

This dissertation showed that the cue-based retrieval operation in agreement processing uses abstract cues and relies on information relevant to the specific dependency in the grammar. In a cue-based retrieval framework of sentence processing, the question arises whether all dependencies have to be established via retrieval (Lewis & Vasishth, 2005; Van Dyke, 2007). If so, encountering a verb would always trigger retrieval of the agreement controller. The other possibility, which I have adopted in this dissertation, is that cue-based retrieval is not necessary when the input is consistent with the parser's prediction. In agreement processing, retrieval of the agreement controller is an error-driven process that only occurs when an agreement violation is detected (Wagers

et al., 2009; Lago et al., 2015). In fact, the results of the experiment discussed in Chapter 2 provide support for the hypothesis that agreement attraction is specifically an error-driven phenomenon. This makes it different from retrieval for dependencies that cannot be predicted. While it is reasonable to predict that there will be a verb that has to agree with the subject, it is not possible for the parser to predict that a sentence will contain a reflexive or VP-ellipsis. The lack of prediction in these cases means that the parser always has to engage in retrieval for these dependencies.

It seems very likely that retrieval that is not error-driven also uses abstract cues and relies on information relevant to the specific dependency in the grammar. However, error-driven retrieval might be different from non-error driven retrieval in regard to type of information that is retrieved from memory is different in error-driven retrieval compared to retrieval for establishing a dependency that is not predictable. The experiments reported here indicate that the cues used in error-driven retrieval are the ones defining the dependency in the grammar. The lack of a structural impact of agreement attraction indicates that it is a low-level rechecking mechanism. For that purpose, it would suffice to retrieve only the number information of the agreement controller to check whether it licenses the verb's number marking. In contrast, the output of the retrieval operation for a reflexive a dependency determines how the reflexive is interpreted. That makes it very different from agreement, which does not have any bearing at all on interpretation and means that the information that has to be retrieved from memory about the antecedent of a reflexive is not limited to its number feature.



## Bibliography

- Acuña-Fariña, J. C. (2009). The linguistics and psycholinguistics of agreement: A tutorial overview. *Lingua*, 119(3), 389-424.
- Acuña-Fariña, J. C. (2012). Agreement, attraction and architectural opportunism. *Journal of Linguistics*, 48(2), 257-295.
- Acuña-Fariña, J. C., Meseguer, E., & Carreiras, M. (2014). Gender and number agreement in comprehension in Spanish. *Lingua*, 143, 108-128.
- Armstrong, A., Bulkes, N., & Tanner, D. (2016). *ERPs show processing similarities of English subject- verb agreement violations in native speakers and L1 Chinese learners*. Talk given at the 35<sup>th</sup> Second Language Research Forum, Teacher's College, Columbia University, 23-24 September 2016.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390-412.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Eds.), *Cognition and the development of language* (pp. 279–362). New York: Wiley.

Bock, K., & Eberhard, K. M. (1993). Meaning, sound and syntax in English number agreement. *Language and Cognitive Processes*, 8(1), 57-99.

Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, 23(1), 45-93.

Bock, K., Eberhard, K. M., & Cutting, J. C. (2004). Producing number agreement: How pronouns equal verbs. *Journal of Memory and Language*, 51(2), 251-278.

Brehm, L., & Bock, K. (2013). What counts in grammatical number agreement? *Cognition*, 128(2), 149-169.

Brehm, L., & Bock, K. (2017). Referential and lexical forces in number agreement. *Language, Cognition and Neuroscience*, 32(2), 129-146.

Champollion, L. (2013). Man and woman: The last obstacle for Boolean coordination. In M. Aloni, M. Franke, & F. Roelofsen (Eds.), *Proceedings of the 19th Amsterdam Colloquium*, 83–90.

Chen, Z., Cao, J., Cao, Y., Zhang, Y., Gu, F., Zhu, G., ... & Cichocki, A. (2008). An empirical EEG analysis in brain death diagnosis for adults. *Cognitive neurodynamics*, 2(3), 257.

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good-enough, underspecified, or shallow language processing. *The Quarterly Journal of Experimental Psychology*, 69(5), 817-828.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, 42(4), 368-407.

Cunnings, I. (2017). Interference in Native and Non-Native Sentence Processing. *Bilingualism: Language and Cognition*, 1-10.

Davies, M. (2008). *The Corpus of Contemporary American English (COCA)*. BYU, Brigham Young University. <http://corpus.byu.edu/coca>

Dempsey, J., Christianson, K., & Tanner, D. (2016). *Misrepresentations of plurality in late processing: evidence from self-paced reading*. Poster presented at the 29<sup>th</sup> CUNY Conference on Human Sentence Processing, University of Florida, 3-5 March 2015.

Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: experimental and modeling evidence. *Journal of Memory and Language*, 69(2), 85-103.

Drummond, A. (2017). Ibex Farm. URL <http://spellout.net/ibexfarm>

Eberhard, K. M. (1997). The marked effect of number on subject–verb agreement. *Journal of Memory and Language*, 36(2), 147-164.

Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making syntax of sense: number agreement in sentence production. *Psychological Review*, 112(3), 531.

Ellis, N. C. (1998). Emergentism, connectionism and language learning. *Language learning*, 48(4), 631-664.

Engelmann, F., Jäger, L. A., & Vasishth, S. (2016). The determinants of retrieval interference in dependency resolution: Review and computational modeling. *Manuscript submitted*.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, 47(2), 164-203.

Ferreira, F., Bailey, K. G., & 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., & Clifton, C. (1986). The independence of syntactic processing. *Journal of memory and language*, 25(3), 348-368.

Franck, J., Vigliocco, G., & Nicol, J. (2002). Subject-verb agreement errors in French and English: The role of syntactic hierarchy. *Language and cognitive processes*, 17(4), 371-404.

Frazier, L. (2008). Processing ellipsis: A processing solution to the undergeneration problem. In *Proceedings of the 26th West Coast Conference on Formal Linguistics* (pp. 21-32). Somerville, MA: Cascadilla Press.

Frazier, L., & Clifton Jr, 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., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, 6(4), 291-325.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14, 178–210.

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, 110(20), 8051-8056.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Haskell, T. R., & MacDonald, M. C. (2003). Conflicting cues and competition in subject–verb agreement. *Journal of Memory and Language*, 48(4), 760-778.

Haznedar, B., & Schwartz, B. D. (1997). Are there optional infinitives in child L2 acquisition. In *Proceedings of the 21st annual Boston University conference on language development* (Vol. 21, pp. 257-68).

Heycock, C., & Zamparelli, R. (2005). Friends and colleagues: Plurality, coordination, and the structure of DP. *Natural language semantics*, 13(3), 201-270.

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, 120(4), 901-931.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3), 346-363.

Humphreys, K. R., & Bock, J. K. (2005). Notional number agreement in English. *Psychonomic Bulletin & Review*, 12, 689–695.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434-446.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, 25(04), 603-634.

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language learning*, 57(1), 1-33.

Jiang, N., Novokshanova, E., Masuda, K., & Wang, X. (2011). Morphological congruency and the acquisition of L2 morphemes. *Language Learning*, 61(3), 940-967.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive psychology*, 21(1), 60-99.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111(2), 228.

Kaan, E. (2002). Investigating the effects of distance and number interference in processing subject-verb dependencies: An ERP study. *Journal of Psycholinguistic Research*, 31(2), 165-193.

Karimi, H., & Ferreira, F. (2016). Informativity renders a referent more accessible: Evidence from eyetracking. *Psychonomic Bulletin & Review*, 23(2), 507-525.

Keung, L., & Staub, A. (2016). *Closest conjunct agreement in English: a comparison with number attraction*. Paper presented at 29th Annual CUNY Conference on Human Sentence Processing, Gainesville, FL.

Koornneef, A., & Reuland, E. (2016). On the shallow processing (dis) advantage: Grammar and economy. *Frontiers in Psychology*, 7, 1-17.

Kush, D., Lidz, J., & Phillips, C. (2015). Relation-sensitive retrieval: evidence from bound variable pronouns. *Journal of Memory and Language*, 82, 18-40

Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement attraction in Spanish comprehension. *Journal of Memory and Language*, 82, 133-149.

Lardiere, D. (1998). Dissociating syntax from morphology in a divergent L2 end-state grammar. *Second language research*, 14(4), 359-375.

Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research*, 25(2), 173-227.

Lau, E., Wagers, M., Stroud, C., & Phillips, C. (2008). Agreement and the subject of confusion. *Talk given at 21st annual CUNY sentence processing conference*. Chapel Hill, NC: University of North Carolina.

Levy, R. (2008). A noisy-channel model of rational human sentence comprehension under uncertain input. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 234-243). Association for Computational Linguistics.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences*, 106(50), 21086-21090.

Lewis, S., & Phillips, C. (2015). Aligning grammatical theories and language processing models. *Journal of Psycholinguistic Research*, 44(1), 27-46.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3), 375-419.

Lim, J. H., & Christianson, K. (2015). Second language sensitivity to agreement errors: Evidence from eye movements during comprehension and translation. *Applied Psycholinguistics*, 36(6), 1283-1315.

Linck, J. A., & Cummings, I. (2015). The Utility and Application of Mixed-Effects Models in Second Language Research. *Language Learning*, 65(S1), 185-207.

Lorimor, H., Jackson, C. N., Spalek, K., & van Hell, J. G. (2016). The impact of notional number and grammatical gender on number agreement with conjoined noun phrases. *Language, Cognition and Neuroscience*, 31(5), 646-661.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, 23(2), 127-150.

Martin, A. E., & McElree, B. (2008). A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *Journal of Memory and Language*, 58(3), 879-906.

Martin, A. E., & McElree, B. (2009). Memory operations that support language comprehension: evidence from verb-phrase ellipsis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(5), 1231.

McCarthy, Co. (2011). Constraining Morphological Variation and Development: Agreement in L2 Spanish. In *Proceedings of the 11th Generative Approaches to Second Language Acquisition Conference (GASLA 2011)*, ed. Julia Herschensohn and Darren Tanner, 76-82. Somerville, MA: Cascadilla.

McCloskey, J. (1991). There, it, and agreement. *Linguistic Inquiry*, 22(3), 563-567.

- McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research*, 29(2), 111-123.
- McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, 48(1), 67-91.
- Mirković, J., & MacDonald, M. C. (2013). When Singular and Plural are Both Grammatical: Semantic and Morphophonological Effects in Agreement. *Journal of Memory and Language*, 69(3), 277–298.
- Parker, D., & Phillips, C. (2016). Negative polarity illusions and the format of hierarchical encodings in memory. *Cognition*, 157, 321-339.
- Patson, N. D. (2014). The processing of plural expressions. *Language and Linguistics Compass*, 8(8), 319-329.\
- Patson, N. D., & Husband, E. M. (2016). Misinterpretations in agreement and agreement attraction. *The Quarterly Journal of Experimental Psychology*, 69(5), 950-971.
- Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, 41(3), 427-456.
- Perkins, K., & Freeman, D. L. (1975). The effect of formal language instruction on the order of morpheme acquisition. *Language Learning*, 25(2), 237-243.

Phillips, C., & Lewis, S. (2013). Derivational order in syntax: Evidence and architectural consequences. *Studies in Linguistics*, 6, 11-47.

Phillips, C., Wagers, M. W., & Lau, E. F. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. *Experiments at the Interfaces*, 37, 147-180.

Pittman, C., & Smyth, R. (2005). The effect of the predicate on agreement error rates. In *Proceedings of the 2005 Canadian Linguistics Association Annual Conference*.

Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second language research*, 16(2), 103-133.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. URL <http://www.R-project.org/>.

Ryding, K.C., (2005). *A Reference Grammar of Modern Standard Arabic*. Cambridge, UK: Cambridge UP.

Sagarra, N., & Ellis, N. C. (2013). From seeing adverbs to seeing verbal morphology. *Studies in Second Language Acquisition*, 35(02), 261-290.

Sharwood-Smith, M. (1986). Comprehension versus acquisition: Two ways of processing input. *Applied linguistics*, 7, 239.

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.

Sorace, A. (2004). Native language attrition and developmental instability at the syntax-discourse interface: Data, interpretations and methods. *Bilingualism: Language and Cognition*, 7(02), 143-145.

Sorace, A. (2011). Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches to bilingualism*, 1(1), 1-33.

Sorace, A., & Serratrice, L. (2009). Internal and external interfaces in bilingual language development: Beyond structural overlap. *International Journal of Bilingualism*, 13(2), 195-210.

Slevc, L. R., & Momma, S. (2015). Noisy evidence and plausibility influence structural priming. In Poster presented at the 21st Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP), Valletta, Malta.

Staub, A. (2009). On the interpretation of the number attraction effect: Response time evidence. *Journal of Memory and Language*, 60(2), 308-327.

Staub, A. (2010). Response time distributional evidence for distinct varieties of number attraction. *Cognition*, 114(3), 447-454.

Tanner, D. (2011). Agreement mechanisms in native and nonnative language processing: Electrophysiological correlates of complexity and interference. University of Washington.

Tanner, D., Nicol, J., Herschensohn, J., & Osterhout, L. (2012). Electrophysiological markers of interference and structural facilitation in native and nonnative agreement processing. In *Proceedings of the 36th Boston university conference on language development* (pp. 594-606).

Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism: Language and Cognition*, 17(02), 277-293.

Tanner, D., Nicol, J., & Brehm, L. (2014). The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of Memory and Language*, 76, 195-215.

Thornton, R., & MacDonald, M. C. (2003). Plausibility and grammatical agreement. *Journal of Memory and Language*, 48(4), 740-759.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language*, 33, 285–318.

Tucker, M. A., Idrissi, A., & Almeida, D. (2015). Representing number in the real-time processing of agreement: self-paced reading evidence from Arabic. *Frontiers in Psychology*, 6: 347.

Vainikka, A., & Young-Scholten, M. (1996). Gradual development of L2 phrase structure. *Second language research*, 12(1), 7-39.

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(2), 407.

Van Dyke, J. A., & Johns, C. L. (2012). Memory interference as a determinant of language comprehension. *Language and Linguistics Compass*, 6(4), 193-211.

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, 49(3), 285-316.

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, 55(2), 157-166.

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of memory and language*, 65(3), 247-263.

Vigliocco, G., Butterworth, B., & Garrett, M. F. (1996). Subject-verb agreement in Spanish and English: Differences in the role of conceptual constraints. *Cognition*, 61(3), 261-298.

Vigliocco, G., Butterworth, B., & Semenza, C. (1995). Constructing subject-verb agreement in speech: The role of semantic and morphological factors. *Journal of Memory and Language*, 34(2), 186.

Vigliocco, G., Hartsuiker, R. J., Jarema, G., & Kolk, H. H. J. (1996). One or more labels on the bottles? Notional concord in Dutch and French. *Language and Cognitive Processes*, 11, 407–442.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61(2), 206-237.

