

ABSTRACT

Title of Dissertation: WHAT COULD GO WRONG? LINGUISTIC ILLU-
 SIONS AND INCREMENTAL INTERPRETATION

Hanna Ellen Muller
Doctor of Philosophy, 2022

Directed by: Professor Colin Phillips
 Department of Linguistics

The systems underlying incremental sentence comprehension are, in general, highly successful — comprehenders typically understand sentences of their native language quickly and accurately. The occasional failure of the system to deliver an appropriate representation of a sentence is therefore potentially illuminating. There are many ways the comprehender’s general success could in principle be accomplished; the systematic pattern of failures places some constraints on the possible algorithms. This dissertation explores two cases of systematic failure, negative polarity illusions and substitution illusions (sometimes called “Moses illusions”) with the goal of identifying the specific circumstances under which the illusion arises, and, as a consequence, the specific constraints placed on possible implementations of linguistic knowledge.

In the first part of this dissertation, I explore the profile of the negative polarity illusion, a case in which a sentence containing an unlicensed negative polarity item and a preceding, but not structurally-relevant licenser is perceived as if it is acceptable, at least in early stages of processing. I consider various proposals for the grammatical knowledge that determines the restricted distribution of negative polarity

items, and possible algorithms for using that grammatical knowledge in real time to process a sentence containing a negative polarity item. I also discuss possible parallels between negative polarity illusions and superficially-similar illusory phenomena in other domains, such as subject-verb agreement. Across sixteen experiments, I show that the profile of the illusion is more restricted than previously thought. Illusions do not always arise when an unlicensed negative polarity item is preceded by a structurally-irrelevant licensor, and the circumstances under which they do arise are quite specific. These findings suggest that the negative polarity illusion may be meaningfully distinct from other illusory phenomena, though this conclusion does not necessarily require stipulating a separate mechanism for every illusion. I discuss the implications of these findings for possible real-time implementations of grammatical knowledge.

In the second part of this dissertation, I turn to the substitution illusion, a case in which a word in a trivia fact is swapped out for another word, making the sentence a world knowledge violation, but comprehenders do not consciously detect the anomalous nature of the sentence. Here I attempt to develop specific and testable hypotheses about the source of the illusion, paying particular attention to how the same mechanism that “fails” in illusion sentences (in that it does not allow the comprehender to detect the anomaly) serves the comprehender well in other circumstances. I demonstrate that the substitution illusion, like the negative polarity illusion, is more restricted than previously thought — some stimuli yield very high illusion rates while others yield very low illusion rates, and this variability appears to be non-random. In seven experiments, I pursue both a correlational approach and an experimental manipulation of illusion rates, in order to narrow the space of possible explanations for the illusion.

These investigations collectively demonstrate that occasional errors in comprehension do not necessarily reflect the use of “shortcuts” in sentence processing, and can be explained by the interaction of the linguistic system with non-linguistic components of the cognitive architecture, such as memory and attention. While neither illusion phenomenon is ultimately fully explained, the research presented here constitutes an important step forward in our understanding of both domains and their broader implications.

WHAT COULD GO WRONG? LINGUISTIC ILLUSIONS AND
INCREMENTAL INTERPRETATION

by

Hanna Ellen Muller

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
2022

Advisory Committee:

Professor Colin Phillips, Chair
Associate Professor Robert Slevc, Dean's Representative
Associate Professor Ellen Lau
Associate Professor Alexander Williams
Professor Philip Resnik

© Copyright by
Hanna Ellen Muller
2022

Preface

The work reported in this dissertation is highly collaborative. Chapter 3 reports on joint work with Iria de Dios Flores and Colin Phillips. Chapter 4 reports on joint work with Iria de Dios Flores, Celeste Joly, and Colin Phillips. Chapter 5 reports on joint work with Iria de Dios Flores, Lalitha Balachandran, and Colin Phillips. Chapter 8 reports on joint work with Shohini Bhattasali, Philip Resnik, and Colin Phillips, and additionally includes analyses from a joint class project with Adam Liter, Masato Nakamura, and Nika Jurov. Chapter 9 reports on joint work with Shohini Bhattasali, Philip Resnik, and Colin Phillips. This work has also been supported by the National Science Foundation (Doctoral Dissertation Research Improvement grant #2141348 and NRT award #1449815) and by ONR MURI award N00014-18-1-2670.

Acknowledgments

This dissertation would not have been possible without the support and thoughtful contributions of many people. First of all, I'd like to thank Colin Phillips, who has been a wonderful mentor to me in many ways. I suspect it's obvious to anyone who reads the following pages that my conversations with Colin have substantially influenced my thinking about psycholinguistics. What may not be as clear, though, is how much fun those conversations have been¹. Colin has always been eager to spend an hour (or four, if he can) talking through a puzzle. Sometimes this leads to some satisfying clarity, and sometimes it leads to the realization that things are in fact much more confusing than either of us realized. But it is always time well spent. It sounds like such a small thing, when I put it on paper, to say that Colin and I have had some good chats over the years, when what I really mean is that it's through these conversations that I have learned what it is to be a scientist. As much as I love a good power analysis, it is this work, the work of thinking carefully through a mysterious pattern — and ideally, thinking out loud with someone who cares just as much about figuring it out — that feels to me like the whole point of what we're doing here. Thank you, Colin, for this and for being such a kind and supportive mentor.

The members of my committee, Bob Slevc, Alexander Williams, Philip Resnik, and Ellen Lau, have also contributed in many ways to this work. Bob has encouraged me to take a broader view of the work, and to think about production as well as comprehension, which I'm very thankful for. Alexander is always a wonderful person to talk to, both for his sense of humor and for his clear thinking. He is also always looking out for the grad students in this department, and we are lucky to have him. I am incredibly grateful for his contributions to the dissertation and for the various classes he has taught. Philip's enthusiasm about research is palpable in every meeting. Much of the work here would not have even begun if not for Philip, since our entire investigation of substitution illusions arose because it turned out to

¹Though perhaps this is clear as well, since Maša tells me she always knows when Colin and I have a meeting because she can hear us laughing from her office around the corner.

be at the intersection of my and Philip's interests. Philip has pushed me to be more open-minded about what research is for and more concrete about what we can do. Ellen has done so much that I am grateful for. I'm consistently astonished by her ability to make me see things that I have been thinking about every day in an entirely new way. I think what I admire most about her is the feeling of wonder and hope that she brings to her work. When Ellen explains how EEG works in her classes, she asks, again and again, *isn't that amazing?!* Hearing her talk about science makes me feel like we are all so incredibly lucky to get to participate in this quest to understand the mind.

I am grateful for the many other people who have taught me during my years at UMD, both through coursework and through our run-ins around the department — Peggy Antonisse, Naomi Feldman, Valentine Hacquard, Norbert Hornstein, Howard Lasnik, Jeff Lidz, and Andrea Zukowski. Thank you especially to Tonia Bleam, who has been my mentor in all things teaching-related. As I write this I realize that I only ever TAed with Tonia for one semester, which is strange given the huge influence she's had on me. That semester was the Spring of 2020, so perhaps it makes sense that it looms large in my memory. Teaching is hard, and teaching during a pandemic was especially hard, but Tonia made me believe it was still worth doing and worth doing well.

Thank you also to Kim Kwok, without whom we would all be lost. A visit to Kim's office is always a delight, both because of her almost superhuman ability to solve the problem that prompted the visit, and because she's just a wonderful person to chat with.

Thank you to the many post-bacs, undergraduates, and high school students who have contributed to this work. Much of the work of implementing the experiments presented here was carried out by amazing research assistants including Andrew Merchant, Jerry Peng, Angela Wu, Celeste Joly, and Maddy Keen. In addition to being incredibly bright and hardworking students, they were also a joy to mentor. Lab managers Maggie Kandel, Bethany Dickerson, Aura Cruz Heredia, and Lalitha Balachandran helped keep everything running and I'm very grateful for their contributions to this community.

I also want to thank all of the wonderful teachers I had before I came to Maryland, who helped me find this path. Itamar Kastner TAed the first linguistics class I ever took, and made it seem very cool. Lucas Champollion and Linmin Zhang hired me as a research assistant, and gave me the opportunity to

see what being a linguist would actually be like. I am so grateful for that opportunity. Dustin Chacón first introduced me to the kind of sentence processing question that this dissertation tries to answer. He is, in many ways, the reason I came to Maryland. Dylan Bumford was uncommonly generous with his time, and spent many hours talking with me about semantics and Haskell and, in one surprisingly rewarding detour, Taylor series. This was both a lot of fun and an amazing exercise in learning stuff because I wanted to know it, and not because it was on a syllabus. And finally, Stephanie Harves is the kind of educator I aspire to be. Both in her classroom and when I wrote my senior thesis, Stephanie was consistently encouraging and kind. Her faith in my ability to do research allowed me to consider the possibility that I might actually be able to do research.

Thinking about the past six years at Maryland, I feel incredibly lucky to have this job, to be a scientist and teacher. There isn't much more I could hope for than to spend my days indulging my own curiosity and encouraging curiosity in others. Perhaps I will be lucky enough to do this for the rest of my working life, but even if not, I'm glad to have done it. Jon likes to sum up this perspective as "we're here for a good time, not a long time". So thank you to all the people who made it such a good time.

First on that list is my fellow post-bacs, aka the Baggettstreet Boys: Jon Burnsky, Jackie Nelligan, Nancy Clarke, and Michaela Socolof. It's hard to sum up what a strange and exciting year it was, but I'm very grateful that these were the people I got to share it with. Thank you for making me go camping in a thunderstorm, for planning and executing several successful Beachwee(k)s, for re-inventing our names many times over², for making fun of my thumbs, for honing our cones, and so much other nonsense. Basically, thank you for the endless silliness, it was exactly what I needed. I am especially grateful that we have stayed in each other's lives as co-workers, cohort-mates, book club members, and friends. I can't wait to see what you all do next.

Thank you also to my cohort, Hisao Kurokami, Yu'an Yang, Adam Liter, and Jackie Nelligan (again). I'm incredibly lucky to have gotten to share this experience with such a talented and lovely group of people. That first year together was such a stressful and uncertain time in so many ways, but I have never felt alone. Thank you for listening when I needed to vent, and for sharing your frustrations with me too.

²My favorites are Jelligan Belligan Nelligan, Hoodlenoodle, and that unpronounceable string of IPA symbols we used for Jon for a while.

Thank you also for the many, many intellectually rewarding conversations we have had over the years. Thinking with you has always helped me think more clearly.

Thank you to the many wonderful grad students that I interacted with in one way or another over the years as classmates and collaborators and friends. First, thank you to all of the grad students who made me want to become a part of this department, especially Zoe Schlueter, Anton Malko, Allyson Ettinger, Lara Ehrenhofer, Nick Huang, Kasia Hitczenko, Max Papillon, Chia-Hsuan Liao, Phoebe Gaston, Paulina Lyskawa, Sigwan Thivierge, Tyler Knowlton, Aaron Doliana, and Mina Hirzel. Thank you all for setting an example for what a kind and supportive and extremely brilliant community of grad students looks like. Thank you also to all of the new friends who have joined me here, especially Masato Nakamura, Craig Thorburn, Alex Krauska, Maša Bešlin, Rosa Lee, Xinchu Yu, Luisa Seguin, Clara Cuonzo, London Dixon, Joselyn Rodriguez, and Katherine Howitt. You all make me wish I didn't have to leave.

I am also extremely grateful to have gotten to meet Iria de Dios Flores during my time at Maryland. Our work together has reliably made me feel excited to do research, and Iria has been a kind and supportive friend since we met six years ago. A number of amazing post-docs have also come through our department during my time here, and I'm especially grateful to know Shohini Bhattacharya, who has been a wonderful collaborator, and Ted Levin, who has continued to be the life of the party even in cardboard form. Thank you to the members of the wonderful working groups I've been a part of, including many of the people mentioned above but also Kathleen Oppenheimer in the amazing stats reading group, and Zach Maher and Tal Ness in the psycholinguistics check in group. Thank you especially to the Bias in Linguistics working group, which has been deeply fulfilling to be a part of.

A few of the grad students I met here have become especially important to me. Thank you to Katherine for repeatedly pulling me out of the dissertation dungeon this past year and for inviting me, the very first time we met, to split the cost of a helicopter to go to Gates of the Arctic National Park together. Thank you also for reminding me, whenever I try to give you unsolicited advice about grad school, that I'm actually "several" years younger than you.

Thank you to Jackie for being beside me in all this. Starting grad school with a friend made it much less scary. Thank you for teaching me how to lift weights, though I am still very bad at it. Thank you

for all the late nights in the first year office talking about grad school (sometimes) and feminism (often). And thank you to all of the people who have shared an office or an office wall with us, and put up with listening to our endless chatting.

Adam has been a wonderful friend and housemate over the last few years. He is also one of the most thoughtful people I have ever known, and whenever I can make myself shut up for long enough to hear what Adam is saying, I'm glad that I did. Thank you for teaching me how to use git. I still don't really use it, but at least I know that it's a directed acyclic graph.

Phoebe and I had already met once before I came to Maryland, when we overlapped at NYU. However, I discovered when I arrived here that Phoebe, tragically, did not remember me. This was somewhat devastating at the time, since I admired the hell out of her. And so our friendship didn't really begin until a while later, when I found myself crying in the department because of some setback that I now cannot remember, but that made me feel like I was absolutely not cut out for this, and Phoebe sent me an email telling me that she was there if I wanted to talk about it, that everybody cries in the department sometimes, and that everybody thinks they have to quit grad school sometimes. She also sent me a numbered list of reasons to believe it was all going to be okay. She has since become one of my dearest friends, and I continue to admire the hell out of her.

Thank you to the many people outside of the linguistics world who have made me feel like a whole person. Thank you to my non-linguist friends, especially Laura Bailey, Michelle Metty, and Taylor Schettino for simply being wonderful friends. Thank you to Linda Muller, Paul Muller, Kathrin Muller, and Darlene Maloney, for all of your love and support, and for giving me opportunities to think about something other than psycholinguistics. Thank you also to the Burnskys — Donna, PJ, Paul, Jaclyn, Colton, Chase, Addie, Ryan, Sarah, Reagan, Brady, and Claire — for welcoming me into your family. I have an amazing network of people and they have all contributed to this journey.

Finally, thank you to my wonderful little family, Calzone and Jon. Callie is, obviously, the sweetest and most perfect cat that has ever existed, and I've loved coming home to her purrs and screams every day for the past five years. Jon, you are my favorite collaborator, my best friend, and my partner in every way. Thank you for so many things, but especially for being the best person to lean on when things went

wrong and the first person to celebrate with me when things went right. Thank you for telling me that I can do this and that it's okay if I don't want to. Thank you for sharing your curiosity with me, and always being down to think through an experiment design over dinner, or wonder together about how bicycles stay upright. Thank you for growing with me through five years of grad school and five years of long distance. I'm so proud of us.

Table of Contents

Preface	ii
Acknowledgments	iii
Table of Contents	ix
List of Tables	xvii
List of Figures	xix
List of Abbreviations	xxi
Chapter 1 Introduction	1
1.1 What are linguistic illusions?	3
1.2 Strategies for explaining illusions	7
1.3 Methodological approaches to the study of linguistic illusions	10
1.4 Dissertation overview	13
Chapter 2 NPI illusions: overview	15
2.1 Illusions of grammaticality	16
2.1.1 Shallow processing	18
2.1.2 Cue-based retrieval	19
2.2 NPI licensing	25
2.2.1 Syntactic accounts	26
2.2.2 Pragmatic accounts	29
2.2.3 Truth conditional accounts	31
2.3 The current study	33
Chapter 3 NPI illusions: the role of licensors	35
3.1 Introduction	35
3.1.1 The grammar of NPIs	36
3.1.2 Previous Accounts of NPI illusions	39
3.1.3 The scalar alternatives hypothesis	42
3.1.4 The present investigation	46
3.2 Contrasting licensor types	47
3.2.1 Experiment 1: offline acceptability	47
3.2.1.1 Participants	48
3.2.1.2 Materials	48

	3.2.1.3	Procedure	49
	3.2.1.4	Analysis	49
	3.2.1.5	Results	50
	3.2.1.6	Discussion	51
3.2.2		Experiment 2: speeded acceptability	52
	3.2.2.1	Participants	52
	3.2.2.2	Materials	52
	3.2.2.3	Procedure	53
	3.2.2.4	Analysis	53
	3.2.2.5	Results	54
	3.2.2.6	Statistical power	54
	3.2.2.7	Discussion	55
3.2.3		Experiment 3: speeded acceptability	56
	3.2.3.1	Participants	56
	3.2.3.2	Materials	57
	3.2.3.3	Procedure	57
	3.2.3.4	Analysis	58
	3.2.3.5	Results	58
	3.2.3.6	Statistical power	58
	3.2.3.7	Discussion	59
3.3		The interpretation of NPI illusions	61
	3.3.1	Experiment 4: interpretation and acceptability	61
		3.3.1.1 Participants	63
		3.3.1.2 Materials	63
		3.3.1.3 Procedure	64
		3.3.1.4 Analysis	65
		3.3.1.5 Results	65
		3.3.1.6 Statistical power	67
		3.3.1.7 Discussion	68
3.4		Non-quantificational negation and scalar meanings	71
	3.4.1	Experiment 5: offline acceptability	72
		3.4.1.1 Participants	73
		3.4.1.2 Materials	73
		3.4.1.3 Procedure	74
		3.4.1.4 Analysis	74
		3.4.1.5 Results	74
		3.4.1.6 Discussion	75
	3.4.2	Experiment 6: speeded acceptability	75
		3.4.2.1 Participants	76
		3.4.2.2 Materials	76
		3.4.2.3 Procedure	76
		3.4.2.4 Analysis	77
		3.4.2.5 Results	77
		3.4.2.6 Statistical power	78
		3.4.2.7 Discussion	78

3.5	General discussion	82
3.5.1	Key findings	82
3.5.2	Effect sizes and statistical power	83
3.5.3	The scalar alternatives hypothesis	84
3.5.4	Alternative explanations for NPI illusions	86
3.5.5	Aligning parsing and grammar	89
3.6	Conclusion	90
Chapter 4	NPI illusions: the role of distance	91
4.1	Introduction	91
4.1.1	NPI licensing	92
4.1.2	Linguistic illusions	96
4.1.2.1	Early accounts of the NPI illusion	98
4.1.2.2	The licensor effect	99
4.1.2.3	The distance effect	101
4.1.3	The present study	103
4.2	Intervening prepositional phrases	104
4.2.1	Experiment 7: speeded acceptability	104
4.2.1.1	Participants	105
4.2.1.2	Materials	105
4.2.1.3	Procedure	106
4.2.1.4	Analysis	107
4.2.1.5	Results	107
4.2.1.6	Statistical power	108
4.2.1.7	Discussion	108
4.3	Intervening verbs	109
4.3.1	Experiment 8: offline acceptability	110
4.3.1.1	Participants	111
4.3.1.2	Materials	111
4.3.1.3	Procedure	112
4.3.1.4	Analysis	112
4.3.1.5	Results	113
4.3.1.6	Discussion	114
4.3.2	Experiment 9: speeded acceptability	115
4.3.2.1	Participants	115
4.3.2.2	Materials	115
4.3.2.3	Procedure	115
4.3.2.4	Analysis	116
4.3.2.5	Results	116
4.3.2.6	Discussion	117
4.4	The licensor effect	118
4.4.1	Experiment 10: speeded acceptability	120
4.4.1.1	Participants	121
4.4.1.2	Materials	121
4.4.1.3	Procedure	122

4.4.1.4	Analysis	122
4.4.1.5	Results	123
4.4.1.6	Discussion	124
4.5	The nature of timing effects and the role of NPI identity	124
4.5.1	Experiment 11: speeded acceptability	127
4.5.1.1	Participants	128
4.5.1.2	Materials	128
4.5.1.3	Procedure	128
4.5.1.4	Analysis	128
4.5.1.5	Results	129
4.5.1.6	Discussion	130
4.6	General discussion	132
4.6.1	Key findings	133
4.6.2	Implications for illusions	135
4.7	Conclusion	140
Chapter 5	NPI illusions: putting the pieces together	141
5.1	Introduction	141
5.1.1	Key findings	141
5.1.1.1	The licenser effect	141
5.1.1.2	The distance effect	142
5.1.1.3	Interpretation	143
5.1.1.4	PPI illusions	143
5.1.2	Theoretical landscape	144
5.1.2.1	Cue-based retrieval	144
5.1.2.2	Pragmatic rescuing	145
5.1.2.3	Scalar alternatives	146
5.1.2.4	Scope miscalculation	147
5.1.2.5	Noisy channel	148
5.1.3	The current study	152
5.2	Scalar alternatives	153
5.2.1	Experiment 12	154
5.2.1.1	Participants	155
5.2.1.2	Materials	156
5.2.1.3	Procedure	156
5.2.1.4	Analysis	157
5.2.1.5	Results	157
5.2.1.6	Discussion	158
5.3	Distance effects	159
5.3.1	Experiment 13	160
5.3.1.1	Participants	161
5.3.1.2	Materials	161
5.3.1.3	Procedure & Analysis	162
5.3.1.4	Results	162
5.3.1.5	Discussion	162

5.3.2	Experiment 14	164
5.3.2.1	Participants	165
5.3.2.2	Materials	165
5.3.2.3	Procedure & Analysis	166
5.3.2.4	Results	167
5.3.2.5	Discussion	167
5.4	Noisy channel	168
5.4.1	Experiment 15	169
5.4.1.1	Participants	172
5.4.1.2	Materials	173
5.4.1.3	Procedure & Analysis	174
5.4.1.4	Results	174
5.4.1.5	Discussion	174
5.4.2	Experiment 16	177
5.4.2.1	Participants	178
5.4.2.2	Materials	178
5.4.2.3	Procedure & Analysis	178
5.4.2.4	Results	179
5.4.2.5	Discussion	180
5.5	Conclusion	181
Chapter 6	NPI illusions: general discussion	183
6.1	Summary of findings	183
6.1.1	Experiments 1 & 2	183
6.1.2	Experiment 3	184
6.1.3	Experiment 4	184
6.1.4	Experiments 5 & 6	185
6.1.5	Experiment 7	186
6.1.6	Experiments 8 & 9	187
6.1.7	Experiment 10	188
6.1.8	Experiment 11	188
6.1.9	Experiment 12	189
6.1.10	Experiment 13	190
6.1.11	Experiment 14	191
6.1.12	Experiment 15	192
6.1.13	Experiment 16	193
6.2	Key generalizations and next steps	194
6.2.1	The licenser effect	195
6.2.2	The distance effect	201
6.2.3	NPI identity	204
6.2.4	Relative clause type	206
6.2.5	Interpretation	208
6.2.6	Timing	209
6.2.7	Tense	211
6.2.8	Summary	214

6.3	Proposed mechanisms	215
6.3.1	Cue-based retrieval	215
6.3.2	Pragmatic rescuing	217
6.3.3	Scalar alternatives	218
6.3.4	Scope miscalculation	220
6.3.5	Noisy channel	225
6.3.6	Local coherence	226
6.4	Towards a plausible mechanism for NPI licensing	226
6.4.1	Processing algorithms for licenser-based grammars of NPIs	227
6.4.2	Processing algorithms for environment-based grammars of NPIs	230
6.4.3	Post-NPI reanalysis	232
6.4.4	NPI illusions in other languages	234
6.5	Broader implications for grammatical illusions	235
Chapter 7	Substitution illusions: overview	237
7.1	Key factors and their influence on illusions	239
7.1.1	Similarity in meaning	241
7.1.2	Similarity in form	252
7.1.3	Presuppositional status	254
7.1.4	Questions versus statements	256
7.2	What could go wrong?	257
7.2.1	Lexical access	262
7.2.2	Identifying concepts at the (sub-)word level	267
7.2.3	Semantic composition	272
7.2.4	Matching sentence meanings to non-linguistic concepts	274
7.2.5	Summary	277
7.3	Other accounts of the illusion	278
7.3.1	Subconscious accommodation	278
7.3.2	Motivation and attention	281
7.3.3	Shallow processing	285
7.4	Other methodological approaches	290
7.4.1	Reading times	290
7.4.2	Neural measures	293
7.4.3	Second language processing	294
7.5	The current investigation	294
7.5.1	Item-wise variability	295
7.5.2	Manipulating ease of access	295
Chapter 8	Substitution illusions: item-wise variability	297
8.1	Experiment 17: variability in the illusion	298
8.1.1	Participants	298
8.1.2	Materials	298
8.1.3	Procedure	298
8.1.4	Analysis	299
8.1.5	Results	301

8.1.6	Discussion	302
8.2	Experiment 18: familiarity	303
8.2.1	Participants	304
8.2.2	Materials	304
8.2.3	Procedure	304
8.2.4	Analysis	305
8.2.5	Results	305
8.2.6	Discussion	305
8.3	Experiment 19: lexical similarity	307
8.3.1	Computational measures of word similarity	308
8.3.2	Correlation with illusion rates	310
8.3.3	Discussion	310
8.4	Experiment 20: language model predictions	312
8.4.1	Computational language models	313
8.4.2	Knowledge check	314
8.4.3	LSTM language model predictions	314
8.4.4	GPT-2 language model predictions	316
8.4.5	Discussion	317
8.5	Experiment 21: cloze	318
8.5.1	Participants	318
8.5.2	Materials	318
8.5.3	Procedure	318
8.5.4	Analysis	319
8.5.5	Results	319
8.5.6	Discussion	321
8.6	Conclusion	322
Chapter 9	Substitution illusions: levels of representation	324
9.1	Experiment 22: declaratives	327
9.1.1	Participants	327
9.1.2	Materials	327
9.1.3	Procedure	327
9.1.4	Analysis	328
9.1.5	Results	328
9.1.6	Discussion	329
9.2	Experiment 23: priming through lexical decision	331
9.2.1	Participants	332
9.2.2	Materials	332
9.2.3	Procedure	335
9.2.4	Analysis	336
9.2.5	Results	337
9.2.6	Discussion	337
9.3	Conclusion	343
Chapter 10	Substitution illusions: general discussion	344

10.1	Summary of findings	344
10.1.1	Experiment 17	344
10.1.2	Experiment 18	345
10.1.3	Experiment 19	345
10.1.4	Experiment 20	345
10.1.5	Experiment 21	346
10.1.6	Experiment 22	346
10.1.7	Experiment 23	347
10.2	Updated theoretical landscape	347
10.3	Broader implications	349
10.3.1	Item-wise variability	350
10.3.2	Prediction	354
Chapter 11	Conclusion	356
11.1	Summary of this dissertation	356
11.2	Conclusions	357
11.3	Next steps	360
Appendix A	Experimental materials	363
A.1	Experiment 1	363
A.2	Experiment 2	367
A.3	Experiment 3	372
A.4	Experiment 4	380
A.5	Experiment 5	398
A.6	Experiment 6	410
A.7	Experiment 7	422
A.8	Experiment 8	431
A.9	Experiment 9	442
A.10	Experiment 10	453
A.11	Experiment 11	462
A.12	Experiment 12	476
A.13	Experiment 13	491
A.14	Experiment 14	499
A.15	Experiment 15	540
A.16	Experiment 16	547
A.17	Experiment 17	552
A.18	Experiment 18	556
A.19	Experiment 21	558
A.20	Experiment 22	560
A.21	Experiment 23	571
Bibliography	577

List of Tables

3.1	Example stimuli for Experiment 1 and Experiment 2	49
3.2	Example stimuli for Experiment 3	57
3.3	Example stimuli for Experiment 4	64
3.4	Example stimuli for Experiment 5 and Experiment 6	73
4.1	Example stimuli for Experiment 7	106
4.2	Example stimuli for Experiment 8 and Experiment 9	112
4.3	Example stimuli for Experiment 10	122
4.4	Example stimuli for Experiment 11	129
5.1	Example stimuli for Experiment 12	156
5.2	Example stimuli for Experiment 13	161
5.3	Example stimuli for Experiment 14	166
5.4	Example stimuli for Experiment 15	173
5.5	Example stimuli for Experiment 16	179
9.1	Example stimuli for Experiment 23	335
10.1	Example stimuli for NPI illusions and substitution illusions	352
A.1	Full experimental stimuli for Experiment 1	367
A.2	Full experimental stimuli for Experiment 2	372
A.3	Full experimental stimuli for Experiment 3	380
A.4	Full experimental stimuli for Experiment 4	398
A.5	Full experimental stimuli for Experiment 5	410
A.6	Full experimental stimuli for Experiment 6	421
A.7	Full experimental stimuli for Experiment 7	431
A.8	Full experimental stimuli for Experiment 8	442
A.9	Full experimental stimuli for Experiment 9	453
A.10	Full experimental stimuli for Experiment 10	462
A.11	Full experimental stimuli for Experiment 11	476
A.12	Full experimental stimuli for Experiment 12	491
A.13	Full experimental stimuli for Experiment 13	499
A.14	Full experimental stimuli for Experiment 14	540
A.15	Full experimental stimuli for Experiment 15	547
A.16	Full experimental stimuli for Experiment 16	552
A.17	Full experimental stimuli for Experiment 17	556
A.18	Full experimental stimuli for Experiment 18	558
A.19	Full experimental stimuli for Experiment 21	560

A.20	Full experimental stimuli for Experiment 22	571
A.21	Full experimental stimuli for Experiment 23	576

List of Figures

3.1	Experiment 1 results	50
3.2	Experiment 2 results	54
3.3	Experiment 3 results	59
3.4	Experiment 4 results — acceptability	66
3.5	Experiment 4 results — interpretation	67
3.6	Experiment 4 results — acceptability, split by interpretation	68
3.7	Experiment 5 results	75
3.8	Experiment 6 results	78
3.9	Experiment 6 results, split by presentation order	81
3.10	Chapter 3 findings, summarised	84
4.1	Experiment 7 results	108
4.2	Experiment 8 results	114
4.3	Experiment 9 results	117
4.4	Experiment 10 results	123
4.5	Experiment 11 results	131
5.1	Experiment 12 results	158
5.2	Experiment 13 results	163
5.3	Experiment 14 results	168
5.4	Experiment 15 results	175
5.5	Experiment 16 results	180
6.1	NPI illusion rates for 13 experiments	195
6.2	Acceptance rates NPI illusion sentences with embedded <i>no</i> versus <i>not</i> for 8 experiments	196
6.3	NPI illusion effect of distance from the licenser	202
6.4	NPI illusion effect of distance from the RC	203
6.5	NPI illusion effect of end-of-sentence position	204
6.6	Effect of distance manipulations on all NPI illusion sentences and their baselines	205
6.7	Experiment 16 results, split by previous trial type	213
7.1	Substitution illusion rates for similar and dissimilar impostors, across studies	250
7.2	The lemma model	263
7.3	The Zöllner illusion	283
8.1	Experiment 17 results	301
8.2	Experiment 17 simulated results	302
8.3	Experiment 18 results	305
8.4	Substitution illusion correlation with familiarity	306

8.5	Semantic similarity in substitution illusions and word2vec	310
8.6	Substitution illusion correlation with similarity	311
8.7	Substitution illusion correlation with position	313
8.8	Substitution illusion correlation with LSTM entropy for filtered data	315
8.9	Substitution illusion correlation with LSTM entropy	315
8.10	Substitution illusion correlation with GPT-2 entropy for filtered data	316
8.11	Substitution illusion correlation with GPT-2 entropy	317
8.12	Experiment 21 results — cloze question completions	320
8.13	Substitution illusion correlation with cloze question completions	320
8.14	Experiment 21 results — cloze question answers	321
8.15	Substitution illusion correlation with cloze question answers	321
8.16	Substitution illusion correlation with all predictors	322
9.1	Experiment 22 results	329
9.2	Substitution illusion rates for questions and declaratives	330
9.3	Experiment 23 results	337
9.4	Experiment 23 results, only recent primes	341
10.1	Comparison of itemwise variability across illusions	351

List of Abbreviations

DE	Downward Entailing
ERP	Event-Related Potential
FCI	Free Choice Item
LTM	Long-Term Memory
MC	main clause
NPI	Negative Polarity Item
ORC	object relative clause
PP	prepositional phrase
PPI	Positive Polarity Item
RC	relative clause
RSVP	rapid serial visual presentation
RT	reaction time
SRC	subject relative clause

Chapter 1 Introduction

Adult native speakers are very good at understanding sentences of their language. To accomplish this, even for very simple sentences, a comprehender makes use of multiple cognitive skills to accomplish critical sub-tasks like encoding linguistic units, constructing dependencies between temporally-distant units, generating expectations, and relating the content of an expression to non-linguistic representations. All of these skills must be deployed in order to efficiently generate a mapping from the perceptual input to a likely speaker meaning. In most comprehension situations, these operations are expertly coordinated, resulting in a representation that is both consistent with the signal and interpretable. The occasional failure of the incremental comprehension system to accurately represent a sentence in a way that aligns with the the comprehender’s knowledge (i.e., “linguistic illusions”) can therefore be a window into the nature of these properties. That is, only a few of the imagineable algorithms for implementing linguistic knowledge will predict misalignment in precisely the circumstances where misalignment arises.

Negative polarity illusions concern the licensing of Negative Polarity Items (NPIs), or words and phrases like *ever*, *any*, *in months/years*, and *lift a finger*, which can only occur in a restricted set of environments. One such environment is within the scope of negation: (1a) is acceptable but (1b), which lacks negation, is unacceptable. We will refer to the NPI *in months* in (1a) as being “licensed” by negation, though this terminology is not intended to imply a particular analysis.

- (1) a. We haven’t left the house in months.
- b. * We have left the house in months.

Prior work has shown that the unacceptability of some sentences with unlicensed NPIs like (2a) is not as immediately apparent as the unacceptability of similar sentences with NPIs like (2b) (though after careful reflection, comprehenders typically conclude that both are unacceptable). This is shown in

various tasks, including speeded judgments of acceptability, reading times, and event related potentials. This initial perception of acceptability for sentences like (2a) is the NPI illusion, first demonstrated by Drenhaus, Saddy, & Frisch 2005 in German, and replicated many times.

- (2) a. * The authors [that no critics have recommended for the award] have ever received acknowledgement for a best-selling novel.
- b. * The authors [that the critics have recommended for the award] have ever received acknowledgement for a best-selling novel.

This pattern raises the question of what the processing of an NPI consists of, such that licensed NPIs can be quickly integrated, and (the majority of) unacceptable NPIs can be quickly detected, but (2a) appears to be acceptable. The phenomenon therefore has the potential to inform our theories about how the incremental sentence comprehension system implements grammatical knowledge in the service of understanding an incoming sentence. In particular, because of the rich literature on the nature of this grammatical knowledge, processing theories can start from a relatively sophisticated level of detail. In our investigation of NPI illusions, we take the critical question to be, what does the process for incorporating an NPI into an in-progress sentence representation look like, such that it yields the immediate impression of unacceptability for sentences like (2b) but not sentences like (2a). This phenomenon is explored in depth in Chapters 2-6.

Chapters 7-10 focus on misalignment with a different kind of knowledge: world knowledge. Comprehenders are able to detect many world knowledge violations in sentences quickly and effortlessly. In “substitution illusions” (sometimes called “Moses illusions”), the detection never occurs at all. For example, the world knowledge violation in (3a) is quickly identified by most comprehenders who read it — or, at least, most comprehenders who actually know that van Gogh cut off his ear not his eye. In contrast, the world knowledge violation in (3b) goes entirely unnoticed for many readers — even those who know that it is blindness, not deafness, that creates the need for Braille. The effect was first demonstrated (using different sentences) by Erickson & Mattson 1981, and has been replicated many times.

- (3) a. How did painter Vincent van Gogh lose his eye during his life?

- b. What is the name of the raised bumps on paper that enable deaf people to read?

It is in principle possible that the substitution illusion is a product of purely non-linguistic systems. For example, if the knowledge that van Gogh lost his ear, or that Braille is used by the blind, cannot be retrieved from memory, comprehenders will fail to detect the world knowledge violation, even if their linguistic representation of the sentence is perfectly accurate. While we cannot necessarily rule out such a possibility, here we focus on the alternative hypothesis that the failed detection of anomalies like (3b) is a consequence of the systems used by a comprehender to generate a sentence-level meaning based on both the stimulus and the comprehender's knowledge. We take the central question to be, what is the process for understanding a sentence, such that the world knowledge violation of (3a) is salient in the resulting representation, but the world knowledge violation of (3b) is not.

Before exploring the details of these phenomena and their possible causes, some clarification regarding what is and is not a linguistic illusion is warranted. This is discussed in section 1.1. We then turn to the possible approaches to explaining illusions in section 1.2. Section 1.3 is an overview of methodological approaches to studying linguistic illusions. And finally, section 1.4 provides a summary of the rest of this dissertation.

1.1 What are linguistic illusions?

We have assumed that NPI illusions and substitution illusions have something in common, at least descriptively — they are illusions. Some brief discussion of what illusions are and are not is therefore valuable. We are interested in linguistic illusions at the sentence level. While illusions also occur in the auditory perception of speech input — for example, the well-known McGurk effect (McGurk & MacDonald 1976, cf. Getz & Toscano 2021), or the perception of epenthetic vowels in consonant clusters that don't occur in one's native language (e.g., Dupoux et al. 1999) — these are beyond the scope of the present work. We will focus exclusively on illusions that might plausibly be attributed to the sentence processing system.

One possible definition of such illusions, which we do not endorse, is that linguistic illusions are

cases where a comprehender gives an initial judgment of a sentence’s acceptability or truth value which misaligns with their slower, more carefully considered judgment. One obvious problem with this definition is that people can change their judgments for a wide variety of reasons, some of which we would be hesitant to call illusions — for example, a native speaker informant might initially judge a sentence acceptable based on their own internal representation of it, but discover upon reflection that the sentence violates a prescriptive rule of the language, and give a different judgment. Defining illusions in a purely data-driven way (i.e., what people say about the sentence) risks treating this situation as equivalent. Relatedly, it is not necessary that a comprehender’s initial perception of the sentence be measured with an explicit acceptability judgment. As we will see in section 1.3, various implicit measures like reading times and ERPs can demonstrate that the early stages of processing for illusion sentences is unlike the early stages of processing for similar baseline sentences in which illusions do not occur.

Thus, we might refine this definition to say that illusions are cases where the comprehender’s initial mental representation of a sentence (whether they state a judgment about that representation or not) misaligns with their slower, more carefully considered judgment. However, this also will not do, as slow judgments are not always a good “gold standard” — some illusions are quite persistent, and even after several minutes of consideration, comprehenders may continue to judge the sentence in a way that reflects the illusion. For example, both comparative illusions (first noted by Montalbetti 1984, and explored by O’Connor 2015; Wellwood et al. 2018) as in (4) and so-called “depth charge” sentences (Wason & Reich 1979; Paape, Vasishth, & von der Malsburg 2020) as in (5) may be judged acceptable or judged to have the illusory meaning, even at quite slow response times.

(4) More people have been to Russia than I have.

(Wellwood et al. 2018:543)

(5) No head injury is too trivial to be ignored.

(Paape, Vasishth, & von der Malsburg 2020:509)

Similarly with NPI illusions, we sometimes find that illusion sentences are judged to be more acceptable than similar ungrammatical baseline sentences, even in untimed tasks. How, then, do we know that

they aren't simply grammatical sentences? Ultimately, the motivation to call such sentences “illusions” is at least in part theoretically motivated — no existing account of the grammar of NPIs predicts that such sentences are grammatical. Moreover, any grammatical theory that would treat illusion sentences as grammatical would be fairly inelegant, since NPI illusion sentences challenge the overwhelming generalization that an NPI cannot be licensed by a licenser that does not take scope over it. We would also need to somehow explain why the kinds of sentences we have so far called illusions (but which would just be grammatical NPI-containing sentences under such an account) are so much less acceptable than other grammatical NPI-containing sentences. That is, since the acceptability of such sentences is intermediate, one can either call them ungrammatical and explain the boosted acceptability through a processing theory or call them grammatical and explain the depressed acceptability through a processing theory.³ The latter is essentially the analysis typically given for multiply center-embedded sentences and garden paths, which we will return to below. But for NPI illusions, and the other illusion phenomena discussed here, we will take for granted that the theories that label these sentences ungrammatical are correct, and seek a processing-based explanation for their initial (possibly fleeting, possibly persistent) acceptability. Note that although this discussion has focused on the grammatical status of NPI-containing sentences, similar arguments can be made about the meanings of illusion sentences like (4) or (5).

What this means for our definition is that illusions are cases where the comprehender's initial mental representation of a sentence (whether they state a judgment about that representation or not) misaligns with the representation that we expect that sentence to be assigned based on the comprehender's knowledge (whether they eventually arrive at that representation or not). One important note about this revised definition is that it is compatible with but does not require treatment of illusions as a kind of “mis-perception” — that is, the comprehender might have a mental representation of the sentence which would be a perfectly appropriate, grammatically-sanctioned representation *of a different sentence*. This is what we might expect by analogy to (a non-technical understanding of) visual illusions — it's as if you're “seeing something that isn't there”. Such an analysis is related to hypotheses in the noisy channel framework, which we discuss in detail in Chapter 5 and Chapter 6. It is also related to an intuitive analysis

³Or consider grammars that are themselves gradient. We will not pursue such approaches here.

of the substitution illusion — we might assume that comprehenders’ mental representation of (3b) is just as if the sentence had said *blind* instead of *deaf*. Importantly, however, these are possible analyses of illusion effects, not definitions of the illusion. A misalignment between an initial representation and the representation that we expect that sentence to be assigned based on the comprehender’s knowledge could just as well be due to the inaccuracy in the deployment of that knowledge. Put simply, it’s possible that the comprehender knows exactly what was said, they just don’t know that it’s ungrammatical (yet). Which variety of analysis is appropriate is, for the most part, an open question, and may have different answers for different illusions.

One further note about our working definition of illusions is that in virtue of its focus on mere misalignment in representations, both the initial impression that an anomalous sentence is appropriate, and the initial impression that an appropriate sentence is anomalous would be included. This is somewhat non-standard in that it categorizes cases like the unacceptability of multiply center-embedded sentences like (6) and the initial unacceptability of garden path sentences like (7) as illusions. This is may not be a deep problem, since the spirit of the explanation offered for such misalignments is similar to what we pursue in more standard illusions — that is, the comprehender attempts to assign a representation to the input that is faithful to both the string and the grammar, but, for some reason likely having to do with the other cognitive systems with which the grammar interacts, fails.

- (6) The prize that the ring that the jeweler that the man that she liked visited made won was given at the fair.

(Miller & Isard 1964:296)

- (7) The horse raced past the barn fell.

(Bever 1970)

The generality of our definition does, however, raise the question of whether linguistic illusions are in any sense a natural class. That is, when we describe some sentences as illusory and others as non-illusory, are we carving nature at its joints? While we cannot give a definitive answer to this question, our intuitions are that we are not. Illusions are a useful category for practical purposes — the comprehension system’s

selective fallibility provides hints to its underlying nature — but we do not expect there to be mechanistic alignment across all illusory phenomena. This is somewhat obvious from just the two illusions explored here: whatever is going wrong in NPI illusions doesn’t seem to be the same as whatever is going wrong in substitution illusions, unless both phenomena are described at a very high level which abstracts away from mechanisms (i.e. “both phenomena involve a representation that has much in common with a non-anomalous representation, and the non-anomalous aspects seem to be missed”). Furthermore, we assume that the processing of illusion sentences involves fundamentally the same mechanisms that are involved in the comprehension of non-illusory sentences, as we discuss in the next section.

1.2 Strategies for explaining illusions

There are essentially two categories of claims made about the source of misalignments between grammatically-sanctioned representations and initial representations: either the initial representation is built by a system that does not make use of the grammar, or the initial representation is built by a system that uses the grammar in conjunction with other (fallible) cognitive systems. By analogy, let us consider some strategies one might adopt to solve a difficult division problem, such as $1323 \div 49$. Mathematically proficient adults know what a solution to such a problem would have to consist of, and likely know at least one explicit algorithm for solving such a problem with pen and paper: long division, but possibly other algorithms, too, like the “area method”⁴. They likely also know some “shortcuts” that will result in solutions that may not be technically correct, but which are close to the correct answer. For example, one might round both numbers, changing the problem into $1300 \div 50$, and arriving at an answer of “approximately 26”.

⁴The area method for division involves conceptualizing a division problem as a problem in which one knows the area of a rectangle and the length of one side of that rectangle, and must determine the length of the other side. This is done by breaking up the unknown side into sub-parts. For example, to solve $1323 \div 49$, we might start by saying the unknown side has a sub-part of length 10 (since $10 \times 49 = 490$, and 490 is less than 1323). The solver can choose any sub-part size to begin with, as long as it doesn’t make the rectangle’s area too big (i.e., bigger than 1323). The remaining area to be accounted for is then $1323 - 490 = 833$. Since 833 is greater than 490 we can assume another subpart of length 10, getting the to-be-accounted-for area down to 343. Now we might draw a subpart of length 5, thus accounting for another $5 \times 49 = 245$ units of area. And finally a subpart of length 2 to account for the last 98 units of area. At the end we add up the subparts of the unknown side, $10 + 10 + 5 + 2$, resulting in the answer to our division problem, 27. The point of this detour is to make explicit for readers who know *only* the long division algorithm that in math, as in the science of sentence processing, there is more than one possible algorithm that can implement the same knowledge. This is effectively the same point made by Marr regarding vision (Marr 1982).

A variety of sentence processing hypotheses have argued, on the basis of comprehender’s occasional failure to represent sentences in a way that aligns with their grammar, that they must be using a system that does not make use of the grammar, at least as a first pass. (That is, based on the math student giving the answer “26”, we assume that they did not solve the problem with long division or the area method or any such algorithm; rather they must have used a shortcut like rounding.) Proposals in this category include the LAST (Late Assignment of Syntax Theory; Townsend & Bever 2001), the “good enough” parsing view (Ferreira, Bailey, & Ferraro 2002)⁵, and various other “shallow processing” claims (e.g., Sanford & Sturt 2002). We will refer to these as “two system” views. To apply such an explanation to NPI illusions, one might argue that the NPI illusion arises because the “first pass” system does not make use of the grammar of NPIs, but uses some shortcut to determine acceptability — for example, perhaps it checks only for the existence of a negative morpheme in the prior string.⁶ This will result in acceptance of illusion sentences like (8) on every trial on which a judgment is given based on only the output of the first pass system.

- (8) * The authors [that no critics have recommended for the award] have ever received acknowledgement for a best-selling novel.

However, such explanations are not the only possible way to make sense of illusions. Extending the division metaphor a bit further, a student might sometimes get a problem wrong even if they know exactly how to do long division. They could have forgotten the first digit of the solution by the time they’ve computed the last one, or mis-retrieved a product from their memorized times tables, or simply copied the question down wrong from the blackboard. Similarly, even when deploying one’s full linguistic knowledge to understand an incoming sentence, problems can arise in the coordination and implementation of the many cognitive systems that are needed for this.

Substantial prior work has demonstrated how the properties of the memory system, and the errors in memory retrieval that are predicted by these properties, can cause linguistic illusions. Early work on

⁵Good enough parsing is not always clearly intended to be such a mechanistic claim. Rather, the hypothesis is sometimes presented as merely a descriptive generalization — that is, people make errors in comprehension — without strong commitments to the underlying cause of that generalization.

⁶The findings presented in Chapter 3, Chapter 4, and Chapter 5 make it clear that this particular hypothesized shortcut for NPI-containing sentences is not tenable, but that is somewhat irrelevant to the current discussion.

both NPI illusions (Vasishth et al. 2008) and agreement attraction (Wagers, Lau, & Phillips 2009) as in (9) blamed this system for the initial perception of acceptability. This type of explanation is explored in greater detail in Chapter 2, but the key idea is that subject-verb agreement (or whichever dependency we’re looking at) is conceptualized as a memory retrieval operation: at the verb, the subject must be identified from among the prior chunks in memory, and related to the verb. Independently motivated properties of the memory architecture and its interaction with linguistic representations (see Lewis & Vasishth 2005) predict that in cases where chunks in memory partially match the search cues, retrieval can still succeed. With some added linking assumptions about how successful retrieval leads to the perception of acceptability, we can then account for (some types of) illusion effects.

(9) * The key to the cells unsurprisingly were rusty from many years of disuse.

(Wagers, Lau, & Phillips 2009:16)

One might worry that while errors in non-linguistic cognitive systems like memory can explain the existence of grammar-judgment misalignments, only a two-system view predicts that illusions are fleeting. In other words, if you’re working with only one system, that system can be flawed but it’s not obvious how it can be both flawed and able to correct its own flaws. A two system view, in which the non-grammatically-bound system can be executed more quickly than the grammatically-bound one, naturally predicts a temporal profile for illusions. We have already noted that not all illusions exhibit this temporal profile. Moreover, modeling evidence from Parker 2019 shows that one-system models in which errors are attributed to the memory architecture can also account for the temporal profile of illusions. However, relatively little is known about the actual reanalysis processes that bring a comprehender from their initial illusion to their ultimate judgment in the cases where illusions are fleeting.

Thus both varieties of explanation are possible: illusions may arise because the system that builds initial representations does not make use of the knowledge that renders the sentence anomalous, or illusions may arise because the systems that implement that knowledge are fallible in specific ways. Among hypotheses in the second category, the fallibility of memory retrievals is the only specific hypothesis that has been seriously considered. If we adopt a broad view of illusions, which includes difficulty in the pro-

cessing of grammatical sentences like center embedding, then we might say memory capacity, in addition to memory retrieval, has been implicated. These are not the only options in this category.

Although both types of explanation can in principle capture illusion data adequately, in what follows we focus primarily on one-system views. That is, we assume that grammatical knowledge and other types of knowledge are available to the comprehender in real time as a sentence is being understood, and we ask what it is about the algorithms for deploying this knowledge that leads to occasional failure. This approach is to some extent theoretically motivated — all else equal, a one-system view is more parsimonious than a two-system view. However, as we will see, there are also empirical reasons to disprefer two-system views. This evidence comes primarily from the specificity of linguistic illusions. Based on only a sentence like (8) above, it may seem reasonable to postulate a shortcut or heuristic for NPI processing like “accept if there’s a negative word in the sentence, reject otherwise.” But as we demonstrate in Chapters 3, 4, and 5, many sentences for which such a heuristic would yield illusions are not, in fact, illusory. In order to capture the actual profile of the illusion, one would need a heuristic more like “accept if there’s a negative quantifier whose scope ended within two words of the NPI or if there’s any licenser that scopes over the NPI, reject otherwise” which is somewhat bizarre in its specificity. Similarly, a first attempt at a heuristic for substitution illusions like “check for conceptual relatedness of the content words” predicts illusions where none exist. We aim to identify processing algorithms that yield illusions in all and only the circumstances where illusions actually arise, and many (though not necessarily all) two-system “shallow processing” accounts are too general for this.

1.3 Methodological approaches to the study of linguistic illusions

We have defined illusions as cases where the comprehender’s initial mental representation of a sentence misaligns with the representation that we expect that sentence to be assigned, based on the comprehender’s knowledge. In the cases we explore here, the illusion sentences are always anomalous, and the initial representation misaligns in that the anomaly does not seem to be apparent. There are essentially two ways to measure whether this misalignment has occurred: asking comprehenders to give explicit judgments of

whether the sentence is anomalous, or measuring whether the sentence gives rise to the processing disruptions that are typical for that type of anomaly. Methods in the first group include judgments of acceptability, meaning, and truth value, and can be speeded or not speeded. Methods in the second group include reading time measurements and Event-Related Potentials (ERPs).

Which of these methods is appropriate obviously depends on the question being asked. In the experiments presented here we are primarily concerned with identifying the profile of the illusion — that is, what are the circumstances under which illusions arise and do not arise. By identifying this profile, we can narrow in on the possible mechanisms that drive the illusion. Part of our goal, then, will be to measure whether illusions occur for a variety of sentence types. Because the processing of an unlicensed NPI is associated with slowdowns in reading and, to a lesser extent, ERP signatures⁷, we could in principle go about identifying the profile of the illusion with one of these methods, asking, for each sentence type of interest, whether the typical error response associated with an unlicensed NPI is reduced.

Instead, we use explicit judgments in almost all studies reported here. There are three key reasons for this. First, acceptability judgments can easily be collected over the internet using services like Amazon Mechanical Turk. This is critical for data collection during a pandemic, for accessing more diverse and representative participant populations than college students, and for studying a number of manipulations fairly quickly (since data collection typically takes only a day or two). Second, binary judgments of acceptability or truth value provide a better signal-to-noise ratio than either of the other methods available. This is critical if one aims to determine, with reasonably good statistical power and reasonably few participants, whether illusions arise for a large number of manipulations. And finally, the linking assumptions underlying acceptability and truth value judgments are somewhat simpler than those underlying reading time and ERP measures.

This of course does not mean that the linking assumptions for explicit judgments of acceptability or truth are *simple*. While it is reasonable to go about determining whether a sentence is initially represented as if it is acceptable by asking “do you find this acceptable? Answer quickly,” when one considers the underlying computations that go into providing an acceptability judgment, it is clearly not

⁷The same is true for world-knowledge-violating lexical substitutions, making this discussion just as relevant to substitution illusions

trivial. A caricature of a mechanism we might sometimes assume is that the comprehender simply feeds the sentence into some internal grammar machine and reads off the answer that it spits out: GRAMMATICAL or UNGRAMMATICAL. Then add some noise to that and we get acceptability. This is wrong for many reasons, not least of which being the fact that grammars don't exist for the purposes of spitting out grammaticality judgments.⁸ But in many discussions of the online licensing of NPIs, something not too different from this seems to be assumed. That is, we often ask questions like "how does the comprehender determine whether the NPI is licensed?" This is a rather strange question in a world where virtually all naturally-occurring NPIs *are* licensed, and knowing that this is so is hardly ever the comprehender's actual goal. It seems more reasonable to assume that there is a process for mapping an NPI-containing sentence onto a structure, and ultimately a meaning, which happens to result in consciously-detectable processing difficulty in cases where an NPI is not licensed (that is, not in a grammatically-generated position). In illusions, then, we need not assume that the sentence was fed into the grammar machine and the answer GRAMMATICAL was mysteriously spat out. Rather, we have a case where the error signal which is usually triggered in the processing of an unlicensed NPI is not triggered or is not as strong. We assume, then that the judgment of acceptability involves evaluation of the error signals that arose over the course of comprehending the sentence, and, through some decision-making procedure, assessing whether those error signals are fatal.

One downside of explicit judgments of acceptability or truth value is that, being sentence-final, they're rather late measures if we wish to tap into comprehenders' initial, incrementally-generated representations. For both NPI illusions and substitution illusions, the anomalous word is typically many words before the end of the sentence, meaning there is much time for re-analysis before the judgment is given. There is a risk, then, that some illusions go undetected by us due to the lateness of our measure. As we have previously noted, relatively little is known about the re-analysis processes that allow comprehenders to recover from illusions. This is an important area for future research.

A second potential issue is that the use of binary judgments of acceptability or truth value may give

⁸This is a bit like saying the visual system exists to tell you whether the lights are on or off. Surely we can use our visual system to do this, and to do it quite well, but the main thing we like to do with our visual system is, when the lights are on, see stuff. Similarly we might say that one of the main things we like to do with our grammar is, when a sentence is grammatical, understand it. The other main thing we like to do with a grammar is, of course, say stuff.

researchers the impression of stochastic mental outcomes where none actually exist. In a binary acceptability judgment task, participants can, obviously, only make one of two choices (good or bad) on each individual trial. Thus of course does not imply that each individual trial the comprehender had either a perfectly-good or perfectly-bad impression of the sentence. Any observed proportion of “good” responses in a binary judgment task is perfectly consistent with a situation in which every single trial has some representation of intermediate acceptability, and noise determines which side of the decision criterion each trial happens to fall on. We in fact do not know whether illusions involve intermediate acceptability on all trials or full acceptability on some trials and full unacceptability on others, and the answer could be different for different illusions.

1.4 Dissertation overview

This dissertation is primarily concerned with furthering our understanding of the causes of two linguistic illusions. Accordingly, it is organized into roughly two parts. Chapters 2 through 6 focus on NPI illusions and Chapters 7 through 10 focus on substitution illusions. In each part, we begin with an overview of the existing literature and key proposals. We present three chapters on NPI illusions and two chapters on substitution illusions in which novel experimental results concerning the profile of the illusion are described. Each part concludes with a discussion of what has been learned. Finally, Chapter 11 summarizes the key contributions of this dissertation.

The key contributions of this dissertation are as follows. Concerning NPI illusions, we report two empirical generalizations which had not previously been reported, which we label the “licensor effect” and the “distance effect”. These generalizations jointly demonstrate that the NPI is a highly specific phenomenon. We additionally present the first investigation of the interpretation that results from an NPI illusion. These three areas of empirical progress allow us to make theoretical progress in our understanding of the mechanisms underlying the illusion. We can confidently rule out a number of proposals based on their inability to capture the licensor effect and the distance effect, including, notably, explanations that attribute NPI illusions and a wide range of other illusions to properties of the memory architecture.

We additionally explore a novel explanation for the NPI illusion, which highlights the incremental generation of scalar alternatives. This hypothesis achieves better (though not perfect) empirical coverage. Our exploration of substitution illusions consists of three main contributions. First, a careful exploration of the cognitive operations that would need to proceed flawlessly in order for substitutions to be detected (i.e. in order for illusions to not occur), provides a set of candidate error points. This gives us a better understanding of both the possible mechanisms underlying the illusion and the ways these hypotheses relate to one another. Empirically, we make progress on two fronts: first, the discovery and exploration of item-wise variability provides a new strategy for exploring this relatively old phenomenon. We effectively import the logic that has yielded much progress in NPI illusion research — that is, reasoning from the specificity of the illusion — into substitution illusion research, and leverage the variability that already exists. Finally, we take the first steps towards testing the theoretical landscape we proposed with an experiment that tests a particular mechanism as the cause of the illusion. In sum, the research presented in this dissertation substantially clarifies the empirical picture for both of these two phenomena, and offers novel hypotheses about the computations that are executed in the process of understanding a sentence more generally.

Chapter 2 NPI illusions: overview

Negative polarity illusions (“NPI illusions”) concern the licensing of NPIs, items like *ever*, *any*, *in months/years*, and *lift a finger*, which can only occur in a restricted set of environments. One such environment is within the scope of a negative operator: (10a) is acceptable but (10b), which lacks negation, is unacceptable.

- (10) a. We haven’t left the house in months.
 b. * We have left the house in months.

Much research has attempted to characterize the natural class of environments in which an NPI can naturally be used. This literature is reviewed in section 2.2, but for now we’ll simplify this picture by saying that NPIs are licensed when they occur in the scope of a negative operator like *not* or *no*. Prior work has shown that the unacceptability of sentences like (11a) is not as immediately apparent as the unacceptability of similar sentences with unlicensed NPIs like (11b) (though after careful reflection, comprehenders typically conclude that both are bad). This is shown in various tasks, including speeded judgments of acceptability, reading times, and ERPs. This initial perception of acceptability for sentences like (11a) is the illusion.

- (11) a. * The authors that no critics have recommended for the award have ever received acknowledgement for a best-selling novel.
 b. * The authors that the critics have recommended for the award have ever received acknowledgement for a best-selling novel.

This pattern has the potential to inform our theories of the real-time implementation of grammatical knowledge, and the interaction between linguistic and non-linguistic systems in sentence comprehension.

As we will see, however, the profile of the illusion is much more restricted than one might suspect based on this example alone. In this dissertation, we are concerned with identifying an algorithm for processing NPI-containing sentences that results in a consciously-detectable error signal for many unlicensed NPIs, including sentences like (11b) but also many others, and a reduced or absent error signal for the unlicensed NPI in (11a). In order to investigate this is, we begin with an overview of the literature on illusions of grammaticality (section 2.1), as well as the literature on the grammar of NPIs (section 2.2). We then summarise the goals of the subsequent chapters on NPI illusions (section 2.3).

2.1 Illusions of grammaticality

NPI illusions are a subcase of illusions of grammaticality, or phenomena in which incrementally generated representations appear to misalign with the representation that the hypothesized grammar would assign to the input. Illusions are sometimes described as cases in which the comprehender’s initial representation misaligns with a slower, more carefully considered judgment. We will not adopt this definition (see section 1.1). Illusions can be quite persistent and we will still consider them illusions, largely for theory-driven reasons.

There are several varieties of grammatical illusion in the literature, the best-known of which is agreement attraction. Most grammatical theories treat both (12a) and (12b) as ungrammatical due to the failed subject-verb agreement between *key* and *were*. Famously, (12b) is more likely to be erroneously produced, is read more quickly and, under time pressure, is judged acceptable more often, compared to (12a) (Bock & Miller 1991; Nicol, Forster, & Veres 1997; Clifton, Frazier, & Deevy 1999; Pearlmutter, Garnsey, & Bock 1999; Wagers, Lau, & Phillips 2009; Patson & Husband 2016; Slevc & Martin 2016; Hammerly, Staub, & Dillon 2019; Schlueter, Parker, & Lau 2019; Lago, Acuña Fariña, & Meseguer 2021; among others). These facts suggest that the ungrammaticality of (12b) is not readily detected.

- (12) a. * The key to the cell unsurprisingly were rusty from many years of disuse.
 b. * The key to the cells unsurprisingly were rusty from many years of disuse

(Wagers, Lau, & Phillips 2009:16)

The feature on which the subject mismatches (and the attractor matches) with the later verb need not be number. Gender attraction has been reported in languages where gender is marked on the verb (e.g. Russian: Slioussar & Malko 2016) as well as languages where gender is marked on a post-verbal predicative adjective (e.g. Spanish: Acuña-Fariña, Meseguer, & Carreiras 2014). Attraction effects have also been found in honorific agreement processing in Korean (Kwon & Sturt 2016).

Anaphora processing may also be subject to illusions of grammaticality (Cunnings & Felser 2013; Patil, Vasishth, & Lewis 2016; Parker & Phillips 2017; Jäger, Engelmann, & Vasishth 2017). Reading times on *herself* are sometimes quicker in sentences like (13b) than sentences like (13a), suggesting that the comprehender doesn't successfully rule out *librarian* as a potential antecedent, contrary to grammatical constraints on binding.⁹ Though note that this pattern does not seem to arise with robust effect sizes for all anaphors with all intrusive antecedents (Nicol & Swinney 1989; Clifton, Frazier, & Deevy 1999; Sturt 2003; Xiang, Dillon, & Phillips 2009; Dillon et al. 2013; Cunnings & Sturt 2014; Sloggett 2017; Malko 2018).

- (13) a. * The strict librarian said that the brief memo reminded herself about the overdue book.
 b. * The strict father said that the brief memo reminded herself about the overdue book.
 (Parker & Phillips 2017:276)

Parker 2022 reports attraction effects in reading times based on voice mismatches (active versus passive) in ellipsis. For example, both (14a) and (14b) are ungrammatical due to the lack of voice parallelism between the ellided clause and its antecedent (i.e., **Jane recruited ... and John was too*). However, reading times at the spillover region for the ellipsis site (*later*) suggest that the ungrammaticality of (14b) is not as immediately disruptive, due to intervening voice-matched clause.

- (14) a. * Jane recruited for the event that the villagers organized, and John was too later in the afternoon.
 b. * Jane recruited for the event that was organized by the villagers, and John was too later in the afternoon.

⁹The logic here relies on the assumption that comprehenders expect *she/her* pronouns to be used for librarians, based on stereotypes.

(Parker 2022:4)

There's an appealing parallel across these cases. Repeatedly we find that in the processing of various dependencies which require (under some analyses) a particular structural relationship between two elements, an intervening but structurally-irrelevant item with the appropriate features can alleviate some of the processing disruption caused by the failed dependency. This pattern has led to an approach that seeks to explain all of them under a common mechanism. Hypotheses that leverage this parallel have the virtue of explaining a wide variety of data points parsimoniously. This can also be couched as a risk, if in fact there is no real commonality in the profile of these phenomena.

2.1.1 Shallow processing

One way to explain the apparent trend across dependencies would be to say that the comprehender's syntactic analysis of the prior context is simply not detailed enough to distinguish between structurally-relevant and structurally-irrelevant elements. This is effectively the explanation we expect under "good enough" or "shallow" hypotheses about sentence comprehension (Ferreira, Bailey, & Ferraro 2002; Sanford & Sturt 2002; among others). Note that hypotheses in this family don't often explicitly address the grammatical illusions described here, but a similar style of explanation is sometimes given for sentences like in (15). Duffy, Henderson, & Morris 1989 demonstrated facilitation in the processing of *cocktails* for both (15a) and (15b) relative to (15c).¹⁰ Ferreira, Bailey, & Ferraro 2002 argue that these findings show that "the semantic representation ... was not detailed enough to distinguish the difference in meaning between the two sentences" (Ferreira, Bailey, & Ferraro 2002:12). That is, the comprehender's representation of *the boy watched the bartender serve the...* and the comprehender's representation of *the boy who watched the bartender served the...* are identical.

- (15) a. The boy watched the bartender serve the cocktails.
 b. The boy who watched the bartender served the cocktails.
 c. The boy saw that the person liked the cocktails.

¹⁰Note that these sentences are not actually among Duffy, Henderson, & Morris's stimuli, though this is the example that Ferreira, Bailey, & Ferraro 2002 provide when referring to the Duffy, Henderson, & Morris 1989 findings.

(Ferreira, Bailey, & Ferraro 2002:12)

A similar argument could be made about the level of detail in the syntactic representation for the grammatical illusion sentences discussed above. At the extreme, if the only representation one had of prior sentential context were a bag of words, there would be no way to distinguish structurally-relevant from structurally-irrelevant prior elements, and illusions of the form discussed here would be rampant. Shallow processing hypotheses are not generally committed to such a complete lack of structure, but the reasoning is effectively the same. However, as we discussed at length in Chapter 1, the existence of errors is not, on its own, evidence of shallowness or “shortcuts” in processing, and, all else equal, a model involving single sentence analyzer is more parsimonious than a model involving two. There is also independent reason to believe that the representation of prior sentential context is not a bag of words (at a minimum, the fact that garden path effects arise demonstrate that this cannot be right), and evidence that such representations are in fact quite richly detailed.

2.1.2 Cue-based retrieval

We might instead try account for the general pattern across illusions as a consequence of failure in another system with which the grammar must interact. The most prominent hypothesis along these lines is the partial-matching hypothesis suggested by Wagers, Lau, & Phillips 2009 for agreement attraction and Vasishth et al. 2008 for NPI illusions. One advantage of such a hypothesis is that it can be explicitly modeled using tools like ACT-R (Anderson et al. 2004).

This hypothesis assumes a cognitive architecture in which everything except the material in the (severely capacity-limited) focus of attention is stored in Long-Term Memory (LTM). This means that any prior element of a sentence that enters into a dependency with a later element must be retrieved from LTM, on the basis of the features with which they were initially encoded. Memory retrieval involves the parallel activation of retrieval cues, which leads to increases in activation for chunks in memory that share those features, until a single chunk reaches high enough activation levels for it to be retrieved. This kind of memory architecture is independently motivated (Lewis & Vasishth 2005; McElree 2006). In the case of subject-verb agreement, nouns are hypothesized to be encoded with features like [+/- plural] and [+/-

subject]. When a plural main verb is encountered, a retrieval is initiated with the [+subject] feature and the [+plural] feature. In the case of sentences like (12b), this retrieval operation yields partial matches with both *key* (which is a subject but not plural) and *cabinets* (which is plural but not a subject). The activation of both of these items therefore increases as the retrieval operation unfolds, but the boost in activation is not as great as it would be if there were a full match.

The same basic mechanism can be applied to NPI illusions. The retrieval is initiated by the NPI instead of a verb, and the relevant cues are those that identify appropriate licensors instead of those that identify appropriate subjects. Vasishth et al. 2008 suggest [+negation] and [+c-command] as retrieval cues for NPI licensing.¹¹ Thus *no authors* in (16a) can be efficiently retrieved and related to the NPI (Vasishth et al. 2008 discuss equivalent German stimuli) because it was encoded with both of these features. Partial matches arise for non-c-commanding negative words and non-negative c-commanding words, as in (16b) — *no critics* matches on the negation cue, but not c-command, whereas *the authors*¹² matches on the c-command cue, but not negation.

- (16) a. No authors [that the critics have recommended for the award] have ever received acknowledgement for a best-selling novel.
- b. * The authors [that no critics have recommended for the award] have ever received acknowledgement for a best-selling novel.
- c. * The authors [that the critics have recommended for the award] have ever received acknowledgement for a best-selling novel.

The fact that partial matches to retrieval cues can occur does not, on its own, explain why illusions occur. Such a theory must be paired with a linking hypothesis that predicts how various cue combinations will lead to a retrieval outcome and translates the retrieval outcome into a judgment of acceptability (or,

¹¹We will return to the question of whether this is the right way to think about NPI licensing, but for now we take it for granted.

¹²It's worth noting that, under many syntactic theories, *the authors* is not actually a constituent in the sentence. It is not entirely clear whether the Vasishth et al. 2008 model gets around this issue by assuming a syntactic representation in which *the authors* is a constituent, or by assuming that being a chunk in memory and being a constituent are not closely related properties. It's also possible that Vasishth et al. had in mind that the actual representation being retrieved is in fact the full DP *the authors that no critics recommended* and they simply referred to this chunk as *the authors* as a shorthand.

for reading time studies, a decision to move on to the next word at some time point). One could imagine many such linking hypotheses, only some of which will predict illusions. For example, it could be that in situations of multiple partial matches, the retrieval operation sometimes succeeds, in that a chunk reaches a high enough level of activation to be retrieved, and sometimes “times out”, resulting in retrieval failure and perceived unacceptability. Or it could instead be that retrieval failure is not an option (i.e., whatever item has the highest activation at some time point is the one that gets retrieved), but any time a memory retrieval operation results in retrieval of an item that only partially matches the retrieval cues, there is an acceptability penalty. Alternatively, it could be that retrieval latencies are the drivers of consciously perceived acceptability, such that the longer it takes for something to be retrieved, the less acceptable the sentence is perceived to be. We consider all of these example linking hypotheses to be plausible. However, the particular linking hypothesis being assumed is often not made explicit by researchers arguing for partial matching explanations for illusions, even though, as the present discussion illustrates, this has clear consequences for whether the existence of partial matches actually leads to illusions of acceptability.

The linking hypothesis that Vasishth et al. 2008 adopt is one in which retrieval always succeeds (in that it always yields some chunk),¹³ and the judgment of acceptability is based on what was retrieved. The assumed acceptability decisions are as follows. For grammatical baseline sentences like (16a), trials in which *no authors* is retrieved are judged acceptable and trials in which *the critics* is retrieved are judged unacceptable. For embedded-negation sentences like (16b), trials in which *the authors* is retrieved are judged unacceptable and trials in which *no critics* is retrieved are judged acceptable. For ungrammatical baseline sentences like (16c), trials in which either *the authors* or *the critics* is retrieved are judged unacceptable.¹⁴ Clearly, the decision of acceptability is based on some process which occurs after retrieval and which evaluates whether the retrieval outcome is consistent with the grammar. Based on the decision outcomes we just summarised, it appears that this grammar-checker is using only one of the two properties that the

¹³This is actually dependent on various parameter settings. Critically, there is a “partial matching” parameter, which, if switched off, leads to failed retrieval in any circumstance where a full match does not exist. However, this parameter setting is not the one researchers are generally interested in, since the possibility of partial matches is central to the explanation.

¹⁴In fact, the mapping from possible retrievals to acceptability judgments is not completely specified in Vasishth et al. 2008. Based on what we have said, condition (16c) should result in judgments of unacceptability on virtually every trial. But in fact the model predicts unacceptability in 76.6 percent of trials for this condition (Vasishth et al. 2008:697). This suggests that there is something that could be retrieved in this condition that would result in acceptability. It is not clear what that something is.

memory retrieval operation was using — the negation property. It is not entirely clear what motivates this choice.¹⁵ If instead the grammar-checker used only the c-command property to evaluate the output of the memory operation (i.e., any time the retrieved chunk actually c-commands the NPI, the sentence is judged to be good, and any time the retrieved chunk fails to c-command the NPI, the sentence is judged to be bad), the predicted acceptability results would be entirely different. Thus the extent to which the model actually accounts for the illusion pattern is in part a consequence of partial matches, but it is also in part a consequence of the implicit assumption that negation is more important than c-command for determining whether an NPI is licensed. Note that these concerns do not arise for Vasisht et al.’s modeling of reading time data, which is based on the model’s retrieval latency, not the retrieval outcome.

The cue-based retrieval model for NPI illusions makes a few assumptions about the nature of the NPI dependency that are worth highlighting, though a fuller exploration of the appropriateness of these assumptions will only be possible after we have reviewed the literature on NPI licensing in section 2.2. First, there is the choice of retrieval cues, [+negation] and [+c-command]. As Kush 2013 describes in depth, there are problems with treating [+c-command] as a cue because c-command is a relational property; no node of a tree can be described as “having the c-command property”, rather every *pair* of nodes in the tree is either in a c-command relation or not.¹⁶ To implement a proxy for c-command, the authors

¹⁵Vasisht et al. justify the model’s acceptability decisions in part based on the fact that the embedded noun is incompatible with the matrix predicate. This is less true for the English stimuli discussed here (it is not unimaginable for a critic to receive acknowledgement for a best-selling novel. Maybe the critic is also a writer. Or maybe their glowing review of a novel is what caused it to be a best-seller.) but very true for the German stimuli they discuss, such as (the German translation of) *A/no pirate who had eaten a/no roast was ever thrifty*. Thus, they claim that it is the fact that a roast cannot be thrifty that leads the “unacceptable” decision for ungrammatical baseline trials in which the embedded noun is retrieved. However, this issue does not seem to factor into acceptability decisions for the embedded-negation condition, since the trials in which *no roast* is retrieved are the trials that are accepted. Even if the actual motivation for the decision rules that were implemented was rooted in intuitions about the acceptability of the noun-predicate pair, the actual decision rules the authors ended up with are effectively just about negation — if a negative phrase was retrieved, the sentence is good, if a non-negative phrase was retrieved the sentence is bad (modulo the mysterious acceptance of ungrammatical baseline sentences, discussed in the previous footnote).

¹⁶There are ways we could try to get out of this problem, the most obvious of which is to effectively curry the two-place predicate into a one-place predicate, and then encode the result on the nodes that require it. That is, one could imagine encoding a representation of node X with a feature for every node that X c-commands (identifying these nodes by their indices, i.e. [+c-commands-node-12], [+c-commands-node-27], [+c-commands-node-142], etc.), and then the search cue at a dependent element at, for example, node 217 would just be [+c-commands-node-217] instead of [+c-command]. However, this could easily become unwieldy — as every new node of the tree is encoded, new c-command relations arise, requiring the retrieval and re-encoding of all prior nodes that c-command the current node. Kush 2013 rules out an “exhaustive encoding” approach like this one on the basis of the time it requires to be executed at each new node. An alternative to encoding features for every node that a given node c-commands is to encode features for every node a given node is c-commanded by. As a new node is encoded, it would be relatively straightforward to identify its c-commanders by simply copying over the c-commanders of the

use the case morphology on the head noun as a cue (in the examples tested, this happens to match up with the licensors that c-command the NPI) but it's clear that this is just a temporary fix and not a plausible hypothesis for how the comprehender could actually retrieve c-commanding licensors in the general case.

[+Negation] is not exactly right either because there are many NPI-licensing environments that are not explicitly negative (e.g. the restrictor of a universal quantifier). In order to correctly account for the licensing of truly grammatical NPIs, the theory requires a feature set that retrieves, with a full match, all and only the true NPI licensors. As we will see in section 2.2, identifying this natural class has proven difficult. Many contemporary theories of the grammar of NPIs therefore emphasize the properties of the environments that contain licit NPIs, not the properties of the items that c-command them. This brings us to the second, deeper issue with the assumptions of the cue-based retrieval model, which is that it treats the constraint on the distribution of NPIs as fundamentally an item-to-item dependency between two elements that are temporally non-adjacent, which must be checked by the incremental sentence comprehender. This is a possible way to conceptualize the restriction on NPIs, but not the only one. As we have previously mentioned, it is tempting to describe the processing of NPIs as a “licensing” operation, but all we actually know is that comprehenders accept sentences like (17a) and reject sentences like (17b). To say that the NPI in (17a) is “licensed” by *haven't* is, to some extent, just a shorthand for describing this fact.

- (17) a. We haven't left the house in months.
 b. * We have left the house in months.

Let's assume, for the sake of simplicity, that we are only concerned with explicit negation and not with other licensing environments. We might still ask, in virtue of what information does a comprehender know, immediately at *in months*, that there is a problem with (17b) but not with (17a)? The memory of an *n't* morpheme need not be the only thing that's different about the comprehender's state when *in* immediately dominating node and adding the current node's sister to the list. Alcocer & Phillips 2012 consider a similar algorithm but do not consider it plausible because each [+c-commanded-by-X] feature would need to be a separate cue. Thus, the parallel activation of all of those cues would lead to rampant partial-match interference, and would guarantee that full matches are virtually impossible. A more general issue is that any explicit encoding of c-command is, in some sense, redundant, since dominance relations are assumed to be encoded on chunks, and c-command relations can always be computed from a fully specified set of dominance relations. Perhaps one might want to do away with dominance relations altogether and encode the relationship between nodes entirely based on c-command. This is beyond the scope of the current investigation.

months is encountered in these two sentences. There is ample evidence that comprehenders do not wait until a sentence is over to begin constructing an interpretation, and negation has serious consequences for the interpretations that are constructed. So, while it could be that the comprehender knows that *in months* in (17a) is appropriate in virtue of the fact that *n't* can be found in memory, it could also be in virtue of the fact that the comprehender is in the middle of understanding a negated sentence.¹⁷ This aspect of the meaning of the sentence could also be tucked away in memory, needing to be retrieved, or it could be already available, in which case the nature of memory retrieval operations becomes irrelevant to NPI licensing.

This does not rule out cue-based retrieval as an explanation for NPI illusions, and it certainly does not have any bearing on the question whether cue-based parallel activation is the right model of memory retrieval in general. At present we merely aim to make explicit the assumptions that go into this model. This is particularly relevant because one of the key theoretical advantages of attributing errors to the memory system is that it allows for a parsimonious mapping between the grammar and the real time comprehension system. If in fact the comprehension system that is assumed has little resemblance to a plausible grammar, this advantage is lost. A thorough exploration of the extent to which the cue-based retrieval model is faithful to the grammars of the other dependencies discussed here (subject-verb agreement, anaphora, and ellipsis) would be valuable but is beyond the scope of this dissertation.

The generality of the cue-based retrieval explanation — as well as the shallow processing explanation mentioned earlier — is an important advantage of the model, but also risks predicting generality where none exists. Although the grammatical illusions described above have some clear commonalities, it is not apparent that all dependencies that fit the description give rise to illusions. Sentences like (18) are in many ways similar to the illusion phenomena we have discussed. Upon encountering *he*, a comprehender has a potential antecedent in their prior representation (*any janitor*) which could bind the pronoun if not for the fact that this element is in a structurally inappropriate position. Kush, Lidz, & Phillips 2015 found

¹⁷This may sound like it presupposes a particular answer to the question of whether the distribution of NPIs is fundamentally a syntactic or semantic (or pragmatic) phenomenon. To some extent it does. If the grammar of NPIs is fundamentally based in the meanings of NPI-containing environments, it is quite natural to expect that it is in virtue of the meaning of the environment that a comprehender detects that an unacceptable NPI is unacceptable. However, even if the grammar of NPIs is fundamentally part of the syntax, this would not rule out the possibility that the comprehender who detects the unacceptability of (17b) does so initially based on the meaning of the clause.

that comprehenders do not appear to consider *any janitor* to be a possible binder of *he*, even at very early stages of processing.¹⁸

- (18) Kathi didn't think any janitor liked performing his custodial duties, but he had to clean up messes left after prom anyway.

(Kush, Lidz, & Phillips 2015:21)

Prior work has shown that, for the various phenomena that do give rise to illusions, the circumstances under which they arise may not be identical. Xiang, Grove, & Giannakidou 2013 showed that pragmatic reasoning ability, as measured by the Autism Quotient, is predictive of an individual's susceptibility to NPI illusions, but not agreement attraction. Parker & Phillips 2016 showed that the amount of intervening material between elements of the dependency impacts NPI illusions, but not agreement attraction. Parker & Phillips 2017 showed that reflexive attraction only arises when multiple search cues align with the intrusive antecedent and misalign with the syntactically licit antecedent, whereas proposals for NPI illusions and agreement attraction suggest that illusions arise with just a single mismatching cue. Variability across phenomena is not straightforwardly predicted under any account that treats them all as consequences of the exact same underlying error. We suspect that a more detailed understanding of the grammar that underlies each of these dependencies, and explication of what an online implementation of such a grammar would consist of, will bring greater clarity to the study of illusions and potentially explain some of the cross-dependency variability.

2.2 NPI licensing

The extensive literature on the grammatical knowledge governing the distribution and interpretation of NPIs engages with several intersecting questions, which we will not be able to do justice to in this brief review. One central goal in this literature is to adequately describe the characteristics that distinguish the contexts in which an NPI can appear from those in which an NPI cannot appear. Because there is

¹⁸Note that this experiment once again relies on stereotyped gender: comprehenders expect *Kathi* to be referred to with *she/her* pronouns and janitors to be referred to with *he/him* pronouns.

such diversity in these contexts (negation, questions, the antecedent of a conditional, the restrictor of a universal quantifier, etc.) this is a non-trivial issue. There is a related hope, among some researchers, for the characterization of NPI-licensing contexts to also suggest an explanation for why the distribution of NPIs is as it is. Somewhat independent of these questions, there is also the question of diversity among NPIs. Strong versus weak NPIs, but also NPIs in different languages, exhibit different licensing constraints. Then there is the question of whether the distribution of Free Choice Items (FCIs) and/or Positive Polarity Items (PPIs) should be explained in a way that is related to the distribution of negative polarity items. And finally there is the question of whether what we’re doing, when we study NPIs, is the work of syntax, semantics, or pragmatics. All of these issues have potential consequences for NPI illusions, but we will focus our attention on the question of the characterization of the contexts in which NPIs appear.

Our ultimate goal will be to define possible processing algorithms which implement possible grammars. There is, of course, not a one-to-one mapping between these — just as the same context free grammar can be implemented by either an Earley parser or a CKY parser, we expect that a single conceptualization of the grammatical knowledge of NPIs may have multiple possible algorithmic implementations. It is also in principle possible that the initial strategy for understanding a sentence (and detecting unacceptability) is divorced from the grammar, as in two-system views. But we will approach the question starting from the assumption that there is a more or less direct mapping between the grammatical knowledge of the constraints on NPIs and the processing of NPI-containing sentences, and deviate from this assumption only as needed. That said, it must be noted that the proposals discussed here were not necessarily intended by their authors to be claims about cognitive processes, and perhaps not even cognitive representations.

2.2.1 Syntactic accounts

Early accounts of NPIs were strictly syntactic in nature, and were formulated in terms of transformational rules that turn a non-NPI (e.g. *some*) into an NPI (e.g. *any*) when it occurs “in construction with” or is c-commanded by negation, or a similar element (Klima 1964). The trick is defining what constitutes

a “similar element” to negation. Clearly, a grammar that allows only explicit negation to license an NPI through c-command would drastically under-generate. Early proposals therefore adopted the (somewhat unsatisfying) approach of labeling all NPI-licensing elements “affective”, and then stating the constraint in terms of c-command by an element with the affective feature (Jackendoff 1969). A precise definition of what constitutes an affective element (independent of NPI-licensing abilities) is somewhat elusive, though. In part because of these difficulties, contemporary approaches to the grammar of NPIs tend to focus on the meanings of the contexts that contain them (and the meanings of NPIs themselves), not the features of the elements that c-command them.

An interesting counter-example comes from Herburger & Mauck 2013, which proposes an entirely syntactic feature-based mechanism for NPI licensing. They argue that the question of why some elements are restricted to certain contexts is not well-answered by an analysis of the meanings of those elements or those contexts, but is largely a historical accident — NPIs become NPIs by chance. The theory requires rules for the propagation of “+” and “–” features through the tree, starting from the licensor. The “+” and “–” features are intended to correspond to upward and downward monotonicity (see the discussion of downward entailment below) but are not themselves the pattern of entailments that the sentence allows — they are just syntactic features. Given differences in the licensing profiles of different quantifiers (e.g. *every* licenses NPIs in its restrictor but not its scope, whereas *no* licenses NPIs in both), licensing is ultimately implemented not by c-command but by dominance by the “–” feature, which has been passed along from node to node.

Although this account is strongly committed to syntactic NPI licensing by a negation feature, it does not align perfectly with the particulars of the feature-based search assumed by Vasishth et al. 2008. Because the syntactic dependency that must be checked is between the NPI and a node that dominates it, the negative word that is the source of the negative feature is ultimately irrelevant. However, a very similar processing algorithm could be implemented with a switch to [+dominates] and [+negative] features, likely without major consequences¹⁹. Of course, other mechanisms are also possible, such as an algorithm which climbs the dominance relations of the tree one by one, until either a negative feature is found (re-

¹⁹Though the same problems persist, such as the fact that [+dominates], like [+c-command], is relational and therefore not obviously a good retrieval cue.

sulting in acceptability) or the root node is reached (resulting in unacceptability). There are many other options, but a key property of such a theory is that the negative feature that licenses the NPI is a syntactic feature that lives on a node of a syntactic tree, and so the determination that an unlicensed NPI is unacceptable would (assuming a tight link between the grammar and the processor that implements it) involve the accessing this feature, not accessing the meaning of the clause or sentence.

If one wishes to maintain a c-command-by-negation analysis of (at least some) NPI licensing, the standard solution to licensing in contexts that are not explicitly negative is to posit a secondary mechanism. In order to account for the acceptability of both NPIs and PPIs in doubly-negated contexts, C. L. Baker 1970 proposes a key role for sentence-level entailments in addition to a syntactic licensing mechanism very similar to the one assumed by Klima 1964 and Jackendoff 1969. Thus an NPI is licensed in the scope of negation, or in any sentence that entails a sentence in which the NPI is within the scope of negation. Linebarger 1987 updates Baker’s theory and proposes that it is negative implicatures not entailments that are critical to the secondary mechanism. Importantly for Linebarger, the secondary mechanism is not actually a grammatical licensing mechanism, but effectively an inference process through which ungrammatical sentences come to be perceived as acceptable. The “rescuing” operation proposed by Giannakidou 2006 plays a similar role, though here it is any proposition “made available” by the global context that can make an unlicensed NPI acceptable.²⁰

Translating such theories into an online processing algorithm would presumably involve two distinct stages through which the acceptability of an NPI-containing sentence is determined. Interestingly, the commitment in some of these proposals to the secondary process being extra-grammatical means that many NPI-containing sentences that are accepted are in fact ungrammatical, suggesting that ungrammaticality in virtue of violating the grammar of NPIs is not, on its own, enough of an error signal to trigger conscious detection of a problem and rejection of the sentence. This could be spelled out in various

²⁰The first-line NPI licensing mechanism in Giannakidou’s theory is not actually c-command by negation, but nonveridicality, which is closely related to downward entailment, which we discuss below. This licensing operation is partly motivated by the goal of capturing the distribution of all polarity sensitive items, including free choice items, under a unified mechanism. Because this licensing mechanism is based on the entailment patterns of a context (technically the scope of a nonveridical operator), and not a c-command relation with a syntactic negation feature, it is importantly different from C. L. Baker’s and Linebarger’s proposals. We mention it here because these proposals have in common a “two-stage” approach to NPI licensing, which has consequences for real-time sentence processing.

ways. One option is to assume a temporal ordering in the processing of NPIs that corresponds to the logical ordering of licensing mechanisms. First, the comprehender attempts to establish a syntactic dependency that licenses the NPI (this might look very similar to the Vasishth et al. 2008 mechanism, in which a search for [+negation] and [+c-command] is executed, or might in principle look like a serial evaluation of c-commanding nodes). Then, if this fails, a strictly-subconscious error signal triggers the deployment of the secondary mechanism, and related propositions (i.e. those entailed by, implicated by, or made available by the sentence) are evaluated. Further details on what this evaluation process consists of, mechanistically, are obviously needed. It is only when this second process fails that the comprehender consciously detects a problem. One consequence of such a model is that, depending on the amount of time the second process requires, there may be some window following the encoding of an (in fact unacceptable) NPI in which the unacceptability has not yet been consciously detected because the secondary operation is still in progress. There does not seem to be clear evidence that this is the case, though we know of no direct attempts to investigate this prediction. Another consequence of the two stage model is that sentences in which NPIs become acceptable in virtue of the second process should, at the earliest stages of processing, look like ungrammatical NPIs, at least in implicit measures, because both of them fail the first attempt at licensing. Xiang, Grove, & Giannakidou 2016 report ERP effects that are somewhat consistent with this prediction, in that they observe a P600 effect for both NPIs in the scope of emotive factives (a rescuing context) and unacceptable NPIs, relative to NPIs in the scope of explicitly negative licensors.²¹ Of course, these predictions do not necessarily hold if there is not a clear temporal ordering between the two processes.

2.2.2 Pragmatic accounts

A separate tradition in the NPI literature more or less abandons the treatment of negation as a “canonical” licensor, since the set of contexts in which NPIs occur is much broader. A critical early observation motivating this approach is that the contexts in which one finds NPIs have something in common, and

²¹However, this finding is complicated by the fact that a reduced N400 is observed for all acceptable NPIs, relative to the unacceptable condition, regardless of the type of licensor. Thus, it is not entirely accurate to say that NPIs in the scope of emotive factives first look like unacceptable NPIs and then look like acceptable ones; rather they first look like acceptable NPIs and then look like unacceptable ones.

it is not that they're all negative. Rather it is that they all are associated with inverted scales of inference, relative to a non-NPI-licensing counterpart (Fauconnier 1975a; Fauconnier 1975b). For example, *Ann did not do x to help* might standardly lead a comprehender to infer that *Ann did not do y to help*, for any *y* that is greater than *x*. Thus, the NPI *lift a finger* can receive its idiomatic meaning of minimal effort in this context, since it occupies the strongest endpoint of the scale, allowing the comprehender to infer all other elements on the scale. Without negation, the scale reverses and the NPI is disallowed. Kadmon & Landman 1993 attempt to unify free choice *any* and NPI *any* under a similar approach. Israel 1997 extends this framework to PPIs and non-minimizer NPIs. Note that under such a theory, the NPI does not enter into any dependency with a prior negative word. Rather, the negative word creates an environment that has the kind of meaning in which the NPI (in virtue of its own meaning) is appropriate. This would suggest that, if hypotheses in this category are correct, an online licensing approach that consists of retrieving a negative word from memory is not merely in need of refinement but is fundamentally the wrong kind of operation, since it operates on the wrong kind of representation. These are fundamentally pragmatic accounts — an NPI is or is not acceptable in a given sentential context purely on the basis of the type of inferences the sentence makes available. Thus, a direct online implementation of such a theory would require that the comprehender access and evaluate those inferences in real time.²² There are various possible implementations of this with respect to the relative timing of when the scalar inferences are inferred. Although the other elements of the scale typically consist of lexical alternatives to the NPI, this doesn't necessarily mean that they're only constructed after the NPI is encountered. Rather, in a natural conversational context a comprehender may be able to infer the relevant scale prior to the NPI, in which case the licensing operation may proceed more quickly or effortlessly. This possibility becomes relevant to the scalar alternatives account of NPI illusions, which we pursue in Chapter 3.

²²Note that “accessing and evaluating” the inferences a context makes available is a much less computationally explicit theory than the hypothesis that NPIs are licensed by cue-based retrieval of a [+c-command] and [+negation] chunk in memory. This makes it difficult to directly compare such hypotheses.

2.2.3 Truth conditional accounts

One final category of theories of NPI licensing focuses on the patterns of logical entailment of the contexts in which NPIs are found. Ladusw 1979 proposed that NPIs are licensed in Downward Entailing (DE) contexts (sometimes called downward monotonic contexts). For example, in negated contexts such as (19), we find that (19a) entails (19b). But without negation, such an entailment does not hold, and instead (20b) entails (20a) (“upward” entailment). Ladusw observed that, for a wide variety of contexts, downward entailment and permitting NPIs go hand in hand.

- (19) a. No men walk.
 b. No men walk slowly.
- (20) a. Some men walk.
 b. Some men walk slowly.

(Ladusw 1979:115)

Subsequent analyses have given different entailment conditions that must be met in order for NPIs to be acceptable, such as nonveridicality (Giannakidou 1998), anti-morphic contexts for strong NPIs (van der Wouden 1997), anti-additivity for strong NPIs (Zwarts 1998), Strawson downward entailment (Von Stechow 1999), downward entailment with/without non-truth-conditional content taken into account (Gajewski 2011). It’s worth noting that although all of these proposals highlight the importance of environment-level inferences, they do not uniformly identify the environment as the locus of licensing. Rather, many are articulated in terms of scope of a lexical item with particular characteristics. For example, Ladusw’s claim is not technically that NPIs are permitted only in DE environments, but that NPIs are permitted only in the scope of DE operators. Homer (2008; 2021) argues that in the few cases where environment-based and operator-based hypotheses make different predictions, environment-based hypotheses better account for the data.

Setting these findings aside, hypotheses in this category could in principle be implemented in either of two ways: an algorithm that seeks a scope-taking operator with the right characteristics, or an algorithm that evaluates the entailments of the local environment. A hypothesis in the first category could be

implemented in an architecture like the cue-based retrieval hypothesis put forth by Vasishth et al. 2008, with some minor adjustments to the cues — instead of [+negative] we might use [+downward-entailing] and instead of [+c-command], [+scope]²³. Such a hypothesis could instead be implemented by an architecture that uses a structurally-guided search to identify only the scope-taking elements, and evaluates those items for their entailment properties. Environment-based licensing hypotheses would require incremental access to entailments of the local environment.

Some research has attempted to test whether something like a processing theory in the second category is possible. Szabolcsi, Bott, & McElree 2008 show that the presence of an NPI doesn't make comprehenders any better at detecting downward entailments, which they take as evidence against the hypothesis that downward entailment is critical to NPI licensing. For example, following either a sentence like (21a), with an NPI, or a comparable sentence like (21b), without an NPI, comprehenders are highly accurate (above 85%) at correctly answering a question like (22).²⁴

- (21) a. Almost no campers have ever had a sunburn or caught a cold.
 b. Almost no campers have had a sunburn or caught a cold.
- (22) Would it be reasonable to say that almost no campers have caught a cold?
- (Szabolcsi, Bott, & McElree 2008:115)

Note, however, that the claim that downward entailment is the licensing condition for NPIs is distinct from the claim that there is a functional explanation (i.e., facilitating inferences) for downward entailment being the licensing condition for NPIs. Moreover, it's possible that the reason NPI processing (and the detection of unacceptable NPIs) typically proceeds so quickly and effectively is because the relevant entailments are computed as part of the normal comprehension processes for a negated sentence, and are therefore already available when the NPI is encountered. Such a theory would predict equally good performance for (21a) and (21b).

Chemla, Homer, & Rothschild 2011 observe a different pattern, finding item-to-item and person-to-person variation in both willingness to make downward inferences and willingness to accept an NPI, and

²³ [+Scope], like [+c-command], is relational and therefore not a good cue, but we are once again setting this concern aside.

²⁴ The inference from “had a sunburn or caught a cold” to “caught a cold” is a downward inference.

that the variability in these judgments is correlated. They suggest that this indicates that it isn't logical downward entailment but *perceived* downward entailment that determines licensing (but see Jacobson 2018 for arguments against this conclusion, and in particular against the notion of “perceived downward entailment”). Thus the evidence for the computation of downward entailments as a processing algorithm is not decisive.

From this brief review of the literature on NPIs, it is clear that a number of deep questions remain unresolved. Proposals for the grammatical knowledge that governs the distribution of NPIs vary in non-trivial ways, ranging from purely syntactic feature-checking accounts to fully pragmatic and semantic accounts that require evaluation of inferences or entailments. Accordingly, there is a wide variety of possible processing algorithms whereby a comprehender could detect the unacceptability of an unlicensed NPI (and construct an appropriate syntactic/semantic/pragmatic representation with a licensed one). Our goal here is not to adjudicate between grammatical theories on the basis of processing evidence, but to use grammatical theories as a starting point for developing plausible accounts of how a comprehender uses a grammar to process a sentence. One takeaway from this discussion is that a retrieval operation targeting a negative word in memory, as has been proposed to account for NPI illusions, is a reasonable approximation of some grammatical theories, but a substantial departure from others. This does not mean this is the wrong account, only that other options remain viable. In our exploration of the profile of the NPI illusion we will be particularly interested in the kinds of NPI-licensing mechanisms that various proposals assume, and their relation to hypothesized grammars.

2.3 The current study

The following three chapters clarify the profile of the NPI illusion by testing key predictions of some hypotheses hypotheses that have been proposed for the cause of NPI illusions. In particular, Chapter 3 presents six experiments which reveal a surprising generalization which we call the “licensor effect” — that is, NPI illusions arise for some embedded licensors but not others. These findings are used to argue for the scalar alternatives hypothesis which was briefly mentioned here but which will be described

in greater detail in section 3.1.3. Chapter 4 explores, in five experiments, the “distance effect” for NPI illusions, which was first reported by Parker & Phillips 2016 — that is, NPI illusions arise for NPIs in some positions in the sentence but not others. We argue that the nature of the distance effect suggest that the online licensing of NPIs is an operation which relates an NPI to some property of the context that contains it, not an operation which relates an NPI to a previously-occurring negative word. Finally, Chapter 5 presents the results of six experiments which substantially complicate the empirical picture, including one which provides evidence against the scalar alternatives hypothesis we argue for in Chapter 3. In Chapter 6 we discuss what these sixteen experiments can collectively tell us about the NPI illusion, and, more generally, the possible algorithms for deploying the grammar of NPIs. We additionally return to the question of possible parallels across grammatical illusion phenomena.

Chapter 3 NPI illusions: the role of licensors

3.1 Introduction

The licensing of NPIs like *any* and *ever* has long been a valuable test case in linguistic theory. In this project we investigate the illusory licensing of NPIs, a subtype of linguistic illusion whose behavior can be equally valuable for understanding real-time interpretation. The NPI illusion is the (typically fleeting) perception of acceptability for ungrammatical sentences with an unlicensed NPI like (23a), in contrast with similar sentences like (23b), whose deviance is immediately detected.

- (23) a. * The bills that no senators voted for have ever become law.
 b. * The bills that the senators voted for have ever become law.

The illusion has been found to be robust across measurements and languages (e.g. Vasishth et al. 2008; Xiang, Dillon, & Phillips 2009; Parker & Phillips 2016). Yet, most existing investigations have studied NPI illusions using quantificational forms of negation, predominantly *no*, but also other quantificational elements such as *few* or *only*, and have neglected the question of the interpretation that results from the illusion. Here we present the results from six experiments, aiming to identify the processing error that underlies this brief deviation from the grammar. Specifically, these experiments investigate the illusory potential of a different type of licensor — namely, non-quantificational negation in the form of *not* and *-n't* — and the interpretation that arises. These results pose problems for current accounts. We suggest an alternative explanation, which builds on grammatical theories of NPI licensing that emphasize the importance of scalar alternatives. Given that we are interested in the online implementation of grammatical knowledge of NPIs, we begin with a brief review of the main linguistic hypotheses about this component of the grammar. We then turn to prior psycholinguistic hypotheses which attempt to explain the NPI

illusion, and their relation to the hypothesized grammatical knowledge.

3.1.1 The grammar of NPIs

Lexical items whose distribution is restricted by the polarity of the context in which they appear are known as polarity items and include PPIs and NPIs. The class of English NPIs includes adverbs like *ever*, *anymore*, *yet* and *in years*, the determiner *any*, noun phrases such as *a red cent* and *a thin dime*, and verb phrase idioms like *lift a finger*, and *have a hope in hell*, among others. Roughly (we will refine this definition later) these elements must occur in the scope of a negative element, and thus, are unacceptable in positive contexts (see (24a) versus (24b)). Note that mere linear precedence of a negative element with respect to the NPI is not enough: the NPI must be within the scope of the negative element — often understood as syntactic c-command (Klima 1964; Laka 1994). For example, in (24c), the negative element *wasn't*, inside the relative clause (RC), is structurally irrelevant to the NPI, resulting in ungrammaticality.

- (24)
- a. No student has ever complained about the coursework.
 - b. *The student has ever complained about the coursework.
 - c. *The student who wasn't in class has ever complained about the coursework.

While the canonical NPI licenser is explicit negation, the class of NPI-licensing environments is in fact much broader. NPIs can be found in questions, comparative structures, the scope of adversative predicates, the antecedent of a conditional, and many other contexts. Identifying the property that these contexts have in common — and accounting for variation across languages and across NPIs within a language — has been a primary focus of research on the grammar of NPIs.

One group of proposals treat NPI licensing as a syntactic relation between the NPI and an overt, c-commanding negative feature (e.g. C. L. Baker 1970; Linebarger 1987). In order to accommodate the acceptability of NPIs in other contexts, a secondary, indirect mechanism is postulated. This allows for an NPI-containing sentence without explicit negation to be accepted in virtue of its close relation to a sentence which does contain negation, as in (25). The details of that relation vary across proposals. For instance, C. L. Baker 1970 emphasizes the importance of logical entailments, while Linebarger

1987 argues that the key relation is that of implicature. Note that the “rescuing” operation of Giannakidou 2006, which relies on any proposition “made available” by the global context is similar in nature, though this secondary mechanism is paired with a primary licensing mechanism that is distinct from the negative-feature-based licensing operation in previous accounts. Translating proposals in this category into processing mechanisms would likely involve postulating two distinct online computations through which the status of an NPI is determined: one that checks the syntactic representation of the sentence for a c-commanding negative element, and another that identifies closely related sentences that could allow for the NPI to be made acceptable. Both such mechanisms have been proposed in the literature on NPI illusions (Vasishth et al. 2008; Xiang, Dillon, & Phillips 2009).

(25) I doubt that you will ever pass the exam = I don’t think that you will ever pass the exam.

A contrasting idea due to Fauconnier 1975a and Fauconnier 1975b treats the distribution of NPIs as only one case study in the broader phenomenon of semantic and pragmatic polarity. The key idea lies in the observation that some sentential contexts are associated with scales in virtue of our world knowledge — if *Ann did not do x to help* and *y* is more effort than *x*, we might typically infer that *Ann did not do y to help* (importantly for Fauconnier, this is not an entailment, merely an implicature). Thus, if we take *lift a finger* to indicate the minimum possible effort, then *Ann did not lift a finger to help* will, in virtue of the associated scale, implicate that *Ann did not do x to help* for all other values of *x*. Flipping the polarity of the sentence has the effect of reversing the scale — if *Ann did x to help* and *y* is more effort than *x*, we cannot infer *Ann did y to help*. Thus, an NPI like *lift a finger*, which indicates the minimum effort, will not carry any implications about the rest of the scale in a scale-reversed (positive) context, and so the NPI loses its idiomatic, quantificational reading in the sentence *Ann lifted a finger to help*. This account predicts that any context in which a polarity item is acceptable will fail to permit the same polarity item if its polarity is reversed.

This key intuition has been pursued in two different approaches to NPI licensing — those that emphasize the meaning of the licenser and those that emphasize the meaning of the NPI. Ladusaw 1979, Ladusaw 1996, and subsequent work in the first category formalized NPI licensing as scope by a DE

operator. “Downward entailment” is illustrated by the inference from (26a) to (26b). Note that this entailment does not hold if the negative operator not is removed from both sentences.

- (26) a. The students have not complained about the coursework.
 b. The students have not complained loudly about the coursework.

Though there is clearly a resemblance between Fauconnier’s scales and Ladusaw’s downward entailment, the theories are not isomorphic. They differ in their focus on pragmatic versus truth conditional aspects of meaning, as well as the locus of licensing. For Fauconnier it is the scalar alternatives themselves that allow the NPI to be successfully interpreted in context, whereas in Ladusaw’s framework NPIs are licensed by a scope relation to a licenser. Thus, an online NPI-licensing mechanism that is faithful to the grammar will require the rapid computation of different types of information under these two hypothesized grammars. The question of what is computed online will be central to hypotheses about what drives illusions.

A separate body of research, also building on Fauconnier’s observation about pragmatic scales, has highlighted aspects of the NPI’s meaningful contribution to the sentence. Kadmon & Landman’s (1993) influential analysis of the NPI *any* proposes that the function of an NPI is to strengthen the claim expressed by the sentence in which it occurs. Subsequent work has elaborated on this idea, focusing on the relevance of subdomain alternatives and scalar inferences in the occurrence of NPIs (e.g. Krifka 1995; Israel 1997; Israel 2011; Chierchia 2006). The central idea of these approaches, following Fauconnier’s initial insight, is that the meanings of NPI-containing sentences correspond to extreme values along a scale of ordered alternatives. Note that there is variability among these approaches with respect to whether these alternatives are contextually-driven or triggered by the NPI itself. We will return to this issue in section 3.1.2 in our discussion of the scalar alternatives hypothesis for NPI illusions.

From this brief presentation of the main theoretical approaches to the grammar of NPI licensing, it is clear that negative polarity phenomena lie at the interface of syntactic, semantic, and pragmatic mechanisms. While we should be cautious about selecting among competing grammatical theories on the basis of sentence processing data, we note that some formulations of the language user’s knowledge lend themselves more to some processing mechanisms. For example, grammatical hypotheses that define licensing

as a syntactic relation between an NPI and the syntactic features of a prior licenser might be straightforwardly implemented as a memory retrieval operation of that prior negative word. In contrast, hypotheses that treat licensing as an operation that relates the NPI-containing sentence to its alternatives would instead lend themselves to a process by which those alternatives are activated and compared. Against this backdrop, we now turn to existing research on the illusory processing of NPIs.

3.1.2 Previous Accounts of NPI illusions

While there is much diversity in extant accounts of the mechanism underlying NPI licensing, the grammatical accounts of NPI licensing explored above all predict that (27a) is grammatical, and (27b) and (27c) are ungrammatical in virtue of the lack of an appropriate licenser for the NPI *ever*. Informal acceptability judgments typically align with this prediction — native speakers accept (27a) and reject (27b) and (27c). Critically, however, (27b) and (27c) are not alike in online measures such as speeded acceptability. Comprehenders sometimes fail to detect the ungrammaticality of sentences like (27b), leading to a grammatical illusion.

- (27) a. No authors [that the critics recommended] have ever received ...
 b. * The authors [that no critics recommended] have ever received ...
 c. * The authors [that the critics recommended] have ever received ...

... acknowledgement for a best-selling novel.

(Parker & Phillips 2016)

The existence of illusion effects in the processing of unlicensed NPIs is an empirically robust phenomenon, both across languages and measurements. It has been replicated using methods such as speeded acceptability judgments (German: Drenhaus, Saddy, & Frisch 2005; English: Xiang, Dillon, & Phillips 2006; Parker & Phillips 2016; de Dios Flores, Muller, & Phillips 2017; Hildebrandt & Husband 2017; Muller, de Dios Flores, & Phillips 2019; Orth, Yoshida, & Sloggett 2021; Korean: Yun, Lee, & Drury 2018), self-paced reading (English: Parker & Phillips 2011; Xiang, Grove, & Giannakidou 2013; Ng &

Husband 2017; Turkish: Yanilmaz & Drury 2018b), eye-tracking (German: Vasishth et al. 2008; English: Orth, Yoshida, & Sloggett 2020a) and event-related potentials (German: Drenhaus, Saddy, & Frisch 2005; English: Xiang, Dillon, & Phillips 2009; Turkish: Yanilmaz & Drury 2018b; Korean: Lee et al. 2018). Yet, the processes that lead to the relative acceptability of ungrammatical illusion sentences are still not well understood. The key question is why a comprehension system equipped with a grammar of NPI licensing appears to not respect these grammatical constraints in initial stages of processing. This framing question forces us to consider not only the assumed grammatical knowledge but also the possible mechanisms for implementing grammatical knowledge in a rapid incremental comprehension system.

One initially appealing hypothesis is that an error in signal detection generates confusion of *never* for *ever*, due to the orthographic and phonological similarities of the two words. This might be spelled out within the noisy channel framework (Levy 2008, among others). Crucially, substituting *never* in place of *ever* would provide a grammatical continuation for NPI illusion sentences. Yet, despite the appealing simplicity of such an account, findings from de Dios Flores 2019 show that continuations with *never* are judged unacceptable in online and offline tasks. This means that if the word *ever* in NPI illusion sentences were being mistaken for *never*, we would expect a penalty in acceptability ratings, rather than a boost, contrary to the NPI illusion pattern.

Setting this issue aside, there are two influential approaches to the NPI illusion: one that conceptualizes NPI licensing as fundamentally a memory retrieval operation and places blame for illusions on properties of the memory architecture (Vasishth et al. 2008), and one that highlights pragmatic inferences such as a “rescuing” operation for NPIs not licensed by explicit negation and places blame for illusions on overzealousness in this system (Xiang, Dillon, & Phillips 2009). We will refer to these as the memory-based hypothesis and the pragmatic rescuing hypothesis, respectively.

Under the memory-based hypothesis, memory retrieval operations are executed via parallel cue-based activation of content-addressable items (Lewis & Vasishth 2005). The licensing of an NPI in real time is re-framed as a problem of retrieving a licenser from the memory store of the preceding sentence fragment. In this model, the successful retrieval of an item in memory is the result of the item’s level of activation and the item’s feature-by-feature match to the retrieval cues. For NPI licensing, [+negation] and [+c-

command] have been suggested as candidate retrieval cues, but other cue combinations are possible. The key factor for explaining illusions is the possibility of multiple partial matches. That is, the representation of *no critics* in an illusion sentence like (27b) results in a match of the [+negation] retrieval cue but not the [+c-command] retrieval cue. The existence of this partial match results in a higher probability of acceptance compared to baseline sentences like (27c). This approach treats NPI licensing as analogous to other kinds of dependency resolution like agreement illusions (Wagers, Lau, & Phillips 2009; Jäger, Engelmann, & Vasishth 2017). Note, however, that the profile of these illusions is not identical to that of NPI illusions (see Parker & Phillips 2016, among others).

An alternative proposal attributes illusions to the over-application of the same kinds of indirect pragmatic licensing mechanisms (discussed in section 3.1.1) that have been proposed to account for the acceptability of NPIs in some contexts that are not explicitly negative (Xiang, Dillon, & Phillips 2009). Under this hypothesis, the use of restrictive relative clauses in NPI illusion sentences is critical, since these modifiers generate contrastive implicatures via Gricean mechanisms (e.g. Sedivy et al. 1999). Importantly, these contrastive implicatures are globally negative — that is, for a sentence like (27b), some other set of authors is inferred to have not received acknowledgement, etc. Thus, the same pragmatic mechanism that allows for NPIs in contexts like the scope of *doubt* (see (25) above) may yield an impression of acceptability for NPI illusion sentences. The claim is not that NPI illusions result from valid negative inferences which would yield full acceptability as in (25), but rather, that the possibility of generating negative inferences results in some momentary pragmatic confusion. An explanation along similar lines is proposed in Mendia, Poole, & Dillon 2018, which suggests that some illusions arise because a covert exhaustive operator (e.g. a silent *only*) is inferred, making this contrastive implicature an entailment. Note that Ng & Husband 2017 report NPI illusions in the context of non-restrictive relative clauses, casting doubt on this account.

A third hypothesis, which we will call the scope miscalculation hypothesis, was initially proposed by de Dios Flores, Muller, & Phillips 2017 and later developed by Orth, Yoshida, & Sloggett 2021. This account treats NPI illusions as a result of the parser’s failure to accurately calculate the scope of negative quantifiers. The proposal appeals to the fact that quantifiers are sometimes compatible with multiple

scope interpretations, and effectively treats the acceptance of the main clause (MC) NPI as merely a side effect of an error in determining the interpretation of a negative quantifier phrase like *no critics* in (27b). That is, if *no critics* is assigned a representation such that it takes scope over the whole sentence, the NPI in the MC can easily be licensed, being within the scope of negation in the comprehender's internal representation of the sentence. We directly address this possibility in Experiment 4, where we present further relevant details of the scope miscalculation hypothesis.

3.1.3 The scalar alternatives hypothesis

Orth, Yoshida, & Sloggett 2021 and our Experiments 1-6 show that the uniformity in the illusion that is predicted by both the memory-based hypothesis and the pragmatic rescuing hypothesis does not exist. Rather, NPI illusions appear to be specific to contexts where the RC expresses a negatively quantified meaning. The scope miscalculation hypothesis discussed above can account for some of this specificity, but, as we will show in what follows, it makes several predictions that are not borne out. This pattern of findings motivates another approach, the scalar alternatives hypothesis, which we propose here.

In order to explain how this hypothesis can account for the NPI illusion, let us first consider how scalar licensing hypotheses could be translated into an online licensing mechanism. Scalar approaches explain the restriction on NPIs' distribution as a consequence of NPI-containing sentences' relation to their alternatives. Such theories are not committed to any particular sequence of operations that compute and evaluate those alternatives, since they are not processing theories. Specifically, they do not address whether the alternatives are generated prior to the NPI or only after the NPI is encountered. Both options are in principle possible. Returning to the example *Ann did not lift a finger to help*, one could imagine a context in which Ann's help (or lack thereof) is under discussion, so that upon hearing *Ann did not...* the comprehender is already able to construct the relevant scale (i.e. Ann's possible actions, ranging from those that help a lot to those that help very little). When the comprehender then encounters *...lift a finger*, the only work that remains to be done is to position the current utterance with respect to those alternatives, which will determine whether the NPI is licensed. In contrast, the same statement, uttered out of the blue, would not allow the comprehender to pre-construct the alternatives. Only after the NPI

is encountered would the relevant alternatives be clear.

Such a licensing mechanism invites a reframing of the illusion — if scalar alternatives are what licenses NPIs, it is not the non-c-commanding negative word that interferes to cause illusions, but the non-local ordered scalar alternatives to the RC. These alternatives, if they are represented at the time the NPI is encountered, would function as a “lure” for the NPI. Potential licensing contexts that do not trigger the pre-construction of NPI-licensing alternatives would not yield illusions under such an account, because the interfering representation (that is, the alternatives themselves) would not exist at the point that the NPI is encountered. A key question is why RC alternatives would interfere at all, in the face of a syntactic representation that clearly places the NPI in the MC. We suspect that this has to do with the NPI’s proximity to the RC, such that the syntactic parse is too new or too uncertain to pull attention away from the RC representation. Parker & Phillips 2016 show that the NPI’s position in the MC is critical to the illusion, making this a plausible avenue, but see Chapter 4 for further discussion. For the present purposes, we are interested in whether the RC can be made more or less of a lure by manipulating the probability that the alternatives are pre-constructed.

Such an investigation requires further elaboration of the circumstances under which alternatives may be pre-constructed. Since illusion sentences are typically presented in isolation, it cannot be conversational context that allows for this, as we suggested for *Ann did not lift a finger*. However, independent work on the processing of negation suggests that encountering a negative word can itself trigger alternatives — this is sometimes framed as inferring a Question Under Discussion (see Tian & Breheny 2016, among others), though we will treat the inference of a QUD and the inference of alternatives (i.e., possible answers to a QUD) as interchangeable. We suspect that the alternatives that are typically inferred for a negative quantifier would be both scalar in nature (in virtue of the use of a quantifier) and appropriately ordered for licensing an NPI (in virtue of the use of negation). In contrast, non-negative quantifiers could trigger scalar alternatives but they would be inappropriately ordered to license an NPI and thus would not yield illusions. Similarly, non-quantificational forms of negation may not, on their own, trigger scalar alternatives and thus would also not yield illusions. Thus, under this account, negative quantifiers create the perfect storm for illusory licensing of NPIs.

The experiments presented here test whether illusions arise for non-quantificational forms of negation like *haven't* and *did not*, in light of the prediction that they should not, due to their non-scalar default interpretations. However, these negative forms are obviously compatible with scalar interpretations (under a hypothesis where all NPI licensing is scalar in nature, and these forms can license NPIs, this must be true), and so some brief motivation of our intuition that they are less likely than quantificational negation to be interpreted in a scalar, NPI-licensing way is warranted. Note, as a first step, that the QUD that is typically attributed to sentences with sentential negation in the negation processing literature is typically a polar one, not a scalar one. For example, a sentence like *the boy does not have an apple* may trigger a QUD like *does the boy have an apple?*. Translating this into an alternatives-based framework, we might say that the alternative set for the sentence contains two elements: *the boy has an apple* and *the boy does not have an apple*. In contrast, a sentence with a quantificational form of negation (e.g., *the boy has no apples*) may instead trigger a quantificational QUD (e.g. *how many apples does the boy have?*) and/or scalar alternatives (e.g., *the boy has no apples*, *the boy has few apples*, *the boy has some apples*, etc.). Scalar alternatives are of course possible for sentences with non-quantificational negation. For example, a sentence like *the boy does not have any apples* may, just like our negative quantifier example, trigger a quantificational QUD (e.g. *how many apples does the boy have?*) and/or scalar alternatives (e.g., *the boy has few apples*, *the boy has some apples*, etc.). But unless the comprehender knows a good deal about how the sentence is likely to unfold, these alternatives will not be available until after an NPI arrives. This question becomes particularly relevant in light of our findings from Experiment 6.

Additional motivation for the intuition that negative quantifiers should be more likely than sentential negation to trigger the pre-construction of negative scalar alternatives comes from a preliminary corpus analysis. The different alternatives (or QUDs) that are triggered by quantificational and non-quantificational negation suggest that the contexts in which these negative forms are used may differ. In particular, we expect that there will be overlap in the types of contexts that make negative quantifiers appropriate (in virtue of the QUDs they answer) and the types of contexts that make NPIs appropriate (in virtue of their licensing conditions). Under our assumptions, non-quantificational negation can be used in a wider variety of contexts, both scalar and non-scalar, and so we expect to see less overlap between

these contexts and NPI-containing contexts.

We conducted a corpus search in order to determine whether there is in fact more overlap between uses of quantificational negation and the NPI *ever* than uses of non-quantificational negation and *ever*. We drew 5000 random instances of sentences containing quantificational negation (*no*) and 5000 random instances of sentences containing non-quantificational negation (*-n't* or *not*) from the Corpus of Contemporary American English (Davies 2008). We coded whether the NPI *ever* occurred anywhere in the scope of the negative element. Environments with *no* were five times more likely to also contain *ever* (59, 1.18%) than environments with *not/-n't* (12, 0.24%). The difference was statistically significant ($X^2(1) = 31.34, p < 0.001$). Note that we do not claim that comprehenders track the surface statistics of the co-occurrence of quantificational negation and *ever* versus non-quantificational negation and *ever*, though such a model would also be consistent with the corpus data. Rather, this pattern is consistent with the idea that quantificational and non-quantificational forms of negation are appropriate in different contexts, and that these contexts align with the appropriateness of the use of NPIs in a non-random way.

One key issue for the scalar alternatives hypothesis is that propositional alternatives that are driven by the lexical alternatives to the negative quantifier are not identical to propositional alternatives that are driven by the lexical alternatives to the NPI. That is, quantifier-driven alternatives to *no student has ever complained about the coursework* may include *some / few / all students have (*ever) complained about the coursework*, whereas NPI-driven alternatives to the same sentence may include *no student has recently / frequently / often complained about the coursework*. Put simply, quantifying over individuals does not provide the same scale as quantifying over event times. However, both quantificational determiners and quantificational adverbials have been shown to be quite flexible in what they quantify over. Quantificational adverbials can sometimes be interpreted as if they quantify over individuals: *Gorillas are usually smart* has an interpretation that can be paraphrased as *most gorillas are smart* (Lewis 1975). Quantificational determiners can sometimes be interpreted as if they quantify over events: *Four thousand ships passed through the lock* has an interpretation that can be paraphrased as *there were four thousand ship-passes-through-lock events* (completed by fewer than 4000 ships that each did multiple passings) (Krifka

1990). There are therefore at least two ways the scale of alternatives could be effectively the same: either the negative quantifier is interpreted as if it quantifies over events or the adverbial NPI is interpreted as if it quantifies over individuals. There is also a third possibility, in which the alternatives over which the NPI-licensing mechanism operates are not lexically-driven propositional alternatives, but alternatives constructed along some ad-hoc scale corresponding to the pragmatic function of the clause. For example, a comprehender may infer that the speaker is using the RC *that no critics recommended* to modify *the authors* in order to communicate that the authors being discussed are at the low end of some quality (or recommendability) scale. These three options are not mutually exclusive — it is possible that some comprehenders, on some trials, interpret the statement and its alternatives in one way and other comprehenders, on other trials, interpret the statement and its alternatives differently.

Summing up, the scalar alternatives hypothesis treats the NPI illusion as a consequence of the persistence of a representation of pre-constructed negative scalar alternatives to the RC, assuming a scalar approach to NPI licensing. Given our assumptions about the alternatives that are typically triggered in the comprehension of *no* versus *not*, we expect that negative scalar alternatives should be less likely to be pre-constructed for *not* and so, under this account, illusions should be less likely to occur for this form of negation.

3.1.4 The present investigation

Here we are primarily concerned with two of the hypotheses sketched above: the scalar alternatives hypothesis and the scope miscalculation hypothesis. Our experiments 1-3 demonstrate that a key prediction which these two hypotheses share — but which neither the memory-based hypothesis nor pragmatic rescuing hypothesis can accommodate — is borne out. That is, we find a contrast between quantificational and non-quantificational forms of negation. Orth, Yoshida, & Sloggett 2021 report similar findings.

Experiment 4 addresses a key prediction of the scope miscalculation hypothesis. Using a sentence comprehension task, we evaluate whether NPI illusion sentences are interpreted in a way that suggests errors in assigning quantifier scope. Finally, Experiments 5 and 6 address a prediction of the scalar alternatives hypothesis. This hypothesis highlights the meanings of NPI-licensing contexts, rather than

the properties of NPI licensors themselves. Thus, these last experiments address the role of meaningful content in the RC other than the licensor.

3.2 Contrasting licensor types

We now turn to the question of whether all non-c-commanding negative words cause NPI illusions. Uniformity in the illusion is a clear prediction of the memory-based hypothesis and pragmatic rescuing hypothesis, whereas other explanations like the scalar alternatives hypothesis and the scope miscalculation hypothesis allow for variability across licensors, and specifically can accommodate contrasts between quantificational and non-quantificational licensors. To test whether the illusion is in fact uniform across licensors, we compared standard illusion sentences like (28a) to sentences like (28b), in which the illusory licensor is non-quantificational.

- (28) a. * The authors [that no critics recommended for the award] have ever received ...
b. * The authors [that the critics did not recommend for the award] have ever received ...
... acknowledgement for a best-selling novel.

Experiment 1 used untimed acceptability ratings in order to verify that our stimuli are appropriate. Experiments 2 and 3 used speeded acceptability judgments to determine whether quantificational and non-quantificational licensors behave similarly with respect to illusions.

3.2.1 Experiment 1: offline acceptability

In order to establish that native speakers' perception of the experimental materials was as expected when given ample time, we conducted an untimed acceptability experiment. The materials used in this and the following experiments were adapted from Parker & Phillips 2016 by adding a fourth condition with non-quantificational negation. Additional minor edits were made to about half of the stimuli in order to improve naturalness, match sentence length across stimuli, and remove modals, due to possible variability in the interaction of different forms of negation with the modal. Based on Parker and Phillips' untimed

acceptability ratings, we expected to obtain a clear pattern of grammatical sensitivity: high acceptability ratings for sentences in which the NPI *ever* is licensed by a c-commanding negation and low ratings for sentence types that lack a structurally-relevant negative word. No differences were expected among the three ungrammatical conditions.

3.2.1.1 Participants

16 US-based native speakers of English participated in this experiment. All participants provided informed consent and they received \$2 as compensation. In this and the following experiments participants were recruited using Amazon Mechanical Turk. In order to ensure that the participants were native speakers of English they were asked to complete a native speaker qualification test (see supplementary materials) and only participants that answered at least 7 out of 9 questions correctly were allowed into the task. We excluded from our analyses workers whose judgments of filler trials did not reliably distinguish between grammatical and ungrammatical fillers, based on a one-sided t-test. All 16 participants met these criteria. The average rating for grammatical fillers was 6.07 out of 7, with a standard deviation of 0.59, and the average rating for ungrammatical fillers was 3.27, with a standard deviation of 0.85.

3.2.1.2 Materials

The experimental materials consisted of 36 sets of items across 4 conditions that varied the presence, location, and type of licenser with respect to the NPI *ever*. This manipulation resulted in the four experimental conditions shown in Table 3.1. Conditions A, B, and D correspond to the standard NPI illusion conditions: grammatical baseline, embedded-*no*, and ungrammatical baseline, respectively. Condition C (embedded-*not*) uses a similar structure to condition B in that it contains a negative word that is structurally irrelevant to the NPI, but in this condition we use non-quantificational negation.

Each participant rated 108 sentences: 36 experimental items and 72 fillers of similar length and complexity. The experimental items were distributed across 4 lists using a Latin Square design and the fillers were the same in each list. Of the 72 total fillers, 42 were constructed to include a range of violations in order to encourage full use of the 1 to 7 scale. The remaining 30 filler sentences were grammatical. Partic-

A. Grammatical baseline	No authors [that the critics recommended] have ever received ...
B. Embedded <i>no</i>	The authors [that no critics recommended] have ever received ...
C. Embedded <i>not</i>	The authors [that the critics did not recommend] have ever received ...
D. Ungrammatical baseline	The authors [that the critics recommended] have ever receivedacknowledgement for a best-selling novel.

Table 3.1: Example stimuli for Experiment 1 and Experiment 2. Brackets indicating the RC boundaries and bold-face font indicating the licenser and NPI were not used in the actual experiment.

ipants completed 6 practice items before beginning the experiment, to ensure that they understood the procedure.

3.2.1.3 Procedure

The sentences were presented using Ibex Farm and the presentation order was randomized for each participant. The instructions asked participants to rate the sentence’s acceptability using a 7-point scale in which 7 was the most acceptable value and 1 the least acceptable. Each sentence was displayed on the screen together with the scale, and participants could take as much time as needed before providing their rating. The task was completed by all participants in less than 20 minutes.

3.2.1.4 Analysis

The results were analyzed using a helmert-coded linear mixed-effects model whose maximal structure was initially built including by-subject and by-item random intercepts and slopes for the experimental conditions. When this model failed to converge, it was reduced according to the recommendations provided by Barr et al. 2013. Further details are included in Supplementary Files. The pairwise comparisons of interest are the following: grammatical baseline versus ungrammatical baseline; grammatical baseline versus embedded-*no*; grammatical baseline versus embedded-*not*; embedded-*no* versus ungrammatical baseline; embedded-*not* versus ungrammatical baseline. The first three comparisons determined whether participants showed sensitivity to the grammaticality manipulation for all three ungrammatical conditions. The last two comparisons determined whether the intended illusion conditions were just as unacceptable as the ungrammatical baseline. We used the emmeans package (Lenth et al. 2018) to extract beta coeffi-

cients and p-values for pairwise comparisons between conditions.²⁵ The comparisons presented here use linear mixed effects models, but a model with similar structure but which treats the dependent variable as ordinal rather than linear yields similar conclusions and can be found in the Supplementary Files.

3.2.1.5 Results

The results from this experiment are presented in Figure 3.1. The model results revealed a clear effect of grammaticality shown by significant differences between the grammatical baseline condition and the other three experimental conditions (grammatical baseline versus embedded-*no*: $\beta=-2.88$, $SE=0.39$, $t=-7.40$, $p<.001$; grammatical baseline versus embedded-*not*: $\beta=-3.49$, $SE=0.37$, $t=-9.35$, $p<.001$; grammatical baseline versus ungrammatical baseline: $\beta=-3.42$, $SE=0.37$, $t=-9.17$, $p<.001$). Furthermore, sentences containing embedded-*no* were rated statistically significantly higher than ungrammatical baseline sentences ($\beta=0.54$, $SE=0.22$, $t=2.39$, $p=.03$), though this effect was numerically small: on average, a 0.54-point difference on a 7-point scale. No differences were observed between sentences containing embedded-*not* and ungrammatical baseline sentences ($\beta=-0.07$, $SE=0.12$, $t=-0.58$, $p=.58$).

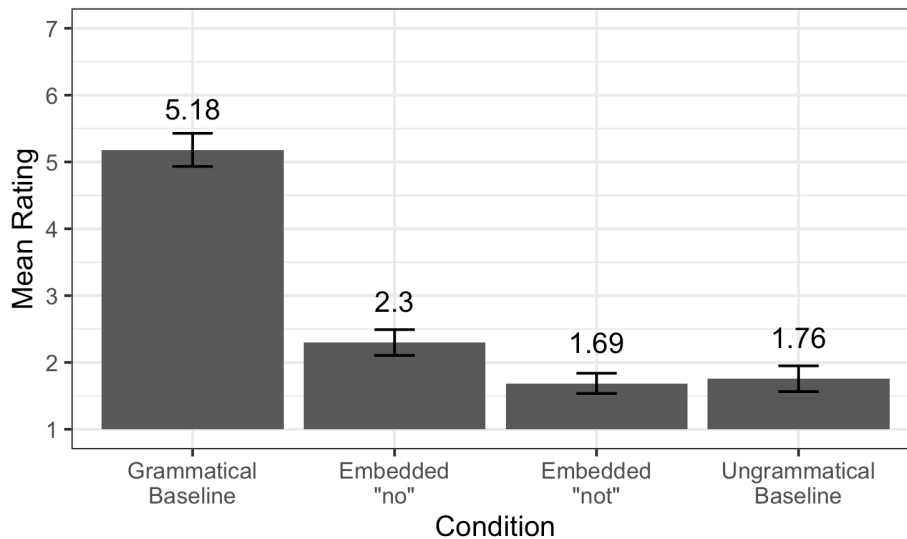


Figure 3.1: Mean ratings for the experimental conditions in Experiment 1. Error bars indicate standard error of the mean across subjects.

²⁵Unless otherwise noted, p-values are not corrected for multiple comparisons, since the critical comparisons were determined a priori.

3.2.1.6 Discussion

The results showed that participants clearly identified the grammatical baseline condition as acceptable and the ungrammatical baseline condition as unacceptable. In addition, the ratings observed for the two conditions containing embedded negative elements were highly degraded relative to the grammatical baseline. In this regard, these results confirm that, in an untimed task, speakers are sensitive to NPI licensing contrasts in our materials.

Nonetheless, we observed a small boost for the embedded-*no* condition when compared to the ungrammatical baseline condition. Note that similar patterns have been observed in other NPI illusion experiments (e.g. Xiang, Dillon, & Phillips 2006; Yanilmaz & Drury 2018b). There are a few possible explanations for this boost. First, the difference in acceptability could, in principle, be due to a difference in grammatical status. Since this would amount to challenging the overwhelming generalization that NPIs can only be licensed when in the scope of a licenser, we do not pursue this possibility. Another possibility is that the conditions differed in their acceptability due to processing factors independent of the NPI. This is plausible, but not especially likely, since the only difference between the embedded-*no* sentences and embedded-*not* sentences was the form of negation. Negative quantifiers are less frequent than non-quantificational negation and so we might expect a small penalty for the embedded-*no* sentence due to this, but instead we find a boost. It is also possible that we are observing an illusion, even in an untimed task. Note that the fact that we allow participants as much time as they need to arrive at a confident judgment does not guarantee that they will take that time. In fact, Mechanical Turk workers are strongly incentivized to complete tasks as quickly as possible. In principle, one could remove the fastest responses from the data set in an attempt to get a better picture of the untimed judgments. However, raw reaction times (RTs) may not be a particularly good indicator of the amount of time a participant spent considering their judgment, since a participant could click to begin a trial, turn to a different task, and then return to our experiment, read the sentence and give their first impression, resulting in a quick judgment with a very slow RT. Thus, the possibility that our “offline” data reflects some speeded judgments cannot be ruled out. Exploratory analyses of RTs can be found in the Supplementary Files. Finally, it is of course possible that we are simply observing noise and the statistically significant ($p=.03$) finding

is a false positive. We are unable to determine definitively whether the cause of the boost is a difference in naturalness, a persistent illusion, or a false positive, but due to the small effect size and the fact that embedded-*no* sentences are overwhelmingly judged unacceptable, it is appropriate to proceed with these items in a speeded-acceptability task.

3.2.2 Experiment 2: speeded acceptability

In this and the following experiment we use a speeded acceptability judgment task to investigate the contrast between embedded quantificational negation and embedded non-quantificational negation in the emergence of illusion effects. Following previous speeded acceptability findings, we expect to replicate the well-documented illusion for sentences containing embedded quantificational negation *no*. The key question here is whether sentences containing embedded non-quantificational forms of negation — in this case, *not* — also generate illusion effects and if these effects are similar to those for quantificational negation.

3.2.2.1 Participants

35 US-based native speakers of English participated in this experiment. All participants provided informed consent and they received \$3 as compensation. We excluded workers who failed to provide a response within 2 seconds for 25% of fillers or more and workers whose judgments of filler trials did not reliably distinguish between grammatical and ungrammatical fillers, based on a chi-squared test. Our goal in using this relatively weak threshold was merely to identify those who were not attending to the task and clicking randomly (note that the native speaker questionnaire was used here too). 4 workers were excluded based on these criteria, resulting in 31 participants in our analysis. The mean filler-trial accuracy of the included participants was 77%.

3.2.2.2 Materials

The materials used in this task were the same 36 sets of experimental items and 72 filler sentences that were used in Experiment 1. Participants saw 3 practice trials before beginning the experiment.

3.2.2.3 Procedure

In the speeded acceptability task, each sentence was displayed word by word at a rate of 400 ms. per word, in the center of the screen, using the RSVP paradigm. At the end of each sentence participants were asked to provide a yes/no button press judgment in response to the question “Was that a good sentence?” within 2 seconds. If participants failed to provide the judgment in time, a message indicated that they were too slow. The dependent measure was the acceptance rate across trials and participants. Although sentence-final judgments are relatively late for probing incremental representations, this method has been reliably used in prior studies of NPI illusions (e.g. Drenhaus, Saddy, & Frisch 2005; Parker & Phillips 2016; Parker 2019; Orth, Yoshida, & Sloggett 2021) as well as other varieties of grammatical illusions. There are at least two main advantages to this method: first, the effect size for illusions is large relative to more implicit measures like reading times, and second, data can be easily collected over the internet. Note that while methods like self-paced reading and eye-tracking while reading are sometimes preferred because they allow for detailed information about the time-course of sentence processing, in this case, timing information is not necessary to answer questions like whether illusions occur. In our task, participants were instructed to read the sentences carefully and judge whether they came across as well-formed English. The task lasted for approximately 30 minutes and the order of presentation for experimental and filler sentences was randomized for each participant.

3.2.2.4 Analysis

Results were analyzed using the same strategy as in Experiment 1. A generalized linear mixed effects model using helmert coding and a logit link function was fit, and pairwise comparisons were computed using emmeans. The critical pairwise comparisons are as follows. The first comparison (ungrammatical baseline versus grammatical baseline) functions as a sanity check to ensure that basic grammaticality effects arise. The second comparison (ungrammatical baseline versus embedded-*no*) replicates the standard illusion effect for embedded *no*. The third comparison (ungrammatical baseline versus embedded-*not*) determines whether a reliable illusion effect arises for embedded *not*. Our final comparison (embedded-*no* — embedded-*not*) determines whether illusion effects for embedded *no* and embedded *not* are the same

magnitude.

3.2.2.5 Results

The results from this experiment are presented in Figure 3.2, which shows the proportion of “yes” responses given to each condition. An effect of grammaticality was observed ($\beta=6.51$, $SE=0.88$, $z=7.40$, $p<.001$), indicating that the grammatical baseline condition was significantly more likely to be judged acceptable than the ungrammatical baseline condition. An effect of embedded-*no* was observed ($\beta=1.29$, $SE=0.38$, $z=3.34$, $p<.001$), replicating the standard illusion effect for negative quantifiers. No effect of embedded-*not* was observed ($\beta=0.39$, $SE=0.32$, $z=1.22$, $p=.22$). The comparison of embedded-*not* and embedded-*no* revealed a significant effect of the type of embedded negation ($\beta=0.90$, $SE=0.37$, $z=2.46$, $p=.01$), indicating that the embedded-*no* condition was significantly more likely to be judged acceptable than the embedded-*not* condition.

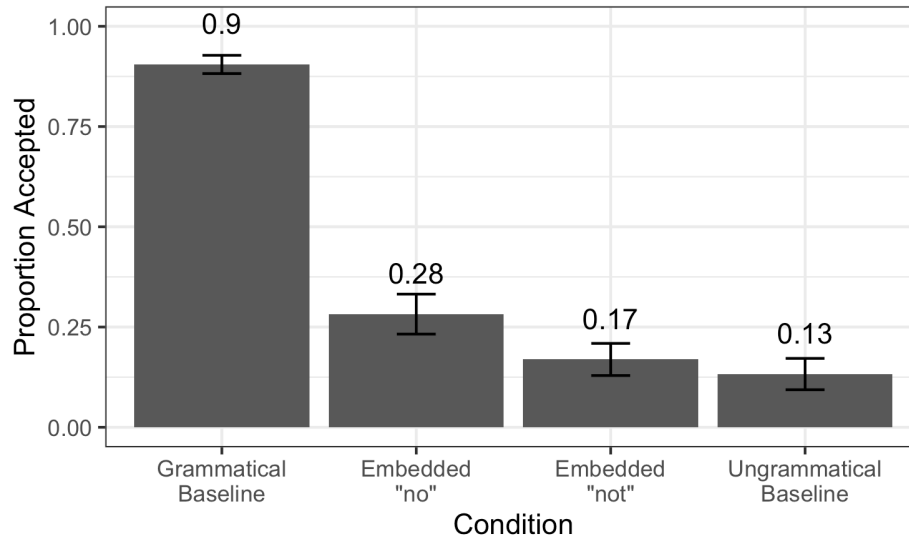


Figure 3.2: Mean percentage of ‘YES’ responses for the experimental conditions in Experiment 2. Error bars indicate standard error of the mean across subjects.

3.2.2.6 Statistical power

Post hoc power analyses were conducted using *simr* (Green & MacLeod 2016). The problems with using an observed effect size to compute power post hoc are well documented (see Goodman & Berlin 1994;

Levine & Ensom 2001; among others). We therefore did not use observed effect sizes, but rather derived an independent estimate of what the illusion effect for embedded-*not* would be, if it existed and if it were similar in magnitude to previously-observed illusion effects for embedded-*no*. We pooled the raw data from the manipulations where illusions were reported from Parker & Phillips 2016 Experiments 2, 4, and 6 (the experiments that used speeded acceptability measures) and Xiang, Grove, & Giannakidou 2013. We fit a logistic mixed effects model to this pooled data set in a similar manner to the analyses of our own data, and computed the lower boundary of a 60% confidence interval (see Perugini, Gallucci, & Costantini 2014) around the meta-analytic estimate of the pairwise comparison between embedded negative quantifiers and the ungrammatical baseline, arriving at an effect size estimate of $\beta=0.95$. We then set the illusion effect for embedded-*haven't* in the present experiment to be equal to this effect and computed power using simulations. The present experiment achieved between 96% and 98% power to detect an illusion for embedded-*not*, assuming an effect size of $\beta=0.95$.

3.2.2.7 Discussion

Experiment 2 replicated the widely attested illusion effect for sentences containing *no* as an embedded licenser. Illusions were not found when the embedded licenser was *not*. Note that there are, effectively, two things we may wish to know about illusions with embedded *not*: whether they occur at all and whether they occur at similar rates to illusions with embedded *no*. The present experiment provides a clear answer to the second question: illusions for embedded *not* are reliably smaller than illusions for embedded *no*, as indicated by the statistically significant contrast between these two conditions. However, regarding the first question, while the null result comparing the embedded *not* condition to the ungrammatical baseline is consistent with a lack of illusions for embedded *not*, null results can of course arise for many reasons and do not indicate evidence for a lack of an effect. Since this experiment was conducted without the aid of a prospective power analysis, additional caution is warranted. Thus, we are able to infer that the illusion for embedded *not* is reliably smaller than the illusion for embedded *no*, but not necessarily zero.

Before discussing the implications of this finding in detail, it is necessary to address two potential con-

findings. First, participants' judgments throughout the task could be influenced by comparisons made across trials. Since our experiment included both quantificational negation and non-quantificational negation conditions, it is in principle possible that illusions for non-quantificational negation were suppressed in our experiment due to cross-trial comparisons. A separate issue has to do with the fact that sentences containing the non-contracted form of non-quantificational negation (*did not* as opposed to *didn't*) might be perceived as less natural. The lower acceptance rate we observed for this condition could be due to that unnaturalness. We addressed both of these issues in Experiment 3.

3.2.3 Experiment 3: speeded acceptability

In order to address the concerns stated above, some small modifications were introduced to the materials and the design. First, the critical illusion conditions (embedded-*no* and embedded-*not*) were tested as a between-subjects factor to avoid possible complications due to cross-trial comparisons. This resulted in two sub-experiments, each of which included the same grammatical and ungrammatical baseline conditions and only one of the embedded-licensor conditions. Second, in the non-quantificational negation condition we used a contracted form to increase naturalness.

3.2.3.1 Participants

49 participants were recruited for this task. The first group (24 participants) completed the sub-experiment with embedded-*no*. The second group (25 participants) completed the sub-experiment with embedded-*haven't*. All participants provided informed consent and received \$3 as compensation. We excluded workers who failed to provide a response within 2 seconds for 25% of fillers or more and workers whose judgments of filler trials did not reliably distinguish between grammatical and ungrammatical fillers, based on a chi-squared test. 1 worker was excluded from our analyses for the *haven't* sub-experiment resulting in 24 participants per sub-experiment (48 participants total). The mean filler-trial accuracy of the included participants was 81%.

A. Gramm. baseline	No authors [that the critics have recommended in their reviews] have ever ...
B. Embedded <i>no</i>	The authors [that no critics have recommended in their reviews] have ever ...
C. Embedded <i>not</i>	The authors [that the critics haven't recommended in their reviews] have ever ...
D. Ungramm. baseline	The authors [that the critics have recommended in their reviews] have ever ...
...received acknowledgement for a best-selling novel.	

Table 3.2: Example stimuli for Experiment 3

3.2.3.2 Materials

The experimental materials used in this task consisted of 36 sets containing the same four experimental conditions as Experiments 1 and 2. The experimental materials used here introduced some minor changes with respect to the previous experiments (see Table 3.2). As already noted above, we used a contracted form of non-quantificational negation to increase naturalness. We also used the present perfect in the RC (where previously we had used the simple past), so that all four conditions had the same structure and the same number of words.

For the *haven't* sub-experiment, we selected the grammatical baseline, ungrammatical baseline, and embedded-*haven't* versions of our 36 items, and created three lists from these, together with 90 filler sentences of similar internal structure, length and complexity. Similarly, we constructed three lists for the *no* sub-experiment from the same set of items, using the embedded-*no* condition. The number of filler items was slightly increased with respect to the previous experiments in order to increase variability within the task. Grammaticality was balanced so that approximately half of the sentences were ungrammatical across the task. Each list had a total of 126 items and participants were randomly assigned to one of the six lists. Participants additionally completed 6 practice items.

3.2.3.3 Procedure

The speeded acceptability procedure followed the same steps as in Experiment 2. The task lasted for approximately 35 minutes.

3.2.3.4 Analysis

A generalized linear mixed effects model using a logit link function and helmert coding, treating sub-experiment and condition as crossed factors, was fit. As with previous experiments, critical comparisons were computed using emmeans. The key comparisons are as follows. Within each sub-experiment, we tested the comparison between the ungrammatical baseline and the grammatical baseline as a sanity check to ensure that basic grammaticality effects arise. We tested the comparison between the ungrammatical baseline and the embedded-negation condition to determine whether illusions arise for each form of negation, *no* and *haven't*. We additionally tested whether an interaction arose between this contrast (i.e., the illusion effect) and sub-experiment (i.e., the effect of the form of embedded negation), which would indicate a difference in the magnitude of illusions.

3.2.3.5 Results

The results are shown in Figure 3.3. An effect of grammaticality was observed for both the embedded-*no* sub-experiment ($\beta=4.38$, $SE=0.58$, $z=7.56$, $p<.001$) and the embedded-*haven't* sub-experiment ($\beta=5.35$, $SE=0.63$, $z=8.45$, $p<.001$), indicating that in both cases the grammatical baseline condition was significantly more likely to be judged acceptable than the ungrammatical baseline condition. We additionally observed an effect of embedded negation for the embedded-*no* sub-experiment ($\beta=1.41$, $SE=0.25$, $z=5.62$, $p<.001$), indicating that the embedded-*no* condition was significantly more likely to be judged acceptable than the ungrammatical baseline condition. The corresponding comparison for the embedded-*haven't* sub-experiment was not statistically significant ($\beta=-0.06$, $SE=0.28$, $z=-0.22$, $p=.83$). We additionally observed a significant illusion by sub-experiment interaction ($\beta=1.47$, $SE=0.37$, $z=3.93$, $p<.001$), indicating that the magnitude of the difference between the embedded-negation condition and the ungrammatical baseline condition was reliably different for the two sub-experiments.

3.2.3.6 Statistical power

Post hoc power analyses were again conducted using simr, and using an adjusted effect size based on data from Parker 2019 and Xiang, Grove, & Giannakidou 2013. The present experiment achieved between

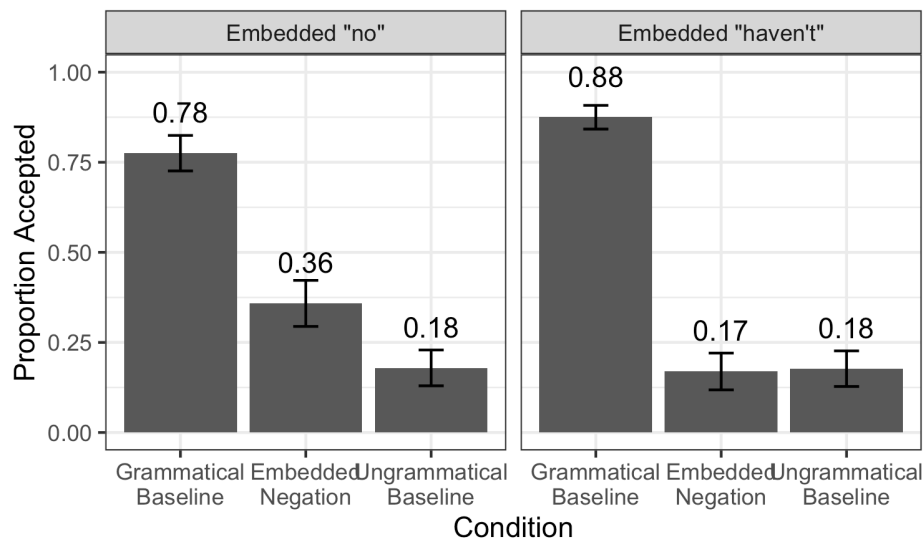


Figure 3.3: Mean percentage of ‘YES’ responses for the experimental conditions in Experiment 3. Error bars indicate standard error of the mean across subjects.

89% and 98% power to detect an illusion for embedded-*haven't*, assuming an effect size of $\beta=0.95$.

3.2.3.7 Discussion

The results from Experiment 3 replicated those from Experiment 2 in all relevant aspects. First, there was a clear illusion effect for sentences containing embedded-*no*. Second, we found a contrast between quantificational and non-quantificational forms of negation. Both experiments also failed to reveal a statistically significant illusion effects for sentences containing embedded non-quantificational negation, but we again note that non-significant findings are compatible with a range of true effects. Since the results from Experiments 2 and 3 align, we conclude that the use of uncontracted negation *not* and the presentation of the two embedded conditions together were not critical to the results of Experiment 2. What is certain so far is that NPI illusions are not general across all forms of embedded negation. We now turn to the four hypotheses presented in section 3.1.2 in light of these findings.

First, the memory-based hypothesis (Vasishth et al. 2008), straightforwardly predicts that the erroneous retrieval of embedded licensors should be uniform across *no*, *not*, and *haven't*, because all of these share the relevant features in their encodings (i.e. [+negation]). Any attempt to adapt the feature set to capture the observed difference (e.g. adding [+quantificational]) will lead to inappropriate predictions

in cases of true licensing, since both *no* and *not/haven't* are perfectly capable of licensing an NPI within their scope. Our findings thus suggest that NPI illusions are not a consequence of erroneous retrievals of a partially feature-matching lexical licenser in memory. Second, recall that the pragmatic rescuing hypothesis (Xiang, Dillon, & Phillips 2009; Xiang, Grove, & Giannakidou 2013) proposes that negative contrastive inferences are responsible for illusions. For example, sentences like (29a), a standard NPI illusion configuration, license the inference in (29b), which, under this hypothesis leads to increased acceptance of NPIs within P. However, a very similar inference is supported by sentences like (30a), which includes non-quantificational negation *haven't* in the RC. Thus, the hypothesis predicts uniformity in the illusion across at least these two licensors, contrary to our findings.

- (29) a. The authors [that no critics recommended] have P
- b. The authors [that some critics have recommended] have NOT P
- (30) a. The authors [that the critics haven't recommended] have P
- b. The authors [that the critics have recommended] have NOT P

The scope miscalculation hypothesis, by contrast, predicts that NPI illusions will be specific to sentences with embedded negative quantifiers, because it is the mis-assignment of quantifier scope that causes the NPI illusion under this account. Note that this hypothesis predicts not only a contrast between embedded-*no* and embedded-*not/haven't* but also a complete lack of illusions for embedded-*not/haven't*. We of course cannot confirm or refute this second prediction, but our findings thus far are compatible with it. The scalar alternatives hypothesis also predicts that illusions will be less likely for non-quantificational forms of negation, as compared to quantificational forms of negation, as was explained in detail in section 3.1.3. Unlike the scope miscalculation hypothesis, this hypothesis does not make strong predictions about whether the illusion rate for embedded-*haven't* should be zero or non-zero, but does predict a contrast with embedded-*no*. Thus, in light of the data presented so far, we find that both the scope miscalculation hypothesis and the scalar alternatives hypothesis make appropriate predictions. We now turn our attention to the interpretation of NPI illusions, in an attempt to discriminate between these hypotheses.

3.3 The interpretation of NPI illusions

While there is substantial prior work manipulating the factors that cause illusions, very little is known about the interpretation a reader comes away with after experiencing an illusion. This is an understandable gap, since probing interpretations is methodologically difficult and it is not always clear that the hypotheses being considered make clear predictions about the interpretation. However, the scope miscalculation hypothesis does in fact make predictions about comprehenders' sentence-final interpretations, and so we tested these predictions in the present experiment.

3.3.1 Experiment 4: interpretation and acceptability

The key claim of the scope miscalculation hypothesis is that NPI illusions arise due to problems in the representation of the scope of the licenser. Under this hypothesis, the licenser is represented as if it takes scope over the whole MC ("wide scope"), not just the RC in which it occurs ("narrow scope"). Once this error has been made, the NPI can be easily licensed by the negative element, which, in the comprehender's internal representation, does take scope over it. Under this hypothesis, non-quantificational negation is invulnerable to this scope error because only quantifiers have such a wide set of possible scope interpretations.

Here we address a straightforward prediction of the scope miscalculation hypothesis: if the acceptance of the main-clause NPI is a consequence of the negative representation of the MC (which is itself a consequence of the quantifier scope error), then the trials on which illusion sentences are accepted should be trials on which the comprehender understands the sentence to be expressing a negative MC. That is, under this hypothesis, after hearing a sentence like (31), comprehenders should, on some trials, come away believing that the authors in question have not received acknowledgment for their novels, because the MC is under the scope of negation in their internal representation. On exactly these negatively-interpreted trials, the comprehender should accept the sentence, under this hypothesis.

- (31) * The authors that no critics have recommended have ever received acknowledgment for a best-selling novel.

Additionally, the trials that are not accepted (recall that NPI illusions occur on 20-50% of trials, across experiments) should be interpreted as if the MC is affirmative — that is, the authors did receive acknowledgement. Experiment 4 tests this predicted correspondence between negatively-interpreted trials and accepted trials with a task in which comprehenders answer two questions on every trial: a comprehension question which probes the polarity of the MC and an acceptability question. It is worth noting that the scope miscalculation hypothesis makes a very strong prediction here — it is not just that illusion rates should be higher for negatively-interpreted trials, but that illusions should arise for all and only the negatively-interpreted trials, because negative interpretations and illusions are simply two ways to probe a single underlying error, the wide-scope representation of the negative quantifier.

Before proceeding to the details of this experiment, we note that the scope miscalculation hypothesis as we have presented it here is underspecified in an important respect — namely, the relative timing of the commitment to the scope of the negative quantifier. Under one version of this hypothesis, which we will label the “early scope-assignment” version, the comprehender stochastically assigns either wide or narrow scope to the quantifier when it is first encountered. The acceptance or rejection of the NPI is then effectively a side effect of this prior commitment. This version of the hypothesis is therefore committed to the occurrence of wide-scope interpretations of the quantifier on a certain proportion of trials regardless of how later parts of the sentence unfold — that is, regardless of whether an NPI even appears. In contrast, under the “late scope-assignment” version of the hypothesis, the comprehender does not commit to a particular scope position for the negative quantifier right away. Both scope assignments can be generated in parallel, and only upon encountering the NPI is there clear evidence favoring the wide-scope interpretation of the quantifier, as this is the scope assignment that can “save” the NPI. Prior work on the assignment of quantifier scope in genuinely ambiguous contexts (e.g., Tunstall 1998; Anderson 2004) has some potential to shed light on which version of the scope miscalculation hypothesis is more promising, though it is not clear to what extent we should expect the assignment of an ungrammatical scope interpretation to function the same way as the resolution of genuine ambiguity. We return to this issue in section 3.3.1.7, but for the present purposes we consider both versions of the scope miscalculation hypothesis to be equally plausible.

We tested an additional prediction in Experiment 4, which is specific to the early scope-assignment version of the hypothesis. Because this version of the hypothesis is committed to wide scope interpretations of the negative quantifier arising prior to, and independent of, the later-arriving NPI, it predicts that globally negative interpretations should arise for sentences with embedded quantifiers even if an NPI never arrives. Thus, Experiment 4 additionally tested sentences like (32), which should receive negative interpretations at rates similar to sentences like (31) under this hypothesis.

- (32) The authors that no critics have recommended have received acknowledgment for a best-selling novel.

To briefly summarize, there are two key predictions to be tested by Experiment 4. First, all versions of the scope miscalculation hypothesis share the prediction that negative interpretations of NPI illusion trials and acceptance of NPI illusion trials should go hand in hand. Second, only the early scope-assignment version of the hypothesis predicts that negative interpretations of the MC should arise just as often for embedded-negative-quantifier sentences with a main-clause NPI as they do for embedded-negative-quantifier sentences without a main-clause NPI.

3.3.1.1 Participants

We recruited 33 participants using Amazon Mechanical Turk. Workers received \$10 for completion of the 45-minute task. We excluded workers whose judgments of filler trials did not reliably distinguish between grammatical and ungrammatical fillers, based on a chi-squared test, as well as those who failed either the “instructions quiz” (see “Procedure” section 3.3.1.3 below) or the “attention checks” (see “Materials” section 3.3.1.2 below). 4 workers were excluded from our analyses resulting in 29 participants. The mean filler-trial accuracy of the included participants was 79% for acceptance judgments and 92% for interpretation judgments.

3.3.1.2 Materials

This experiment probed interpretations by asking participants directly about the asserted content of the MC. The comprehension questions for experimental trials were always written such that a “no” answer

With <i>ever</i>	
A. Main clause <i>no</i>	No authors [that the critics recommended] have ever received ...
B. Embedded <i>no</i>	The authors [that no critics recommended] have ever received ...
C. Embedded <i>didn't</i>	The authors [that the critics didn't recommend] have ever received ...
D. Without negation	The authors [that the critics recommended] have ever received ...
...acknowledgment for a best-selling novel.	
Without <i>ever</i>	
E. Main clause <i>no</i>	No authors [that the critics recommended] have received ...
F. Embedded <i>no</i>	The authors [that no critics recommended] have received ...
G. Embedded <i>didn't</i>	The authors [that the critics didn't recommend] have received ...
H. Without negation	The authors [that the critics recommended] have received ...
...acknowledgment for a best-selling novel.	

Table 3.3: Example stimuli for Experiment 4

indicated that the participant believed that the MC assertion was under the scope of negation and a “yes” answer indicated that the participant believed that the MC assertion is not under the scope of negation. The 36 item sets from Experiment 3 were used to generate the eight conditions shown in Table 3. This included the four conditions with *ever* with various licensor positions, as well as four conditions which are identical except that the NPI *ever* has been omitted. These four NPI-free conditions allow us to determine whether interpretations corresponding to erroneous quantifier scope assignment arise independent of the NPI. We also included 72 fillers, which were balanced both for “yes” and “no” interpretation responses and for grammaticality. Before beginning the experiment, participants judged six practice trials. We also included eight attention check trials, which were intended to identify participants who were clicking answers randomly without reading the sentences. For example, the sentence for one attention check trial read “After this sentence, please choose ‘no’ as the answer.” It was then followed by only one question, “What is the answer?”. Participants who responded incorrectly to two or more attention check trials were excluded from our analyses (see “Participants” section 3.3.1.1 above).

3.3.1.3 Procedure

Participants viewed sentences presented one word at a time with rapid serial visual presentation, as in Experiments 2 and 3. At the conclusion of each sentence, participants first gave an untimed binary judgment of acceptability, and then were asked a comprehension question about the MC. For example, the

sentences in Table 2 were all followed by the question “Have the authors received acknowledgement for a novel?”. The instructions were careful to instruct participants to interpret definite descriptions as referring to the same set of individuals as was mentioned in the sentence (i.e. *the authors* in the question refers to the same authors that were discussed in the sentence), so that the questions could be stated without the inclusion of RCs in the questions themselves. The possible responses were “yes”, “no”, and “I can’t answer”. The “I can’t answer” option was to be used in cases where the participant found the sentence not only ungrammatical but uninterpretable. There was no time limit for responses. Because this task was somewhat more complex than previous experiments, we also included a brief quiz following the instructions to ensure that participants had read and understood the instructions correctly. For example, participants were asked “When should you choose ‘I can’t answer?’” and had to select from the options (A) “Whenever I don’t have an opinion on the question”, (B) “Whenever the sentence was ungrammatical”, or (C) “When I can’t determine what the sentence meant”. The correct answer was displayed after participants made their selection. Participants who responded incorrectly to two or more instructions quiz questions were excluded from our analyses (see “Participants” section 3.3.1.1 above).

3.3.1.4 Analysis

Trials receiving an “I can’t answer” comprehension question response were removed from our analyses (both acceptability data and comprehension data). This amounted to 82 trials across our eight experimental conditions (7.8% of the total number of experimental trials). Results were analyzed using logistic mixed effects models, as in Experiments 2 and 3. For analyses of comprehension data, we are interested in the probability of a negative interpretation, so responses were coded as 1 in the case that comprehenders gave a “no” response and 0 in the case that comprehenders gave a “yes” response. Experimental manipulations were helmert coded and critical comparisons were extracted using emmeans.

3.3.1.5 Results

Overall results are shown in Figure 3.4 and Figure 3.5. We focus first on the NPI illusion data — that is, acceptance rates for the four conditions that included *ever*. An effect of grammaticality was observed

($\beta=4.85$, $SE=0.69$, $z=6.99$, $p<.001$), indicating that the main-clause-*no* condition²⁶ was significantly more likely to be judged acceptable than the without-negation condition²⁷. An effect of embedded-*no* was observed ($\beta=1.98$, $SE=0.50$, $z=3.93$, $p<.001$), replicating the standard illusion effect for negative quantifiers. No effect of embedded-*didn't* was observed ($\beta=0.19$, $SE=0.52$, $z=0.37$, $p=.71$). An effect of the type of embedded negation was also observed ($\beta=1.79$, $SE=0.47$, $z=3.82$, $p<.001$), indicating that embedded-*no* sentences were more likely to be accepted than embedded-*didn't* sentences.

To evaluate the key prediction of the early scope-assignment version of the hypothesis, we ask whether there is an impact of the presence of *ever* on negative interpretation rates for embedded-*no* sentences. A model of the same eight conditions, but which treated negative interpretation rates, instead of acceptance rates, as the dependent variable was fit. Focusing on only the negative interpretation rates for these two conditions (embedded-*no* with *ever* versus embedded-*no* without *ever*), the model revealed a statistically significant effect of the presence of *ever* ($\beta=3.90$, $SE=0.55$, $z=7.13$, $p<.001$). This finding indicates that embedded-quantifier sentences containing *ever* were in fact substantially more likely to be interpreted negatively than embedded-quantifier sentences not containing *ever*.

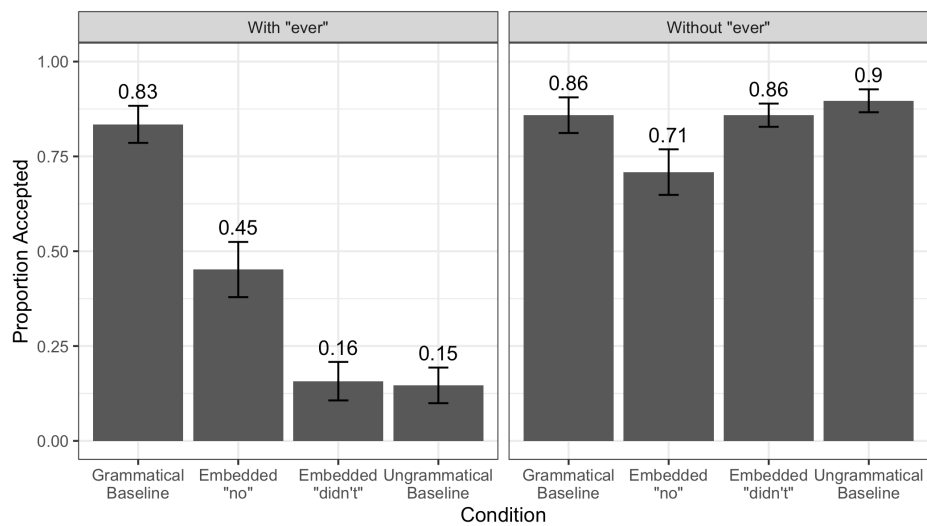


Figure 3.4: Mean percentage of ‘YES’ responses for the experimental conditions in Experiment 4 for the acceptability judgment task. Error bars indicate standard error of the mean across subjects.

²⁶This is equivalent to what was called, in other experiments, the “grammatical baseline” condition.

²⁷This is equivalent to what was called, in other experiments, the “ungrammatical baseline” condition.

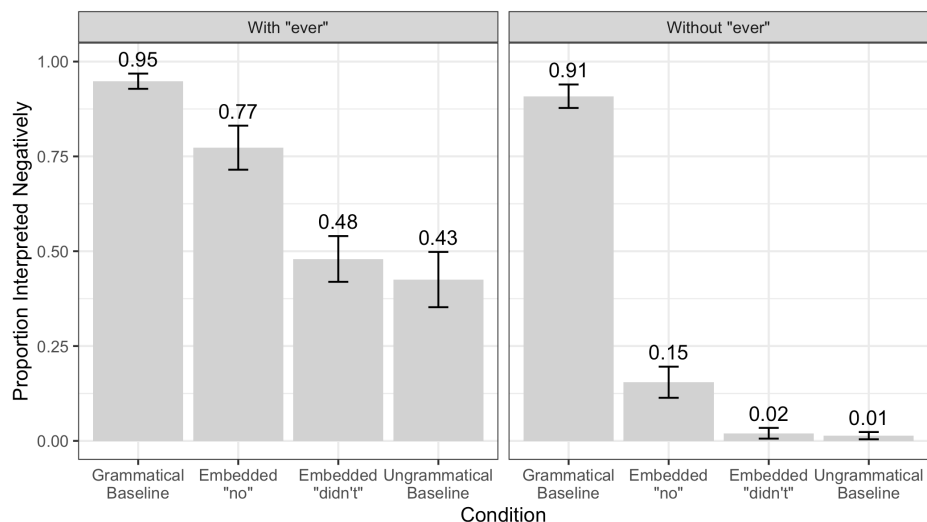


Figure 3.5: Mean percentage of ‘NO’ responses for the experimental conditions in Experiment 4 for the interpretation question. Error bars indicate standard error of the mean across subjects.

Lastly, we evaluate whether globally negative interpretations and NPI acceptance go hand-in-hand as all versions of the scope hypothesis predict. We evaluate this prediction by focusing just on the three standard illusion conditions (that is, conditions with *ever*, looking at just the grammatical baseline, ungrammatical baseline, and embedded-*no* conditions), and splitting the data by interpretation (“yes” versus “no” answers to comprehension questions). The relevant conditions are displayed in Figure 3.6. The illusion effect (ungrammatical baseline versus embedded-*no*) did not significantly interact with interpretation (“yes” comprehension question response versus “no” response) ($\beta=0.56$, $SE=1.10$, $z=0.50$, $p=0.61$). Thus we do not find evidence that illusions are specific to negatively-interpreted trials.

3.3.1.6 Statistical power

Post hoc power analyses were again conducted using *simr*. We computed power to detect an illusion for embedded-*didn't*, although this was not the primary aim of the present experiment. We again used an adjusted effect size based on data from Parker & Phillips 2016 and Xiang, Grove, & Giannakidou 2013. The present experiment achieved between 54% and 60% power to detect an illusion for embedded-*haven't*, assuming an effect size of $\beta=0.95$. We additionally computed power for the key comparison tested by this experiment: the correspondence between negative interpretation and acceptance. The predicted effect

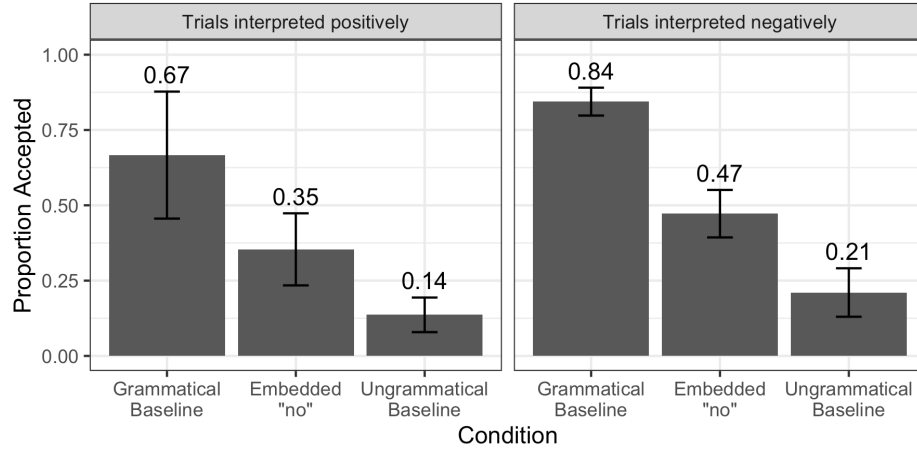


Figure 3.6: Mean percentage of ‘YES’ responses for the experimental conditions in Experiment 4 in the acceptability judgment as a function of interpretation: positive (left) and negative (right). Error bars indicate standard error of the mean across subjects.

size for the acceptance-by-interpretation interaction under the quantifier scope account was computed by simulating a dataset of the same size and structure as ours but in which the true acceptance rate for the negatively-interpreted embedded-*no* trials is as high as the acceptance rate for grammatical baseline trials, and the true acceptance rate for the positively-interpreted embedded-*no* trials is as low as the acceptance rate for ungrammatical baseline trials. This resulted in a very large predicted effect size, estimated at $\beta=4.20$. The present experiment achieved between 88% and 94% power to detect an interaction between the illusion effect for embedded-*no* and trial-level interpretation, assuming an effect size of $\beta=4.20$.

3.3.1.7 Discussion

Experiment 4 was designed to evaluate whether the interpretation of NPI illusion sentences is consistent with the predictions of the scope miscalculation hypothesis. In short, our findings demonstrated the following: (a) previous findings of an illusion for embedded-*no* and reduced or absent illusions for embedded-*didn't* were replicated, (b) while there are some negative interpretations for sentences with embedded quantifiers, these are much less frequent (15% of trials) for sentences without NPIs than those with NPIs (77% of trials), and (c) the NPI illusion does not appear to be specific to negatively-interpreted trials.

Findings (b) and (c) present challenges for the scope miscalculation hypothesis. Both the early- and late-scope-assignment versions of the hypothesis predict that, among NPI illusion sentences, the trials that are accepted should be the ones that are interpreted negatively and the trials that are rejected should be the ones that are interpreted positively, contrary to finding (c). It is important to acknowledge that finding (c) is a null effect, since we fail to observe a statistically significant interaction. However, the hypothesis predicts not only that acceptance rates should be *higher* for negatively-interpreted trials, but that *all* negatively-interpreted trials should be accepted (when in fact 47% of such trials were accepted) and *no* positively-interpreted trials should be accepted (when in fact 35% of such trials were accepted). Thus, a statistical interaction is a low bar, relative to the predictions of the hypothesis, but the data do not support even this. In order to further explore this issue, we directly tested whether acceptance rates for negatively-interpreted illusion trials are at ceiling and acceptance rates for positively-interpreted illusion trials are at floor. We operationalized “ceiling” as the acceptance rates for the grammatical baseline condition and “floor” as the acceptance rates for the ungrammatical baseline condition since it is of course unrealistic to expect acceptance rates at 100% or 0% for any condition in this type of task. We find that the negatively-interpreted illusion trials were accepted significantly less than the grammatical baseline ($\beta=3.19$, $SE=0.94$, $z=3.40$, $p=.001$) and the positively-interpreted illusion trials were accepted significantly more than the ungrammatical baseline ($\beta=1.72$, $SE=0.75$, $z=2.29$, $p=.044$)²⁸. This casts serious doubt on the possibility that acceptance and interpretation align in the way the scope miscalculation hypothesis predicts.

There is one other consideration regarding the lack of correspondence between interpretation and acceptance for illusion trials. Since the comprehension question was presented after the acceptability judgment, it is in principle possible that comprehenders “recovered” from their illusions by the time they encountered the comprehension question. This would mean that the lack of alignment arises because the two questions actually probe entirely different representations. However, we think this is unlikely. If comprehenders have truly recovered from the illusion by the time they provide judgments for comprehension questions, they should, at that point, correctly recognize that the illusion sentence does not have main-clause negation, it has only negation within the RC. In other words, after recovering from the illu-

²⁸Reported p-values are adjusted for multiple comparisons using the Bonferroni correction for 2 comparisons.

sion, embedded-*no* sentences should look just like the embedded-*didn't* sentences. But comprehension question responses at this point revealed numerically large differences between these conditions (77% versus 48% negative interpretations). It is therefore unlikely that comprehenders have simply recovered from the illusion by the time they answer the second question.

Turning to the other key prediction tested here, recall that the early-scope-assignment version of the hypothesis claims that NPI illusions are merely a side effect of earlier commitments. However, finding (b) demonstrates that globally negative interpretations are far rarer in the absence of an NPI. We therefore do not consider the early scope-assignment version of the hypothesis a plausible explanation for the NPI illusion. Note that there may be independent reasons to disfavor the late-scope-assignment version of the hypothesis. Because this account claims that wide scope of the quantifier is pursued only in an attempt to license the NPI once it is encountered, a PPI in the same MC position would never be subject to interference from embedded negative quantifiers. The presence of a PPI would indicate that narrow scope is the best interpretation of the quantifier, which is exactly consistent with the grammatically-sanctioned narrow-scope representation. However, reported findings from Orth, Yoshida, & Sloggett 2020a contradict this prediction, showing illusions of ungrammaticality for main-clause PPIs preceded by embedded negative quantifiers. This makes the late scope-assignment version less promising.

There may, however, be a third option, in which a wide-scoping representation of the quantifier is generated late (i.e., at the NPI) but not specifically because doing so would *license* the NPI. Rather, the wide-scope representation would be entertained purely because it would *influence* the NPI — whether that influence is to make the sentence better or worse could be irrelevant. This seems to be what Orth, Yoshida, & Sloggett 2021 have in mind, since they state that encountering the NPI is what triggers quantifier raising under their hypothesis, but they also state that their account predicts PPI illusions. While this version of the hypothesis fares better than the early-scope assignment version with respect to finding (b) in the present experiment, and fares better than the late-scope assignment version with respect to PPI illusions, it is equally unsuccessful in making sense of finding (c) in the present experiment. That is, all quantifier scope hypotheses incorrectly predict that negative interpretations and acceptance of NPI illusion sentences should go hand in hand, which did not occur in Experiment 4.

In addition to the critical findings from Experiment 4 with respect to the scope miscalculation hypothesis, there are some notable empirical generalizations suggested by these data which are informative for NPI illusion research more broadly. First, the fact that we replicate the illusion even when there are comprehension questions should put to rest concerns that illusions in acceptability tasks reflect a certain kind of “shallow processing” in which comprehenders fail to represent sentences at all levels of representation. Second, this is the first exploration, to our knowledge, of the interpretation of NPI illusion sentences. It is clear that NPI illusion stimuli are overwhelmingly interpreted as if the MC is negative, but even positively-interpreted trials are subject to the illusion. While the mere polarity of the MC is of course a very coarse-grained assessment of the interpretation of a sentence, this is a useful starting point for future work.

3.4 Non-quantificational negation and scalar meanings

While the findings from Experiment 4 present evidence against the scope miscalculation hypothesis, the scalar alternatives hypothesis does not make very strong predictions for the kinds of interpretations we tested. In Experiments 5 and 6 we directly evaluate predictions of the scalar alternatives hypothesis. Specifically, we investigate whether illusions arise when the embedded negative word *haven’t* co-occurs with the NPI *any* as in (33)²⁹.

- (33) * The critics [that haven’t recommended any authors of alternative genres] have ever objected to mainstream literary trends.

Recall that the scalar alternatives hypothesis assumes an online NPI-licensing mechanism that hews closely to scalar accounts of the grammar of NPIs. A key ingredient for NPI illusions, under this account,

²⁹ An anonymous reviewer points out that our stimuli containing *haven’t* followed by *any* in the RC and *ever* in the MC appear initially similar to so-called *parasitic licensing* constructions in Dutch (see Hoeksema 2007 for details). However, we know of no evidence that parasitic licensing is a component of the English grammar and moreover, even if it were, it would not lead to the acceptability of sentences like these. Under Hoeksema’s account (applied to our stimuli), the mechanism responsible for parasitic licensing would allow a [Neg] feature to spread from the negative word *haven’t* to the NPI *any*, such that then *any* could license NPIs that it c-commands. But, of course, *any* does not c-command *ever*, and so even with the spreading of the [Neg] feature, the sentence is ungrammatical and should be rejected. We thus do not pursue this apparent parallel further.

is the ability of negative quantifiers to evoke an appropriate scale before *ever* is even read. However, non-quantificational forms of negation like *not/haven't/didn't* are not incompatible with scalar meanings, they just do not trigger them by default, prior to the NPI. If one could create a context in which a clause containing *haven't* receives a scalar interpretation prior to *ever*, illusions would be possible under this account. The inclusion of NPI *any* inside the RC is intended to achieve this. If all NPI licensing is scalar in nature, the RC in (33) would have to be interpreted in a scalar way in order to license *any*, thus providing the key ingredient that makes it a lure for the main-clause NPI *ever*. This hypothesis therefore predicts a contrast between (33) and (34), with illusions being more likely for (33).

- (34) * The critics [that haven't recommended the authors of alternative genres] have ever objected to mainstream literary trends.

Turning to the scope miscalculation hypothesis, this account succeeds in capturing the contrast between *no* and *not/haven't/didn't* by attributing NPI illusions to problems in the assignment of quantifier scope. Since *haven't* is non-quantificational, it is not subject to these problems and so illusions are expected to be categorically impossible with this licenser. Thus, under this account, sentences like (33) and (34) are predicted to be similarly unacceptable and invulnerable to illusions, and only negative quantifier sentences like (35) should give rise to illusions.

- (35) * The critics [that have recommended no authors of alternative genres] have ever objected to mainstream literary trends.

Thus, the two hypotheses make clearly contrasting predictions. If NPI illusions are a consequence of scope miscalculation, NPI illusions should be equally impossible for (33) as for (34). If instead, illusion magnitude varies as a function of the alternatives evoked in the RC, (33) should yield illusions despite its use of non-quantificational negation *haven't* as an embedded licenser.

3.4.1 Experiment 5: offline acceptability

We first conducted an offline rating study to establish that the materials are appropriate. The materials used in this experiment were adapted from our stimuli from Experiments 1-4. We expected to obtain

A. Gram. baseline	No critics [that have recommended any authors of alternative genres] have ever ...
B. Embedded <i>no</i>	The critics [that have recommended no authors of alternative genres] have ever ...
C. Embedded <i>not...any</i>	The critics [that haven't recommended any authors of alternative genres] have ever ...
D. Embedded <i>not</i>	The critics [that haven't recommended the authors of alternative genres] have ever ...
E. Ungram. baseline	The critics [that have recommended the authors of alternative genres] have everobjected to mainstream literary trends.

Table 3.4: Example stimuli for Experiment 5 and Experiment 6

a clear pattern of grammatical sensitivity and no differences among the four ungrammatical conditions regardless of the presence and type of a structurally irrelevant licenser.

3.4.1.1 Participants

15 US-based native speakers of English participated in this experiment. All participants provided informed consent and they received \$9 as compensation. We excluded workers whose judgments of filler trials did not reliably distinguish between grammatical and ungrammatical fillers, based on a one-sided t-test, as well as workers who failed the “attention checks” (see “Materials” below). 1 worker was excluded from our analyses, resulting in 14 participants. Among the included participants, the average rating for grammatical fillers was 5.81 out of 7, with a standard deviation of 0.53, and the average rating for ungrammatical fillers was 3.66, with a standard deviation of 0.82.

3.4.1.2 Materials

The experimental materials for this and the following tasks consisted of 40 sets of 5 items. The need for a condition like (33) which contains *haven't* followed by *any*, required that we convert our stimuli from object relative clauses (ORCs) to subject relative clauses (SRCs). Note that merely changing the clause type and nothing else would result in strange meanings for many stimuli (for example, our standard example, *the authors that no critics recommended* would become *the authors that recommended no critics*, which is inconsistent with world knowledge). This required further modification to improve naturalness and plausibility. A sample set of the five experimental conditions is shown in Table 3.4.

Each participant was asked to rate 130 sentences: 40 experimental items and 90 fillers of similar length and complexity. The experimental items were distributed across five lists using a Latin Square design and

the fillers were the same in each list. Participants completed two practice items before beginning the task. We again included eight attention check trials, randomly interspersed through the experiment. For example an attention check trial read “For this sentence, please choose six as the answer.” Participants who answered two or more attention check trials incorrectly were excluded from our analyses (see “Participants” section 3.4.1.1 above).

3.4.1.3 Procedure

The offline acceptability procedure followed the same steps as in Experiment 1.

3.4.1.4 Analysis

Results were analyzed the same way as with Experiment 1. We again present the results from linear mixed effects models, but ordinal model results yield similar conclusions and can be found in the Supplementary Files.

3.4.1.5 Results

The results from this experiment are presented in Figure 3.7. Linear mixed effects models revealed a clear effect of grammaticality shown by significant differences between the grammatical baseline condition and the other four experimental conditions (grammatical versus embedded-*no*: $\beta=-2.43$, $SE=0.37$, $t=-6.48$, $p<.001$; grammatical versus embedded-*haven't-any*: $\beta=-2.69$, $SE=0.37$, $t=-7.29$, $p<.001$; grammatical versus embedded-*haven't*: $\beta=-2.86$, $SE=0.37$, $t=-7.75$, $p<.001$; grammatical versus ungrammatical baseline: $\beta=-2.79$, $SE=0.37$, $t=-7.55$, $p<.001$). We again compared the embedded-negation conditions to the ungrammatical baseline and again found a small (on average 0.36 points on a 7-point scale) but statistically significant boost for embedded-*no* ($\beta=0.36$, $SE=0.17$, $t=2.06$, $p=.048$). This comparison did not reach statistical significance in the ordinal regression analysis. We find no such difference for embedded-*haven't-any* ($\beta=0.10$, $SE=0.15$, $t=0.64$, $p=.54$) or embedded-*haven't* ($\beta=-0.07$, $SE=0.15$, $t=-0.45$, $p=.65$).

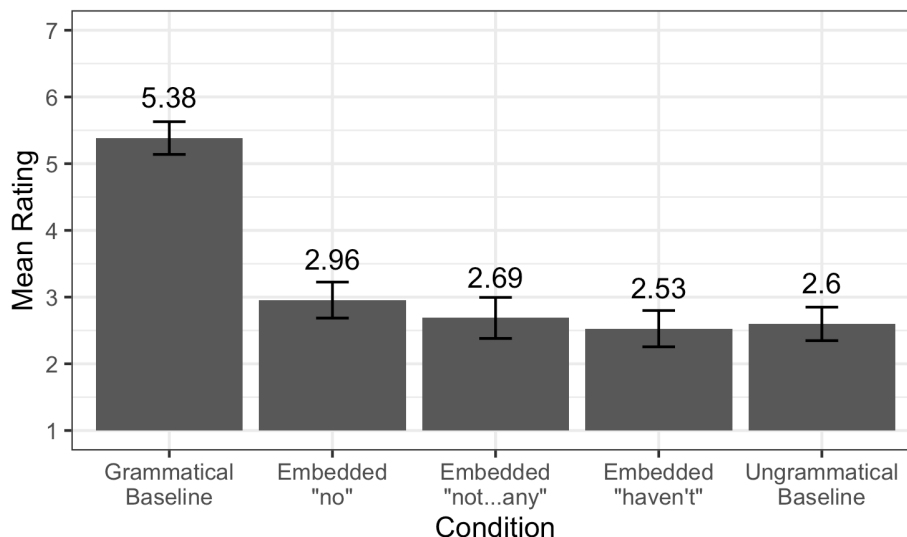


Figure 3.7: Mean ratings for the experimental conditions in Experiment 5. Error bars indicate standard error of the mean across subjects.

3.4.1.6 Discussion

The main objective of Experiment 5 was to confirm the grammatical status of the experimental materials. The results show that participants clearly identify the grammatical baseline condition as acceptable and the ungrammatical baseline condition as unacceptable. In addition, the ratings observed for the three conditions containing non-c-commanding negative elements were highly degraded relative to the grammatical baseline. In this regard, these results confirm that speakers are sensitive to NPI licensing contrasts in our materials. Note that we again observe a numerically small but statistically significant boost in ratings for embedded-*no*, even in an offline task (see discussion of Experiment 1).

3.4.2 Experiment 6: speeded acceptability

We used speeded acceptability measures with the goal of determining whether sentences containing embedded-*haven't* followed by *any* inside the RC yield illusions. The scope miscalculation hypothesis predicts that the *haven't-any* condition should pattern with the *haven't* condition and fail to yield illusions, because both conditions are missing the critical ingredient for illusions: a negative quantifier. In contrast, the scalar alternatives hypothesis predicts that the *haven't-any* condition should pattern with the *no* condi-

tion and yield illusions because the critical ingredient is the existence of pre-constructed scalar alternatives.

3.4.2.1 Participants

195 US-based native speakers of English participated in this experiment. Note that we increased the sample size for this experiment. This was done for two reasons. First, the offline data in Experiment 5 suggest that the switch to SRC may result in less clear judgments of acceptability — that is, even the grammatical and ungrammatical baseline conditions were judged closer to the middle of the scale. Thus, a larger sample was necessary to maintain adequate statistical power for a measurement with increased noise. Second, because the critical finding is now not only whether illusions arise for each of three embedded-negation conditions, but how those conditions compare to each other, the effect size of interest may be smaller (as corroborated by pilot data). All participants provided informed consent and they received \$6 as compensation. We excluded workers who failed to provide a response within 2 seconds for 25% of fillers or more, workers whose judgments of filler trials did not reliably distinguish between grammatical and ungrammatical fillers, based on a chi-squared test, and workers who failed the “attention checks”. 40 workers were excluded based on these criteria, resulting in 155 participants. The mean filler-trial accuracy of the included participants was 80%.

3.4.2.2 Materials

The materials used in this task were the same 40 sets of experimental items and 90 filler sentences that were used in Experiment 5. Participants completed two practice trials before beginning the experiment. We additionally included eight attention check trials, as in Experiments 4 and 5. Participants who answered two or more attention check trials incorrectly were excluded from our analyses (see “Participants” section).

3.4.2.3 Procedure

The speeded acceptability procedure was identical to Experiments 2 and 3.

3.4.2.4 Analysis

As with previous experiments, a generalized linear mixed effects model using helmert coding and a logit link function was fit, and pairwise comparisons were computed using emmeans. The critical pairwise comparisons once again compare the ungrammatical baseline condition to each of the other four conditions to determine, first, whether basic grammaticality effects arise, and then whether illusions arise for embedded-*no*, embedded-*haven't*, and embedded-*haven't-any*, respectively. In addition, the three embedded-negation conditions were compared directly to one another to determine whether differences in the magnitude of the illusion arise.

3.4.2.5 Results

The results from this experiment are presented in Figure 3.8, which shows the percentage of “yes” responses given to each condition. An effect of grammaticality was observed ($\beta=4.58$, $SE=0.28$, $z=16.74$, $p<.001$), indicating that the grammatical baseline condition was significantly more likely to be judged acceptable than the ungrammatical baseline condition. An effect of embedded-*no* was observed ($\beta=0.63$, $SE=0.14$, $z=4.65$, $p<.001$), replicating the standard illusion effect for negative quantifiers. A statistically significant but numerically small effect of embedded-*haven't* was also observed ($\beta=0.32$, $SE=0.12$, $z=2.71$, $p=.007$). Note that this is inconsistent with our findings in Experiments 2, 3, and 4 which did not reveal statistically reliable illusions for sentences with embedded-*not/-n't*. A statistically significant and again small effect of embedded-*haven't-any* was also observed ($\beta=0.46$, $SE=0.13$, $z=3.43$, $p<.001$).

Comparisons between embedded-negation conditions revealed the following. The embedded-*no* condition was significantly more likely to be judged acceptable than the embedded-*haven't* condition ($\beta=0.31$, $SE=0.13$, $z=2.41$, $p=.02$), replicating previous findings. However, we identified no statistically significant differences between either embedded-*haven't-any* and embedded-*haven't* ($\beta=-0.13$, $SE=0.12$, $z=-1.16$, $p=.25$) or embedded-*haven't-any* and embedded-*no* ($\beta=0.17$, $SE=0.13$, $z=1.32$, $p=.19$). That is to say, although embedded-*no* and embedded-*haven't* differed from one another, embedded-*haven't-any* was numerically intermediate and statistically not distinguishable from either one.

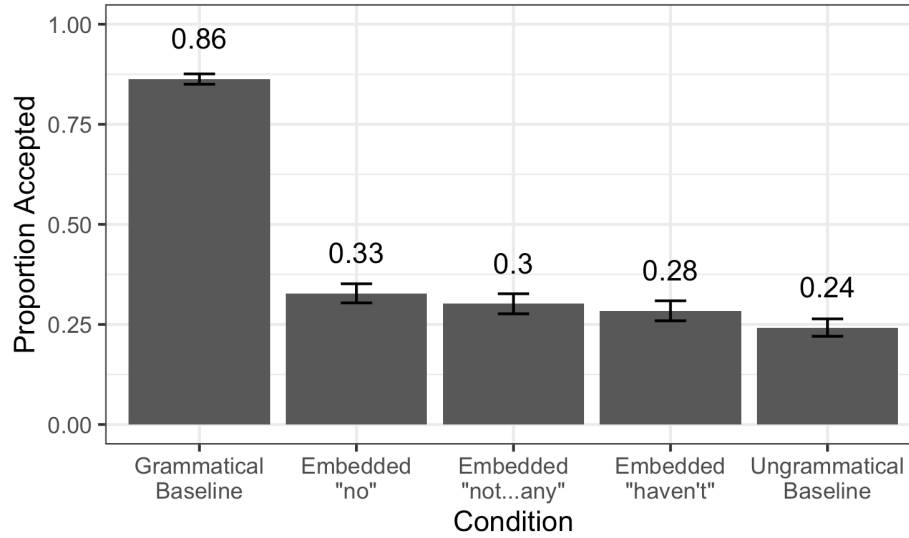


Figure 3.8: Mean percentage of ‘YES’ responses for the experimental conditions in Experiment 6. Error bars indicate standard error of the mean across subjects.

3.4.2.6 Statistical power

Note that the sample size for the present experiment was increased relative to previous experiments because pilot data indicated that the basic illusion effect may be smaller for NPI illusion sentences with SRCs, and because we wanted to be able to determine whether embedded-*haven't-any* patterned with embedded-*no* or with embedded-*haven't* even if it was numerically intermediate (though note that we did not achieve this second aim). Post hoc power analyses were again conducted using *simr*, and using an adjusted effect size based on data from Parker & Phillips 2016 and Xiang, Grove, & Giannakidou 2013. The present experiment achieved between 96% and 100% power to detect an illusion for embedded-*haven't*, or embedded-*haven't-any*, assuming an effect size of $\beta=0.95$.

3.4.2.7 Discussion

In Experiment 6, we aimed to determine whether the contrast in illusion rates between embedded-*no* and embedded-*not/-n't* in prior experiments was due to the categorical impossibility of illusions with non-quantificational embedded licensors, as the scope miscalculation hypothesis predicts, or instead due to a lower probability of inferred scalar meanings for *not/-n't* in the absence of an NPI, as the scalar alternatives

hypothesis predicts. In order to address this, we measured illusion rates for both the previously-tested forms of negation, *no* and *haven't*, as well as a novel condition containing *haven't* paired with NPI *any* inside the RC, which should lead to scalar interpretations of the RC given our assumptions. While we successfully replicated three important contrasts — the grammaticality effect, the illusion for embedded-*no*, and the contrast between embedded-*no* versus embedded-*haven't* — we additionally found, to our surprise, that all three embedded negation conditions yielded statistically reliable illusions, and embedded-*haven't-any* could not be statistically distinguished from either of the other two embedded negation conditions.

It is particularly noteworthy that we observed an illusion for embedded-*haven't* in this experiment, since this was not found for embedded-*not*, embedded-*haven't*, or embedded-*didn't* in Experiments 2, 3, and 4. There are three logically possible ways to explain this discrepancy in findings: (a) the true state of the world is that there are no illusions for embedded-*not/-n't* and the observed effect in Experiment 6 was a false positive, (b) the true state of the world is that there are illusions for embedded-*not/-n't* and the observed null effects in Experiments 2, 3, and 4 were false negatives, or (c) illusions sometimes arise for embedded-*not/-n't*, and they are sensitive to some factor that we inadvertently manipulated between Experiments 2-4 and Experiment 6. We address each of these possibilities in turn, paying particular attention to the consequences of these possibilities for the scalar alternatives hypothesis and the scope miscalculation hypothesis.

If the observed illusion for embedded-*not/-n't* in Experiment 6 is a false positive, the take-aways are clear: we observed an illusion for embedded-*haven't-any*, contrary to the predictions of the scope miscalculation hypothesis and consistent with the predictions of the scalar alternatives hypothesis. It is of course in principle possible that the observed illusion for embedded-*haven't-any* is also a false positive, in which case the scope miscalculation hypothesis fares better. However, unlike the illusion for embedded-*haven't*, which misaligns with previous findings, there is no independent reason to suspect that the illusion for embedded-*haven't-any* is a false positive, and so we do not pursue this further.

In order to address the possibility that the lack of illusions for embedded-*not/-n't* in Experiments 2, 3, and 4 were false negatives, we conducted additional power analyses. We estimate the effect size of the

illusion for embedded-*not/-n't* at $\beta=0.3$, as this was the observed effect size in Experiment 6. Experiments 2, 3, and 4 had only between 15% and 20% power to detect an effect size of $\beta=0.3$. Note that this does not mean that Experiments 2-4 were simply “under-powered experiments”. They were designed to detect illusion effects of the size typically observed for embedded-*no* (or, in the case of Experiment 4, to detect an entirely different effect), for which the expected β is 0.95, and for this they were adequately powered (see sections 3.2.2.6, 3.2.3.6, and 3.3.1.6 above). However, these findings are consistent with the possibility that there are illusions for embedded-*not/-n't* with a small ($\beta=0.3$) effect size. As discussed above, a key difference between the scope miscalculation hypothesis and the scalar alternatives hypothesis is in whether illusions are predicted to be categorically impossible for non-quantificational licensors, or simply less likely. Under the scope miscalculation hypothesis, NPI illusions are a side effect of scope errors, which can only arise with quantificational licensors. In contrast, under the scalar alternatives hypothesis, the illusion rate can be turned up or down as a function of the likelihood of activating scalar alternatives in the RC. Thus, if there are small ($\beta=0.3$) illusions for embedded-*not/-n't* and larger ($\beta=0.6$ to $\beta=1.0+$) illusions for embedded-*no*, this would again be consistent with the scalar alternatives hypothesis and inconsistent with the scope miscalculation hypothesis.

Lastly, illusions for embedded-*not/-n't* could have arisen in Experiment 6 due to some inadvertently manipulated factor. One notable change was the shift to SRCs instead of ORCs in Experiment 6. While we think this difference is a potentially important one, it is not clear what mechanism would impact judgments of SRCs in precisely this way. A more promising candidate for the critical change is the inclusion of the *haven't-any* condition. It is possible that over the course of the experiment, participants adjust their interpretations based on other stimuli they have seen. Under the scalar alternatives hypothesis one such adjustment might be to interpret clauses containing *haven't* in a scalar way — that is, participants may begin to pre-construct scalar alternatives already at *haven't* because of the relatively high likelihood, within the experiment, that an NPI is coming. If this were the case, we would expect the earliest trials to show a cleaner pattern of results - illusions for embedded-*no* and embedded-*haven't-any* but not for embedded-*haven't*. The results from Experiment 6 for only the very first (non-filler) trial of the experiment are shown in Figure 3.9. Because trial order is randomized for each participant, selecting each participant's first trial

yields a roughly balanced, though small, dataset. We also plot the very last non-filler trial for each participant. While this is of course a post-hoc exploratory analysis, it does appear to be the case that in early trials, illusions arose for embedded-*no* and embedded-*haven't-any* but not for embedded-*haven't*. In late trials, all ungrammatical conditions, even the ungrammatical baseline, were accepted about equally, regardless of the presence of embedded negation. Caution is warranted in interpreting these findings, since we are looking at only a very small subset of the data, and there are many researcher degrees of freedom in this analysis. It is possible that cross-trial comparisons influenced judgments in a manner that the scalar alternatives hypothesis can accommodate, though we cannot definitively say whether this was the cause of the surprising illusion for embedded-*haven't* observed here.

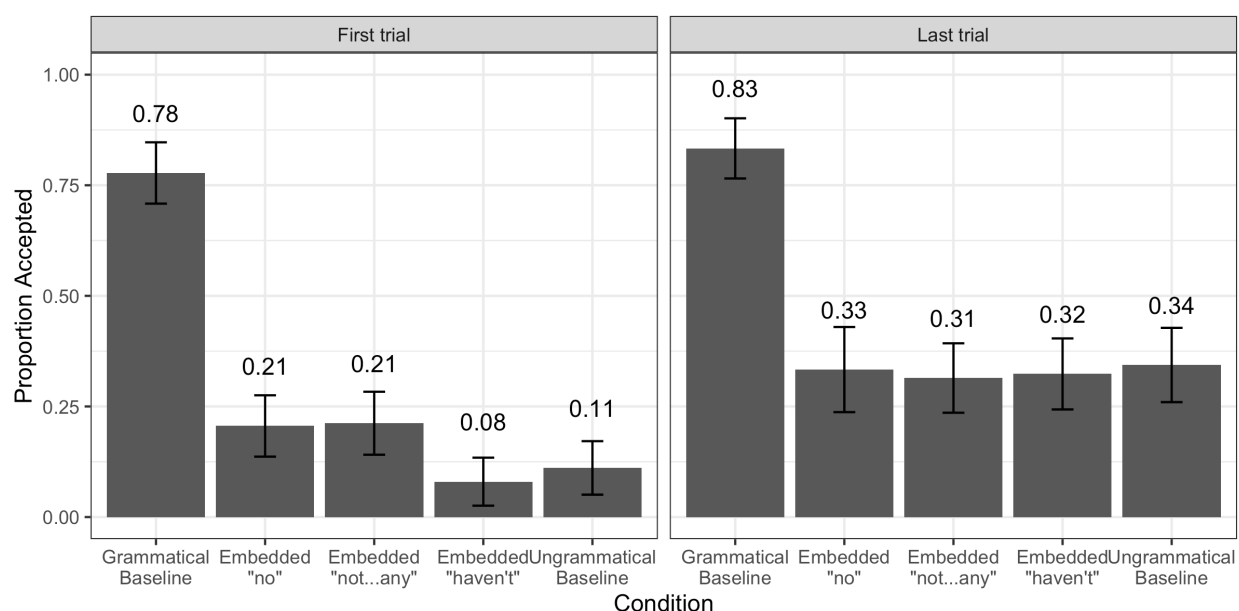


Figure 3.9: Mean percentage of ‘YES’ responses for the experimental conditions the very first non-filler trial and very last non-filler trial of Experiment 6. Error bars indicate standard error of the proportion.

In sum, while we cannot determine the cause of the sudden appearance of an illusion for embedded-*haven't* in Experiment 6 with certainty, we find that a number of possible explanations for this finding are more consistent with the claims of the scalar alternatives hypothesis than with those of the scope miscalculation hypothesis.

3.5 General discussion

3.5.1 Key findings

The evidence presented here provides a strong case for narrowing down the existing range of hypotheses for NPI illusions. Experiments 2, 3, and 4 established a clear contrast in the illusion profile for sentences containing embedded-*no* such as (36a) compared to sentences containing embedded-*not/-n't* such as (36b), as has also been shown in Orth, Yoshida, & Sloggett 2021. Based on these three experiments, it appears that illusions arise routinely for embedded negative quantifiers, but are reduced or absent for embedded non-quantificational forms of negation.

- (36) a. * The authors [that no critics recommended for the award] have ever received acknowledgement for a best-selling novel.
- b. * The authors [that the critics did not recommend for the award] have ever received acknowledgement for a best-selling novel.

Experiment 4 additionally investigated the sentence-final interpretation of illusion sentences by asking participants for both an acceptability judgment and a response to a question like “Did the authors receive acknowledgements for their novels?” following sentences similar to (36a) and (36b). These findings reveal that illusions arise for both positively-interpreted and negatively-interpreted trials, at rates that are not statistically distinguishable. Furthermore, although NPI-illusion sentences like (36a) are often understood as expressing a globally negative meaning, these interpretations arise very rarely in the absence of an NPI. The interpretive error is therefore not a likely cause of problems in processing the NPI, but rather a consequence.

Experiment 6 measured illusion rates for sentences like (37) in order to determine whether illusions are categorically impossible for non-quantificational embedded licensors (*haven't* in (37b) and 37c) or if the illusion rate can be turned up or down as a function of the alternatives inferred at the level of the clause. The results tentatively suggest that clause-level meanings play an important role, though we acknowledge that the surprising illusion for embedded-*haven't* in this experiment complicates the interpretation of the

results.

- (37) a. * The critics [that have recommended no authors of alternative genres] have **ever**...
- b. * The critics [that haven't recommended any authors of alternative genres] have **ever**...
- c. * The critics [that haven't recommended the authors of alternative genres] have **ever**...

...objected to mainstream literary trends.

Thus, the key critical empirical contributions are the following: (a) non-quantificational negation in the form of *not* or *-n't* yields few or no illusions, (b) illusions are not specific to negatively interpreted trials, (c) erroneous globally-negative interpretations are not independently established prior to the NPI, and (d) aspects of the RC meaning beyond just the presence of a negative word appear to play a role in illusions.

3.5.2 Effect sizes and statistical power

Post hoc power analyses based on independent estimates of effect sizes were computed for all speeded acceptability studies presented here. To review, Experiments 2, 3, and 6 each achieved approximately 90% power to detect an illusion for embedded-*not* assuming the effect size for embedded-*not/-n't* is equal to .95, a value which was computed by taking the lower bound of a 60% confidence interval around a meta-analytic estimate of the illusion effect size for embedded-*no* using speeded acceptability data from Parker & Phillips 2016 and Xiang, Grove, & Giannakidou 2013. (Experiment 4 had lower power to detect this effect but was not primarily designed to test this.) However, our findings from Experiment 6 suggest that a smaller but non-zero illusion effect may exist (though see section 3.4.2.7 for other explanations). This of course does not undermine our claim that there is a contrast between embedded-*no* and embedded-*not/-n't* with respect to illusion rates, which is an effect that was statistically significant in all of our speeded acceptability experiments (Experiments 2, 3, 4, and 6). In order to provide a more complete picture of the range of possible illusion effect sizes, we additionally provide the 95% confidence interval for the illusion effects for embedded-*no*, embedded-*not/-n't*, and embedded-*not/-n't...any*, for each of our experiments as well as for the contrast between embedded-*no* and embedded-*not/-n't* (Figure 3.10).

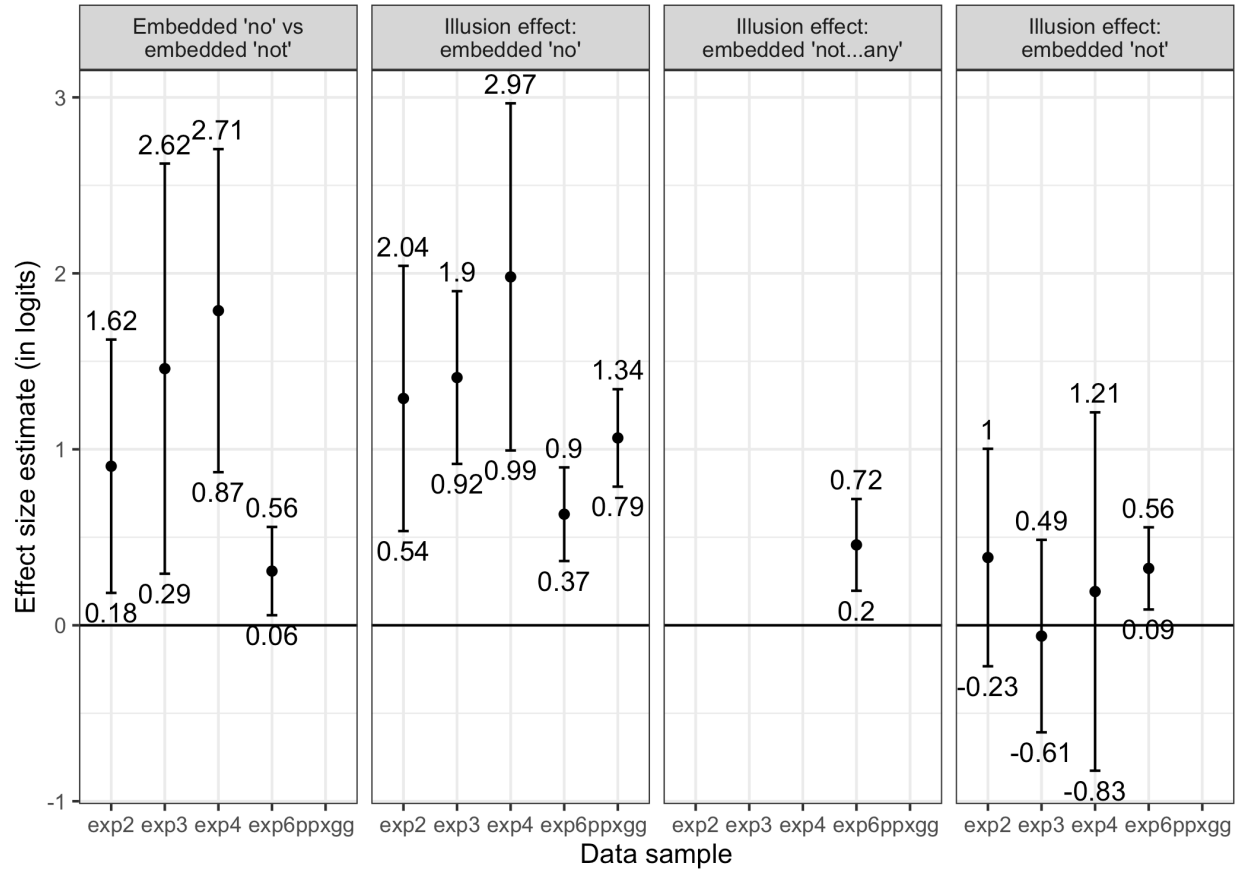


Figure 3.10: 95-percent confidence intervals for the contrast between embedded-*no* and embedded-*not* and for illusion effect sizes for embedded-*no*, embedded-*not*, and embedded-*not-any*. “ppxgg” refers to pooled data from Parker & Phillips 2016 and Xiang, Grove, & Giannakidou 2013

3.5.3 The scalar alternatives hypothesis

We propose a novel explanation for the NPI illusion, namely the scalar alternatives hypothesis. This hypothesis assumes that the online computation of NPI licensing does not consist of the retrieval of a negative word from memory, but rather the integration of the NPI meaning into a clause-level meaning, which is only successful when the appropriate negatively-ordered scalar alternatives are available. The key error, then, is the failure to rapidly inhibit alternatives to the RC prior to the NPI (see Chapter 4 for further exploration of the narrow window of vulnerability). This hypothesis attributes the contrast between *no* and *not/n’t* as embedded licensors to differences in the probability that scalar alternatives to the RC will be pre-constructed. It additionally predicts that when embedded-*not/n’t* is contained in a

scalar-alternatives-evoking clause it should behave similarly to *no* with respect to illusions, which we find tentative support for in Experiment 6. We note, however, that while the hypothesis does not make especially strong predictions with respect to sentence-final interpretation, our Experiment 4 finding that majority of illusion trials are interpreted as if the MC is negative is surprising under this account. If the problem resulting in the erroneous acceptance of the NPI is that the RC and its alternatives interfere with the NPI, it is then not clear why the MC's meaning is so drastically altered. We suspect that this sentence-final globally-negative interpretation is a result of reanalysis processes that attempt to reconcile an acceptable NPI with a syntactic parse that clearly places the NPI in the MC. Very little is known about the processes that take place from the point the NPI is encountered until a sentence-final decision is made, and more work is clearly needed to determine whether this is a plausible explanation for the interpretation findings we obtain.

There are a few other findings regarding NPI illusions that warrant mention here, with respect to the scalar alternatives hypothesis. Importantly, the hypothesis treats the NPI licensing dependency as one between an NPI and a linguistic context that contains it, not as a dependency between an NPI and negative lexical licenser which must c-command it. Because containment is an inherently local relation, there is a natural alignment between this framework and the finding that NPIs that are farther away from a potential licensing context are not be subject to illusions (see Parker & Phillips 2016 and Chapter 4). Additionally, under this hypothesis, the illusion is not driven specifically by the NPI's search for a licenser *per se*, but by the presence of NPI-licensing alternatives at the point when the NPI is encountered. Thus, because the claim is about a fundamentally representational problem and not a problem with retrieval of prior material specifically for the purposes of licensing an NPI, the hypothesis straightforwardly predicts illusions for any item that is sensitive to the same kind of scalar negative representations — including illusions of ungrammaticality for PPIs (see Orth, Yoshida, & Sloggett 2020a). Also of interest is the reported finding that NPI illusions are more robust in comprehenders with better pragmatic reasoning abilities, as measured by the Autism-spectrum Quotient (Xiang, Grove, & Giannakidou 2013). Such a trend is consistent with any account of the illusion that treats the erroneous acceptance of the NPI as a consequence of pragmatic inferences that speakers may be more or less likely to make. This includes both the prag-

matic rescuing account proposed by Xiang, Grove, & Giannakidou and the scalar alternatives account proposed here.

Finally, there is an apparent discrepancy between our findings and existing work in Turkish (Yanilmaz & Drury 2018b) and Korean (Yun, Lee, & Drury 2018) which find clear illusions for non-quantificational forms of negation. We suspect that the key factor is that these languages are head-final and so the NPI appears before its licensor. If we maintain a scalar approach to NPI licensing, it would likely be the NPI itself that triggers the construction of such a representation in these languages. This would trigger a strong prediction for an upcoming negative word, which, for unknown reasons, seems to be satisfied by a negative word in an irrelevant position. Critically, there is no reason why the quantificational or non-quantificational status of that negative word should matter in such a language, since under the scalar alternatives hypothesis quantificational status only matters because of the consequences for the pre-construction of scalar alternatives. In sum, we conclude that the scalar alternatives hypothesis, in addition to predicting many of the key contrasts presented here, can accommodate several key generalizations in the existing literature on NPI illusions.

3.5.4 Alternative explanations for NPI illusions

We additionally considered three other explanations for the NPI illusion. The memory-based hypothesis (Vasishth et al. 2008) attributes illusions to partial matches in memory between a set of search cues ([+negative] and [+c-command], though other feature sets are possible) and the features encoded on chunks in memory representing prior words and phrases. While this hypothesis is appealing in its generality, it cannot explain the observed contrast between *no* and *not/-n't*, since these licensors would be encoded with the same [+negation] feature which should allow them to interfere with NPI processing.

Similarly, the pragmatic rescuing hypothesis (Xiang, Dillon, & Phillips 2009; Xiang, Grove, & Giannakidou 2013) seems to predict uniformity across embedded licensors, contrary to our findings. This hypothesis attributes illusions to erroneous inferences from (38a) to (38c). However, as we noted in section 3.2.3.7 above, these inferences seem equally available for RCs containing non-quantificational negation as in (38b). What is needed for this account to capture the contrast between quantificational

and non-quantificational negation is some further specification of the circumstances under which (38c) may be inferred. Note that the hypothesis independently needs such an elaboration, since even the ungrammatical baseline sentences used in NPI illusion experiments (39a) could license negative contrastive inferences that would license an NPI in P (39b). Thus, it cannot be that any restrictive RC will do.

- (38) a. The authors [that no critics recommended] have P
- b. The authors [that the critics haven't recommended] have P
- c. The authors [that some/the critics have recommended] have NOT P
- (39) a. The authors [that the critics recommended] have P
- b. The authors [that the critics have NOT recommended] have NOT P

Xiang, Dillon, & Phillips 2009 do not address this issue at length, but do suggest that “speakers may be more likely to generate such inferences if the contrasting referents are made very salient in the discourse. Negative quantifiers can do exactly this” (Xiang, Dillon, & Phillips 2009:53). If they are correct in asserting that negative quantifiers are critical, this could in principle explain the contrast between embedded-*no* and embedded-*not*/*-n't*. To support the idea that negative quantifiers make contrasting referents salient, Xiang, Dillon, & Phillips cite work showing that following a downward entailing quantifier, comprehenders readily accept reference with a pronoun to the complement set of the quantifier (Sanford, Moxey, & Paterson 1996). For example, they in (40) can be understood as referring to the many football fans who did not go to the game, rather the referent of *few of the football fans* — i.e., those who did go.

- (40) Few of the football fans went to the game. They watched it on TV instead.
- (Sanford, Moxey, & Paterson 1996)

However, it is not clear how this mechanism is relevant to the issues with the pragmatic rescuing hypothesis. If we apply the mechanism explored by Sanford and colleagues to NPI illusion sentences like (41), we may expect reference to the (many) critics who declined to recommend the authors (i.e., the complement set to the downward-entailing quantifier-phrase *few critics*) in a following sentence. The mechanism does not generate inferences to any other set of authors, only to other sets of critics, and it

does not generate inferences with globally negative meanings. While Xiang, Dillon, & Phillips may be correct in asserting that negative quantifiers have a special status under their theory, this is not sufficiently spelled out and the reference to Sanford, Moxey, & Paterson 1996 does not clearly pertain to the issue. Thus, we do not find a clear prediction of a contrast between quantificational and non-quantificational forms of negation under the pragmatic rescuing hypothesis.

(41) The authors [that few critics recommended] have P

Finally, we consider three variants of a hypothesis that attributes NPI illusions to problems in correctly assigning quantifier scope, which was suggested by both de Dios Flores, Muller, & Phillips 2017 and Orth, Yoshida, & Sloggett 2021. The core claim of the scope miscalculation hypothesis is that the negative quantifier is represented as if it takes scope over the entire MC, and given this scope assignment, the NPI can be licensed. Non-quantificational forms of negation would not give rise to illusions due to their more limited scope possibilities, thus capturing the observed contrast between embedded-*no* and embedded-*not/-n't*. Note, however, that negative quantifiers are also quite limited in their scope possibilities (Liu 1990), and so the wide-scope representation that gives rise to the illusion requires a parsing procedure that is willing to temporarily ignore some scope-related constraints of the grammar. Why the scope constraints of negative quantifiers are violable but the scope constraints of other forms of negation are not violable is not obvious.

Two relevant variants of this hypothesis are one in which scope assignment is early, such that on some trials a wide-scope representation has already been established prior to the NPI, and a version in which multiple scope configurations are considered in parallel until at least the NPI position, at which point a choice is made, taking the NPI as evidence for a wide-scope representation. If scope assignment occurs prior to the NPI, we would expect to see interpretive consequences for the MC regardless of whether an NPI is eventually encountered, contrary to our findings in Experiment 4. Under both versions of the hypothesis (as well as the late, but not error-driven version pursued by Orth, Yoshida, & Sloggett 2021), we would expect the NPI illusion trials that are interpreted negatively to be the ones that are accepted and the trials that are interpreted positively to be the ones that are rejected. This is not what we observe in Experi-

ment 4. Both versions of the hypothesis also predict that illusions are impossible for non-quantificational forms of embedded negation, contrary to our findings in Experiment 6.

3.5.5 Aligning parsing and grammar

We have addressed a number of competing grammatical hypotheses that aim to account for the distribution of NPIs in natural language, as well as a number of competing processing hypotheses that aim to account for the error profile of the comprehension of NPI-containing sentences. While we do not aim to choose between grammatical accounts on the basis of processing data, some brief discussion of the possible alignment between grammar and parsing is warranted. It is not logically necessary that the representations that guide incremental interpretation are identical to those that are licensed by the mental grammar. Notable two-system views include “quick-and-dirty” parsing strategies (Bever 1970) and the “good enough hypothesis” (Ferreira, Bailey, & Ferraro 2002). While it may be tempting to treat grammatical illusions as an obvious case of two-system processing, Parker 2019 argues from modeling evidence with agreement attraction showing that changes over time in the perceived grammatical status of an incoming sentence are expected under both two-system and one-system accounts. In the case of NPIs we have argued for an account of illusions which relies on the timing of the computation of scalar alternatives, invoking a particular style of grammatical explanation — namely, one in the spirit of Fauconnier 1975a. Assuming that a scalar approach to NPI licensing is a reasonable grammatical theory, we have shown that a direct online implementation of such a theory is not only tenable but can make sense of comprehender’s occasional failure to provide judgments that align with their grammar. We see this as a more parsimonious style of explanation than those that account for illusions only by assuming an online NPI-licensing mechanism that is fairly removed from the grammar. For example, framing licensing as a search for an item in memory with a [+negation] and [+c-command] feature forces a disconnect between the online licensing mechanism and the grammar of NPI licensing, which must account for the many NPI licensors that are not explicitly negative. The appeal of such a theory of illusions is of course the generality across processing phenomena that it offers (i.e., by explaining agreement attraction and NPI illusions as by-products of the same memory architecture), which an account such as ours cannot

achieve. That is, the scalar alternatives hypothesis makes no predictions about agreement attraction whatsoever. Since illusion phenomena are clearly diverse and seem to be sensitive to different factors, this shift away from uniformity in explanation across illusions and towards uniformity between the grammar and the real-time comprehension system may ultimately be appropriate.

3.6 Conclusion

This work proposes a novel explanation NPI illusions, which emphasizes the relevance of real-time semantic interpretation — specifically, the availability of scalar alternatives. This claim is largely motivated by the finding that illusions are robust for negative quantifiers but reduced or absent for non-quantificational forms of embedded negation. Importantly, this finding calls into question the previously promising hypothesis that a wide range of linguistic illusions can be explained by the properties of the memory architecture. While it is clear that memory systems are critical to language comprehension, it appears that the re-framing of NPI licensing as merely the retrieval of a prior lexical item in memory is both unfaithful to hypothesized grammars of NPIs and inconsistent with our results. We additionally review some competing hypotheses including the pragmatic licensing hypothesis and scope miscalculation hypothesis and find that they do not predict the error profile nor the interpretation patterns that we observe. The scalar alternative hypothesis proposed here allows for NPI illusions to result from the same detailed operations that are deployed during routine NPI dependency resolution, which we consider a useful shift in our thinking about how grammatical knowledge is deployed by the parser.

Chapter 4 NPI illusions: the role of distance

4.1 Introduction

The failure of the sentence comprehension system to construct an accurate parse under specific circumstances, as in the case of linguistic illusions, has proven to be a useful window into the mechanisms that underlie this system in general. Here we investigate the selectivity of illusions of NPI licensing, specifically exploring the role of the position of the NPI relative to an interfering representation. Much prior work has shown that the unacceptability of sentences with unlicensed NPIs is sometimes not apparent in early stages of processing for sentences like (42a). In contrast, the unacceptability of (42b) is readily detected.

- (42) a. * The bills that no senators voted for have ever become law.
 b. * The bills that the senators voted for have ever become law.

This pattern of findings is often summarized as an intrusion effect in which the representation of the negative word in the RC somehow interferes with the processing of the NPI in the MC, leading to an illusion of acceptability. However, it has so far been unclear whether it is in fact the negative word itself that interferes or if instead the representation of the entire negated context (i.e. the RC) interferes. This has consequences for whether we think of NPI licensing more generally as an operation that relates an NPI to an individual word in the sentence or as an operation that relates an NPI to the properties of the context that contains it. Our present aim is to identify the nature of the representation that interferes — a negative word or a negative clause — by leveraging the previously-reported “distance effect” on NPI illusions. We additionally explore related questions concerning the nature of the distance effect — that is, why the relevant interfering representation becomes more or less accessible at various points in the

sentence. This work serves the broader goal of identifying the mechanisms underlying both the NPI illusion and typical, successful NPI processing. In order to approach this question, we begin with some background on the grammar of NPI licensing and the nature of linguistic illusions.

4.1.1 NPI licensing

NPIs encompass a wide range of syntactic categories, and are defined by their distribution. Sentences like (43a) are routinely judged by comprehenders to be highly degraded, whereas the introduction of negation in sentences like (43b) dramatically improves their acceptability. Furthermore, the source of the unacceptability of (43a) can be located specifically in the word *ever*, since when this word is omitted, the sentence is perfectly acceptable. Based on these types of patterns, *ever* and similar items are labelled “Negative Polarity Items” since they are sensitive to the polarity of the context that contains them, and they are acceptable in negative contexts (in contrast with PPIs, which are unacceptable in negative contexts).

- (43) a. I think I’ve (**ever*) been here before.
 b. I don’t think I’ve (*ever*) been here before.

In this work, we explore how an incremental sentence comprehender uses knowledge of the grammar of NPIs to construct representations of NPI-containing sentences. We use the selective failure of this system — the NPI illusion — as a window into these processes. First, though, we discuss some proposals about the nature of this grammatical knowledge. Different formulations of the grammatical constraint on the distribution of NPIs lend themselves to different expected weaknesses in the online implementation of this knowledge.

The key issue we discuss here with respect to the grammar of NPIs is the question of what kind of representation makes the NPI in (43b) acceptable. At first blush, a reasonable answer is that it is the negative word *don’t* that makes the NPI acceptable. Accordingly, the negative word is often called the “licensor” of the NPI. An alternative hypothesis locates the domain of licensing in the representation of the context that contains the NPI, which must have an appropriate kind of meaning (i.e., a negative meaning). The particular word(s) that created that meaning are, under this type of hypothesis, not directly relevant.

We note that this is by no means the only (or even the primary) question addressed by the literature on the grammar of NPIs. Important work has emphasized the differences between NPIs across languages and within a language, including so-called “strong” and “weak” NPIs. However, for the purposes of our investigation of the time course of NPI illusions, the domain of licensing is central and so we focus our attention on this aspect of the grammar.

NPI licensing hypotheses fall into three broad categories: (i) those that focus on a critical syntactic relationship between a negative word and the NPI, often with an additional pragmatic licensing mechanism based on negative implicatures; (ii) those that focus on a critical pattern of entailments of the context containing the NPI, which hold regardless of whether the NPI is actually included in the sentence; and (iii) those that focus on the NPI-containing sentence’s meaning, as compared to pragmatic alternatives to the sentence. There are of course proposals that do not align with any of these categories, such as Postal’s suggestion that NPIs are in fact negative concord items (Postal 2005) and Barker’s scope licensing hypothesis (Barker 2018), but most accounts of NPI licensing fall into one of these categories.

Within the first category of hypotheses, the earliest proposals for the grammar of NPIs were formulated in terms of transformational grammar rules that govern the use of NPIs instead of non-NPI lexical items (e.g. *any* instead of *some*) when they occur “in construction with” or are c-commanded by a negative word (Klima 1964; Jackendoff 1969). These hypotheses highlight a syntactic relationship between a negative word and an NPI. They achieve better empirical coverage by formulating the constraint in terms of “affective” elements rather than negative elements. The word “reluctant” in (44) for example, is in the “affective” class under these hypotheses, despite not being explicitly negative. This broadening of the set of so-called “licensors” allows the theory to capture the acceptability of sentences like (44), which contains the NPI *anything*.

(44) John was reluctant to read anything about the war.

(Jackendoff 1969:220)

A key challenge, however, is the problem of defining the class of licensors. Simply stating that all “affective” items are licensors is unsatisfactory if our only diagnostic for whether an item is “affective” is whether it licenses NPIs. In response to this difficulty, subsequent work has backtracked this expansion

of the licensing mechanism — such that only explicitly negative words may license NPIs through a direct c-command relation — and has postulated a secondary mechanism to account for the acceptability of sentences like (44). The nature of this secondary mechanism varies across proposals. C. L. Baker 1970, in an exploration of the licensing properties of doubly-negated sentences, proposes a key role for sentence-level entailments in addition to a syntactic licensing mechanism. Linebarger 1987 updates C. L. Baker’s theory and proposes that it is negative implicatures rather than entailments that are critical to the secondary licensing mechanism. Giannakidou’s (2006) “rescuing” operation plays a similar role, though for her it is any proposition “made available” by the global context that can license the NPI, making this proposal more general than Linebarger’s implicature-based mechanism. (It should be noted that Giannakidou’s hypothesis also differs in the nature of the primary licensing mechanism.) These hypotheses have an intuitive appeal in their ability to straightforwardly predict the difference in acceptability between sentences like (43a) and sentences like (43b), though it has proven difficult to pinpoint the precise nature of a secondary licensing mechanism. We note that the cue-based retrieval model that Vasishth et al. propose for the processing of NPIs (Vasishth et al. 2008), which uses [+negation] and [+c-command] as retrieval cues, is essentially a direct online implementation of a grammar in this category, but without an implemented secondary mechanism for licensing by entailment/implicature/inference. We explain this model in greater detail in section 4.1.2.1.

In what would become a significant influence on subsequent hypotheses, Ladusw 1979 proposed that the licensing constraints on NPIs are best described by the entailment patterns of the contexts that contain them. Specifically, Ladusw proposed downward entailment — that is, entailment from sets to subsets — as the critical property. Subsequent analyses have proposed different entailment conditions that must be met in order for NPIs to be acceptable, such as nonveridicality (Giannakidou 1998), anti-morphic contexts for strong NPIs (van der Wouden 1997), anti-additivity for strong NPIs (Zwarts 1998), Strawson downward entailment (Von Stechow 1999), or downward entailment with or without non-truth-conditional content taken into account (Gajewski 2011). The key difference between hypotheses in this group and those discussed above is the emphasis on clause-level properties. A transparent online implementation of such a hypothesis, then, would involve access to and evaluation of these properties. How-

ever, one could also re-state such a hypothesis as an item-to-item dependency, by treating the lexical item that gives rise to the relevant property as a “licensor” which must be retrieved, and defining the class of “licensors” as the DE (or nonveridical, or anti-morphic, etc.) operators. In fact, the actual mechanism Ladusw originally proposes is one in which NPIs are licensed by a scope relation with a DE operator, though Homer 2008 and Homer 2021 argue that in cases where operator-based and context-based hypotheses make different predictions, context-based hypotheses better account for the data.

Another family of approaches to NPI licensing focuses not on the entailments of the clause that contains the NPI, but rather on the NPI-containing sentence’s meaning relative to its pragmatic alternatives. The main idea in this third category of hypothesis is that NPIs, for the most part, refer to endpoints of scales — for example, *lift a finger* indicates investing some minimal amount of effort and *ever* indicates some maximal time span. Fauconnier 1975a proposed that in virtue of these properties, NPI-containing sentences have a particular relation to their scalar alternatives. For example, *Mary hasn’t ever been to Paris* expresses a stronger or more informative claim than *Mary hasn’t been to Paris in the last three weeks*. This strength relation no longer exists when negation is omitted. For Fauconnier it is the uninformativity of the NPI-containing positive sentence that makes it unacceptable. In later work, Israel 1997 extends this framework to PPIs and non-minimizer NPIs and Kadmon & Landman 1993 attempt to unify FCI *any* and NPI *any* under a similar approach. This third category of hypotheses is incompatible with the treatment of the negative word as a “licensor” per se because it is truly the properties of the entire clause or even the entire sentence that allow the NPI to be licensed. The fact that a particular word gave rise to those properties, and that that word may have a c-command or scope relation to the NPI, is incidental. A processing theory that is faithful to such a grammar, then, would need to involve the incremental generation of pragmatic alternatives at the clause or sentence level, and evaluation of these alternatives with respect to whether they are scalar and appropriately ordered. For further details on one possible implementation of such a grammar, see Chapter 3.

In this brief review we have seen that existing proposals for the nature of the grammatical constraints that govern the distribution of NPIs differ considerably. The key dimension of variability that we have highlighted here is the unit of representation that the NPI is in a dependency with: while many early

proposals identified the negative word itself as a “licensor” (making NPI-licensing an item-to-item dependency), other work has emphasized the properties of the context that contains the NPI (an item-to-context dependency). In what follows, we address a similar question, but with respect to the nature of online sentence processing. That is, we aim to determine whether the real-time licensing of NPIs is an operation that relates NPIs to negative words that precede them or an operation that relates NPIs to the contexts that contain them. Note that these questions are not identical, as real time licensing operations might not be fully isomorphic to grammatical constraints. But we consider it likely that they are related, and can therefore inform one another. We now turn to the online sentence processing background that motivates this work.

4.1.2 Linguistic illusions

Grammatical theories of NPI licensing like the ones discussed above aim to identify the nature of the linguistic knowledge that makes sentences like (45a) acceptable and sentences like (45b) unacceptable. A shared prediction across all of these hypotheses is that sentences like (45c) should also be unacceptable. For hypotheses that focus on the role of the negative word, this is because the appropriate structural relationship (c-command or scope) between items does not obtain. For hypotheses that focus on the role of the surrounding context, this is because the NPI is not contained within the negated RC.

- (45)
- a. No bills that the senators voted for have ever become law.
 - b. * The bills that the senators voted for have ever become law.
 - c. * The bills that no senators voted for have ever become law.

For traditional acceptability judgments as well as untimed Likert acceptability ratings, this prediction is generally borne out. Native speakers who have the time and motivation to carefully consider (45c) conclude that it is unacceptable. Importantly, however, in speeded measures, the ungrammaticality of (45c) is less apparent to comprehenders, and these sentences are judged acceptable much more frequently than sentences like (45b). First demonstrated by Drenhaus, Saddy, & Frisch 2005, the NPI illusion is the contrast between (45b) and (45c) in early stages of processing. This contrast exists in speeded judgments

of acceptability (German: Drenhaus, Saddy, & Frisch 2005; English: Xiang, Dillon, & Phillips 2006; Parker & Phillips 2016; de Dios Flores, Muller, & Phillips 2017; Hildebrandt & Husband 2017; Muller, de Dios Flores, & Phillips 2019; Orth, Yoshida, & Sloggett 2020a; Korean: Yun, Lee, & Drury 2018), as well as in self-paced reading (English: Parker & Phillips 2011; Parker & Phillips; Xiang, Grove, & Giannakidou 2013; Ng & Husband 2017; Yanilmaz & Drury 2018a) eye-tracking (German: Vasishth et al. 2008; English: Orth, Yoshida, & Sloggett 2020b) and event-related potentials (German: Drenhaus, Saddy, & Frisch 2005; English: Xiang, Dillon, & Phillips 2009; Turkish: Yanilmaz & Drury 2018b; Korean: Lee et al. 2018). A key question is what the real-time NPI licensing mechanism consists of, such that it fails in exactly these cases.

NPI illusions are one example from a broader class of phenomena in which initial judgments misalign with stored knowledge. Agreement attraction is another such case, in which the ungrammaticality of (46a) is not as readily detected as the ungrammaticality of (46b), as can be shown in production tasks, speeded acceptability tasks, and reading times (Bock & Miller 1991; Nicol, Forster, & Veres 1997; Clifton, Frazier, & Deevy 1999; Pearlmutter, Garnsey, & Bock 1999; Wagers, Lau, & Phillips 2009; Patson & Husband 2016; Slevc & Martin 2016; Hammerly, Staub, & Dillon 2019; Schlueter, Parker, & Lau 2019; Lago, Acuña Fariña, & Meseguer 2021; among others)

- (46) a. * The key to the cabinets are rusty.
 b. * The key to the cabinet are rusty.

(Bock & Miller 1991:56)

Note, however, that not all linguistic illusions involve an irrelevant element which intervenes in the resolution of a syntactic dependency. For example, in the substitution illusion (often called the “Moses illusion”), comprehenders respond *two* to questions like (47), even if they know that it was Noah, not Moses, in the story about an ark, and even if they know that their task is to identify word substitution errors (Erickson & Mattson 1981). We therefore might not expect linguistic illusions to constitute a natural class at the level of mechanism.

- (47) How many animals of each kind did Moses bring on the ark?

We now turn to some existing proposals for the mechanism underlying the NPI illusion, as well as some important empirical generalizations that call into question the viability of previous explanations.

4.1.2.1 Early accounts of the NPI illusion

One prominent variety of hypothesis for the NPI illusion attributes the phenomenon to the properties of the memory architecture (Vasishth et al. 2008). This hypothesis leverages independently-motivated properties of memory systems, specifically the idea that memory retrievals are executed through parallel activation of retrieval cues. For any item in memory that was encoded with features that match these retrieval cues, activation should increase as a consequence of the increased activation of the features. Once an item’s activation (which is determined by both its prior activation state and the boost it received from the activation of retrieval cues) crosses some threshold, it is retrieved. Vasishth et al. propose such a mechanism for the online licensing of NPIs, such that a lexical licenser in memory can be retrieved according to a set of retrieval cues. They propose [+c-command] and [+negation] as the cue set, but note that this could easily be adapted to align more closely with any grammatical proposal that treats the negative word as a licenser — for example, a cue set such as [+scope] and [+DE] would combine this proposal with Ladusw’s (1979) hypothesis.

An important caveat is that relational properties like c-command and scope are not easily translated into a feature encoding system like the one assumed here. Intuitively, this can be understood by noting that “c-command” is not a property that holds of some nodes of a tree and not others, but rather a relation that holds between some pairs of nodes and not others. Setting this concern aside, the critical successful prediction of this hypothesis lies in the treatment of partial matches to retrieval cues. That is, because of noise in the system, an item in memory that matches some but not all retrieval cues will sometimes reach threshold anyway; thus, the non-c-commanding licenser *no* inside the RC in (45c) will sometimes be retrieved, leading to the illusion of acceptability. The same mechanism, with different feature combinations, has been proposed as an explanation for agreement attraction (Wagers, Lau, & Phillips 2009), and the framework can easily be extended to other phenomena like illusion effects in anaphora processing (Jäger et al. 2020). The potential of a single mechanism to account for numerous phenomena makes this

hypothesis appealingly general. However, it is not obvious that these illusions have identical profiles, an issue we return to below. Note also that the treatment of NPI licensing as a memory retrieval operation in which a licenser is sought based on its features aligns nicely with grammatical hypotheses in which NPIs' distribution is dictated by their syntactic relation to a negative word, but is less straightforwardly compatible with hypotheses in which the locus of licensing is the context that contains the NPI. That is, while it is straightforward to extend retrieval-based hypotheses such that a larger unit (e.g., a clause instead of a word) is retrieved, under context-based licensing assumptions, the representation that licenses the NPI is in fact not a prior unit in memory, but the representation that is currently under construction. Moreover, the kinds of properties that are thought to be relevant at the context level are aspects of the meaning of a clause, and it remains to be seen how something like a group of propositions that are ordered according to a particular scale could be encoded as a feature on an individual chunk in memory.

An alternative explanation for the NPI illusion comes from Xiang, Dillon, & Phillips 2009, who propose that the impression of acceptability for sentences like (45c) is due to over-application of a secondary NPI licensing mechanism like the ones proposed by C. L. Baker 1970, Linebarger 1987, and Giannakidou 2006. Recall that these hypotheses treat licensing by a syntactic relation with explicit negation as the primary way NPIs are licensed, whereas acceptance of NPI-containing sentences without negation is attributed to a secondary mechanism, called “rescuing” in Giannakidou’s framework. Under this hypothesis, NPI illusion sentences are erroneously “rescued” due to the contrastive implicatures triggered by the use of a restrictive RC. However, as is discussed at length in Chapter 3, such a mechanism risks being too general, in that any sentence with a restrictive RC, including the ungrammatical baseline sentences like (45b), could trigger the relevant implicatures. Because of this limitation, we do not explore this hypothesis further.

4.1.2.2 The licenser effect

Two important empirical generalizations inform our understanding of the conditions that give rise to NPI illusions. The first of these is a contrast between standard illusion sentences like (48a) and similar sentences like (48b). Note that both of these sentences involve a negative word inside of an RC.

- (48) a. * The bills [that no senators voted for] have ever become law.
 b. * The bills [that the senators didn't vote for] have ever become law.

Despite the obvious parallels between these sentences, de Dios Flores, Muller, & Phillips 2017, Orth, Yoshida, & Sloggett 2021 and Chapter 3 all find that illusions are substantially reduced or absent for (48b), relative to (48a). That is, sentences with sentential negation in the RC are consistently rejected, typically at equal rates to ungrammatical baseline sentences. We refer to this contrast as the “licensor effect”. Importantly, this finding is unexpected under both hypotheses sketched above. For the cue-based retrieval hypothesis, any element with the same features (i.e. any non-c-commanding negative word) should yield illusions. Similarly, for the pragmatic rescuing hypothesis, any restrictive RC should yield illusions. Accordingly, both Orth, Yoshida, & Sloggett 2021 and Chapter 3 propose alternative explanations for the illusion, which capture the observed contrast. Orth, Yoshida, & Sloggett argue that the scope-taking properties of quantifiers are central to the NPI illusion. It is well-documented that quantified phrases can be interpreted in positions other than their surface position, though we note that negative quantifiers do not typically allow for this flexibility. Under Orth, Yoshida, & Sloggett’s hypothesis, the negative quantified phrase *no senators* is erroneously interpreted as if it takes scope over the entire MC and, as a result of this prior error, the subsequent MC NPI can be licensed. Non-quantificational forms of negation are not prone to this interpretive error under this hypothesis, because sentential negation lacks the scope flexibility that quantifiers sometimes demonstrate. In contrast, Chapter 3 proposes that it is not the individual lexical licensors that drive this difference but rather their consequences for the interpretation for the entire RC. Specifically, they suggest that negative quantifiers like *no* may evoke scalar alternatives (i.e. *the bills that few senators voted for*, *the bills that many senators voted for*, etc.) whereas verbal negation like *didn't* may, in the absence of further context, evoke merely binary alternatives (i.e. *the bills that the senators did vote for*). Assuming a grammatical hypothesis like Fauconnier’s (1975) proposal, scalar pragmatic alternatives are critical to NPI licensing; thus the interpretation of the RC in (48a) creates a vulnerability that is not triggered for (48b), under this hypothesis.

While this contrast is important for understanding the NPI illusion, and will become relevant to our experiments in what follows, it is worth noting that this finding does not, on its own, bear on the

present question. Recall that we are interested in whether the NPI illusion involves erroneous licensing by a non-c-commanding negative word or erroneous licensing by a non-local negative context, in order to determine whether online NPI licensing in general constitutes an item-to-item dependency or an item-to-context dependency. Sentences like (48a) and (48b) obviously use different negative words, but the RCs that contain them also express different meanings (a fact which is central to the explanation for the contrast proposed in Chapter 3). In order to determine whether it is negative c-commanding words or negative local contexts that license NPIs, we must manipulate these representations (or their accessibility) independently. The distance effect provides an opportunity to do this.

4.1.2.3 The distance effect

The second key generalization regarding the selectivity of the NPI illusion concerns the position of the NPI, an effect first reported by Parker & Phillips 2016. A critical comparison which illustrates this effect is given in (49). Parker & Phillips found that while sentences like (49a) give rise to robust illusions, sentences like (49b) yield no detectable illusions. That is, they are consistently rejected, at rates similar to those for ungrammatical baseline sentences. In both cases the NPI occurs in a position outside the RC, where it cannot be licensed, but illusions only arise when the NPI is relatively early in the sentence. It is important to clarify that the distance effect is distinct from the contrast between speeded and untimed judgments, though both effects are concerned with the relative timing of illusion-related processes.

- (49) a. * The journalists [that no editors recommended for the assignment] ever thought that the readers would understand the complicated situation.
- b. * The journalists [that no editors recommended for the assignment] thought that the readers would ever understand the complicated situation.

(Parker & Phillips 2016:328)

Parker & Phillips summarize their findings as demonstrating that an increase in the distance between the negative word *no* and the NPI *ever* has the effect of “turning off” the illusion, though note that we will question this characterization momentarily. They offer two additional demonstrations of the distance

effect, exemplified by the contrasts in (50) and (51). First, the comparison in (50) demonstrates that even just one additional intervening word can turn off the illusion, as the post-verbal NPI *any* in (50b) is not subject to illusions, whereas the pre-verbal NPI *ever* in (50a) yields clear illusions in their data. Assuming that it is the distance from negative word itself that matters, it is surprising that five intervening words and four intervening words would have such different effects. Parker & Phillips therefore infer that not all intervening words are created equal, such that intervening verbs are in some sense special.

- (50) a. * The authors [that no critics recommended] have ever received acknowledgment for a best-selling novel.
- b. * The authors [that no critics recommended] have received any acknowledgment for a best-selling novel.

(Parker & Phillips 2016:325)

Parker & Phillips additionally demonstrate a contrast between (51a), where they find illusions, and (51b), where they find none. This finding is critical for the inference that the distance effect is in fact a distance effect per se, rather than an effect that is specific to post-verbal NPIs. That is, while a verb may be a particularly good intervener, it is not strictly necessary for the illusion to be turned off — the contrast between (51a) and (51b) shows that intervening parentheticals can have the same impact.

- (51) a. * As the editors mentioned, the authors [that no critics recommended for the assignment] have ever received a pay raise.
- b. * The authors [that no critics recommended for the assignment] have, as the editor mentioned, ever received a pay raise.

(Parker & Phillips 2016:331)

Based on these findings, the authors conclude that the accessibility of the interfering negative word declines throughout the sentence after it has been encountered, though this decline is not necessarily linear, since some intervening words like verbs seem to have a particularly strong effect. Importantly, their framing presupposes that the interfering representation is that of the negative word *no*. However, the data pattern they report is equally well described as follows: an increase in the distance between the

context in which NPIs are licensed — the entire RC — and the NPI *ever* has the effect of turning off the illusion. Because the stimuli in this experiment uniformly manipulated the position of the NPI in the post-RC region, and not the content of the RC itself, these two factors — distance to the negative word and distance to the negative context — are perfectly confounded. But by de-confounding them, as we attempt to do in the current study, we may determine whether it is in fact the distance to the negative word or the distance to the negative context that matters and, by extension, whether it is the negative word or the negative context that interferes. One consequence of this shift in the framing of the distance effect is that, the contrast between (50) and (51) is not an effect of four versus five intervening words, but of one versus two intervening words (since we count from the RC edge). In this case, verbs might still be especially impactful interveners, as Parker & Phillips suggest, or it could be that a lone intervening auxiliary like *have* is especially easy to ignore.

In the current study, we begin by assuming that the accessibility of the interfering representation declines monotonically but not necessarily linearly, and ask what the interfering representation consists of (a negative word or a negative clause). Note that some results presented in Chapter 5 may call into question the assumption of monotonicity — that is, comprehenders may become vulnerable to illusions again at later points in the sentence — though these findings are inconclusive.

4.1.3 The present study

Here we investigate the nature of the interfering representation in NPI illusions, specifically asking whether it is the non-c-commanding negative word that interferes or the non-local negative context. We approach this question primarily through manipulations of the previously-reported distance effect, through six experiments using a mixture of untimed acceptability and speeded acceptability measures. In the interest of brevity, we will refer to the distance between the negative word and the NPI as the “negation-NPI” distance, and to the distance between the negative context (which is always an RC) and the NPI as the “RC-NPI” distance. Our investigation also engages with the question of the nature of the change in accessibility of the interfering representation — that is, whether this representation becomes monotonically less accessible with the passage of time, or monotonically less accessible as a function of structural distance

to the NPI, or if the accessibility of representation changes in a non-monotonic way, such that after being “turned off” it can “turn back on” later in the sentence. While this latter question is critical to a full understanding of the NPI illusion, we begin by adopting the simplifying assumption, following Parker & Phillips 2016, that the accessibility of the interfering representation declines monotonically with the passage of time.

Experiment 7 evaluates a straightforward prediction of the hypothesis that the interfering representation is the negative word, by adding material inside the RC, thereby increasing the negation-NPI distance while holding constant the RC-NPI distance. We do not find an impact of such material. Experiment 8 and Experiment 9 directly compare the consequences of increasing the negation-NPI distance and those of increasing the RC-NPI distance. We find that only RC-NPI distance influences illusion rates. Experiment 10 then investigates whether the negative word could still be what interferes, but its accessibility declines not with time but with structural distance. We address this using a modification of the licenser effect discussed above, and find no support for the hypothesis that it is structural distance to the negative word that matters.

Experiment 11 was designed to establish whether the RC representation lingers because the comprehender does not have sufficient time to de-activate it or because the comprehender encounters nothing in the MC to pull their attention away. We additionally explore whether NPI illusions are specific to the NPI *ever*, or can occur equally for other NPIs such as *any* when the distance effect is controlled for. We find that illusions for *any* are possible, and while distance effects are robust, it remains unclear what drives the decline in susceptibility. This issue is an important area for future research.

4.2 Intervening prepositional phrases

4.2.1 Experiment 7: speeded acceptability

As a first attempt to identify the relevant distance for turning off illusions, we evaluated the impact of intervening prepositional phrases (PPs) within the RC for NPI-illusion type sentences. For example, we compared sentences like (52a) and (52b). Note that similar sentences to these, both with and without

PPs, have been shown to yield illusions in prior work. For example, Parker & Phillips 2016 Experiment 2 did not include PPs within the RC, whereas their Experiment 4 did include them, and both experiments found NPI illusions. However, we know of no direct comparison of these sentence types.

- (52) a. * The authors [that no critics have recommended in their reviews] have ever received acknowledgment for a best-selling novel.
- b. * The authors [that no critics have recommended] have ever received acknowledgment for a best-selling novel.

Critically, the licenser-NPI distance is seven words in (52a) and four words in (52b), whereas the RC-NPI distance is one word in both cases. Thus, a contrast in illusion rates between these two sentence types would constitute clear evidence in favor of the hypothesis that it is negation-NPI distance that matters, and, by extension, that it is the negative word itself that interferes. Equal illusion rates are clearly predicted by the hypothesis that it is RC-NPI distance that matters, but null effects are of course difficult to interpret. We address these concerns further in section 4.2.1.7.

4.2.1.1 Participants

46 US-based native speakers of English participated in this experiment. All participants provided informed consent and they received \$6 as compensation. In this and the following experiments participants were recruited using Amazon Mechanical Turk, and were asked to complete a native speaker qualification test; only participants that answered at least 7 out of 9 questions correctly were allowed into the task. We excluded workers who failed to provide a response within 2 seconds for 25% of fillers or more and workers whose judgments of filler trials did not reliably distinguish between grammatical and ungrammatical fillers, based on a chi-squared test. 3 workers were excluded based on these criteria, resulting in 43 participants in our analysis. The mean filler-trial accuracy of the included participants was 80%.

4.2.1.2 Materials

The experimental materials consisted of 36 items across 6 conditions that varied the position and presence of the negative quantifier crossed with the presence of a PP within the RC. A representative item,

With prepositional phrases	
A. Grammatical baseline	No surgeons [that the patients have consulted about the operation] have ever ...
B. Embedded negation	The surgeons [that no patients have consulted about the operation] have ever ...
C. Ungrammatical baseline	The surgeons [that the patients have consulted about the operation] have ever ...
...expressed dissatisfaction with the hospital staff.	
Without prepositional phrases	
D. Grammatical baseline	No surgeons [that the patients have consulted] have ever ...
E. Embedded negation	The surgeons [that no patients have consulted] have ever ...
F. Ungrammatical baseline	The surgeons [that the patients have consulted] have ever ...
...expressed dissatisfaction with the hospital staff.	

Table 4.1: Example stimuli for Experiment 7

including all six conditions, is shown in Table 1. Both conditions with negation in the MC (conditions A and D) were expected to be judged acceptable, and these served as grammatical baselines. Both conditions without negation (conditions C and F) were expected to be judged unacceptable, and these served as ungrammatical baselines. The key question was whether the conditions with embedded negation (conditions B and E) yield equal illusion rates, relative to their respective ungrammatical baselines, or if instead the illusion rate is reduced when the PP is included, resulting in an interaction.

In this and all following experiments, the items were distributed in a Latin Square design and each participant was randomly assigned to a list. In addition to 36 experimental items, each participant judged the same 72 filler items. These items included a mix of grammatical and ungrammatical sentences, including 24 fillers whose structures closely matched the experimental items but without the inclusion of an NPI. This was done to ensure that participants were not able to anticipate a sentence’s grammatical status before the end of the sentence based on frequent patterns. Participants additionally completed two practice trials before beginning the experiment.

4.2.1.3 Procedure

In this and all following experiments, experimental trials and fillers were presented in a randomized order for each participant. Each sentence was displayed word by word at a fixed rate of 400 ms per word, in the center of the screen. At the end of the sentence participants were asked “Was that a good sentence?” and had to provide a “yes” or “no” judgment within 2 seconds. If participants failed to respond in time, the

trial ended automatically and a message indicated that they were too slow. The task lasted approximately 30 minutes.

4.2.1.4 Analysis

Results were analyzed using logistic mixed effects models, fitting the maximal random effects structure first, which included random intercepts and slopes for both participants and items. When models failed to converge, the random effects structure was simplified following recommendations from Barr et al. 2013. Models used were fit using helmert coding, then we used the emmeans package (Lenth et al. 2018) to extract beta coefficients and p-values for critical pairwise comparisons and interactions. The key comparisons are as follows. First, we expect to find pairwise differences between grammatical baseline and ungrammatical baseline sentences, for both the with-PP and without-PP conditions. The key question, then, is whether the embedded-negation conditions for both the with-PP and without-PP conditions differ from their respective ungrammatical baselines (i.e., whether illusions occur for both) and, critically whether the illusion effect interacts with the presence or absence of a PP.

4.2.1.5 Results

The results are shown in Figure 4.1. An effect of grammaticality was observed for both the with-PP ($\beta=4.06$, $SE=0.47$, $z=8.65$, $p<.001$) and without-PP ($\beta=4.78$, $SE=0.49$, $z=9.77$, $p<.001$) conditions, indicating that the grammatical baseline conditions were significantly more likely to be judged acceptable than the corresponding ungrammatical baseline conditions. An effect of embedded-negation was observed for both the with-PP ($\beta=1.31$, $SE=0.32$, $z=4.16$, $p<.001$) and without-PP ($\beta=1.82$, $SE=0.34$, $z=5.42$, $p<.001$) conditions, indicating the embedded-negation conditions were significantly more likely to be judged acceptable than the corresponding ungrammatical baseline conditions — that is, reliable NPI illusions occurred in both cases. We did not observe a significant PP by embedded-negation interaction ($\beta=-0.51$, $SE=0.34$, $z=-1.49$, $p=.14$).

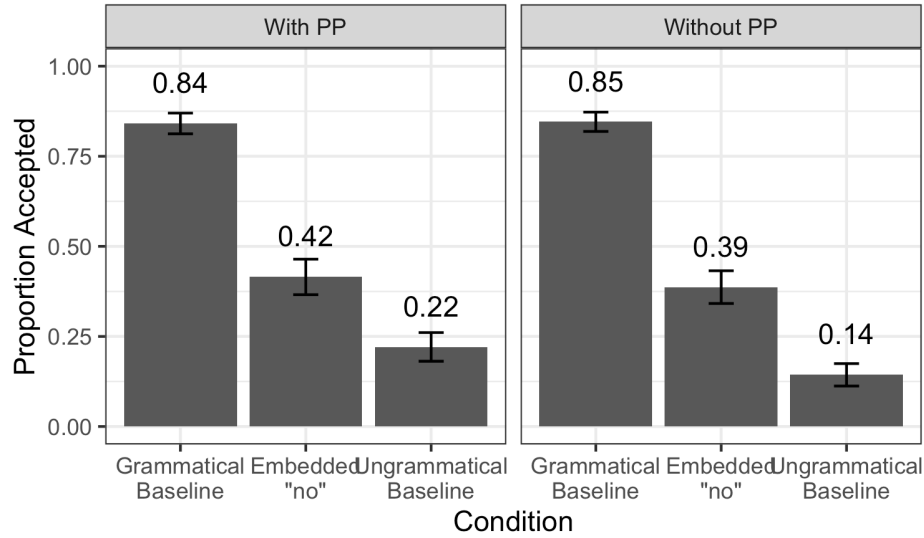


Figure 4.1: Mean percentage of ‘YES’ responses for the experimental conditions in Experiment 7. Error bars indicate standard error of the mean across subjects.

4.2.1.6 Statistical power

Post hoc power analyses were conducted using *simr* (Green & MacLeod 2016). The critical comparison is the interaction between the illusion effect (embedded negation versus ungrammatical baseline) and distance (with-PP versus without-PP). We estimated the predicted effect size using the three speeded acceptability experiments conducted in Parker & Phillips 2016, all of which found a distance effect for NPI illusions. The raw data from these three experiments was pooled. We then fit a logistic mixed effects model and computed the lower boundary of a 60% confidence interval around the meta-analytic estimate of the critical interaction (see Perugini, Gallucci, & Costantini 2014), arriving at an effect size estimate of $\beta=-1.01$. We then set the equivalent interaction in the present experiment to be equal to this effect and computed power using simulations. The present experiment achieved between 71% and 94% power to detect the interaction between the illusion effect and distance, assuming an effect size of $\beta=-1.01$.

4.2.1.7 Discussion

The findings from Experiment 7 demonstrate a clear illusion effect for sentences with and without a PP in the RC, and no statistically reliable interaction. Thus, we do not see evidence that an additional three

words between the negative word and the NPI can turn off the illusion. In principle, it is possible that licensor-NPI distance does in fact matter, but the impact is more gradient than was previously reported, such that only a very small decrease in the illusion occurs when a PP is added, which this experiment failed to detect. Note, however, that Parker & Phillips 2016 Experiment 2 found that the illusion disappeared with only one additional intervening word, making this explanation less likely. Alternatively, the lack of a contrast observed in Experiment 7 could be accommodated by a hypothesis in which negation-NPI distance matters, but only certain types of added material are able to impact the relevant representations. Along these lines, Parker & Phillips 2016 suggest that verbs may trigger a re-encoding of prior content. In order to address these concerns and determine definitively whether it is negation-NPI distance or RC-NPI distance that matters for NPI illusions, we designed Experiment 8 and Experiment 9 to directly compare the same intervening words inside and outside the RC.

4.3 Intervening verbs

Although Experiment 7 revealed no evidence that increases in negation-NPI distance can impact NPI illusions, this null finding could be explained by either a gradient impact of distance, such that the effect of the added PP was too small to detect, or a distance effect that is sensitive to not only the amount of intervening material but the type of intervening material, such that a PP is not the right kind of intervening content. We therefore designed Experiment 8 and Experiment 9 in an attempt to match both the type and, to the extent possible, quantity of intervening material inside and outside the RC. Thus, we compared sentences like those in (53).

- (53)
- a. * The surgeons [that no patients trusted] have ever prescribed experimental treatments.
 - b. * The surgeons [that no patients trusted to heal injuries] have ever prescribed experimental treatments.
 - c. * The surgeons [that no patients trusted] have healed any injuries with experimental treatments.

The comparison between (53a) and (53c) is analogous to the contrast reported in Parker & Phillips 2016 Experiments 1-3. Note that there are some differences between our stimuli and the ones used by Parker & Phillips (see section 4.3.1.2 below), but the key parallel is that we compare NPIs *ever* and *any*, positioned only one word apart. Parker & Phillips 2016 report a contrast between these sentence types, with illusions arising only for sentences like (53a). As discussed in section 4.1.2.3, this finding is expected under both accounts of the distance effect, since the added main verb increases both the negation-NPI distance and the RC-NPI distance. Critically, however, predictions diverge for sentences like (53b). Similar to the logic of Experiment 7, we reason that if it is the RC-NPI distance that matters, the added material in (53b) relative to (53a) should have no impact, but if it is the licensor-NPI distance that matters, this added material should reduce or eliminate the illusion. Unlike in Experiment 7, the added material includes a verb — in fact, it is precisely the same verb as that which immediately precedes *any* in (53c). Thus, concerns about the type of intervening material should not arise. Similarly, concerns about the number of intervening words are accounted for by the fact that the intervening material in (53b) is even longer than that in (53c). That is, if one added word (“healed”) in (53c) is enough to measurably reduce the illusion, then three added words (“to heal injuries”) in (53b) should also be sufficient, under the hypothesis that it is the licensor-NPI distance that matters. In sum, we expected to replicate Parker & Phillips 2016 and find a contrast between sentences like (53a) and sentences like (53c). The key question was whether (53b) would pattern with (53a) and yield illusions or with (53c) and yield no illusions.

4.3.1 Experiment 8: offline acceptability

The materials we developed for Experiment 9 deviate from previously-tested NPI illusion materials in a few ways (see section 3.1.2), making it important to validate the assumption that these novel conditions and their corresponding baselines are judged as expected in an offline task. Thus, in Experiment 8 we collected untimed Likert acceptability judgments for the sentence types discussed above, along with their corresponding grammatical and ungrammatical baselines.

4.3.1.1 Participants

27 US-based native speakers of English participated in this experiment. All participants provided informed consent and they received \$9 as compensation. We excluded workers whose judgments of filler trials did not reliably distinguish between grammatical and ungrammatical fillers, based on a one-sided t-test, as well as participants who answered two or more “attention check” trials incorrectly (see section 4.3.1.2). 1 participant was excluded based on these criteria. Among the 26 included participants, grammatical fillers were rated on average 4.85 out of 7 with a standard deviation of 0.55, and ungrammatical fillers were rated on average 2.73 out of 7 with a standard deviation of 0.81.

4.3.1.2 Materials

Experimental materials consisted of 36 items in nine conditions, as shown in Table 2. These conditions cross the standard illusion contrasts (i.e., an embedded negation condition, and its corresponding grammatical and ungrammatical baseline sentences) with the distance manipulation. In the “short” conditions (A-C), the NPI was *ever* and the RC contained no added material. Thus, both the negation-NPI and the RC-NPI distance can be considered “short”, or at least short enough to yield illusions. In the “long RC” conditions (D-F), the NPI was *ever* and the RC contained an adjunct with a verb, thus lengthening (relative to conditions A-C) the negation-NPI distance but leaving the RC-NPI distance unchanged. Finally, in the “long MC” conditions (G-I), the RC had no added material but the NPI was *any*, positioned after the main verb. Thus, both the negation-NPI distance and the RC-NPI distance were lengthened, relative to the “short” conditions. Note that because we matched the intervening verbs in conditions D-F exactly to the intervening verbs in conditions G-I, it was necessary to create different post-NPI spillover regions, to improve naturalness. We know of no prior research showing that the content of the spillover region influences illusion rates.

Each participant additionally rated 90 filler sentences, 39 of which contained a range of grammatical errors to encourage full use of the rating scale. Note that the fillers slightly over-represent grammatical sentences because 6 out of our 9 experimental conditions are ungrammatical; thus, the experiment as a whole is balanced. Participants also completed 8 “attention check” trials which were randomly inter-

Short	
A. Grammatical baseline	No surgeons [that the patients trusted] have ever ...
B. Embedded <i>no</i>	The surgeons [that no patients trusted] have ever ...
C. Ungrammatical baseline	The surgeons [that the patients trusted] have ever ...
...prescribed experimental treatments.	
Long RC	
D. Grammatical baseline	No surgeons [that the patients trusted to heal injuries] have ever ...
E. Embedded negation	The surgeons [that no patients trusted to heal injuries] have ever ...
F. Ungrammatical baseline	The surgeons [that the patients trusted to heal injuries] have ever ...
...prescribed experimental treatments.	
Long MC	
G. Grammatical baseline	No surgeons [that the patients trusted] have healed any injuries ...
H. Embedded negation	The surgeons [that no patients trusted] have healed any injuries ...
I. Ungrammatical baseline	The surgeons [that the patients trusted] have healed any injuries ...
...with experimental treatments.	

Table 4.2: Example stimuli for Experiment 8 and Experiment 9

spersed in the experiment. For example, an attention check trial read “For this sentence, please choose one as the answer.” Participants who answered more than one of these checks incorrectly were excluded from our analyses (see section 4.3.1.1 above).

4.3.1.3 Procedure

Participants were instructed to rate each sentence’s acceptability using a 7-point scale in which 7 was the most acceptable value and 1 the least acceptable. Sentences were displayed on the screen together with the scale, and participants were not under time pressure to provide a response. The experiment lasted about 45 minutes.

4.3.1.4 Analysis

The results were analyzed using a helmert-coded linear mixed-effects model whose maximal structure was initially built including by-subject and by-item random intercepts and slopes for the experimental conditions. When this model failed to converge, it was reduced according to the recommendations provided by Barr et al. 2013. Further details are included in Supplementary Files. For each of the distance configurations, we aimed to establish that the grammatical baseline sentences were significantly more acceptable

than either of the ungrammatical conditions. We additionally expected that the ungrammatical sentences within each distance configuration would be equivalent in their acceptability, though we note that small but reliable “offline illusions” have occasionally been observed for NPI illusions. We used the emmeans package (Lenth et al. 2018) to extract beta coefficients and p-values for pairwise comparisons between conditions.³⁰ The comparisons presented here use linear mixed effects models, but a model with similar structure but which treats the dependent variable as ordinal rather than linear yields similar conclusions and can be found in the Supplementary Files.

4.3.1.5 Results

The results from Experiment 8 are shown in Figure 4.2. A main effect of grammaticality was observed for both the comparison of the grammatical baseline to the ungrammatical baseline ($\beta=2.01$, $SE=0.25$, $t=-7.91$, $p<.001$) and the comparison of the grammatical baseline to the embedded-negation condition ($\beta=1.99$, $SE=0.25$, $t=7.84$, $p<.001$). These findings indicate that, averaging across distance configurations, both the embedded-negation and ungrammatical baseline conditions were judged less acceptable than the grammatical baseline conditions. Planned follow up comparisons yielded the same patterns for the short conditions (grammatical versus ungrammatical: $\beta=2.60$, $SE=0.29$, $t=9.13$, $p<.001$; grammatical versus embedded-negation: $\beta=2.62$, $SE=0.29$, $t=9.22$, $p<.001$), the long MC conditions (grammatical versus ungrammatical: $\beta=1.65$, $SE=0.29$, $t=5.80$, $p<.001$; grammatical versus embedded-negation: $\beta=1.96$, $SE=0.29$, $t=6.88$, $p<.001$), and the long RC conditions (grammatical versus ungrammatical: $\beta=1.78$, $SE=0.29$, $t=6.25$, $p<.001$; grammatical versus embedded-negation: $\beta=1.39$, $SE=0.29$, $t=4.88$, $p<.001$). We additionally observed no main effect of embedded negation relative to the ungrammatical baseline ($\beta=0.02$, $SE=0.13$, $t=0.15$, $p=.88$). Planned follow up comparisons yielded similar patterns for the short conditions ($\beta=-0.02$, $SE=0.18$, $t=-0.13$, $p=.90$) and the long MC conditions ($\beta=-0.31$, $SE=0.18$, $t=-1.70$, $p=.09$). However, in the long RC conditions there was a small but statistically significant effect of embedded negation ($\beta=0.39$, $SE=0.18$, $t=2.16$, $p=.03$).

³⁰Unless otherwise noted, p-values are not corrected for multiple comparisons, since the critical comparisons were determined a priori.

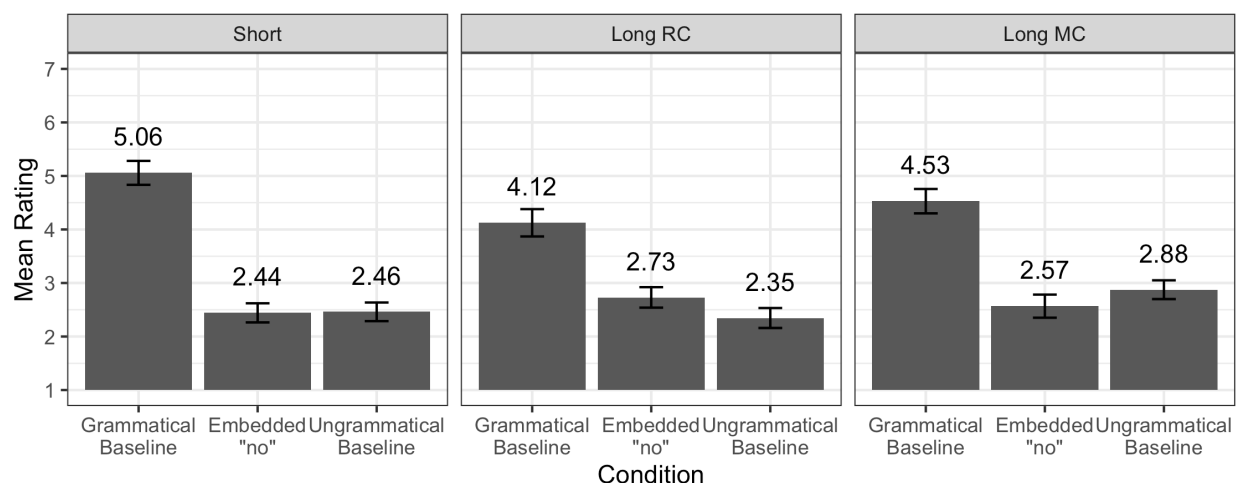


Figure 4.2: Mean percentage of ‘YES’ responses for the experimental conditions in Experiment 8. Error bars indicate standard error of the mean across subjects.

4.3.1.6 Discussion

Experiment 8 aimed to establish, using an untimed acceptability rating task, that the items designed for Experiment 9 are appropriate with respect to grammatical status and overall naturalness. Critically, we expected all three grammatical baseline conditions to be judged on the whole acceptable and all six ungrammatical conditions (three ungrammatical baseline conditions and three embedded negation conditions) to be judged on the whole unacceptable. This is the pattern that we observed in Experiment 8, with the caveat that a small but statistically reliable contrast between the embedded negation condition and the ungrammatical baseline condition emerged in the long RC distance configuration. However, due to the small magnitude of this effect (only .38 points on a seven-point scale) and previous reports of “offline illusions” in the NPI illusions literature (citations), we do not consider this finding to be a cause for concern. We additionally note that while the ratings for grammatical baseline sentences are not at ceiling (average ratings were 5.06 for the short baseline condition, 4.12 for the long negation-NPI, short RC-NPI condition, and 4.53 for the long baseline condition), these ratings are not surprising for sentences of this length and complexity.

4.3.2 Experiment 9: speeded acceptability

Having established that the stimuli are appropriate, we turn to the question of whether illusions arise in each of the distance configurations discussed in section 4.3.1.2, using a speeded acceptability task. Recall that we expected, following Parker & Phillips 2016, that the short conditions would yield clear illusions and the long MC conditions would yield a clear lack of illusions. The critical question was whether the long RC conditions would also yield illusions, indicating that the relevant distance for turning off the illusion is the RC-NPI distance, or would fail to yield illusions, indicating that the relevant distance is the negation-NPI distance.

4.3.2.1 Participants

45 US-based native speakers of English participated in this experiment; however, due to a server error only data from 44 participants was recorded. All participants provided informed consent and they received \$8 as compensation. We excluded workers who failed to provide a response within 2 seconds for 25% of fillers or more and workers whose judgments of filler trials did not reliably distinguish between grammatical and ungrammatical fillers, based on a chi-squared test, as well as participants who answered two or more “attention check” trials incorrectly. All 44 workers met these criteria. The mean filler-trial accuracy of the included participants was 81%.

4.3.2.2 Materials

The experimental items and fillers were identical to those used in Experiment 8. Participants additionally judged two practice trials before beginning the experiment, and eight attention check trials which were randomly interspersed during the experiment.

4.3.2.3 Procedure

The procedure was identical to that of Experiment 7. The task lasted approximately 50 minutes.

4.3.2.4 Analysis

As in our analysis of the Experiment 7 data, results were analyzed using logistic mixed effects models, fitting the maximal random effects structure first, and simplifying the random effects structure as needed. We again used emmeans package to extract beta coefficients and p-values for critical pairwise comparisons and interactions. The key comparisons are as follows. First, we expect to find pairwise differences between grammatical baseline and ungrammatical baseline sentences, for all three distance configurations. We additionally expect to find reliable illusions (a contrast between the embedded-negation condition and the ungrammatical baseline condition) for the short distance configuration and a lack of illusions for the long-MC distance configuration, and an interaction between these, replicating Parker & Phillips 2016. The key question is whether we find reliable illusions for the long-RC distance configuration, and whether this effect interacts with either the comparison between long-RC conditions and long-MC conditions or the comparison between long-RC conditions and short conditions.

4.3.2.5 Results

The results from Experiment 9 are shown in Figure 4.3. We observe significant grammaticality effects for all three distance configurations (short: $\beta=4.72$, $SE=0.43$, $z=10.95$, $p<.001$; long-RC: $\beta=3.91$, $SE=0.39$, $z=9.91$, $p<.001$; long-MC: $\beta=3.58$, $SE=0.39$, $z=9.30$, $p<.001$) indicating that the grammatical baseline sentences were judged acceptable more often than the ungrammatical baseline sentences in all three cases. Focusing on just the short and long-MC conditions which parallel Parker & Phillips 2016, we find a significant illusion effect (i.e. a contrast between the embedded negation condition and the ungrammatical baseline) for the short distance configuration ($\beta=0.98$, $SE=0.33$, $z=2.93$, $p=.003$) but not for the long-MC distance configuration ($\beta=-0.517$, $SE=0.32$, $z=-1.63$, $p=.10$), replicating Parker & Phillips 2016.

In the long-RC distance configuration we also found statistically significant illusions ($\beta=1.13$, $SE=0.32$, $z=3.50$, $p<.001$). We also found a statistically significant interaction between illusion magnitude (embedded negation versus ungrammatical baseline) and the location of intervening material (long-RC versus long-MC) ($\beta=1.65$, $SE=0.42$, $z=3.92$, $p<.001$). We did not observe a statistically significant interaction between illusion magnitude and the presence or absence of intervening material in the RC (short versus

long-RC) ($\beta=0.15$, $SE=0.43$, $z=0.36$, $p=0.72$).

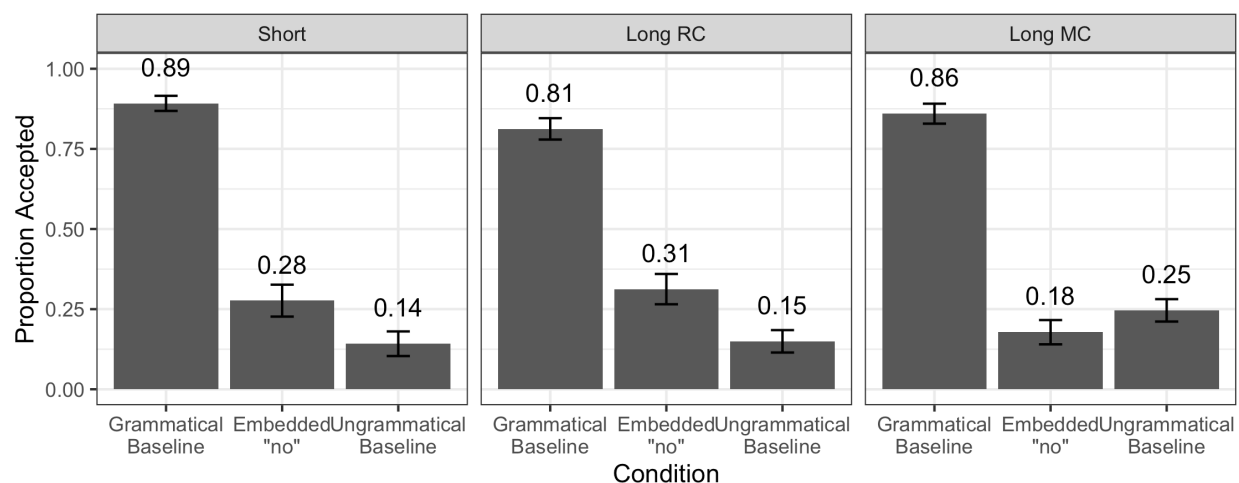


Figure 4.3: Mean percentage of ‘YES’ responses for the experimental conditions in Experiment 9. Error bars indicate standard error of the mean across subjects.

4.3.2.6 Discussion

Experiment 9 was designed to directly test whether added material inside the RC has the same impact on illusion rates as added material outside the RC. Under a hypothesis where negation-NPI must be short in order for illusions to occur, both of these two types of added material should reduce illusion rates. In contrast, under a hypothesis where RC-NPI distance must be short in order for illusions to occur, only added material outside the RC is predicted to have the effect of reducing illusion rates. Here we observed clear and statistically reliable illusion patterns for both the short distance conditions and the long RC conditions, in which material is added inside the RC, but we found no illusions for the long MC conditions, in which material is added outside the RC. Importantly, statistical interactions support these generalizations. These findings suggest that it is the RC-NPI distance that matters for NPI illusions, consistent with accounts in which it is the entire negated RC that interferes with the processing of the MC NPI.

These findings are inconsistent with an account of the NPI illusion in which it is the negative word itself that interferes, and the accessibility of this word declines with time. However, there are at least three

ways in which these findings can be accommodated without adopting the hypothesis that it is the entire RC representation that interferes. The first of these concerns the nature of the decline in accessibility. We have so far assumed, following Parker & Phillips 2016 that the relevant representation becomes less accessible with the passage of time. However, it is in principle possible that the negative word’s accessibility is instead a function of the number of intervening nodes between it and the NPI in the syntactic representation of the sentence. This would explain why added material inside the RC, which does not structurally intervene between the negative word and the NPI, has no impact on the illusion rate in Experiment 7 and Experiment 9. This possibility is addressed in Experiment 10. The second possibility is that the lack of illusions for post-verbal *any* is not due to distance effects at all, but instead due to a categorical impossibility of NPI illusions for the NPI *any*. In other words, under this explanation, NPI illusions might more accurately be called “*ever* illusions”. If the lack of illusions for *any* is not a distance effect, this raises the possibility that distance effects operate on a longer timescale than we assumed, in which case it may be that none of the experiments presented so far introduce enough added material to see distance effects. The possibility that illusions are impossible for *any* is addressed in Experiment 11. We find that this is not the case — rather, illusions do arise for *any* when *any* is sufficiently close to the RC. Finally, it is possible that the accessibility of the interfering representation is modulated in unexpected and perhaps non-monotonic ways throughout the remainder of the sentence, in which case the disappearance of illusion effects for some NPI positions would not be diagnostic of the interfering representation at all. This possibility is not especially likely in light of the evidence provided by Parker & Phillips 2016 showing that these effects are, in fact, distance effects, but see Chapter 5 for further discussion.

4.4 The licensor effect

The results of Experiment 7 and Experiment 9 suggest that NPI illusions are unaffected by the distance between the negative word and the NPI and that this distance can be manipulated with no impact on illusion rates as long as the distance from the RC edge to the NPI is short. However, another possible interpretation of the distance effect is that the relevant distance may be the distance from the licensor

if the critical computation is over structural distance, not time elapsed. If we imagine an online NPI-licensing mechanism which, upon encountering an NPI, traverses the constructed tree structure from child node to parent node and parent node to child node until a c-commanding negative word can be found — or, in the case of illusions, a non-c-commanding negative word is sometimes mistakenly found — then we might expect that nodes that are farther away in the syntactic structure may be less likely to be accessed. The intervening material added inside the RC in Experiment 7 and Experiment 9 do not intervene structurally between the licenser and the NPI and so, under this hypothesis, they should have no effect. Of course, modifications to this hypothesis are needed to account for the fact that a true NPI licenser can be many, many nodes away from the NPI it licenses, but we set these concerns aside for the moment.

In light of recent findings regarding the types of licensors that yield illusions, a structure-based explanation becomes particularly appealing. Both Orth, Yoshida, & Sloggett 2021 and Chapter 3 report a contrast in the illusion profile of sentences like (54a) and (54b). Negative quantifiers like *no* yield robust illusions while non-quantificational negation like *didn't* or *not* does not.

- (54) a. * The authors [that no critics recommended for the award] have ever received acknowledgement for a best-selling novel.
- b. * The authors [that the critics did not recommend for the award] have ever received acknowledgement for a best-selling novel.

While Orth, Yoshida, & Sloggett explain this effect as a consequence of the scope-taking properties of quantifiers and Chapter 3 explain it as a consequence of the inferred alternatives to negative clauses, it is also in principle possible that the effect is in fact due to the differing structural positions of *no* and *not*. Under the structure-mediated hypothesis sketched above, the online NPI licensing mechanism would need to traverse more nodes to reach *not* in (54b) than to reach *no* in (54a), if we take the NPI as the starting position. The possibility of explaining both the licenser effect and the distance effect under a single mechanism makes this hypothesis worthy of further consideration. This possibility is addressed by Experiment 10.

4.4.1 Experiment 10: speeded acceptability

In order to determine whether the difference in the structural position of an embedded negative quantifier as compared to embedded verbal negation underlies the reported licensor effect — and, by extension, whether there are structure-based distance effects on NPI illusions — we tested whether this contrast exists for SRCs. SRCs allow us to position a negative quantifier on an embedded object, instead of an embedded subject, thus placing the quantifier even lower than embedded verbal negation (see (55)). Note that we used *very few* as the form of quantificational negation, rather than *no*, following Xiang, Dillon, & Phillips 2009. See section 4.4.1.2 for details.

- (55) a. * The critics [that have recommended very few authors of alternative genres] have ever
 objected to mainstream literary trends.
- b. * The critics [that haven't recommended the authors of alternative genres] have ever ob-
 jected to mainstream literary trends.

If the cause for the reported lack of illusions for (54b) is the structurally low position of *not*, we should find a lack of illusions for both (55a) and (55b)), since both *very few* in (55a) and *haven't* in (55b)) are sufficiently structurally low to be inaccessible under this hypothesis. If instead the cause of the contrast in for sentences like (54a) and (54b) is a difference in scope taking properties (as is suggested in Orth, Yoshida, & Sloggett 2021) or scalar interpretations (as is suggested in Chapter 3), we should find illusions for (55a) but not (55b)). Note that both Orth, Yoshida, & Sloggett 2021 and Chapter 3 used SRCs in at least some conditions of some experiments. However, none of these experiments can on their own definitively answer the present question. Orth, Yoshida, & Sloggett compared embedded quantifiers in SRCs to embedded quantifiers in ORCs in their Experiment 11, finding a null result. This finding is consistent with the generalization that embedded negative quantifiers cause illusions regardless of their structural position, but we are cautious about reasoning from null results. Chapter 3 also used SRCs in Experiments 5 and 6, but found a small but statistically reliable illusion effect for embedded sentential negation, contrary to previous findings. This effect may be due to interference from other conditions in the experiment, and so we again cannot be certain. Thus, the present experiment fills a gap in prior

literature by determining whether the licensor effect reported in prior work extends to SRCs as predicted by both the scalar alternatives (Chapter 3) and scope miscalculation (Orth, Yoshida, & Sloggett 2021) hypotheses, or if instead this effect is a consequence of a more general structural distance effect.

4.4.1.1 Participants

41 US-based native speakers of English participated in this experiment. All participants provided informed consent and they received \$3 as compensation. We excluded workers who failed to provide a response within 2 seconds for 25% of fillers or more and workers whose judgments of filler trials did not reliably distinguish between grammatical and ungrammatical fillers, based on a chi-squared test. 3 workers were excluded based on these criteria. The mean filler-trial accuracy of the included participants was 83%.

4.4.1.2 Materials

Experimental materials consisted of 36 items in eight conditions, four of which are shown in Table 4.3 (The other four conditions were identical to these except that the NPI *ever* was removed. They functioned as filler items and are discussed below.) These conditions correspond to the standard grammatical and ungrammatical baseline sentences (conditions A and D, respectively, in Table 4.3), along with an embedded negative-quantifier condition and an embedded non-quantificational-negation condition (conditions B and C, respectively, in Table 4.3). One significant departure from previous studies is the use of SRCs, which is motivated by our discussion in section 4.4.1 above. Although our 36 items used the same items as used in the previous three experiments as a starting point, changes to the content of both the RC and spillover region were necessary to maintain naturalness (changing an ORC to a SRC changes who did what to whom). A second important change is the use of *very few* as a quantificational licensor in both the grammatical baseline condition and the embedded-negative-quantifier condition. Most prior studies have used the quantifier *no* here. We implemented this change due to concerns that *no* in object position is somewhat unnatural, possibly because of competition with the meaning-equivalent construction *didn't ... any* (for example, *The critic recommended no authors* is slightly awkward compared to *The*

A. Gramm. baseline	Very few critics [that have recommended the authors of alternative genres] have ever ...
B. Embedded <i>very few</i>	The critics [that have recommended very few authors of alternative genres] have ever ...
C. Embedded <i>not</i>	The critics [that haven't recommended the authors of alternative genres] have ever ...
D. Ungramm. baseline	The critics [that have recommended the authors of alternative genres] have ever ...
...objected to mainstream literary trends.	

Table 4.3: Example stimuli for Experiment 10

critic didn't recommend any authors). *Very few* does not share this property (at least not to the same extent), and one prior study (Xiang, Dillon, & Phillips 2009) found that illusions also arise for *very few* as an intrusive licenser in embedded subject position. No other notable changes were made to the items.

Each participant additionally rated 36 filler sentences, 20 of which contained a range of grammatical errors. As noted above, the experimental items also included four grammatical conditions not shown in Table 3, which correspond to the same sentences but with *ever* removed. These conditions functioned as fillers and ensured that participants could not predict that an NPI would arise (nor predict the grammatical status of the sentence) based on structure of the beginning of the sentence. This achieved the same purpose that is served by the sheer number and diversity of fillers in previous experiments. It also allowed us to ensure that the items were sufficiently natural and acceptable independent of the contribution of the (unlicensed) NPI. That is to say, they achieved the same purpose that is served by the untimed acceptability norms in previous experiments.

4.4.1.3 Procedure

The procedure was identical to that of Experiment 7 and Experiment 9. The task lasted approximately 30 minutes.

4.4.1.4 Analysis

As with Experiment 7 and Experiment 9, we analyzed the data using helmert-coded logistic mixed effects models and extracted pairwise comparisons using emmeans. The critical comparisons are as follows. We expect to find reliable differences in acceptance rates for the grammatical baseline and ungrammatical baseline conditions. We additionally expect to replicate Chapter 3 and Orth, Yoshida, & Sloggett 2021 and expect not to find a reliable contrast between embedded-*haven't* and the ungrammatical baseline

(i.e., a lack of illusions for embedded-*haven't*) The key questions are whether we find a reliable contrast between the embedded-*very-few* condition and the ungrammatical baseline (i.e., illusions for embedded-*very-few*) and whether we find a reliable contrast between the two forms of embedded licensor.

4.4.1.5 Results

The results from Experiment 10 are shown in Figure 4.4. An effect of grammaticality was observed ($\beta=4.13$, $SE=0.62$, $z=6.70$, $p<.001$), indicating that the grammatical baseline condition was significantly more likely to be judged acceptable than the ungrammatical baseline condition. An effect of embedded-*very-few* was observed ($\beta=0.69$, $SE=0.28$, $z=2.45$, $p=.01$), indicating a statistically reliable illusion for embedded *very-few* in SRCs. No such effect was observed for embedded-*haven't* ($\beta=-0.27$, $SE=0.29$, $z=-0.92$, $p=.36$). Furthermore, follow-up analyses using the embedded-*very-few* condition as a baseline revealed a significant contrast with embedded-*haven't* ($\beta=-0.96$, $SE=0.29$, $z=-3.31$, $p<.001$), indicating that illusion rates for embedded-*very-few* in SRCs are reliably higher than illusion rates for embedded-*haven't* in SRCs.

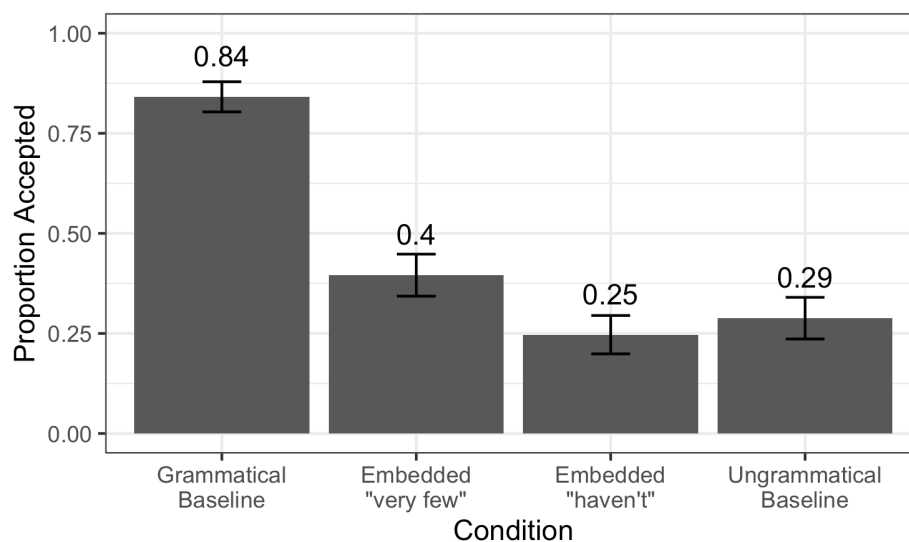


Figure 4.4: Mean percentage of 'YES' responses for the experimental conditions in Experiment 10. Error bars indicate standard error of the mean across subjects.

4.4.1.6 Discussion

In Experiment 10 we evaluated the possibility that the distance effects observed here and in Parker & Phillips 2016 are better understood as consequences of the structural distance between an embedded licensor and an NPI, as measured by the number of nodes traversed to reach one from the other in a syntactic tree. A structural-distance-based hypothesis would capture both distance effects like those reported here and licensor effects like those reported by Orth, Yoshida, & Sloggett 2021 and Chapter 3 under the same mechanism. In Experiment 10 we investigated the licensor effect in SRCs, such that quantificational licensors were lower in the RC structure (i.e., farther from the NPI by number of nodes) than non-quantificational licensors. Thus, a structural-distance-based hypothesis would predict that illusion rates for quantificational licensors should be no higher than illusion rates for non-quantificational licensors (which are typically estimated at or near zero) in SRCs. Instead, we again found robust illusions for embedded quantificational licensors and no illusions for embedded non-quantificational licensors. These findings suggest that structural distance is not the underlying factor driving the licensor effect. In brief, our results from the first four experiments collectively demonstrate that manipulations of the position of the NPI with respect to the edge of the RC significantly influence illusion rates, but we see no evidence that manipulations of the position of the NPI with respect to the negative word — either by adding material to the RC or by moving the negative word around within the RC — has the same influence. This pattern of findings strongly suggests that the representation that interferes with NPI processing is a representation that spans the entire RC, not merely the representation of the licensor. It does not, however, tell us why this representation interferes in the first place. We turn to this issue in Experiment 11.

4.5 The nature of timing effects and the role of NPI identity

The findings from Experiment 7, Experiment 8, Experiment 9, and Experiment 10 collectively suggest that the relevant distance for NPI illusion distance effects is the distance between the RC and the NPI. This pattern is predicted under an account in which NPI licensing is an operation which relates the NPI to a local context, and near-local contexts can sometimes interfere. It is not straightforwardly predicted

by hypotheses in which NPI licensing is an operation which relates the NPI to a negative c-commanding word. In fact, the fact that distance effects arise at all is somewhat surprising under item-based licensing accounts, since such accounts would need to allow for licensing by a negative word that could in principle be very far away. In contrast, context-based licensing is an inherently local relation — the NPI must be contained within the context that licenses it. An open question, however, is why near-containment by the licensing context appears to be sufficient for licensing in NPI illusions. In other words, it is not clear what happens (or fails to happen) in the moments between the end of the RC and the NPI (i.e. at *have* in (56)). One way to frame this effect, which we pursue here, is to assume that the representation of the RC lingers briefly. This could in principle arise for either of two reasons.

(56) * The surgeons [that no patients consulted] have ever ...

First, the representation could remain active for a few words after the end of the RC because the computations that would need to be executed in order for that representation to be suppressed have not had time to run yet. For example, consider all that would need to be done for a comprehender to know, prior to *ever*, that the meaning of the RC is no longer directly relevant. It is only at the MC auxiliary *have* that the comprehender can even know that the RC is over, since the sentence might have instead continued *The surgeons that no patients consulted about the operation scheduled for tomorrow....* So, only after reading *have* and executing all lexical access procedures to identify *have* in the mental lexicon can the comprehender even construct a syntactic representation in which the current position is in the MC, not the RC. Then, only as a consequence of this syntactic parse can the comprehender begin the work of shifting the semantic and pragmatic representations under construction to the MC. In terms of the actual comprehension goals, this is a substantial shift — the comprehender is no longer engaged in the work of identifying the subset of surgeons under discussion, but rather predicating something of those surgeons. This might be particularly hard since the subset of surgeons under discussion is not a particularly intuitive subset and comprehenders will not have particularly strong expectations about what will be predicated of those surgeons. But it is only once this shift has taken place that the representation of the meaning of the RC (and the aspects of meaning that are relevant for NPI licensing) will be made unavailable. In a task with uncontrolled presentation rates (i.e. rapid serial visual presentation (RSVP) or

listening to speech) there is no guarantee that this operations will be completed before *ever* is read. Even in natural reading, if processing is cascaded, downstream semantic operations could still be in progress when *ever* is encountered. Thus, one possible explanation for why the RC representation lingers is that the computations required to suppress this representation have not had time to complete.

An alternative explanation for why the RC representation lingers is that the MC content that is encountered between the RC edge and the NPI has failed to pull attention away from the RC representation. That is, the only intervening word is an auxiliary, *have*, which carries substantial syntactic consequences but very little information about the content of the message being expressed by the sentence. Thus, a comprehender may continue to attend to the RC representation for lack of new contentful material to attend to.

There is also a third option for why the RC interferes, along the lines of what Parker & Phillips 2016 propose regarding the distance effect for pre-verbal *ever* as in (56) versus post-verbal *any* as in (57). That is, some types of material (in particular, lexical verbs), may trigger a change in the encoding of the subject, which makes the RC representation less available. It is not entirely clear what this re-encoding process consists of. One possibility is that instead of maintaining a linguistic representation of the subject noun modified by an RC, the comprehender transitions into a representation of the referent (i.e., the surgeons, regardless of the linguistic expression used to identify them). With the linguistic content of the RC discarded, the processing of the NPI would presumably no longer be subject to interference from the RC. Such a hypothesis may make some interesting predictions about other subsequent operations that require access to the internal structure of the linguistic representation of the subject, but this is beyond the scope of the present study.

(57) * The surgeons [that no patients trusted] have healed any injuries ...

We focus on the first two possibilities here, which we label “timing-based” and “content-based” explanations for the lingering RC representation. These accounts make sense of the contrast between *ever* in (56), which is subject to illusions, and *any* in (57), which is not, in different ways. Under timing-based accounts, the two intervening words in (57) provide enough time for the relevant computations that suppress the RC representation to be executed. Under content-based accounts, the existence of an inter-

vening verb that pulls attention away from the RC representation. Thus by independently manipulating the number of intervening words and whether or not the intervening words include a verb, we can test these competing claims.

4.5.1 Experiment 11: speeded acceptability

In Experiment 11, we manipulated the number of intervening words and the presence of an intervening verb orthogonally by either removing the auxiliary in sentences like (58a), resulting in a condition in which only one word intervenes, and that word is a lexical verb, as in (58b), or adding a modal to sentences like (59b), resulting in a condition in which two words intervene, but neither is a lexical verb, as in (59a)

- (58) a. * The surgeons [that no patients trusted] have shown any appreciation for the hospital staff.
b. * The surgeons [that no patients trusted] showed any appreciation for the hospital staff.
- (59) a. * The surgeons [that no patients trusted] would have ever shown appreciation for the hospital staff.
b. * The surgeons [that no patients trusted] have ever shown appreciation for the hospital staff.

This experiment served an additional purpose of potentially revealing illusions for the NPI *any*. That is, under timing-based explanations for the persistence of the RC, the condition corresponding to (58b) should reveal illusions. Such an outcome would not only bear on the question of why the RC representation lingers, but also address concerns about whether NPI illusions are specific to the NPI *ever*, an issue that arises in our interpretation of the findings from Experiment 9. In brief, if NPI illusions are only possible for *ever*, and impossible for *any* regardless of position — that is, if NPI illusions would more accurately be labeled “*ever* illusions” — then the pattern observed in Experiment 9 would not bear on the question of the distance effect. Thus, finding an illusion for (58b) would rule out a potential confound in Experiment 9, making us more confident in our interpretation of those findings.

4.5.1.1 Participants

72 US-based native speakers of English participated in this experiment. All participants provided informed consent and they received \$9 as compensation. We excluded workers who failed to provide a response within 2 seconds for 25% of fillers or more and workers whose judgments of filler trials did not reliably distinguish between grammatical and ungrammatical fillers, based on a chi-squared test, as well as participants who answered two or more “attention check” trials incorrectly. 68 participants met these criteria. The mean filler-trial accuracy of the included participants was 82%.

4.5.1.2 Materials

Experimental materials consisted of 36 items in 12 conditions, as shown in Table 4.4. These conditions constitute a 2x2x3 design, crossing the number of intervening words (one or two), with the presence of an intervening verb (verb or no verb), with the standard illusion contrasts (grammatical baseline, embedded negation, and ungrammatical baseline). Participants additionally rated the same 90 filler sentences and 8 attention check sentences that were used in previous experiments.

4.5.1.3 Procedure

The procedure was identical to that of Experiment 7, Experiment 9, and Experiment 10. The task lasted approximately 30 minutes.

4.5.1.4 Analysis

As with Experiment 7, Experiment 9, and Experiment 10, data were analyzed using helmert-coded generalized linear mixed effects model with a logit link function, first fitting the maximal random effects structure and then simplifying as needed to achieve convergence. Critical comparisons were derived using emmeans. The key question was which of two possible two-way interactions would arise: the interaction between the number of intervening words (one versus two) and the illusion effect (embedded negation versus ungrammatical baseline) or the interaction between the presence of an intervening verb (verb versus no verb) and the illusion effect (embedded negation versus ungrammatical baseline). Prospective

One intervening word, no intervening verb	
A. Grammatical baseline	No surgeons [that the patients trusted] have ever shown appreciation ...
B. Embedded negation	The surgeons [that no patients trusted] have ever shown appreciation ...
C. Ungrammatical baseline	The surgeons [that the patients trusted] have ever shown appreciation ...
...for the hospital staff.	
One intervening word, including intervening verb	
A. Grammatical baseline	No surgeons [that the patients trusted] showed any appreciation ...
B. Embedded negation	The surgeons [that no patients trusted] showed any appreciation ...
C. Ungrammatical baseline	The surgeons [that the patients trusted] showed any appreciation ...
...for the hospital staff.	
Two intervening words, no intervening verb	
D. Grammatical baseline	No surgeons [that the patients trusted] would have ever shown appreciation ...
E. Embedded negation	The surgeons [that no patients trusted] would have ever shown appreciation ...
F. Ungrammatical baseline	The surgeons [that the patients trusted] would have ever shown appreciation ...
...for the hospital staff.	
Two intervening words, including intervening verb	
G. Grammatical baseline	No surgeons [that the patients trusted] have shown any appreciation ...
H. Embedded negation	The surgeons [that no patients trusted] have shown any appreciation ...
I. Ungrammatical baseline	The surgeons [that the patients trusted] have shown any appreciation ...
...for the hospital staff.	

Table 4.4: Example stimuli for Experiment 11

power analyses using *simr* indicated that with 72 participants we would achieve between 88% and 92% power to detect either of these effects, assuming an effect size similar to what was observed in Parker & Phillips 2016 Experiment 2 (specifically we used the lower bound of a 60% confidence interval around the observed interaction effect size, as suggested by Perugini, Gallucci, & Costantini 2014). Note that although the experiment design makes it possible to test the three-way interaction (number of intervening words \times presence of an intervening verb \times illusion effect), we expected to find that *either* the number of intervening words *or* the presence of an intervening verb would matter (and not both), and so we did not intend to test this three-way interaction and did not conduct power analyses for it.

4.5.1.5 Results

The results from Experiment 11 are shown in Figure 4.5. A main effect of grammaticality was observed ($\beta=3.86$, $SE=0.25$, $z=15.25$, $p<.001$), indicating that the grammatical baseline condition was significantly more likely to be judged acceptable than the ungrammatical baseline condition, averaging across the four

distance configurations. A main effect of embedded negation was observed ($\beta=1.09$, $SE=0.18$, $z=6.07$, $p<.001$), indicating a statistically reliable illusion, averaging across the four distance configurations. However neither the interaction between the illusion effect and the number of intervening words ($\beta=0.33$, $SE=0.30$, $z=1.10$, $p=0.27$), nor the interaction between the illusion effect and the presence or absence of an intervening verb ($\beta=0.44$, $SE=.30$, $z=1.47$, $p=0.14$) was significant. Pairwise comparisons revealed statistically significant illusions in three of the four distance configurations: one word, no verb ($\beta=1.30$, $SE=0.32$, $z=4.07$, $p<.001$); two words, no verb ($\beta=1.32$, $SE=0.31$, $z=4.22$, $p<.001$); one word including a verb ($\beta=1.21$, $SE=0.31$, $z=3.89$, $p<.001$); but not two words including a verb ($\beta=0.54$, $SE=0.31$, $z=1.71$, $p=0.09$). We did test the three-way interaction, although, as noted in section 4.5.1.4, the experiment was not specifically designed to have high power to detect this interaction. It was not significant ($\beta=0.70$, $SE=0.59$, $z=1.18$, $p=0.24$).

4.5.1.6 Discussion

In Experiment 11 we aimed to determine whether the RC interferes with the NPI when not enough time has passed since the RC edge or when not enough meaningful content has been encountered since the RC edge. We addressed this question by independently manipulating the number of intervening words between the RC edge and the NPI (one versus two) and the presence or absence of an intervening verb between the RC edge and the NPI. Note that two of the four distance configurations generated by crossing these factors have been previously tested: it is well established that illusions *do* arise when only one word intervenes and that word is not a main verb, and it is well established that illusions *do not* arise when two words intervene and one of them is a main verb. Thus we expected to replicate these patterns and determine whether the two novel distance configurations suggested that the critical element for making the illusion go away is the addition of a second word or the inclusion of a verb.

We found reliable illusions for all but one of the four configurations generated by crossing these factors: the two-word, including a verb condition, which had been previously shown to not yield illusions. The three other distance configurations all revealed illusions of the typical effect size ($\beta=1.30$, $\beta=1.32$, $\beta=1.21$). Thus it does not appear to be the case that the illusion goes away whenever one adds a second

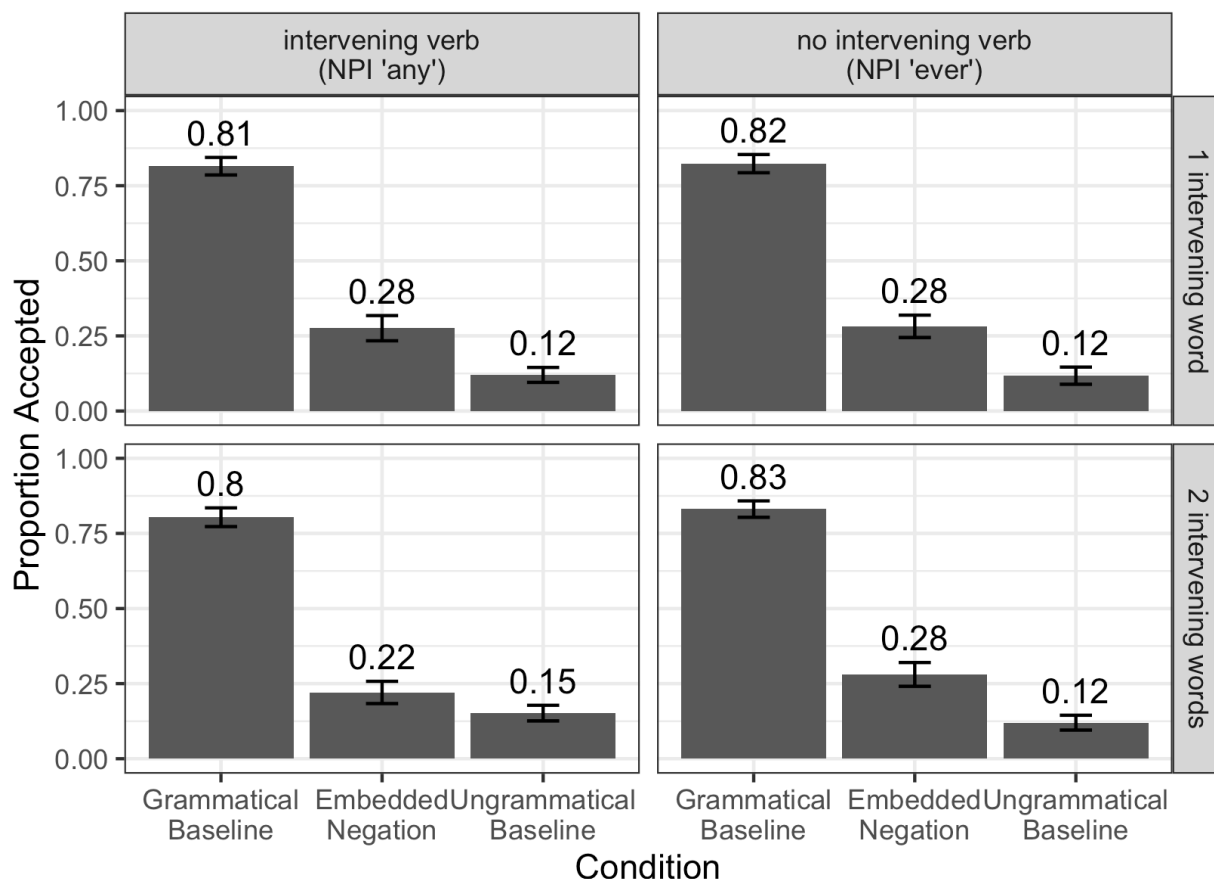


Figure 4.5: Mean percentage of 'YES' responses for the experimental conditions in Experiment 11. Error bars indicate standard error of the mean across subjects.

word, nor does it seem to be the case that the illusion goes away whenever one introduces an intervening verb — rather, it is only the *combination* of these factors that makes the illusion go away.

It is important to note that the three-way interaction that would definitively show that it is only when these factors are both present that the illusion is diminished was not statistically significant. The findings here are therefore also compatible with a generalization in which both the addition of a word and the inclusion of a verb have small, additive effects on the illusion rate, such that we were unable to detect either effect when it was the *only* thing pushing the illusion rate down, but when they *both* were present, illusions become rare enough as to not be statistically detectable. Given the uncertainty in these findings, it is difficult to draw strong mechanistic conclusions about why it is that RC representations sometimes linger.

Framing this finding in terms of the timing-based and content-based explanations for the persistence of the RC representation, it appears that neither explanation is fully correct. If the timing-based explanation were correct, we should not have found illusions for the two-intervening-words-no-verb condition. If the content-based explanation were correct, we should not have found illusions for the one-intervening-word-including-verb condition. It could be that both factors matter (i.e. the probability that the RC has been suppressed increases with time and with the addition of meaningful content), or it could be that the temporal dissociation of the significant parsing operations that occur at the RC edge and the significant meaningful update that occurs at the verb allow these effects to combine super-additively, shutting down the RC representation in a way that is difficult to achieve when only one of those operations has occurred (as in the no-verb conditions), or when they must both occur simultaneously (as in the one-intervening-word-including-verb condition). It is also of course possible that neither conceptualization of the cause of the lingering RC representation (nor their combination) is correct.

A second important takeaway from Experiment 11 is that illusions are possible with the NPI *any*. This is a critical finding, since many explorations of the distance effect in NPI illusions (including Experiment 9) rely on the use of both *ever* and *any* in order to finely manipulate the position of the NPI, in a way that neither NPI on its own allows. If it turned out that NPI illusions in fact never arise with *any*, this would undermine the conclusions drawn from such investigations. Thus the discovery that illusions can occur with *any* as long as a short RC-NPI distance is achieved is reassuring.

4.6 General discussion

The present study aimed to illuminate the nature of the online computation of NPI licensing through the lens of the distance effect for NPI illusions. We reason that if licensing consists of computing a relation between an NPI and the properties of the context that contain it, distance effects should arise when the distance between the NPI-licensing context and the NPI is manipulated. If instead licensing consists of computing a relationship between an NPI and a negative word or phrase that c-commands or takes scope over it, distance effects (if they arise at all) should arise when the distance between the licensing word or

phrase and the NPI is manipulated. Thus, we aimed to clarify the profile of the distance effect for the NPI illusion with respect to three key dimensions of uncertainty. First, we asked whether the relevant landmark from which distance is computed is the negative word itself or the negated context (i.e., the RC). Second, we asked whether the distance effect can be reformulated as an effect of the number of intervening syntactic nodes between the NPI and the relevant landmark. And finally, we asked whether the distance effect reflects changes that occur with the mere passage of time or if instead a verb or some other content word is critical. We note that there remains an open question of whether the distance effect is truly an effect of monotonically diminishing vulnerability (either gradually or suddenly) to illusions or if instead illusion vulnerability fluctuates both up and down throughout the region after the relevant landmark. We have thus far not presented evidence that bears on this third question, and have simply assumed, following Parker & Phillips 2016 that vulnerability to illusions only decreases. We present evidence in Chapter 5 that calls this assumption into question, but this is beyond the scope of the present investigation.

4.6.1 Key findings

There are three significant empirical contributions made by the present study. The first concerns the relevant landmark for the distance effect for NPI illusions — that is, whether increased distance from the negative word to the NPI or increased distance from the RC edge to the NPI matters for illusion rates. For the purposes of addressing this question we adopt the assumption that the nature of the distance effect is that vulnerability monotonically decreases with time following the relevant landmark. In Experiment 7 we addressed this question by inserting PPs within the RC, after the negative word. If the relevant landmark is the negative word itself, this manipulation yields an increase of three words in the critical distance and so we expect some decrease in the illusion rate. If instead the relevant landmark is the RC edge, the manipulation yields no change in the critical distance, and therefore should have no impact on illusion rates. We found no evidence of an influence of PPs on illusion rates. This finding is consistent with the generalization that the RC edge is the relevant landmark, but we are cautious in drawing conclusions from this experiment alone, as the finding is a null effect of a two-way statistical interaction.

Experiment 8 and Experiment 9 investigated this question further by adding the same material in-

side the RC for some conditions and outside the RC for other conditions, within the same experiment. Importantly, the content of the added material was largely overlapping for the two added distances. We again found that added material within the RC, which increases the distance to the negative word but leaves the distance to the RC edge unchanged, had no detectable impact on illusion rates. In contrast, we replicate Parker & Phillips’s (2016) finding that added material outside the RC, which increases both the distance to the negative word and the distance to the RC edge, had the effect of “turning off” the illusion. We take these findings to suggest that the RC edge is in fact the relevant boundary for the distance effect.

Experiment 10 addressed the possibility that it is structural distance, not mere time passing that leads to a change in illusion vulnerability. By re-framing the impact of added material as a manipulation of the number of intervening nodes, an account where the relevant landmark is the negative word becomes plausible once again. That is, the evidence from Experiment 7, Experiment 8, and Experiment 9 can be re-interpreted as showing that added material that increases the number of intervening nodes between the licenser and the NPI can turn off the illusion, but added material that increases the time that passes between the presentation of these items without changing the number of intervening nodes in the syntactic structure has no impact on illusion rates. Such a hypothesis would also account for previously reported licenser effects, since the licensors that are compared occupy different syntactic positions. Experiment 10 showed that the licenser effect still arises (i.e. illusions for negative quantifiers, no illusions for verbal negation), even in SRCs where the number of intervening nodes between licenser and NPI is greater for the quantifier condition than the verbal negation condition. Thus, a re-framing of the distance effect in terms of structural distance does not successfully account for the data. We note however that the question of the relevant landmark (the negative word or the RC edge) and the question of the type of distance that matters (intervening syntactic nodes or the passage of time) are orthogonal and so there are in principle four possible combinations of these generalizations. Experiment 7, Experiment 8, and Experiment 9 ruled out the combination in which it is the negative word that matters and distance is the passage of time, whereas Experiment 10 ruled out the combination in which it is the negative word that matters and distance is structural. But we cannot at present say whether a structural or temporal version of the RC-distance effect is appropriate.

Finally, Experiment 11 tested whether the possibility of interference from the RC is particularly sensitive to the presence of an intervening main verb. This question has consequences for how we conceive of the cause of the RC's interference with the processing of the NPI — that is, does this representation linger because the comprehender has not yet executed the computations that allow the representation to be de-activated, or does it linger because no new content has been encountered to pull attention away? We addressed this by crossing the number of intervening words between the RC and the NPI (one versus two) with the presence or absence of an intervening main verb. We found that neither of these factors on its own is enough to turn off the illusion completely (and in fact we see no clear evidence that either of them has any effect on their own), but when they combine, illusions disappear. This experiment also addressed a potential confound in Experiment 8 and Experiment 9: the type of NPI used. In the manipulations where the illusion was turned off, we not only increased the distance but also used the NPI *any* instead of *ever*. This was necessary in order to precisely manipulate the distance, since there are no NPIs that can be freely placed in any position in the sentence. The theoretical importance of these findings make it valuable to determine whether illusions are categorically impossible with the NPI *any*. Experiment 11 revealed that illusions are in fact possible for the NPI *any* and so this confound cannot be responsible for the findings in Experiment 8 and Experiment 9.

4.6.2 Implications for illusions

Having established some key generalizations regarding the nature of the distance effect on NPI illusions, we turn to some existing hypotheses about the mechanisms underlying the illusion, and, more generally, the mechanisms underlying real-time NPI licensing. Before exploring the specific hypotheses that have been discussed, we note that the generalizations reported here point us toward expecting an explanation of a certain shape. That is, given our finding that NPI illusions are sensitive to the distance between the RC and the NPI, we believe the language in which we typically describe the NPI illusion deserves an update. Literature on NPI illusions previously described the effect as a pattern in which a negative word interferes with the processing of an NPI which cannot be licensed by it. It now seems more appropriate to describe the phenomenon as one in which a negated clause interferes with the processing of an NPI

which is not contained within it. Thus, we expect that the space of plausible explanations for the illusion should move toward explanations in which NPI licensing is an operation in which the NPI is related to a context, and the critical error is the use of the RC context in this operation.

This is of course not a strictly necessary property of a plausible hypothesis for the NPI illusion, since it remains possible that the dependency does in fact relate the NPI to the negative word and the distance effect is explained by some independent factor. We can imagine two possible revisions to a mechanism that is committed to licensing by negative words which would account for the observed distance effects. The first is to stipulate that the activation of a negative word does not decay as long as the incoming content is still within the scope of negation. Such a stipulation would guarantee that any distance effects that arise are sensitive to the edge of the RC (i.e., where the negative word's scope ends), not the position of the negative word itself. Thus, an NPI that is encountered immediately after the RC edge may access the still-very-active negative word, whereas an NPI encountered a few words later would not be able to access that same negative word, since the decay process has begun. While we cannot definitively rule out such an account, we note that there is a lack of corroborating evidence in independent literature on the processing of scope-taking elements. For example, Kush, Lidz, & Phillips 2015 demonstrated that a pronoun just one word away from the edge of the scope domain of a quantifier cannot be bound by that quantifier, and these effects are found even in early eye-tracking measures (suggesting there is no illusion of binding). Thus comprehenders seem to be quite good at quickly shutting down the representation of a scope-taking element when its scope ends. Of course, one could argue that the binding of pronouns relies on different operations than the licensing of NPIs. In fact, we believe this to be true. But it is rather surprising to claim that NPI licensing is the operation that relies on the retrieval of a prior word in memory whereas pronoun binding relies on some other operation. Thus, while it is in principle possible that the activation of a negative word begins to decline only when its scope ends, we do not find this account especially compelling. The other option if one wishes to maintain a licensing-by-a-negative-word account is roughly what Parker & Phillips 2016 propose — namely that some critical change in the encoding of the subject noun phrase occurs at some point after the RC, rendering the internal components of that noun phrase inaccessible. However, in light of our findings from Experiment 11, we suspect such a point will

prove difficult to define. It is clear that encountering a main verb, which Parker & Phillips 2016 suspected may be the key point, is neither necessary nor sufficient for the illusion to disappear. Moreover, findings from Chapter 5 suggest that the illusion might “turn back on” for NPIs much later in the sentence — a pattern which is wholly inconsistent with the re-encoding explanation.

Having considered these possibilities, we believe that the NPI illusion is a case in which negated *clause* interferes with the processing of an NPI which is not contained within it. We now turn to the existing hypotheses as to the cause of the illusion. As discussed in section 4.1.2.1, two key proposals in the literature on NPI illusions have been the cue-based retrieval hypothesis (Vasishth et al. 2008) and the pragmatic rescuing hypothesis (Xiang, Dillon, & Phillips 2009). We first discuss how these hypotheses fare with respect to the evidence presented here.

The cue-based retrieval hypothesis is committed to a mechanism in which NPI licensing is operationalized as a memory retrieval of a prior chunk based on a set of cues, potentially [+c-command] and [+negation]. As stated, this hypothesis predicts that illusions will occur for any configuration in which a set of elements constructed prior to the NPI have these properties, and so it predicts uniformity of the illusion throughout all sentence regions following the negative word.³¹ Thus, this hypothesis requires stipulating something like one of the two revisions described above — either the activation of the negative word does not decline within its scope, or the representation of the subject is re-encoded at some later point. While we cannot rule out (these amended versions of) the cue-based retrieval explanation, we think that especially in light of the licenser effect reported by de Dios Flores, Muller, & Phillips 2017, Orth, Yoshida, & Sloggett 2020a, and Chapter 3, the hypothesis is on the whole an unsatisfactory explanation for the NPI illusion. What at first appeared to be a key advantage of the cue-based retrieval framework — its generality, which enables it to account for a number of distinct phenomena under the same mechanism — has become a liability in light of the accumulating evidence that NPI illusions are in fact not very general, but instead arise only under very specific circumstances.

³¹Technically, the hypothesis is not committed to *uniform* illusion rates for NPIs anywhere in the sentence. There is a decay parameter in the Vasishth et al. 2008 model, which could be fine-tuned such that a version of the distance effect arises. However, this distance effect would be relative to the negative word itself, and we have shown that this is the wrong characterization of the actual distance effect. Turning up the decay parameter also risks predicting severely degraded acceptability for sentences in which an NPI is truly licensed but the negative word is quite distant.

The pragmatic rescuing hypothesis attributes illusions to contrastive implicatures generated as a result of use of restrictive RCs. While this explanation does not obviously predict any impact of distance on illusion rates, we might expect that if distance were to matter it would be distance from the RC, not distance from the licensor. This is because under this hypothesis the RC plays a critical role in triggering the illusion. Thus, one might modify this hypothesis such that the contrastive implicature triggered by a particular item or structure becomes less accessible following the presentation of that item or structure. We know of no evidence suggesting this (or ruling it out), making this a possible avenue for future research. However, we do not expect this to be the most productive path to pursue, given the concerns with this hypothesis discussed in section 4.1.2.1 and in Chapter 3. That is, while it is certainly advantageous with respect to the distance effect that this hypothesis makes a key player out of the RC, it in fact centers the RC to such an extent that illusions are predicted even when the RC is not negated, contrary to prior findings. We therefore do not discuss this possibility further.

In light of the licensor effect summarized in section 4.1.2.2, Orth, Yoshida, & Sloggett 2021 and Chapter 3 propose two alternative accounts of the NPI illusion, to which we now turn our attention. Recall that Orth, Yoshida, & Sloggett attribute the NPI illusion to an error in the computation of the scope of a negative quantifier. After this error has occurred, the NPI can be licensed by the wide-scoping quantifier under whatever mechanism one prefers for the online licensing of NPIs (c-command by a negative word, scope by a DE-operator, containment in a DE clause, etc.). It is not obvious why, under this account, there should be a distance effect at all. One possibility is that the initial error in the assignment of quantifier scope is re-evaluated as the sentence continues, and so the later in the sentence the NPI arrives the more likely it is that this initial error has been corrected, resulting in a robust representation that is no longer vulnerable to illusions. We consider this explanation plausible, but note that it is more straightforwardly compatible with a distance effect that treats the negative word itself as the relevant landmark. If the problem in scope assignment occurred at the negative quantifier itself, it should be possible to correct this error at any point after the negative quantifier. Instead it seems that the representation is unchanged throughout the RC but begins to change rapidly at the end of the RC. This is unexpected and requires stipulation that re-analysis of quantifier scope assignment begins only at the MC.

Lastly, we consider the possibility that scalar alternatives to the RC, which are evoked on the basis of the form of negation used inside the RC (as well as other factors such as the broader context) interfere with the processing of the MC NPI. This hypothesis builds on grammatical approaches to NPIs which treat the environment and its scalar alternatives as the locus of licensing. Thus we would expect that the online resolution of this dependency would involve relating the NPI to a context, not an individual lexical licenser. In that sense, the present findings are consistent with the hypothesis — the NPI must be contained within a certain kind of representation, and so it is only when the NPI is *nearly* contained by the appropriate representation that the illusion arises.

It remains unclear, however, why this representation persists at all. Note that this question turns the question of the distance effect on its head — it is not mysterious that the RC is irrelevant to late-arriving NPIs, but that it is still so relevant to early-arriving ones. One possibility is that the inference that the RC alternatives are no longer relevant is dependent upon a number of steps which may simply not be fully executed at the point when the NPI is encountered. First the comprehender must identify the word that follows the RC (typically *have*); then they must infer the syntactic structure that this word requires and must construct the appropriate syntactic parse (that is, they must detect that the RC has ended); then finally as a result of the syntactic closure of the RC the comprehender can begin to suppress the meaningful representation of the RC, including its pragmatic alternatives. If these steps proceed in a cascaded manner beginning at the point when *have* is fixated, it is possible that they have not terminated at the point the NPI is encountered, and so the alternatives to the RC are still active. An alternative explanation for the brief persistence of the RC alternatives is that a single auxiliary simply does not carry enough interpretive consequences to pull attention away from the RC representation (whereas a main verb might). Experiment 11 explored these possibilities and found that neither factor on its own caused the illusion to go away. Thus we consider this an important area for future research.

4.7 Conclusion

Across six experiments using untimed acceptability and speeded acceptability measures we explored the distance effect for NPI illusions, first reported by Parker & Phillips 2016. We find that the distance effect is better understood as an effect of added material between the RC edge and the NPI, rather than an effect of added material between the negative word and the NPI. Accordingly, it is appropriate to describe the illusion itself as a phenomenon in which a negated clause representation interferes with the processing of the NPI, whereas previously it was described as a phenomenon in which a negative word interferes with the processing of the NPI. One important implication of this work is that it suggests that the online licensing of NPIs (even in non-illusory contexts) involves forming a dependency between the NPI and the licensing context, rather than a dependency between the NPI and the negative word.

Chapter 5 NPI illusions: putting the pieces together

5.1 Introduction

The findings presented in Chapter 3 and Chapter 4 make a strong case for re-considering the nature of real-time NPI licensing, moving away from models that treat licensing as the retrieval of a negative word in memory. Moreover, we have argued that the findings suggest that a scalar alternatives based approach better accounts for the illusion. Here we briefly review the evidence presented so far, as well as some alternative accounts of the illusion. We then turn our attention to five experiments which somewhat complicate the picture.

5.1.1 Key findings

Chapter 3 and Chapter 4 collectively present 11 experiments, demonstrating two key generalizations about the selectivity of the NPI illusion: the licensor effect and the distance effect. Chapter 3 additionally presents the first investigation, to our knowledge, of the interpretation comprehenders arrive at following an illusion. Here we additionally highlight one other piece of evidence that bears on the NPI illusion: a possible illusion of *ungrammaticality* for PPIs in the same position that hosts an NPI in typical illusion experiments, which was demonstrated by Orth, Yoshida, & Sloggett 2020a. Although this last finding has not yet been replicated, we think it has important implications for the possible explanations for NPI illusions.

5.1.1.1 The licensor effect

The licensor effect, explored primarily in Chapter 3, is the finding that while NPI illusions regularly arise when the form of embedded negation is a negative quantifier (e.g., *no critics*), they do not regularly arise

when the form of embedded negation is simple sentential negation (e.g., *the critics didn't*). Importantly, this contrast arises regardless of whether sentential negation is contracted (*didn't*) or not contracted (*did not*) (see Experiment 2 and Experiment 3), regardless of whether quantificational negation is strictly negative (*no*) or some other downward-entailing quantifier (*very few*) (see Experiment 10), and regardless of whether the clause housing the embedded negation is a SRC or ORC (see Experiment 6 and Experiment 10). Precisely which difference between quantificational and non-quantificational licensors drives this effect is not obvious, but we have argued that differences with respect to scalar alternatives are a promising candidate. Consistent with this, we find that illusions may arise for sentential negation when the appropriate alternatives are triggered prior to the main-clause NPI (Experiment 6), though these data are not especially clear.

5.1.1.2 The distance effect

The licensor effect, explored primarily in Chapter 3, is the finding that the NPI must be positioned close to the edge of the RC (within one to three words) in order for the illusion to occur. This effect was first demonstrated by Parker & Phillips 2016, though it was not clear from their findings whether the NPI had to be positioned close to the RC or close to the negative word. Our findings from Experiment 7 and Experiment 9 demonstrate clearly that it is the RC that matters. Moreover, the fact that the licensor effect discussed above is not sensitive to the distinction between ORCs (which place the negative quantifier high in the clause) and SRCs (which place the negative quantifier lower) shows that the distance effect cannot be recast as an effect of the number of intervening nodes between the negative word and the NPI. If the distance had been about the distance to the negative word, we might have asked why a distance effect exists at all — that is, why would the interfering negative word become *less* able to interfere? But given that the distance effect is in fact about the distance to the RC, this question takes a different form — why does the representation of the negative clause persist as long as it does? In other words, since NPIs must be *contained* by the clauses that license them, it is reasonable that an NPI that is very far from the clause would not be mistaken for being contained by it. What is more surprising is that an NPI that is very near to the clause seems to be mistaken for being contained by it despite unambiguous evidence from the

syntactic parse indicating that this is not the case. In Experiment 11 we explored possible reasons for why the RC representation may linger, though the evidence from this experiment was not entirely conclusive.

5.1.1.3 Interpretation

In Experiment 4 we investigated the interpretation of NPI illusion sentences using a task in which comprehenders answered both an acceptability judgment and sentence-final comprehension question. This experiment revealed that the sentence structures that are typically used in NPI illusion research (i.e., an embedded negative quantifier and a MC NPI) are overwhelmingly interpreted as expressing a negative proposition — that is, participants’ answers suggest they believed that the MC was in the scope of negation. Importantly, these interpretations do not arise as frequently for sentences with just an embedded negative quantifier (and no NPI), suggesting that they are not interpretive errors driven by problems in quantifier scope processing, but rather interpretive errors driven by problems in NPI processing. Moreover, illusions occur at similar rates for negatively-interpreted and positively-interpreted trials.

5.1.1.4 PPI illusions

Lastly, we note findings from Orth, Yoshida, & Sloggett 2020a showing illusions of ungrammaticality for PPIs in the same position that we place NPIs for NPI illusions. For example, they measure acceptability for sentences like in (60). The PPI *still* is “anti-licensed” in (60a) since it is within the scope of negation. In (60b) the negative quantifier is inside the RC and therefore has no impact on the grammatical status of the main-clause PPI, yet Orth, Yoshida, & Sloggett 2020a report degraded acceptability for this condition, relative to the grammatical baseline in (60c).

- (60) a. *No hunter [that the fisherman trusted with a secret] will still shoot a bear with a bow.
 b. The hunter [that no fisherman trusted with a secret] will still shoot a bear with a bow.
 c. The hunter [that the fisherman trusted with a secret] will still shoot a bear with a bow.
(Orth, Yoshida, & Sloggett 2020a:1)

Note that this experiment used an untimed Likert acceptability rating task, which is not typical for illusion research. Reading time data revealed trends toward illusions, but no significant effects. Thus we interpret these findings with some caution. If the PPI illusion is robust, however, this would have serious implications for how we think about the NPI illusion. Specifically, any hypothesis which treats the illusion as a consequence of error-driven processes or processes that specifically seek to license the NPI would not be able to account for PPI illusions.

5.1.2 Theoretical landscape

We now turn our attention to the handful of explanations for the NPI illusion which have been suggested in the literature, paying specific attention to these hypotheses' ability to account for the four patterns of findings discussed above.

5.1.2.1 Cue-based retrieval

The cue-based retrieval explanation for NPI illusions, proposed by Vasishth et al. 2008, treats the illusion as a consequence of properties of the noisy memory architecture with which the grammar must interact. Specifically, NPI licensing is conceptualized as a search through memory for an item with [+negation] and [+c-command], and the illusion arises when an item that partially matches these features is selected, due to the partial match plus noise. The hypothesis does not predict the licenser effect and in order to accommodate these findings would have to stipulate the existence of retrieval cues that distinguish negative quantifiers from other forms of negation. This is undesirable because in cases of true c-command, all forms of negation are equally good licensors. The hypothesis also does not predict the distance effect, and in order to accommodate these findings would have to stipulate either a substantial change in the encoding of the subject at some point after the RC, or a mechanism through which the activation of the negative word is maintained at a constant level until the end of the RC, at which point it rapidly decays. These possibilities are discussed in greater detail in section 4.6.2. The cue-based retrieval mechanism does not make any particular prediction about sentence-final interpretations, and so the evidence from Experiment 4 does not weigh in favor of or against this hypothesis. And finally, the hypothesis does not clearly

predict the existence of PPI illusions, since PPIs should not trigger the same search for a c-commanding negative word that NPIs trigger. It is worth noting that it is not clear how such a framework would begin to account for even normal PPI licensing, since PPIs do not require that something exist in the prior representation in some particular position, but rather require that a negative word *not* exist in *any* c-commanding position. How cue-based retrieval operations could be used to evaluate such a constraint remains to be seen. In sum, the cue-based retrieval framework is unable to account for the specific profile of the NPI illusion, and in particular fails to account for the two most empirically robust findings: the licenser effect and the distance effect. We therefore do not think that this is the right explanation. Importantly, this does not mean that the memory architecture is not as described, nor that memory operations are irrelevant to NPI licensing. This is further clarified in Chapter 6.

5.1.2.2 Pragmatic rescuing

The pragmatic rescuing hypothesis, proposed by Xiang, Dillon, & Phillips 2009 and Xiang, Grove, & Giannakidou 2013, attributes the NPI illusion to over-zealousness of a secondary NPI-licensing mechanism sometimes called “rescuing”. The basic idea of rescuing is that some NPIs are made acceptable not by a direct licensing relation to a negative word in the sentence, but in virtue of a negative inference that the sentence makes available. This mechanism is intended to explain the acceptability of NPIs in the scope of words that are not explicitly negative, like *doubt*. NPI illusion sentences also make negative inferences available, due to their use of restrictive RCs. Thus the rescuing operation, when applied to NPI illusion sentences, would yield increased acceptability.³² However, as we discussed in section 3.5.4, because it is the use of restrictive RCs that drives the illusion, and not the embedded negative word, the hypothesis risks predicting illusions even for ungrammatical baseline sentences. For the same reason, the hypothesis does not predict the licenser effect — the hypothesis does not straightforwardly distinguish between the illusion potential of RCs based on their content. As for the distance effect, it is not clear, under this hypothesis, why the position of the NPI would matter at all, but one way to accommodate this finding

³²It’s not clear why the rescuing operation wouldn’t yield *full* acceptability of illusion sentences, but the authors are explicit in stating that they don’t intend for this grammatical mechanism to actually apply in these cases. We explore this issue further in Chapter 6.

the sense that the hypothesis clearly treats the whole RC as the interfering representation, and so NPIs that are closer to this representation should be more vulnerable to interference.³³ The hypothesis also predicts that PPI illusions should arise — if the negative RC representation lingers, it should be able to interfere with the processing of any negation-sensitive element that occurs soon after. Thus, the scalar alternatives hypothesis accounts for some of the key generalizations regarding the profile of the illusion quite well.

One problem for this hypothesis, however, is the sentence-final interpretation that NPI illusion sentences receive. If the illusion is explained by a processing problem whereby the RC representation lingers long enough that the NPI in the MC is interpreted with respect to that representation instead of the MC representation, there should be no impact on the MC interpretation. But Experiment 4 clearly demonstrated that the MC is interpreted inaccurately in illusion sentences. One possible way to make sense of this failed prediction is to consider the processing that occurs between the moment when the NPI is encountered and the moment when an acceptability (or interpretation) judgment is given. Perhaps the comprehender (consciously or subconsciously) detects a discrepancy between, on one hand, the syntactic representation of the sentence, which places the NPI in the MC, and, on the other, the initial impression of the NPI as an acceptable form. One way to reconcile these representations is to infer that the MC was in fact negative. Relatively little is known about the processes that take place between the NPI and the sentence-final judgment, making it difficult to determine whether such an explanation is viable.

5.1.2.4 Scope miscalculation

The key motivation underlying the scope miscalculation hypothesis, proposed by de Dios Flores, Muller, & Phillips 2017 and Orth, Yoshida, & Sloggett 2021, is the licensor effect itself — that is, since NPI illusions appear to be largely specific to negative quantifiers, one might suspect that the problem is not with the NPI but with the quantifier. Because quantifiers in other contexts are known to exhibit more flexibility in their scope-taking possibilities than other scope-taking elements, including sentential negation, this approach has some appeal. The basic idea is that if the scope of the quantifier is misanalyzed such

³³However, it is not currently obvious why the RC representation lingers at all.

that the comprehender’s internal representation has the quantifier scoping over the NPI, then the processing of the NPI — under any account of the online licensing of NPIs — will result in the impression of acceptability. Importantly, however, the negative quantifiers used in NPI illusion sentences do not have the possibility of taking scope outside of the RCs that contain them. Thus, the scope miscalculation hypothesis allows us to say that the parser is entirely faithful to the grammar of NPI licensing, but at the cost of saying that it is not faithful to the grammar of quantifier scope. It also, for some reason, *is* faithful to the grammar of the scope of other elements, like sentential negation. This of course doesn’t make the hypothesis wrong — it may well be that the parser deviates from the grammar in precisely this way. But more work is clearly needed to understand why this would be the case.

Setting this concern aside, it is clear that the hypothesis can account for the licenser effect, since this is the finding that motivates the proposal. The finding from Experiment 6 that NPI illusions may be possible with non-quantificational licensors is unexpected under this account, but recall that the data pattern in this experiment was not entirely clear. The scope miscalculation hypothesis also does not predict a distance effect, unless scope is, for some reason, re-calculated at some later point in the sentence. Perhaps most importantly, hypotheses in this group predict a strong correspondence between globally negative interpretations and acceptance of illusion sentence. We found in Experiment 4 that this correspondence does not exist. Furthermore, some but not all versions of the hypothesis predict PPI illusions; but, for the most part, those that can make sense of the PPI illusion also make incorrect predictions about the interpretation of sentences with embedded negative quantifiers but no NPI. We further explore the possible versions of the hypothesis in Chapter 6, but for the present purposes we may simply note that the hypothesis fails to predict some of the key generalizations about NPI illusions.

5.1.2.5 Noisy channel

Lastly, we consider variants of the noisy channel hypothesis as an explanation for the NPI illusion. While the noisy channel hypothesis (Levy 2008) often comes up in discussions of NPI illusions we know of no written work explicitly arguing for such an account. The basic idea underlying the noisy channel framework is that a perceived sentence is a noisy signal of the speaker’s message — that is, there may be

distortions to the signal, both on the producers' and comprehenders' part. Because of this possibility, the comprehender considers not only the input as it is perceived but also similar strings that may have been the intended signal but were mis-perceived due to noise.

Applying such a mechanism to NPI illusions, there are two ways the comprehender might arrive at a representation that appears acceptable through minor edits to the input: either the NPI is not what it seems to be or the negative word is not where it seems to be. The first version was briefly considered (and rejected) in Chapter 3. To review, one might suspect that the comprehender considers the possibility that the NPI *ever* was in fact the orthographically- and phonologically-similar word *never*. Since *never* does not have any licensing restrictions, the edit from *ever* to *never* (i.e. the edit from (62a) to (62b)) should result in improved acceptability. However, since the availability of this edit is in no way related to the presence of a negative word in the RC, it should be just as possible for ungrammatical baseline sentences like (63a) to be edited into *never*-versions like (63b). Thus the hypothesis risks predicting equal acceptability for illusion sentences and ungrammatical baseline sentences. Since we define the illusion as the contrast between these conditions, this hypothesis is not promising. Even more worryingly, de Dios Flores 2019 demonstrated that sentences like (62b) are actually *less* acceptable than sentences like (63b). It appears that multiple negations within the same sentence incur some processing cost, even if they do not scope over the same material. Thus, if the noisy channel hypothesis, applied to NPI illusions, is a claim that comprehenders sometimes mis-perceive *ever* as *never*, the hypothesis actually predicts an *anti*-illusion — a boost for the ungrammatical baseline above the illusion condition. This is obviously undesirable.

- (62) a. The authors [that no critics have recommended] have **ever** received acknowledgment for a best-selling novel.
- b. The authors [that no critics have recommended] have **never** received acknowledgment for a best-selling novel.
- (63) a. The authors [that the critics have recommended] have **ever** received acknowledgment for a best-selling novel.

- b. The authors [that the critics have recommended] have **never** received acknowledgment for a best-selling novel.

However, this is not the only possible account of the NPI illusion in the spirit of the noisy channel framework. The other possibility is that the comprehender's representation of the negative word is subject to edits. Critically, there is no small orthographic change that could make the negative word a good licenser for the NPI, since it is not this word's identity but its location that makes it a bad licenser. Thus, one might argue that the comprehender entertains a representation of the string in (64a) that is more like (64b), as one of the many possible underlying intended signals that has been distorted by noise.

- (64)
- a. The authors [that **no** critics have recommended] have ...
 - b. **No** authors [that the critics have recommended] have ...

Such an edit is obviously not in line with Levy's original proposal that the probability of a possible analysis is related to its Levenshtein edit distance from the actual string (roughly, the number of characters that must be changed to get from one to the other). It is also not in line with subsequent noisy channel accounts, which exclusively focus on deletions and insertions (Gibson, Bergen, & Piantadosi 2013). But there may be independent reasons to believe that a "shuffling" of the words in memory is possible. Potter & Lombardi 1990 argue that verbatim recall of sentences involves "regeneration of the sentence from a conceptual representation, using words that have been recently activated. A key claim is that the activated lexical items are unordered" (Potter & Lombardi 1990:633). Of course, these claims are about the representation of a sentence after it has been fully encoded. So it does not necessarily follow that the representation of a sentence as it's being read involves any kind of unordered set of lexical items. It is also clear that one could extend this reasoning too far — the representation of prior sentential context cannot simply be a bag of words that can be rearranged in whatever way makes upcoming material grammatical. But it is possible that the comprehender assigns a higher probability to analyses that involve the same lexical items in a different order than those that involve completely different lexical items.

Such a theory would account for the interpretation data quite naturally — comprehenders believe the sentence to be negative because, following the NPI, the analyses in which the negative quantifier is

moved to the sentential subject become more probable, thus making it more probable that the MC is within the scope of negation. The lack of correspondence between negative interpretations and illusions is somewhat surprising. This might be explained if both the version in (64a) and the version in (64b) are maintained (with differing probabilities) until the end of the sentence, and later judgments of acceptability and meaning are made based on different representations.

The hypothesis might also be able to account for the licenser effect, if the edit from (65a) to (65b) is less probable than the edit from (64a) to (64b).

- (65) a. The authors [that the critics **haven't** recommended] have ...
 b. The authors [that the critics have recommended] **haven't** ...

This might be because *have* was the most recent word at the point when the NPI is encountered, and so any edit to this word is considered less probable than edits to words that are farther back in memory. However, this relates to a key issue for the hypothesis: the distance effect. If the illusion is due to the possibility that the first few words of the sentence were in a different order than they appeared, it is somewhat surprising that at a later point in the sentence these distortions become *less* probable. Extending the logic that we're applying to (65), it seems that very recent information is very clearly encoded, and as the beginning of the sentence becomes more remote, we might expect the comprehender to become less confident that it is being remembered accurately, and therefore more vulnerable to distortions. This would predict the opposite of the observed distance effect — more illusions for NPIs that occur later in the sentence.

Alternatively, in order to account for the actual distance effect, one might argue that as the sentence progresses, the comprehender finds increasing evidence that the top-probability analysis of the sentence (i.e., the veridical one) is a good analysis, and so its probability increases while the probability of other analyses (including the one in which *no* has moved) decreases. This is possible, but it predicts a gradient effect of distance, starting from the negative word. As we demonstrated at length in Chapter 4, the distance effect is insensitive to the position of the negative word, and only sensitive to the distance between the edge of the RC and the NPI. This conceptualization of temporal changes in the edits that are con-

sidered is also directly in conflict with the explanation for the licenser effect we just considered. To some extent, the licenser effect and the distance effect are in conflict under the noisy channel hypothesis.

Finally, the PPI illusion is not well-explained by a noisy channel model. At the point when the PPI is encountered, the comprehender would, under this hypothesis, have a probability distribution over representations of both (64a) and (64b), among many others, with the highest probability assigned to (64a). Upon encountering a main-clause PPI, the evidence favoring this analysis increases and so its probability should rise, while the probability assigned to (64b) should decrease. By the end of the sentence, the veridical representation is well in the lead, so it is puzzling that the vanishingly small probability assigned to (64b) would have any influence on acceptability judgments. Moreover, there are always minor edits that one could make to a sentence that would result in blatant ungrammaticality — we do not expect that the existence of these possible edits should result in degraded acceptability.³⁴

5.1.3 The current study

The five experiments presented in this chapter substantially complicate the empirical picture. All of the findings presented here were unexpected, and should be replicated in future work. They are important specifically because they are so surprising under many accounts.

We have reviewed four key empirical generalizations concerning the NPI illusion — the licenser effect, the distance effect, the interpretation, and the PPI illusion — as well as five proposed explanations for the cause of the illusion — cue based retrieval, pragmatic rescuing, scalar alternatives, scope miscalculation, and the noisy channel framework. It is clear that none of these hypotheses perfectly accounts for the data, though some fare better than others. In particular, the findings from Chapter 3 and Chapter 4 seem to favor the scalar alternatives hypothesis. A key piece of evidence favoring this hypothesis over others comes from the finding that illusions may be possible for non-quantificational licensers if the appropriate scalar meanings can be achieved in other ways. Recall that this finding from Experiment 6 was not especially robust.

³⁴Or, if they do, they degrade the acceptability of every single grammatical sentence, because every single grammatical sentence is just one edit away from being ungrammatical.

Thus, Experiment 12 aimed to verify this generalization with a more direct manipulation of scalar meanings. However, the predictions of the scalar alternatives hypothesis were not borne out. In Experiments 13 and 14 we turn to the distance effect. Experiment 13 aimed to demonstrate that illusions are possible for *any* when it is positioned close to the RC, ruling out possible confounds in previous studies of the distance effect³⁵. While we do find illusions with *any*, we also find a surprising lack of distance effects — that is, illusions arise even for NPIs placed very far from the RC. Experiment 14 attempts to replicate this surprising finding with different items, but fails to do so because of baseline issues. In Experiment 15 we turn to the predictions of the noisy channel hypothesis, specifically asking if the licenser effect goes away when both types of licenser can be “shuffled” into a position from which they would license the NPI. The results of this experiment are not clearly in line with the predictions of the noisy channel account, but once again interpretation of the observed patterns is difficult — we find no clear illusions in this experiment for embedded *no*. Experiment 16 attempts to rule out a potential confound in this experiment, namely the use of past tense in the MC. We found, to our surprise, that tense seems to matter. These disparate findings do not tell a clear story as to the cause of the illusion but they are surprising enough that future work should follow up on them.

5.2 Scalar alternatives

The scalar alternatives hypothesis explains the licenser effect as a consequence of the kinds of alternatives different negative forms evoke. Under this hypothesis, illusions arise not because of interference from any particular lexical item, but because of interference from NPI-licensing scalar alternatives. *No* and *haven't* differ in their capacity to cause illusions only insofar as they are more or less likely to lead the comprehender to infer such alternatives prior to the presentation of an NPI. The hypothesis therefore makes a strong prediction: if scalar alternatives are guaranteed to have been generated before the end of the RC, there should be interference regardless of the form of negation that was used. Experiment 6 provided tentative support for this prediction, but the illusion effect sizes were small across the board, and a surprising illusion was detected for *haven't* even in the absence of scalar alternatives.

³⁵Experiment 11 also demonstrated that illusions are possible for *any* but this was not the primary purpose of that study.

5.2.1 Experiment 12

Recall that Experiment 6 tested illusion rates for sentences like those in (66). The key idea was that if the critical difference between *no* and *haven't* is the clause-level scalar alternatives they evoke, then (66b) should be well-matched to (66a) (having essentially the same clause-level meaning), should evoke the same alternatives, and should yield equal illusion rates. In contrast, if some inherent property of quantifiers (such as their scope-taking properties) is responsible for the licenser effect, no illusions are expected for (66b).

- (66)
- a. * The critics [that have recommended no authors of alternative genres] have ever objected to mainstream literary trends.
 - b. * The critics [that haven't recommended any authors of alternative genres] have ever objected to mainstream literary trends.
 - c. * The critics [that haven't recommended the authors of alternative genres] have ever objected to mainstream literary trends.

This experiment revealed numerically small but statistically significant illusions for all three conditions. Although the illusions for (66c) were reliably smaller than those for (66a), (66b) was intermediate and could not be statistically distinguished from either (66a) or (66c). This is obviously not what is expected if illusions are a consequence of the scope-taking properties of quantifiers (in which case no illusions should arise for either (66c) or (66b)). These findings are *compatible* with the scalar alternatives account, but not exactly the pattern we expected.

While the design in Experiment 6 was useful because the meanings of (66b) and (66a) are well-matched, this required the use of SRCs which may have smaller illusion effect sizes in general³⁶. The key thing for the scalar alternatives hypothesis is just that there's an NPI in the RC. Since all NPI licensing is scalar under this hypothesis, the comprehender must generate scalar alternatives (which will be appropriately ordered in virtue of the negative word) at the RC NPI, and so those alternatives should be available to

³⁶This might be because it's a bit odd to put *no* in object position. If that's the case, it's not that the illusion is smaller in SRCs, it's that the acceptability of the sentence as a whole is degraded making the illusion look smaller. We will return to this issue.

interfere with the NPI in the MC. Thus in Experiment 12 we used more typical ORC sentences as in (67a), which have been known to yield strong illusions, and simply added NPIs to the RC as in (67b). This design has the added benefit that the same NPIs can be added to sentences with *no* as the intrusive licenser, thus allowing us to directly test whether the licenser contrast disappears when the likelihood of inferring scalar alternatives is matched.

- (67) a. The authors [that the critics haven't recommended in their reviews] have ever received acknowledgment for a best-selling novel.
- b. The authors [that the critics haven't recommended in any of their reviews] have ever received acknowledgment for a best-selling novel.
- (68) a. The authors [that no critics have recommended in their reviews] have ever received acknowledgment for a best-selling novel.
- b. The authors [that no critics have recommended in any of their reviews] have ever received acknowledgment for a best-selling novel.

The predictions are as follows. Under the scalar alternatives hypothesis, we expect that illusions will arise whenever scalar alternatives have been inferred prior to the main-clause NPI — that is, in (68b), (68a), and, critically, (67b), but not (67a).

5.2.1.1 Participants

31 US-based native speakers of English participated in this experiment. All participants provided informed consent and they received \$9 as compensation. We excluded workers who failed to provide a response within 2 seconds for 25% of fillers or more and workers whose judgments of filler trials did not reliably distinguish between grammatical and ungrammatical fillers, based on a chi-squared test. 4 workers were excluded based on these criteria, resulting in 27 participants in our analysis. The mean filler-trial accuracy of the included participants was 77%.

Normal	
A. Grammatical baseline	No authors [that the critics have recommended in their reviews] have ever ...
B. Embedded <i>no</i>	The authors [that no critics have recommended in their reviews] have ever ...
C. Embedded <i>haven't</i>	The authors [that the critics haven't recommended in their reviews] have ever ...
D. Ungrammatical baseline	The authors [that the critics have recommended in their reviews] have ever ...
...received acknowledgement for a best-selling novel.	
With NPIs	
E. Grammatical baseline	No authors [that the critics have recommended in any of their reviews] have ever ...
F. Embedded <i>no</i>	The authors [that no critics have recommended in any of their reviews] have ever ...
G. Embedded <i>haven't</i>	The authors [that the critics haven't recommended in any of their reviews] have ever ...
H. Ungrammatical baseline	The authors [that the critics have recommended in any of their reviews] have ever ...
...received acknowledgement for a best-selling novel.	

Table 5.1: Example stimuli for Experiment 12

5.2.1.2 Materials

The experimental materials consisted of 32 sets of items across 8 conditions that varied the presence, location, and type of licenser with respect to the NPI *ever*, crossed with the presence or absence of NPIs in the RC. This manipulation resulted in the experimental conditions shown in Table 5.1. Conditions A, B, C, and D correspond to the standard NPI illusion conditions, including the embedded-*haven't* condition that has been shown many times to not yield illusions. Conditions E, F, G, and H parallel these conditions, but with the addition of various NPIs inside the RC, which we refer to as the “scale-inducing” manipulation. Different NPIs were used for different items (but always the same NPIs for the different conditions of the same item). This manipulation meant that the number of words was not always perfectly matched across conditions.

Each participant rated 90 sentences: 39 experimental items and 51 fillers of similar length and complexity. Participants also saw eight “attention check” trials. The experimental items were distributed across 8 lists using a Latin Square design and the fillers were the same in each list. Participants completed 2 practice items before beginning the experiment, to ensure that they understood the procedure.

5.2.1.3 Procedure

The sentences were presented using PCIBex and the presentation order was randomized for each participant. Each sentence was displayed word by word at a rate of 400 ms. per word, in the center of the screen,

using the RSVP paradigm. At the end of each sentence participants were asked to provide a yes/no button press judgment in response to the question “Was that a good sentence?” within 2 seconds. If participants failed to provide the judgment in time, a message indicated that they were too slow. The dependent measure was the acceptance rate across trials and participants. The task was designed to be completed in 30-45 minutes.

5.2.1.4 Analysis

The results were analyzed using a helmert-coded generalized linear mixed effects model using a logit link function. The maximal structure was initially built including by-subject and by-item random intercepts and slopes for the experimental conditions. When this model failed to converge, it was reduced according to the recommendations provided by Barr et al. 2013. We used the emmeans package (Lenth et al. 2018) to extract beta coefficients and p-values for pairwise comparisons between conditions.

5.2.1.5 Results

The results from this experiment are presented in Figure 5.1, which shows the proportion of “yes” responses given to each condition. An effect of grammaticality was observed ($\beta=3.07$, $SE=0.41$, $z=7.46$, $p<.001$), indicating that the grammatical baseline conditions were significantly more likely to be judged acceptable than the ungrammatical baseline conditions, averaging over the scale-inducing manipulation. An effect of embedded-*no* was observed ($\beta=0.94$, $SE=0.30$, $z=3.12$, $p=.01$), replicating the standard illusion effect for negative quantifiers, averaging over the scale-inducing manipulation.

No effect of embedded-*haven't* was observed ($\beta=-0.11$, $SE=0.31$, $z=-0.36$, $p=.98$). Critically, there was no interaction between this factor and the scale-inducing manipulation ($\beta=0.25$, $SE=0.53$, $z=0.48$, $p=.63$), and pairwise comparisons revealed no significant illusion effects for either the normal ($\beta=0.01$, $SE=0.41$, $z=0.04$, $p=.97$) or NPI-containing ($\beta=-0.24$, $SE=0.41$, $z=-0.59$, $p=.56$) embedded-*haven't* sentences.

The comparison of embedded-*haven't* and embedded-*no* revealed a significant effect of the type of embedded negation ($\beta=1.06$, $SE=.31$, $z=3.44$, $p=.003$), indicating that the embedded-*no* conditions were

significantly more likely to be judged acceptable than the embedded-*haven't* conditions, averaging over the scale-inducing manipulation. Critically, this effect did not interact with the scale-inducing manipulation ($\beta=0.21$, $SE=0.50$, $z=0.41$, $p=.68$), and pairwise comparisons revealed significant differences between the two forms of negation for both the normal ($\beta=1.16$, $SE=0.40$, $z=2.93$, $p=.003$) and NPI-containing ($\beta=0.95$, $SE=0.40$, $z=2.39$, $p=.02$) conditions.

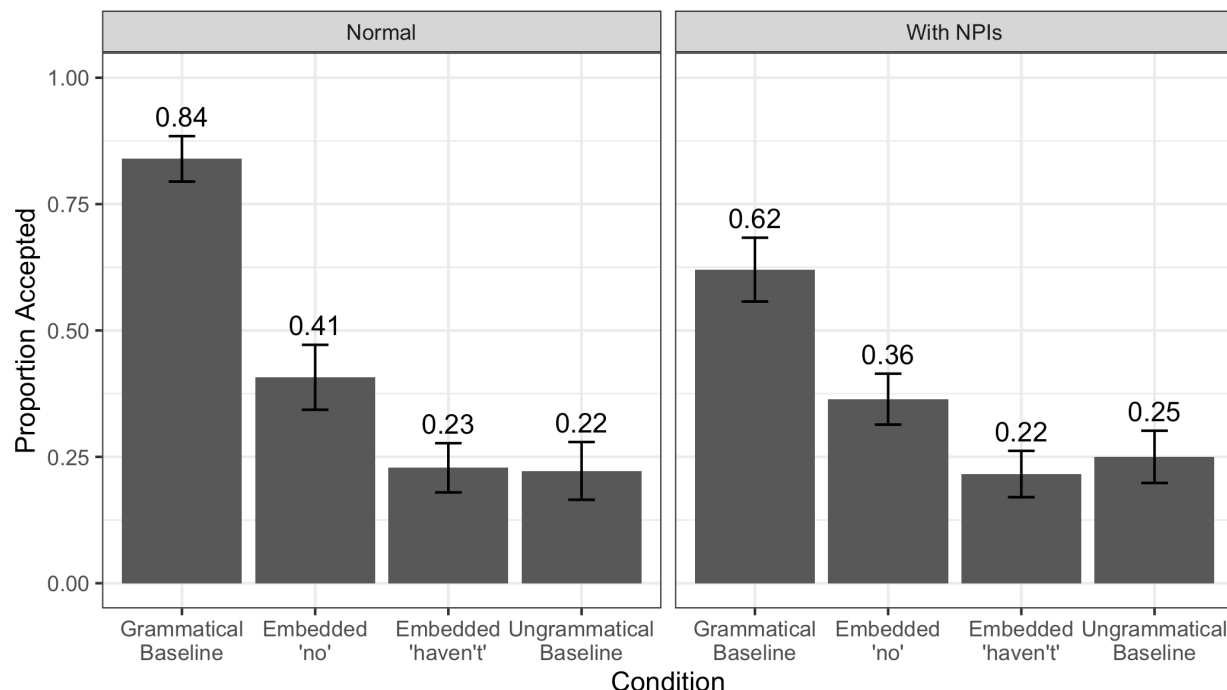


Figure 5.1: Mean percentage of ‘YES’ responses for the experimental conditions in Experiment 12. Error bars indicate standard error of the mean across subjects.

5.2.1.6 Discussion

The results obtained in Experiment 12 pose a clear problem for the scalar alternatives hypothesis. This hypothesis made a clear prediction that illusions should arise for embedded *haven't* when the appropriate alternatives are achieved through other parts of the clause, and this prediction was not borne out. We find no illusions for *haven't*, even when the RC contains additional NPIs. One might in principle worry that the lack of a statistically significant illusion is due to lack of power, but such concerns are made less plausible by the significant contrast between the two forms of negation. For both the normal and NPI-

containing sentences, we find a clear contrast between embedded-*no* and embedded-*haven't*. Under the scalar alternatives hypothesis, this contrast should have been neutralized for the sentences with additional NPIs. Thus the failed prediction is instantiated by both a null effect that was predicted to exist (the illusion for embedded-*haven't* when NPIs were added to the RC) and a significant effect that was predicted not to exist (the contrast between embedded-*no* and embedded-*haven't* when NPIs were added to the RC).

One option to make sense of this finding under this hypothesis is to say that an illusion *did* occur for *haven't* paired with RC NPIs, but reanalysis processes that occur between the NPI and the sentence-final judgment lead to the ultimate perception that the sentence is ungrammatical. This is obviously stipulative. The implications of this finding are discussed in greater detail in Chapter 6, but for now we simply conclude that the evidence in favor of the scalar alternatives hypothesis is not as strong as it once seemed.

One other observation worth noting is that the grammaticality effect is quite reduced for the sentences with additional NPIs — the grammatical baseline sentences were accepted only about 62% of the time, whereas grammatical baselines for NPI illusions are typically accepted around 80% of the time. This is rather surprising, since negative quantifiers are generally said to license NPIs in both their restrictor and their scope, so it is not obvious why the addition of NPIs in the restrictor should have this effect. Our intuition, upon inspecting the items, is that this may have something to do with the use of the present perfect in the RC: *No authors [that the critics recommended in any of their reviews]* sounds a bit better than *No authors [that the critics have recommended in any of their reviews]*. We don't know why this would be the case but it may be an interesting area for future work.

5.3 Distance effects

We now turn our attention to the distance effect. As has been noted several times already, much of our reasoning about the distance effect is based on contrasts in illusion rates for NPIs that are only one word apart. But since most NPIs are not so flexible with their word order, such fine-grained comparisons re-

quire comparing different NPIs — typically *ever* versus *any*. While this comparison is reasonable given the constraints on the stimuli, it is important to verify that the observed differences are not due to inherent differences between these NPIs. Experiment 11 is the only demonstration, to our knowledge, that illusions are possible with the NPI *any*. Thus in Experiment 13 we aimed to re-establish this fact using NPI-containing prepositional phrases like *at any time* and *in any way*. These phrases allow greater flexibility in the positioning of the NPI and thus allow us to demonstrate both that illusions arise for *any* and, potentially, that the same distance effect arises for these phrases.

5.3.1 Experiment 13

The aim in Experiment 13 was to determine whether *ever* and *any* give rise to similar illusions when positioned similarly close to the RC. A secondary aim was to replicate the distance effect using *any*. This we compared stimuli like in (69). Technically *any* is one word farther from the RC than the NPI *ever*, but since this is only the preposition *at* (or, for other items, *in*), we think this is unlikely to matter in light of the findings from Experiment 11. Rather, we will in general consider the entire *any*-containing PP to be an NPI. We expect illusions to arise at equal rates for (69a) and (69b), since both position the NPI close to the RC. We additionally expect to find a lack of illusions for (69c), since the NPI *any* is much farther from the RC. Note that these expectations are not hypothesis-driven predictions, but simply an extension of the best generalization over previously-observed distance effects (that is, that NPIs that occur close to the RC are subject to illusions and NPIs that occur later are not).

- (69)
- a. *The surgeons [that no patients consulted] have ever suggested unnecessary operations.
 - b. *The surgeons [that no patients consulted] have, at any time, suggested unnecessary operations.
 - c. *The surgeons [that no patients consulted] have suggested unnecessary operations at any time.

<i>Ever</i> , close	
A. Gramm. baseline	No surgeons [that the patients consulted] have ever suggested unnecessary operations.
B. Embedded <i>no</i>	The surgeons [that no patients consulted] have ever suggested unnecessary operations.
C. Ungramm. baseline	The surgeons [that the patients consulted] have ever suggested unnecessary operations.
<i>Any</i> , close	
D. Gramm. baseline	No surgeons [that the patients consulted] have, at any time , suggested unnecessary operations.
E. Embedded <i>no</i>	The surgeons [that no patients consulted] have, at any time , suggested unnecessary operations.
F. Ungramm. baseline	The surgeons [that the patients consulted] have, at any time , suggested unnecessary operations.
<i>Any</i> , far	
G. Gramm. baseline	No surgeons [that the patients consulted] have suggested unnecessary operations at any time .
H. Embedded <i>no</i>	The surgeons [that no patients consulted] have suggested unnecessary operations at any time .
I. Ungramm. baseline	The surgeons [that the patients consulted] have suggested unnecessary operations at any time .

Table 5.2: Example stimuli for Experiment 13

5.3.1.1 Participants

93 US-based native speakers of English participated in this experiment. All participants provided informed consent and they received \$6 as compensation. We excluded workers who failed to provide a response within 2 seconds for 25% of fillers or more and workers whose judgments of filler trials did not reliably distinguish between grammatical and ungrammatical fillers, based on a chi-squared test. 4 workers were excluded based on these criteria, resulting in 89 participants in our analysis. The mean filler-trial accuracy of the included participants was 81%.

5.3.1.2 Materials

The experimental materials consisted of 27 sets of items across 9 conditions that varied the presence and location of the licenser, crossed with the presence and type of NPI in the MC. These manipulations resulted in the experimental conditions shown in Table 5.2. Conditions A, B, and C correspond to the standard NPI illusion conditions with *ever* (i.e., embedded-*no*, and the corresponding grammatical and ungrammatical baselines). Conditions D, E, and F use the same licenser manipulation, but use the NPI *any* housed the prepositional phrase *at any time* (for about half of the items, this was *in any way*). Conditions G, H, and I use the same *any*-PP, but position this phrase sentence-finally. The experiment used the same 90 fillers as Experiment 12.

5.3.1.3 Procedure & Analysis

The procedure was identical to Experiment 12. The results were again analyzed with logistic mixed effects models, using the same analysis strategy as Experiment 12.

5.3.1.4 Results

The results from this experiment are presented in Figure 5.1, which shows the proportion of “yes” responses given to each condition. An effect of grammaticality was observed ($\beta=4.45$, $SE=0.30$, $z=15.04$, $p<.001$), indicating that the grammatical baseline conditions were significantly more likely to be judged acceptable than the ungrammatical baseline conditions, averaging over the three NPI-distance configurations (*ever*-close, *any*-close, and *any*-far). An effect of embedded-*no* was observed ($\beta=1.00$, $SE=0.16$, $z=6.22$, $p<.001$), replicating the standard illusion effect for negative quantifiers, averaging over the three NPI-distance configurations.

Focusing on only the conditions with NPIs close to the RC (*ever*-close and *any*-close), we did not observe an interaction between NPI identity and the size of the illusion effect ($\beta=0.57$, $SE=0.33$, $z=1.76$, $p=.08$), and follow up pairwise comparisons revealed significant illusions for both *ever* ($\beta=1.39$, $SE=0.26$, $z=5.40$, $p<.001$) and *any* ($\beta=0.82$, $SE=0.23$, $z=3.51$, $p<.001$). Focusing on only the conditions with *any* (*any*-close and *any*-far), we did not observe an interaction between NPI position and the size of the illusion effect ($\beta=0.01$, $SE=0.33$, $z=0.03$, $p=.97$), and follow up pairwise comparisons revealed significant illusions for both *any*-close, as was mentioned above ($\beta=0.82$, $SE=0.23$, $z=3.51$, $p<.001$), and *any*-far ($\beta=0.80$, $SE=0.27$, $z=3.02$, $p=.003$).

5.3.1.5 Discussion

There are two important components of our findings from Experiment 13. The first is that we find, once again, that illusions are possible for *any*, making this confound an implausible explanatory variable in research investigating the distance effect. The second notable finding is that Experiment 13 revealed illusions for NPIs in a sentence final position. This is wholly inconsistent with the generalizations concerning the distance effect that we have previously considered. In Chapter 4 we asked whether it is the

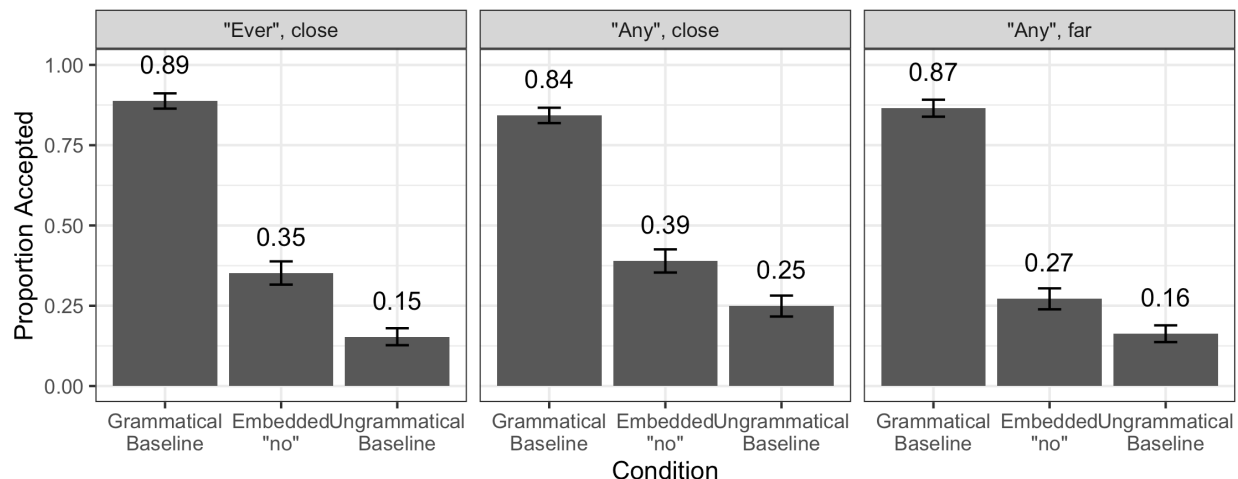


Figure 5.2: Mean percentage of ‘YES’ responses for the experimental conditions in Experiment 13. Error bars indicate standard error of the mean across subjects.

distance from the negative word to the NPI or the distance from the RC to the NPI that must be short for illusions to occur. The present findings seem to suggest that it is neither — rather, in Experiment 13, it appears as if there is no distance effect at all. Of course, distance effects have been demonstrated several times, in both our experiments and those presented in Parker & Phillips 2016, so we are hesitant to abandon this generalization altogether.

One possible explanation for the surprising illusion for late *any*-PPs is that this result is simply a false positive. Thus we aim to replicate the pattern (with different stimuli) in Experiment 14.

Another possibility is that *any* is not subject to the same kind of distance effects as *ever*, perhaps because of its dual life as a FCI. That is, the late-*any* items may be accepted not because comprehenders believe that the NPI is licensed but because they believe the free choice use is licensed. If this were the case, though, we would expect the ungrammatical baseline to be accepted just as often, since free choice readings shouldn’t depend on the presence of negation. We therefore don’t consider this explanation especially likely.

Finally, it may be that the distance effect is not monotonic in the way that was previously assumed. All previous demonstrations of the distance effect have compared NPIs that occur very soon after the edge of the RC to NPIs that occur a few words later, in more or less the “middle” of the MC. This experiment

is the first to test sentence-final NPIs³⁷, and the first to find illusions for NPIs that are far from the RC. This raises the possibility that illusion vulnerability fluctuates throughout the MC, being first high, then low, then high again. This might be spelled out in terms of fluctuations in the availability of the RC representation — at first, its recency makes it highly active, then attention is directed to other sentence processing operations, then “sentence-final wrap-up” effects bring the RC back into the comprehender’s focus of attention. Experiment 14 was in part designed to determine whether fluctuating vulnerability is the right way to think about this effect.

5.3.2 Experiment 14

Given the surprising illusion for late-*any* in Experiment 13, one of the goals of Experiment 14 was to simply replicate the effect. A second goal was to replicate one of the demonstrations of the distance effect from Parker & Phillips 2016. They demonstrated distance effects in three ways. One of these experiments has been replicated — the *ever-any* contrast, which was part of our Experiment 9. To our knowledge, no replication of the other two contrasts, illustrated in (70) and (71), has been attempted. We chose to focus our attention on (71) because we also wanted to replicate our Experiment 13 findings in the same experiment, and the parentheticals in (70) combine awkwardly with *at any time* and *in any way*.

- (70) a. * As the editors mentioned, the authors [that no critics recommended for the assignment] have ever received a pay raise.
b. * The authors [that no critics recommended for the assignment] have, as the editor mentioned, ever received a pay raise.

(Parker & Phillips 2016:331)

- (71) a. * The journalists [that no editors recommended for the assignment] ever thought that the readers would understand the complicated situation.
b. * The journalists [that no editors recommended for the assignment] thought that the readers would ever understand the complicated situation.

³⁷Technically, the NPI *any* is the second-to-last word, but it’s part of a sentence-final PP, and we are assuming that the entire PP is polarity sensitive.

(Parker & Phillips 2016:328)

Finally, a third goal of Experiment 14 was to determine whether there is in fact a pattern of fluctuating vulnerability to illusions over the post-RC region of the sentence, by testing the same NPI in three positions of the same sentence. Given these goals, we used items like those in (71) as a starting point, and included a set of conditions intended to directly replicate this contrast. We also included versions of these items using *at any time* and *in any way* in place of *ever* in both the early (as in (71a)) and sentence-medial (as in (71b)) positions. And finally we included a set of conditions that used *at any time* and *in any way* sentence-finally.

5.3.2.1 Participants

43 US-based native speakers of English participated in this experiment. All participants provided informed consent and they received \$6 as compensation. We excluded workers who failed to provide a response within 2 seconds for 25% of fillers or more and workers whose judgments of filler trials did not reliably distinguish between grammatical and ungrammatical fillers, based on a chi-squared test. 2 workers were excluded based on these criteria, resulting in 41 participants in our analysis. The mean filler-trial accuracy of the included participants was 77%.

5.3.2.2 Materials

The experimental materials consisted of 45 sets of items across 15 conditions. We used the materials from Parker & Phillips 2016 Experiments 4 and 5 as a starting point, but because they had only 36 items and the present experiment has 15 conditions, we added 9 new stimuli that used the same structures. The 15 conditions consisted of five triples of conditions, where each triple consists of an embedded-negation sentence and a corresponding grammatical and ungrammatical baseline. The 5 versions of these triples essentially crossed the identity of the NPI (*ever* or *any*) with the position of the NPI (short distance, long distance, or sentence-final), but one with one of the cells of this 2x3 missing: sentence-final *ever*. Because these sentences are not particularly natural, and they do not address any of the three goals we laid out above, we excluded them. One other noteworthy change is that we added *have* between the RC and NPI

Short distance		
A. Gram. baseline	<i>ever</i>	No journalists [that the editors recommended for the assignment] have ever argued that the readers would understand the complicated situation.
B. Embedded <i>no</i>	<i>ever</i>	The journalists [that no editors recommended for the assignment] have ever argued that the readers would understand the complicated situation.
C. Ungram. baseline	<i>ever</i>	The journalists [that the editors recommended for the assignment] have ever argued that the readers would understand the complicated situation.
D. Gram. baseline	<i>any</i>	No journalists [that the editors recommended for the assignment] have in any way argued that the readers would understand the complicated situation.
E. Embedded <i>no</i>	<i>any</i>	The journalists [that no editors recommended for the assignment] have in any way argued that the readers would understand the complicated situation.
F. Ungram. baseline	<i>any</i>	The journalists [that the editors recommended for the assignment] have in any way argued that the readers would understand the complicated situation.
Long distance		
G. Gram. baseline	<i>ever</i>	No journalists [that the editors recommended for the assignment] argued that the readers would ever understand the complicated situation.
H. Embedded <i>no</i>	<i>ever</i>	The journalists [that no editors recommended for the assignment] argued that the readers would ever understand the complicated situation.
I. Ungram. baseline	<i>ever</i>	The journalists [that the editors recommended for the assignment] argued that the readers would ever understand the complicated situation.
J. Gram. baseline	<i>any</i>	No journalists [that the editors recommended for the assignment] argued that the readers would in any way understand the complicated situation.
K. Embedded <i>no</i>	<i>any</i>	The journalists [that no editors recommended for the assignment] argued that the readers would in any way understand the complicated situation.
L. Ungram. baseline	<i>any</i>	The journalists [that the editors recommended for the assignment] argued that the readers would in any way understand the complicated situation.
Sentence-final		
M. Gram. baseline	<i>any</i>	No journalists [that the editors recommended for the assignment] argued that the readers would understand the complicated situation in any way.
N. Embedded <i>no</i>	<i>any</i>	The journalists [that no editors recommended for the assignment] argued that the readers would understand the complicated situation in any way.
O. Ungram. baseline	<i>any</i>	The journalists [that the editors recommended for the assignment] argued that the readers would understand the complicated situation in any way.

Table 5.3: Example stimuli for Experiment 14

in all six short-distance conditions. This was done because of a concern that *at any time* and *in any way*, when positioned immediately after the RC, might be parsed as part of the RC. About half of the items used *in any way* and half used *at any time*. A full set of conditions for one item can be found in Table 5.3. We used the same 90 fillers as in previous experiments.

5.3.2.3 Procedure & Analysis

The procedure was identical to Experiments 12 and 13. The results were again analyzed with logistic mixed effects models, using the same analysis strategy as Experiments 12 and 13.

5.3.2.4 Results

The results from this experiment are presented in Figure 5.1, which shows the proportion of “yes” responses given to each condition.

We focus first on only the six conditions with *ever*, which aimed to directly replicate Parker & Phillips 2016 Experiment 4. An effect of grammaticality was observed ($\beta=3.12$, $SE=0.37$, $z=8.42$, $p<.001$), indicating that the grammatical baseline conditions were significantly more likely to be judged acceptable than the ungrammatical baseline conditions, averaging over the two distance configurations (short distance and long distance). An effect of embedded-*no* was observed ($\beta=0.64$, $SE=0.28$, $z=2.31$, $p=0.02$), replicating the standard illusion effect, averaging over the two distance configurations. There was no statistically significant interaction between this effect and the distance manipulation ($\beta=0.82$, $SE=0.48$, $z=1.71$, $p=0.09$), but follow up pairwise comparisons revealed statistically significant illusions for the short distance ($\beta=1.05$, $SE=0.39$, $z=2.73$, $p=0.006$) but not the long distance ($\beta=0.23$, $SE=0.35$, $z=0.67$, $p=0.50$) conditions.

Turning to the nine conditions with *any*, we observed a dramatic increase in the number of ungrammatical baseline trials that were accepted for both the long-distance and sentence-final conditions (41% and 49%, respectively). The grammatical and ungrammatical baselines in illusion experiments serve as a sanity check, demonstrating that the acceptability of the sentences is as expected when illusions are not at issue. Because these distance configurations failed this sanity check, we did not analyze them further and we are unable to draw strong conclusions on the basis of these data.

5.3.2.5 Discussion

One important finding from Experiment 14 is the qualitative replication of Parker & Phillips’s (2016) reported distance effect for *ever* in embedded versus unembedded positions. That is, while the critical interaction was not statistically significant, we found statistically significant illusions for *ever* when it was close to the RC but not when it was farther away. Since this experiment was conducted without a prospective power analysis, we are hesitant to make much of the non-significant interaction ($p=.09$).

However, this experiment did not achieve two of its main goals: a replication of Experiment 13’s find-

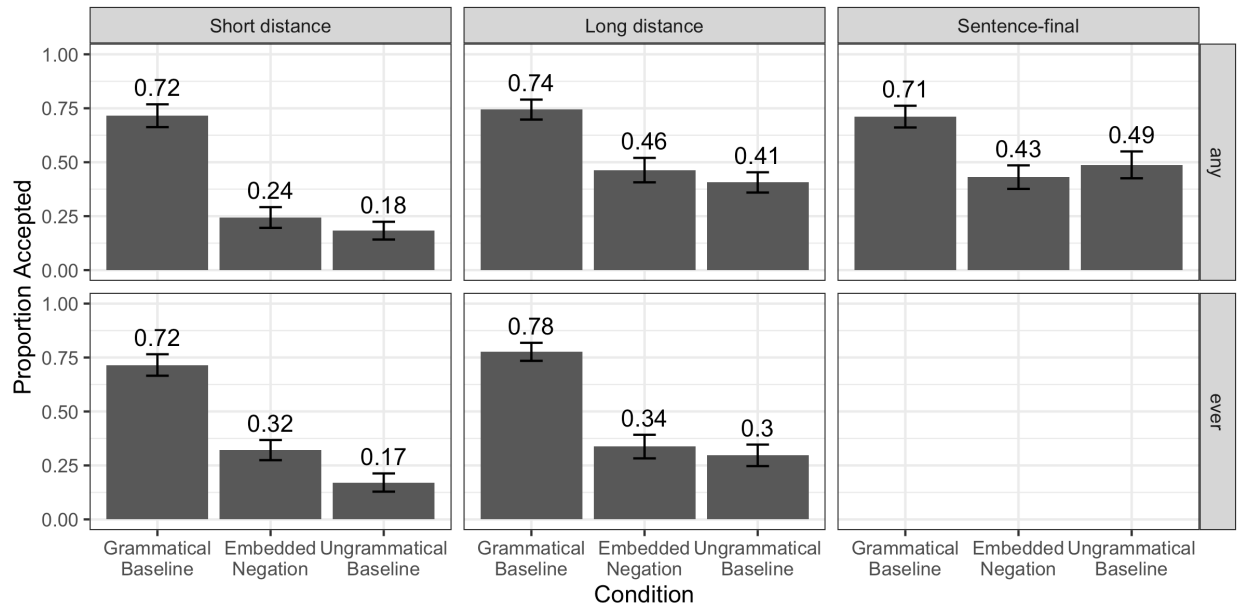


Figure 5.3: Mean percentage of ‘YES’ responses for the experimental conditions in Experiment 14. Error bars indicate standard error of the mean across subjects.

ing that illusions are possible for sentence-final *any*, and a full depiction of the timecourse of vulnerability using the same NPI at three points in the same sentence. It failed in these aims because of a failed sanity check — both the long-distance and sentence-final conditions revealed concerning high rates of acceptance for ungrammatical baseline sentences. We suspect this is because of the availability of a free choice reading of *any*. In particular, the use of a modal (*would*) in the embedded clause may have encouraged free choice readings. In any case, we are unable to draw any conclusions about the questions that were raised by Experiment 13. We return to this issue in Chapter 6.

5.4 Noisy channel

We now turn our attention to the predictions of the noisy channel explanation for NPI illusions, as presented in section 5.1.2.5. Recall that under this hypothesis the NPI illusion arises because the comprehender maintains a representation of the prior context in which the words are “shuffled” so that the negative word does in fact c-command the NPI. Non-quantificational licensors like *didn’t* don’t give rise to illusions, under this account, because there’s nowhere to shuffle them to (or, rather, the only candidate

position was just one word ago, and we must stipulate that the comprehender is more confident about the identity of extremely recent stimuli). Thus we expect, under this hypothesis, that if the sentential context were to provide a better “landing site” for sentential negation — that is, a position that both c-commands the NPI and is remote enough to be vulnerable — illusions would arise. This is the key prediction that we tested in Experiment 15.

5.4.1 Experiment 15

In order to make quantificational negation and non-quantificational negation equally subject to illusions under the noisy channel hypothesis, we embedded typical illusion sentences with both forms of negation under neg-raising verbs, as in (72). Under the noisy channel hypothesis, (72a) should be subject to illusions because the comprehender represents the sentence as (73a) (among many other representations maintained in parallel, with a probability distribution), and this representation allows the NPI to be licensed. Similarly, a comprehender reading (72b) should represent it as (73b) (among others), allowing the NPI to be licensed. Note that the comparison is not perfectly parallel, because of the use of *do*-support for sentential negation — that is, (73b) is not merely a re-shuffling of the orthographic words in (72b). However, perfectly matching the words would require using auxiliaries for all three clauses, and we felt these sentences were unnatural.³⁸ Thus the manipulation should just make illusions more likely for *didn’t* than usual, though not necessarily as likely as illusions for *no*. We used proper names for the highest subject so as to avoid providing even more options for positions for the negative quantifier to move to, and because

³⁸What is required depends a little bit on the particulars of the noisy channel hypothesis we adopt. If the idea is that two words can be *swapped*, then we would have needed sentences like *Mary had thought that the authors that hadn’t recommended the critics had ever ...*, so that the first *had* and *hadn’t* could be swapped (the second *had* needs to be there to separate the NPI from the RC). These sentences are a bit odd, so we wanted to avoid them. If instead the idea is that a single word can be moved to a different location (and we assume, as in Levy 2008 that deletions and insertions are also possible) then stimuli like *Mary believed that the authors that the critics didn’t recommend had ever...* are sufficient. *Didn’t* moves to the matrix clause, the *-ed* on *believe* is deleted and an *-ed* is added to *recommend*. This is also fine if morphemes can move on their own (instead of *-ed* being deleted and inserted, it moves). This is made slightly less elegant by the use of *thought* as one of the matrix verbs, since the edit from *think* to *thought* could be greater, depending on your model of edit distance. However, we opted to use *think* anyway because we wanted to use exclusively neg-raising verbs as the matrix predicates. Note that non-neg-raising verbs can still license lower NPIs, but they are generally more limited (see Gajewski 2007), and we wanted to avoid this issue. What this means is that even if the present experiment fails to show illusions for *didn’t*, there are ways for a noisy channel model to remain in play. This is ultimately just a consequence of underspecification in (our version of) the model. If instead we adopt a very specific metric for the probability of edits, like the versions assumed by Levy 2008 or Gibson, Bergen, & Piantadosi 2013, this issue does not arise. Those models do not predict basic NPI illusions in the first place.

the proper name versions were, based on our impressions, less taxing to process.

- (72) a. * Mary thought that the authors [that no critics recommended] had ever written a best-selling novel.
- b. * Mary thought that the authors [that the critics didn't recommend] had ever written a best-selling novel.
- (73) a. * Mary thought that no authors [that the critics recommended] had ever written a best-selling novel.
- b. * Mary didn't think that the authors [that the critics recommended] had ever written a best-selling novel.

This experiment also evaluated another possible prediction of the noisy channel hypothesis. We have observed before that NPI illusions appear to have a smaller effect size with SRCs (see Experiment 6 and Experiment 10) than with ORCs, though we have not systematically investigated this contrast. One possible explanation that we considered is that acceptance rates for SRC illusion sentences are suppressed because of the awkwardness of using quantificational *no* in object position.³⁹ However, another possibility is raised by the noisy channel hypothesis. In ORCs, the embedded negation is part of a clausal subject. This could make it more confusable with the sentential subject, since it may share more features with it. In SRCs, in contrast, the embedded negation is part of the RC's object, potentially making edited version in which negation c-commands the NPI by swapping positions with the sentential subject determiner less probable. A similar version of this account in which it is mere linear proximity, not shared features, that drives this effect is also possible.⁴⁰ Thus we additionally tested SRCs as in (74) in Experiment 15.

- (74) a. * Mary thought that the authors [that admired no critics] had ever written a best-selling novel.

³⁹We don't have an analysis of why this is slightly less acceptable, we only note that it is. It could have something to do with the availability of *didn't+any* to convey the same meaning, but it's not obvious why a speaker or comprehender would have this preference. In fact, under some theories of negation processing — “two stage” theories (Clark & Chase 1972; Fischler et al. 1983) — comprehenders should prefer for the negative word to arrive as late as possible in the string, since it has to be maintained in memory until everything else about the sentence has been processed, and then negation can be applied. This is beyond the scope of the present work. Note also that Experiment 10 used *very few* instead of *no* specifically to avoid this issue.

⁴⁰As this discussion highlights, there are many free parameters in the noisy channel hypothesis we consider here

- b. * Mary thought that the authors [that didn't admire the critics] had ever written a best-selling novel.

There is one additional benefit of testing the SRC-ORC comparison, which relates to a hypothesis that we have not previously addressed. The idea is that NPI illusions arise because of the detection of a locally coherent string. We know of no formal presentation of such a hypothesis, but it often comes up in discussions of NPI illusions. The basic idea of local syntactic coherence, proposed by Tabor, Galantucci, & Richardson 2004, is that comprehenders generate a full parse by creating small units of syntactic structure for substrings of the input, and later try to combine these units. This contrasts with parsing algorithms in which the structure assigned to an incoming word must be consistent with the structure assigned to everything to the left of it. Thus, local coherence accounts predict that comprehenders will initially represent *the player tossed a frisbee* in (75) as a clause in which *the player* is the subject of the verb phrase *tossed a frisbee*, which is inconsistent with the globally-coherent representation, in which *tossed a frisbee* is a reduced relative clause. Tabor, Galantucci, & Richardson 2004 present evidence that comprehenders do pursue such an analysis, though note that Levy 2008 offers a different explanation for these effects, in the noisy channel framework.

(75) The coach smiled at the player tossed a frisbee by the opposing team.

(Tabor, Galantucci, & Richardson 2004:1)

Applying this idea to NPI illusions, while comprehending a string like (76a), comprehenders might attempt to construct an analysis of the substring in (76b). Since the NPI is licensed in such a representation, this may give comprehenders an impression of acceptability for the string as a whole. The local coherence account explains the licensor effect because the equivalent substring in (77b) for sentences with embedded *didn't* is not locally coherent.

(76) a. * The authors [that no critics recommended] have ever ...

b. No critics recommended have ever ...

(77) a. * The authors [that the critics didn't recommend] have ever ...

b. * The critics didn't recommend have ever ...

We have generally been skeptical of such an account because of the existence of NPI illusions in other languages in which the locally-coherent analysis does not arise for the particular stimuli used (e.g., German, see Drenhaus, Saddy, & Frisch 2005), and because the locally-coherent substring in (76b) contains a reduced RC, which is well known to be a difficult structure. A local coherence account would have to say that comprehenders pursue this analysis even though it's generally dispreferred. However, getting rid of the reduced RC should make the locally-coherent analysis even more appealing to a comprehender. Thus, SRCs like (78a), which have locally-coherent substrings like (78b), should be especially vulnerable to illusions.

- (78) a. * The authors [that admired no critics] have ever ...
 b. No critics have ever ...

The local coherence hypothesis therefore predicts that embedded-*no* illusion sentences with SRCs should yield, if anything, *more* illusions than embedded-*no* illusion sentences with ORCs. The present experiment allows us to directly test whether this is the case.

Thus the key predictions are as follows. Under all accounts we expect to find illusions for embedded-*no* in ORCs and SRCs. Under (some versions of) the local coherence hypothesis, we expect the effect size to be larger for SRCs than ORCs. Under (some versions of) the noisy channel hypothesis, we expect the effect size to be larger for ORCs than SRCs. Additionally, under the noisy channel hypothesis, we expect to find illusions for embedded-*didn't* in this experiment (due to the embedding under neg-raising verbs), for both clause types, though not necessarily the same effect size as illusions for embedded-*no*.

5.4.1.1 Participants

100 US-based native speakers of English participated in this experiment. All participants provided informed consent and they received \$6 as compensation. We excluded workers who failed to provide a response within 2 seconds for 25% of fillers or more and workers whose judgments of filler trials did not reliably distinguish between grammatical and ungrammatical fillers, based on a chi-squared test. 12 workers were excluded based on these criteria, resulting in 88 participants in our analysis. The mean filler-trial accuracy of the included participants was 83%.

Object relative clause	
A. Grammatical <i>no</i>	Mary thought that no authors [that the critics recommended] had ever ...
B. Grammatical <i>didn't</i>	Mary didn't think that the authors [that the critics recommended] had ever ...
C. Embedded <i>no</i>	Mary thought that the authors [that no critics recommended] had ever ...
D. Embedded <i>didn't</i>	Mary thought that the authors [that the critics didn't recommend] had ever ...
E. Ungrammatical	Mary thought that the authors [that the critics recommended] had ever ...
...written a best-selling novel.	
Subject relative clause	
F. Grammatical <i>no</i>	Mary thought that no authors [that admired the critics] had ever ...
G. Grammatical <i>didn't</i>	Mary didn't think that the authors [that admired the critics] had ever ...
H. Embedded <i>no</i>	Mary thought that the authors [that admired no critics] had ever ...
I. Embedded <i>didn't</i>	Mary thought that the authors [that didn't admire the critics] had ever ...
J. Ungrammatical	Mary thought that the authors [that admired the critics] had ever ...
...written a best-selling novel.	

Table 5.4: Example stimuli for Experiment 15

5.4.1.2 Materials

The experimental materials consisted of 20 sets of items across 10 conditions that varied the presence, type, and location of the licenser, crossed with the type of embedding clause (SRC or ORC). These manipulations resulted in the experimental conditions shown in Table 5.4. Conditions A and B are both grammatical baselines, but use different forms of negation (*no* and *didn't*, respectively). Conditions C and D are the illusion conditions, with embedded negation (*no* and *didn't*, respectively). Condition E is the ungrammatical baseline, with no negation. Conditions F-J parallel these manipulations, but with SRCs instead of ORCs. All conditions had the illusion sentence embedded under a neg-raising verb, with a proper name as the subject. Six different neg-raising verbs were used across the items. One other change relative to previous NPI illusion experiments is that we used the past perfect in the clause containing the NPI, whereas previously we had used either the present perfect (other experiments, including some of those reported in Parker & Phillips 2016, used the simple past). This was done in order to achieve the most natural pairing of tenses between clauses. We did not expect the tense change to matter. We also simplified and shortened the post-NPI region of many items in order to keep the sentences from becoming overly long and complicated.

The experiment additionally used 60 filler sentences, half of which were grammatical. The fillers used a mix of syntactic structures, including a number of the same name-verbed-that-S structure used in the

experimental stimuli, so that these items would not stand out.

5.4.1.3 Procedure & Analysis

The procedure was identical to Experiments 12-14. The results were again analyzed with logistic mixed effects models, using the same analysis strategy as Experiments 12-14.

5.4.1.4 Results

The results from this experiment are presented in Figure 5.4, which shows the proportion of “yes” responses given to each condition. Descriptively, we once again have some sanity check issues. The ungrammatical baseline sentences were accepted fairly often (36% of trials and 40% of trials for ORCs and SRCs, respectively) and we see no trends toward illusions for embedded-*no*. Because of these issues we do not subject the results to statistical analysis.⁴¹

5.4.1.5 Discussion

The results of Experiment 15 are obviously unexpected. Recall that we expected, under all hypotheses, to find illusions for embedded-*no* in both SRCs and ORCs. This was not found in Experiment 15.⁴² Here we consider some possible explanations for the observed patterns.

One possibility is that embedding the sentences under neg-raising verbs with proper names as subjects resulted in some general confusion. Obviously the addition of another layer of embedding makes the sentences more complex and potentially more difficult to process. We did simplify and shorten the post-NPI content of many items, relative to previous stimuli, in an attempt to counteract this increase in

⁴¹Note that the present results make it clear that we have not been very precise in our definition of the sanity checks for NPI illusion results. That is, we don’t have a strict cutoff for how many accepted ungrammatical baseline trials is too many. Intuitively, if comprehenders are at or around chance (50%) on these trials, it seems inappropriate to call the condition an ungrammatical baseline. However, acceptance rates for these conditions are never all the way at 0%. We expect that comprehenders will judge a number of trials incorrectly in a speeded acceptability task, simply because the words are presented in RSVP, so if the comprehender just blinks at the wrong moment they could miss a critical word and mis-analyze the sentence. So we routinely see ungrammatical baselines around 20% (and grammatical baselines around 80%) and we don’t worry about them. The space around 30-40% is a bit murkier.

⁴²Not only is there no significant effect (we did not run statistical tests), which might be explained by lack of power. Rather, there is no trend toward an illusion. Of course, this doesn’t make power concerns irrelevant — as Vasishth et al. 2018 demonstrated, low-powered experiments yield, on average, less accurate point estimates of population-level mean differences. Replication of these findings is therefore important before we draw any strong conclusions

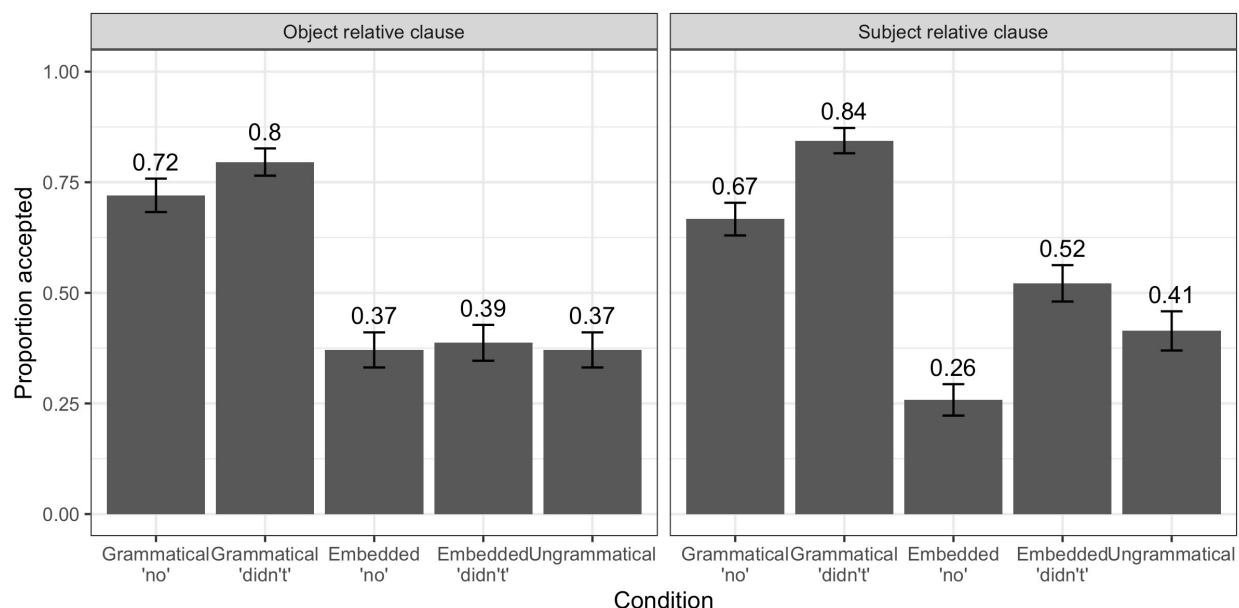


Figure 5.4: Mean percentage of ‘YES’ responses for the experimental conditions in Experiment 15. Error bars indicate standard error of the mean across subjects.

complexity. But it is of course possible that those efforts were not enough. The added complexity could have the effect of reducing acceptability across the board (which does not seem to have happened) or moving judgments toward chance across the board, due to guessing (which may have happened, since the ungrammatical baselines were accepted more than expected and some of the grammatical baselines were accepted less than expected). Why this would affect different conditions differently is not clear. It is also technically possible that it was specifically the simplification of the post-NPI content was responsible for the change in illusion patterns relative to previous studies, though we see no clear mechanism that would drive this.

Another issue is the relative oddness of *no* in object position, which we had previously mentioned as a possible explanation for depressed illusion rates in SRCs. That is, possibly due to a preference for a *didn't+any* construction, clauses like *the authors admired no critics* are a bit odd. It's possible that we underestimated the influence of this preference, and any condition in which the negative quantifier *no* could be replaced with *didn't+any* was perceived as less acceptable. Embedding the sentences under neg-raising verbs effectively makes it so that all of the sentences with *no* have a *didn't+any* competitor. Thus each of the four conditions with *no* in (79a), (80a), (81a), and (82a), may have been judged unacceptable

some of the time because comprehenders thought they should have been the versions in (79b), (80b), (81b)⁴³, and (82b). Note that (82a) actually has two *didn't+any* options: (82b) and (82c). To us, the contrast in acceptability between these pairs is not very large, but this difference may have been amplified in the experiment because the instructions we provide for how to give acceptability judgments is to think about whether they sound “like something a native speaker would say”. It’s possible that this biases participants to give judgments that do not merely evaluate the sentence on its own merits, but try to determine whether the sentence is the optimal way to express the thought (making “competitors” like the *didn't+any* versions more influential).

- (79) a. ? Mary thought that no authors [that the critics recommended] had ...
 b. Mary didn’t think that any authors [that the critics recommended] had ...
- (80) a. ? Mary thought that no authors [that admired the critics] had ...
 b. Mary didn’t think that any authors [that admired the critics] had ...
- (81) a. ? Mary thought that the authors [that no critics recommended] had ...
 b. ? Mary didn’t think that the authors [that any critics recommended] had ...
- (82) a. ? Mary thought that the authors [that admired no critics] had ...
 b. ? Mary didn’t think that the authors [that admired any critics] had ...
 c. Mary thought that the authors [that didn’t admire any critics] had ...

It may be that both of these factors matter — that is, the general difficulty of multi-clause sentences drives all conditions towards chance, while competition with *didn't+any* drives all *no*-containing conditions down. This would explain some of what we see, but mysteries remain.

One final possibility is the change in the MC tense. We did not expect this change to matter, but since it is another thing that was changed between this experiment and previous ones, it must be considered. We don’t have a clear mechanistic hypothesis about why the past perfect would yield different illusion results than the present perfect.

⁴³To us, (81b) actually isn’t so good. Maybe this is related to the fact that this condition was one of the only ones whose raw acceptance rate was not dramatically skewed in the present experiment (35% for an illusion sentence is pretty typical).

5.4.2 Experiment 16

As a first step toward making sense of the Experiment 15 results, we tested whether the stimuli were in some important way different from NPI illusion stimuli we had used in the past. In the discussion section above, we identified three changes that make these stimuli different: the addition of the extra clause containing a proper name and a neg-raising verb under which the typical NPI illusion sentence is embedded, the simplification of the post-NPI region, and the use of the past perfect. We had two ideas for why the added clause might matter (i.e., general complexity issues, and a dispreference for sentences with *no* due to competition with other possible forms), but no clear intuitions about why changing the post-NPI region or the tense would matter. Experiment 16 aimed to test which of these three things was actually responsible for the surprising findings in Experiment 15. We focused our attention on only the ORC stimuli, for two reasons: first, there are more previous studies using ORCs than SRCs, so the expectation that Experiment 15 should have revealed illusions is stronger, and second, the ORC findings from Experiment 15 are in some sense simpler than the SRC findings from Experiment 15 (in that we see no clear illusions at all, whereas the SRC conditions yielded a surprising illusion for *didn't* and something like an *anti*-illusion for *no*).

In Experiment 16 we tested ORC illusion sentences without extra clause, and compared sentences in the present perfect to sentences in the past perfect. The stimuli for Experiment 16 were exactly the stimuli from Experiment 15 (that is, each item had the exact same content words), except that we trimmed off the matrix clause at the beginning (*Mary thought that*), and we used only the ORC versions. Thus, if the reason we saw no clear illusions for embedded-*no* ORCs in Experiment 15 was because of the added matrix clause, we should see normal illusions in Experiment 16, since the matrix clause has been removed. If instead the reason we saw no clear illusions was because of the changes to the post-NPI region, Experiment 16 should replicate this finding, and again show no clear illusions. Finally, we used conditions with both the present perfect and past perfect, so that if the reason we saw no clear illusions was because of the tense, we should find evidence for this through the comparison of the two tenses. Importantly, those three factors were the only ways the stimuli from Experiment 15 were different from previous NPI illusion studies. So unless the surprising findings from Experiment 15 are due to random chance, one of

these three factors (or a combination of them) must be responsible.

5.4.2.1 Participants

64 US-based native speakers of English participated in this experiment. All participants provided informed consent and they received \$6 as compensation. We excluded workers who failed to provide a response within 2 seconds for 25% of fillers or more and workers whose judgments of filler trials did not reliably distinguish between grammatical and ungrammatical fillers, based on a chi-squared test. 0 workers were excluded based on these criteria, resulting in 64 participants in our analysis. The mean filler-trial accuracy of the included participants was 85%.

5.4.2.2 Materials

The experimental materials consisted of 20 sets of items that matched the 20 items from Experiment 15. Conditions A, B, C, and D were identical to the ORC items from Experiment 15 (conditions A-E in Table 5.4), except that we removed the matrix clause (e.g., *Mary thought that / Mary didn't think that*). Removing the matrix clause makes the grammatical *didn't* condition identical to the ungrammatical baseline condition (because these conditions only differed in the matrix clause), so what had been 5 conditions in Experiment 15 becomes 4 conditions in Experiment 16. We added 4 conditions in which the MC was in the present perfect instead of the past perfect. These manipulations resulted in the experimental conditions shown in Table 5.5. The fillers used in Experiment 16 were the same fillers that we had used in Experiment 15.

5.4.2.3 Procedure & Analysis

The procedure was identical to Experiments 12-15. The results were again analyzed with logistic mixed effects models, using the same analysis strategy as Experiments 12-15.

Past perfect	
A. Grammatical baseline	No authors [that the critics recommended] had ever ...
B. Embedded <i>no</i>	The authors [that no critics recommended] had ever ...
C. Embedded <i>didn't</i>	The authors [that the critics didn't recommend] had ever ...
D. Ungrammatical	The authors [that the critics recommended] had ever ...
...written a best-selling novel.	
Present perfect	
E. Grammatical baseline	No authors [that the critics recommended] have ever ...
F. Embedded <i>no</i>	The authors [that no critics recommended] have ever ...
G. Embedded <i>didn't</i>	The authors [that the critics didn't recommend] have ever ...
H. Ungrammatical	The authors [that the critics recommended] have ever ...
...written a best-selling novel.	

Table 5.5: Example stimuli for Experiment 16

5.4.2.4 Results

The results from this experiment are presented in Figure 5.5, which shows the proportion of “yes” responses given to each condition. An effect of grammaticality was observed ($\beta=6.56$, $SE=0.64$, $z=10.20$, $p<.001$), indicating that the grammatical baseline conditions were significantly more likely to be judged acceptable than the ungrammatical baseline conditions, averaging over the tense manipulation. An effect of embedded-*no* was observed ($\beta=1.64$, $SE=0.34$, $z=4.80$, $p<.001$), replicating the standard illusion effect for negative quantifiers, averaging over the tense manipulation. Pairwise comparisons revealed significant illusion effects for both the past perfect ($\beta=1.69$, $SE=0.44$, $z=3.86$, $p<.001$) and present perfect ($\beta=1.59$, $SE=0.43$, $z=3.68$, $p<.001$) embedded-*no* sentences.

No effect of embedded-*didn't* was observed ($\beta=0.57$, $SE=0.33$, $z=1.71$, $p=.32$). We did not find a statistically significant interaction between this factor and the tense manipulation ($\beta=0.56$, $SE=0.57$, $z=0.98$, $p=.33$), and pairwise comparisons revealed that illusion effects were not significant for either the past perfect ($\beta=0.84$, $SE=0.44$, $z=1.93$, $p=.05$) or present perfect ($\beta=0.29$, $SE=0.44$, $z=0.65$, $p=.51$) embedded-*didn't* sentences. The comparison of embedded-*didn't* and embedded-*no* revealed a significant effect of the type of embedded negation ($\beta=1.07$, $SE=0.32$, $z=3.39$, $p=.004$), indicating that the embedded-*no* conditions were significantly more likely to be judged acceptable than the embedded-*didn't* conditions. This effect did not significantly interact with the tense manipulation ($\beta=0.46$, $SE=0.50$, $z=0.91$, $p=0.36$), and pairwise comparisons revealed significant differences between the two forms of negation for both the

past perfect ($\beta=0.84$, $SE=0.39$, $z=2.16$, $p=0.03$) and present perfect ($\beta=1.30$, $SE=0.42$, $z=3.14$, $p=0.002$) conditions.

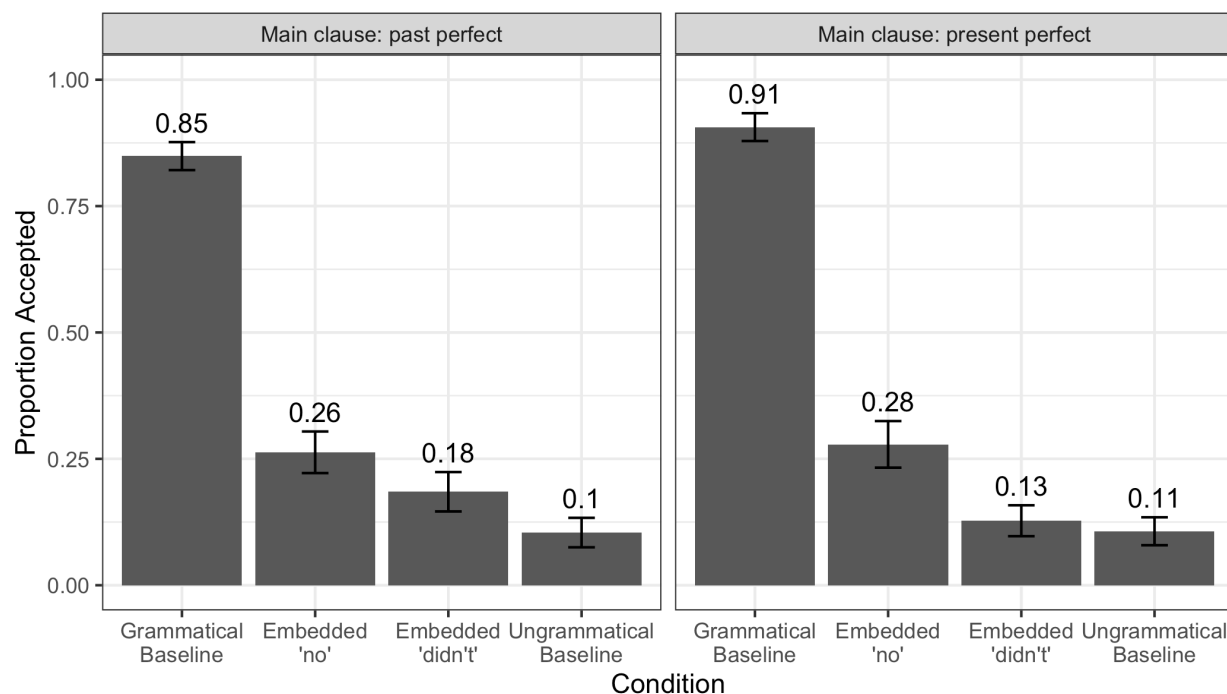


Figure 5.5: Mean percentage of ‘YES’ responses for the experimental conditions in Experiment 16. Error bars indicate standard error of the mean across subjects.

5.4.2.5 Discussion

Experiment 16 was designed to identify the cause of the surprising pattern observed in Experiment 15 with respect to NPI illusions with ORCs — that is, a lack of illusions for embedded *no*, possibly due to inflated acceptance of the ungrammatical baseline. Experiment 16 teased apart the three properties that made Experiment 15 different from other NPI illusion experiments: the added clause, the shorter spillover region, and the tense of the MC. Since Experiment 16 revealed illusions for embedded-*no* for both past perfect and present perfect sentences, and Experiment 16 had the same short spillover region as Experiment 15, we can be confident that this was not the main issue. It appears that at least some of the problem with Experiment 15 was the added clause. This is unsurprising, since we had identified at

least two ways this added clause might distort judgments: a shift toward guessing because of the added complexity, and a penalty for sentences with *no* because of competition with *didn't+any*.

This experiment also revealed, to our surprise, a trend toward illusions for embedded *didn't* when the MC was in the past perfect, though this trend was not significant, with a p-value of 0.05. These findings are of course perfectly compatible with an explanation in which this trend is purely due to random chance. However, it is worth noting that prior to this experiment, a pilot experiment which tested only the past-perfect conditions, revealed qualitatively the same pattern.⁴⁴ This makes us somewhat more inclined to believe that this trend is not purely due to random sampling. Nonetheless, replication of this effect is obviously needed. We discuss the potential interpretation and implications of this finding in section 6.2.7.

5.5 Conclusion

Here we presented the results from five experiments which in some ways clarify the profile of the NPI illusion, and in some ways complicate the empirical picture. The key findings are as follows. First, Experiment 12 attempted to induce illusions for embedded *didn't* by guaranteeing the generation of NPI-licensing scalar inferences prior to the NPI, as is predicted to be critical to the illusion under the scalar alternatives hypothesis. We found no effect on illusion rates. Experiment 13 aimed to clarify the distance effect and the non-effect of NPI identity (*ever* versus *any*) on illusions. While we did find clear illusions for both *ever* and *any* when positioned close to the RC, we also found illusions for *any* in a (nearly) sentence-final position. Experiment 14 followed up on this effect but did not clarify matters, due to problems with the baseline conditions, possibly because of free choice readings for *any*. Experiment 15 tested the predictions of the noisy channel hypothesis, but did not clearly support or challenge the predictions of the hypothesis due to a surprising *lack* of illusions for embedded-*no*. Finally, Experiment 16 aimed to identify the key property of the Experiment 15 stimuli that caused the surprising results, and revealed, to our surprise, an trend toward an illusion for embedded-*didn't* when the MC was in the past perfect, but

⁴⁴Specifically, this pilot experiment found 10% acceptance for the ungrammatical baseline, 20% for the embedded *didn't* condition, 24% for the embedded *no* condition, and 88% for the grammatical baseline. The experiment had 16 subjects.

not when it was in the present perfect.

These findings are, on the whole, puzzling for any account of the NPI illusion that we considered in section 5.1.2. A few findings, however, are worth noting specifically because of their consequences for particular candidate explanations. First, negative scalar alternatives are not sufficient to trigger NPI illusions (Experiment 12). This is a failed prediction of the scalar alternatives hypothesis, and calls into question the viability of this explanation. Second, quantifiers do not appear to be necessary to trigger NPI illusions (Experiment 16). This is a failed prediction of the quantifier scope hypothesis. Thus, the two hypotheses which appeared most plausible in light of the findings from Chapter 3 and Chapter 4 are no longer clearly able to account for the profile of the illusion. It is worth noting that the two earliest proposals for the cause of the illusion, the cue-based retrieval account and the pragmatic rescuing account, do not fare any better in light of these data. The noisy channel hypothesis remains in principle plausible, since the experiment designed to test its predictions (Experiment 15) was more or less uninterpretable. In Chapter 6 we discuss the theoretical landscape in greater detail, and consider the conclusions that can be drawn from this investigation.

Chapter 6 NPI illusions: general discussion

Chapters 3 through 5 presented a series of experiments testing the profile of the NPI illusion. Here we review these findings and their implications for possible explanations of the illusion. The aim of this chapter is to identify both the empirical contributions of the present work and the theoretical progress that can be made as a result. As will become clear, a number of hypotheses that appeared plausible at the outset are more or less unable to account for the NPI illusion's rather specific profile. We consider adjustments to these hypotheses, as well as a back-to-the-drawing-board approach in which possible paths forward are more firmly rooted in what is known about the grammar of NPIs.

6.1 Summary of findings

Here we provide example stimuli and a brief description of key findings for each of the 16 experiments presented in Chapters 3, 4, and 5. In general, we provide only the ungrammatical baseline and embedded-negation examples, since these are the stimuli that we look to to determine if illusions arise.

6.1.1 Experiments 1 & 2

Experiment 2 directly compared illusion rates in a speeded acceptability task for two forms of embedded negation, as in (83a) and (83b). Experiment 1 provided verification that the grammatical status of stimuli used in Experiment 1 was as expected (i.e., (83a), (83b), and (83c) are all ungrammatical, and were all given low ratings in Experiment 1).

- (83) a. * The authors [that **no** critics recommended] have **ever** received acknowledgement for a best-selling novel.

- b. * The authors [that the critics did **not** recommend] have **ever** received acknowledgement for a best-selling novel.
- c. * The authors [that the critics recommended] have **ever** received acknowledgement for a best-selling novel.

Experiment 2 revealed reliable illusions for sentences with embedded *no*, but not for sentences with embedded *not*, and found that acceptance rates for sentences with embedded *no* and embedded *not* were reliably different from one another.

6.1.2 Experiment 3

Experiment 3 replicated the basic findings of Experiment 2 with slightly different stimuli and a between-subjects design for the type of embedded negation.

- (84)
- a. * The authors [that **no** critics have recommended in their reviews] have **ever** received acknowledgement for a best-selling novel.
 - b. * The authors [that the critics **haven't** recommended in their reviews] have **ever** received acknowledgement for a best-selling novel.
 - c. * The authors [that the critics have recommended in their reviews] have **ever** received acknowledgement for a best-selling novel.

We again found reliable illusions for sentences with embedded *no* but not for sentences with embedded non-quantificational negation (*haven't*), and found reliable differences between the two forms of negation.

6.1.3 Experiment 4

Experiment 4 tested the interpretations comprehenders infer for NPI illusion sentences, and for the constructions that give rise to illusions but without the NPI, as in (86). Each trial presented the sentence in RSVP (as in speeded acceptability tasks) and included both an untimed, binary acceptability judgment, and an untimed, binary comprehension question, as in (87).

- (85) a. * The authors [that **no** critics recommended] have **ever** received acknowledgement for a best-selling novel.
- b. * The authors [that the critics **didn't** recommend] have **ever** received acknowledgement for a best-selling novel.
- c. * The authors [that the critics recommended] have **ever** received acknowledgement for a best-selling novel.
- (86) a. * The authors [that **no** critics recommended] have received acknowledgement for a best-selling novel.
- b. * The authors [that the critics **didn't** recommend] have received acknowledgement for a best-selling novel.
- c. * The authors [that the critics recommended] have received acknowledgement for a best-selling novel.
- (87) Have the authors received acknowledgement for a novel?

The acceptability judgment data from Experiment 4 once again revealed reliable NPI illusions when the form of embedded negation was *no* but not when it was non-quantificational negation (*didn't*). The comprehension question data revealed substantial rates of globally negative interpretations (“no” answers to comprehension questions) for illusion sentences as in (85a) (77%), but negative interpretations were much rarer for embedded-*no* sentences without *ever* as in (86a) (15%). We also found that, among sentences like (85a), illusions are not unique to trials that are interpreted negatively — acceptance rates are elevated relative to the ungrammatical baseline (85c), regardless of whether the trial receives a “yes” or “no” answer to the comprehension question.

6.1.4 Experiments 5 & 6

Experiment 6 tested whether illusions are possible for non-quantificational negation when other aspects of the RC meaning (i.e. the availability of scalar alternatives) are matched to the meaning of the embedded-*no* sentences that typically give rise to clear illusions. Experiment 5 provided verification that the gram-

mathematical status of stimuli used in Experiment 6 was as expected. These experiments, unlike previous experiments, used SRCs.

- (88)
- a. * The critics [that have recommended **no** authors of alternative genres] have **ever** objected to mainstream literary trends.
 - b. * The critics [that **haven't** recommended the authors of alternative genres] have **ever** objected to mainstream literary trends.
 - c. * The critics [that **haven't** recommended **any** authors of alternative genres] have **ever** objected to mainstream literary trends.
 - d. * The critics [that have recommended the authors of alternative genres] have **ever** objected to mainstream literary trends.

We again found reliable illusions for embedded-*no* sentences such as (88a), though the illusion effect size was somewhat smaller than previous experiments. We again found a contrast between *no* and *haven't* ((88a) versus (88b)), though the effect size was again smaller than previous experiments. We found reliable illusions for embedded-*haven't...any* sentences such as (88c), but again with a smaller effect size. This numerically-intermediate condition could not be statistically distinguished from either (88a) or (88b). We also found, for the first time, reliable illusions for sentences with embedded *haven't* such as (88b), though, again, the effect size was smaller than previously reported illusions.

6.1.5 Experiment 7

Experiment 7 compared illusion rates for NPI illusion stimuli with and without a prepositional phrase intervening between the embedded negative word *no* and the NPI. The prepositional phrases were always inside the RC.

- (89)
- a. * The surgeons [that **no** patients have consulted about the operation] have **ever** expressed dissatisfaction with the hospital staff.
 - b. * The surgeons [that the patients have consulted about the operation] have **ever** expressed dissatisfaction with the hospital staff.

- (90) a. * The surgeons [that **no** patients have consulted] have **ever** expressed dissatisfaction with the hospital staff.
- b. * The surgeons [that the patients have consulted] have **ever** expressed dissatisfaction with the hospital staff.

We found reliable illusions for both sentences with intervening PPs and sentences without intervening PPs, and we did not find a statistically significant interaction which would suggest a reduction in the illusion.

6.1.6 Experiments 8 & 9

Experiment 9 compared the influence of intervening material inside and outside the RC. Both (92a) and (93a) have a longer *no*-to-NPI distance than (91a). But only (93a) has a longer RC-to-NPI distance than (91a). The experiment thus dissociates two possible analyses of the distance effect. Experiment 8 provided verification that the grammatical status of stimuli used in Experiment 9 was as expected.

- (91) a. * The surgeons [that **no** patients trusted] have **ever** prescribed experimental treatments.
- b. * The surgeons [that the patients trusted] have **ever** prescribed experimental treatments.
- (92) a. * The surgeons [that **no** patients trusted] have healed **any** injuries with experimental treatments.
- b. * The surgeons [that the patients trusted] have healed **any** injuries with experimental treatments.
- (93) a. * The surgeons [that **no** patients trusted to heal injuries] have **ever** prescribed experimental treatments.
- b. * The surgeons [that the patients trusted to heal injuries] have **ever** prescribed experimental treatments.

Experiment 9 found reliable illusions for both sentences with no added material like (91a) and sentences with added material inside the RC like (93a) but not for sentences with added material between

the RC and the NPI like (92a). The statistical analyses also indicated significant interactions in the illusion effects, suggesting that the reduction in the illusion for sentences like (92a) was reliable.

6.1.7 Experiment 10

Experiment 10 tested whether illusions are more likely for embedded quantifiers than non-quantificational forms of negation when the quantifier is lower in the RC than non-quantificational negation. This required the use of SRCs. The experiment de-confounds the licensor effect from possible structural distance effects.

- (94)
- a. * The critics [that have recommended **very few** authors of alternative genres] have **ever** objected to mainstream literary trends.
 - b. * The critics [that haven't recommended **very few** authors of alternative genres] have **ever** objected to mainstream literary trends.
 - c. * The critics [that have recommended the authors of alternative genres] have **ever** objected to mainstream literary trends.

Experiment 10 again found reliable illusions for quantificational licensors (*very few*) but not for non-quantificational licensors (*haven't*), and found reliable differences between these two conditions. We take these findings to suggest that previously observed licensor effects are not a consequence of differences in the structural position of the licensors being compared.

6.1.8 Experiment 11

Experiment 11 tested whether the key element causing the disappearance of illusions for sentences like (92a) (see Experiments 8 & 9) is the addition of a second intervening word between the RC and the NPI or the addition of an intervening lexical verb between the RC and the NPI. We orthogonally manipulated the number of intervening words (one as in (95a) and (96a), or two as in (97a) and (98a)) and the presence or absence of an intervening verb (present in (96a) and (98a), absent in (95a) and (97a)) to test which of these was responsible.

- (95) a. * The surgeons [that **no** patients trusted] have **ever** shown appreciation for the hospital staff.
- b. * The surgeons [that the patients trusted] have **ever** shown appreciation for the hospital staff.
- (96) a. * The surgeons [that **no** patients trusted] showed **any** appreciation for the hospital staff.
- b. * The surgeons [that the patients trusted] showed **any** appreciation for the hospital staff.
- (97) a. * The surgeons [that **no** patients trusted] would have **ever** shown appreciation for the hospital staff.
- b. * The surgeons [that the patients trusted] would have **ever** shown appreciation for the hospital staff.
- (98) a. * The surgeons [that **no** patients trusted] have shown **any** appreciation for the hospital staff.
- b. * The surgeons [that the patients trusted] have shown **any** appreciation for the hospital staff.

This experiment revealed reliable illusions for (95), (96), and (97), but not (98). This pattern suggests that it is the combination of multiple words and a verb that causes the illusion to disappear. However, note that the critical interaction that would demonstrate this was not significant. Another important takeaway from Experiment 11 is that illusions are possible for both *ever* and *any* when positioned sufficiently close to the RC.

6.1.9 Experiment 12

Experiment 12 tested again whether illusions can be induced for embedded non-quantificational negation when this form of negation is used in a way that evokes scalar alternatives. This experiment attempts to conceptually replicate the key findings of Experiment 6, which were not especially robust. The experiment therefore compared quantificational and non-quantificational negation in sentences where they were paired with additional NPIs, as in (99), and sentences where they were not, as in (100).

- (99) a. * The authors [that **no** critics have recommended in **any** of their reviews] have **ever** received acknowledgement for a best-selling novel.
- b. * The authors [that the critics **haven't** recommended in **any** of their reviews] have **ever** received acknowledgement for a best-selling novel.
- c. * The authors [that the critics have recommended in **any** of their reviews] have **ever** received acknowledgement for a best-selling novel.
- (100) a. * The authors [that **no** critics have recommended in their reviews] have **ever** received acknowledgement for a best-selling novel.
- b. * The authors [that the critics **haven't** recommended in their reviews] have **ever** received acknowledgement for a best-selling novel.
- c. * The authors [that the critics have recommended in their reviews] have **ever** received acknowledgement for a best-selling novel.

Experiment 12 revealed illusions for embedded *no* in both types of sentences, and did not identify statistically reliable illusions for embedded *haven't* in either type of sentence. The contrast between embedded *no* and embedded *haven't* was significant in both cases. Thus we do not find that scalar RC meanings are sufficient to trigger NPI illusions for non-quantificational forms of negation.

6.1.10 Experiment 13

In Experiment 13 we aimed to replicate the finding from Experiment 11 that showed that illusions are possible for *any*. We did this with prepositional phrases containing *any* like *in any way* and *at any time*, which can be positioned both early in the MC, in the same position as *ever*, and later in the MC, sentence-finally.

- (101) a. * The surgeons [that **no** patients consulted] have **ever** suggested unnecessary operations.
- b. * The surgeons [that the patients consulted] have **ever** suggested unnecessary operations.
- (102) a. * The surgeons [that **no** patients consulted] have, **at any time**, suggested unnecessary operations.

- b. * The surgeons [that the patients consulted] have, **at any time**, suggested unnecessary operations.
- (103)
- a. * The surgeons [that **no** patients consulted] have suggested unnecessary operations **at any time**.
 - b. * The surgeons [that the patients consulted] have suggested unnecessary operations **at any time**.

We found reliable illusions for both *ever* and *any* when positioned close to the RC as in (101) and (102). Surprisingly, we also found reliable illusions for *any* positioned sentence-finally as in (103). This is not consistent with the previously demonstrated distance effect.

6.1.11 Experiment 14

In Experiment 14 we tested whether the surprising illusion for sentence-final *any*-phrases replicates with a different set of stimuli. At the same time, we aimed to replicate Parker & Phillips's (2016) finding that NPI illusions disappear for *ever* positioned later after the RC, but not sentence-finally, and to determine whether this pattern extended to *any*-phrases. Thus we tested illusions for both *any* and *ever* positioned soon after the RC (see (104a) and (105a)), both *any* and *ever* positioned later in the sentence (see (106a) and (107a)), and only *any* positioned sentence-finally (see (108a)).

- (104)
- a. * The journalists [that **no** editors recommended for the assignment] have **in any way** argued that the readers would understand the complicated situation.
 - b. * The journalists [that the editors recommended for the assignment] have **in any way** argued that the readers would understand the complicated situation.
- (105)
- a. * The journalists [that **no** editors recommended for the assignment] have **ever** argued that the readers would understand the complicated situation.
 - b. * The journalists [that the editors recommended for the assignment] have **ever** argued that the readers would understand the complicated situation.

- (106) a. * The journalists [that **no** editors recommended for the assignment] have argued that the readers would **in any way** understand the complicated situation.
- b. * The journalists [that the editors recommended for the assignment] have argued that the readers would **in any way** understand the complicated situation.
- (107) a. * The journalists [that **no** editors recommended for the assignment] have argued that the readers would **ever** understand the complicated situation.
- b. * The journalists [that the editors recommended for the assignment] have argued that the readers would **ever** understand the complicated situation.
- (108) a. * The journalists [that **no** editors recommended for the assignment] have argued that the readers would understand the complicated situation **in any way**.
- b. * The journalists [that the editors recommended for the assignment] have argued that the readers would understand the complicated situation **in any way**.

The sentences with *ever* replicated findings from Parker & Phillips 2016. We found reliable illusions when *ever* was close to the RC but not when it was positioned later in the sentence. The findings for *any* were uninterpretable due to baseline problems — sentences like (106b) and (108b), which are ungrammatical baselines, were accepted in close to half of trials. We suspect this was due to free choice readings for *any* following a modal.

6.1.12 Experiment 15

Experiment 15 tested whether the licenser effect can be attributed to differences between *no* and *didn't* with respect to the possible positions to which the licenser could be moved in the comprehender's mental representation, which would result in a licensed NPI. Thus we embedded sentences with embedded *no* and embedded *didn't* under a neg-raising verb, so that both forms of negation could in principle be “shuffled” to an earlier position in the sentence from which they would license the NPI (i.e. *Mary thought that no authors...* or *Mary didn't think that the authors...*). We tested both ORCs and SRCs.

- (109) a. * Mary thought that the authors [that **no** critics recommended] had **ever** written a best-selling novel.
- b. * Mary thought that the authors [that the critics **didn't** recommend] had **ever** written a best-selling novel.
- c. * Mary thought that the authors [that the critics recommended] had **ever** written a best-selling novel.
- (110) a. * Mary thought that the authors [that admired **no** critics] had **ever** written a best-selling novel.
- b. * Mary thought that the authors [that **didn't** admire the critics] had **ever** written a best-selling novel.
- c. * Mary thought that the authors [that admired the critics] had **ever** written a best-selling novel.

The findings from this experiment were largely uninterpretable due to baseline problems — both ungrammatical baseline sentences were accepted more than expected. In addition, neither embedded-*no* condition showed clear illusions (and in fact (110a) was accepted *less often* than the ungrammatical baseline (110c)). We do see a trend toward illusions for embedded *not* in SRCs only, but we are hesitant to draw any conclusions from this.

6.1.13 Experiment 16

Experiment 16 aimed to identify the problem with the stimuli used in Experiment 15, but revealed a surprising new effect. The ORC items from Experiment 15 were tested, without the additional neg-raising verb and subject. Experiment 16 directly compared these items, which used the past perfect in the MC as in (111), to equivalent sentences which used the present perfect in the MC as in (112), as all prior studies described here had done.

- (111) a. * The authors [that **no** critics recommended] had **ever** written a best-selling novel.
- b. * The authors [that the critics **didn't** recommend] had **ever** written a best-selling novel.

- c. * The authors [that the critics recommended] had **ever** written a best-selling novel.
- (112)
- a. * The authors [that **no** critics recommended] have **ever** written a best-selling novel.
 - b. * The authors [that the critics **didn't** recommend] have **ever** written a best-selling novel.
 - c. * The authors [that the critics recommended] have **ever** written a best-selling novel.

This experiment revealed, as expected, reliable illusions for embedded *no* for sentences in both the present perfect and the past perfect. It also revealed a surprising trend toward illusions for embedded *didn't* only for sentences in the past perfect. This trend was not significant, with a p-value of 0.05. Note, however, that a small pilot study of just the past perfect sentences found a qualitatively similar pattern.

6.2 Key generalizations and next steps

There are a few generalizations that can be made based on the pattern of findings described here. The first is simply that the NPI illusion is a very robust effect. In 13 speeded acceptability studies that included a condition with an embedded quantifier and an NPI close to the RC, statistically reliable illusions were observed in all but one (Experiment 15, which had other issues). The average acceptance rate for the illusion condition and the ungrammatical baseline condition for each of these 13 experiments is shown in Figure 6.1.

We additionally computed an estimate of the effect size across these 13 experiments, using a Bayesian mixed effects model with random intercepts and slopes for both items and subjects, nested within experiment. This model was fit to a dataset of more than 16,000 trials across three conditions (ungrammatical baseline, illusion, and grammatical baseline). The illusion effect size (i.e., the contrast between the ungrammatical baseline and the illusion condition) was estimated at $\beta=1.06$ (SE=.23), which can be understood as an odds ratio of 2.89. That is, the odds of acceptance for illusion sentences is 2.89 times the odds of acceptance for ungrammatical baseline sentences. Future explorations of the NPI illusion could use this estimate for prospective power analyses.

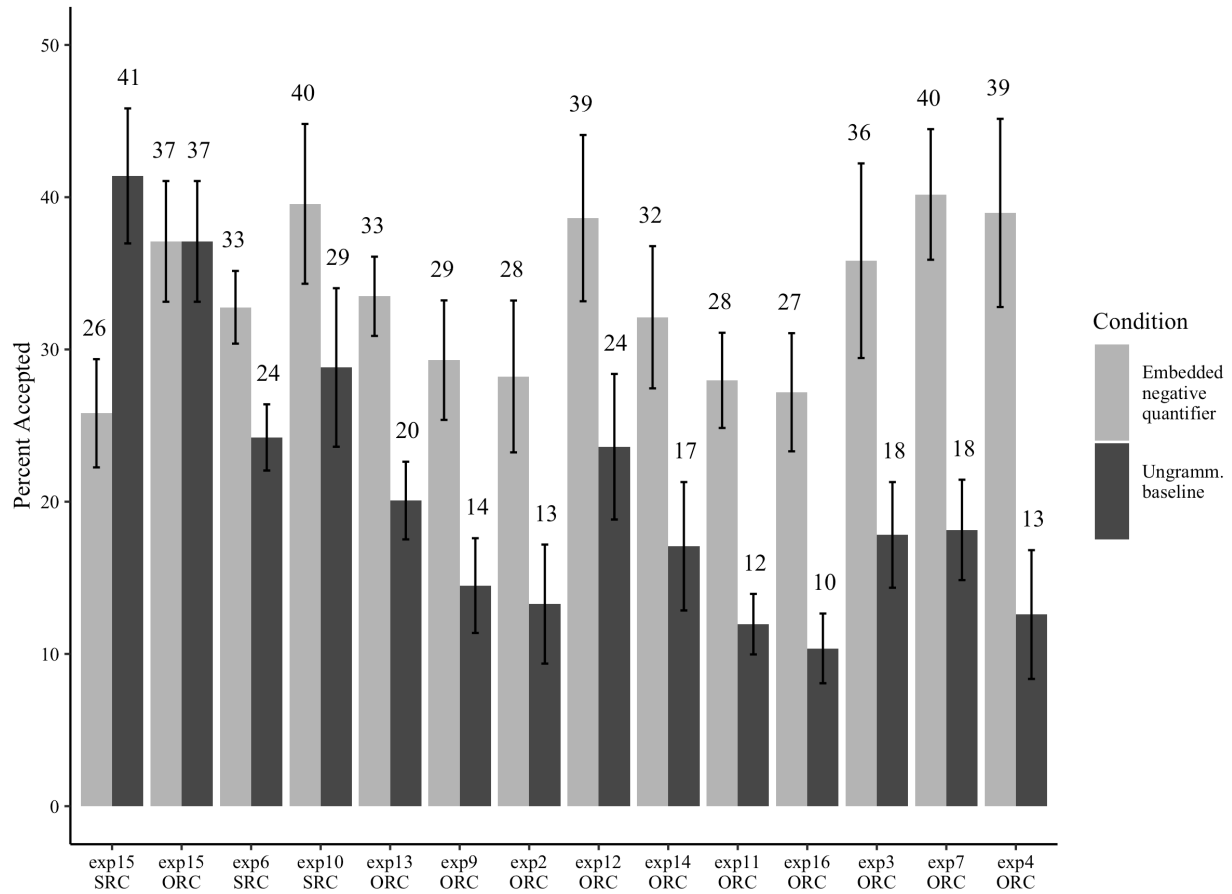


Figure 6.1: Mean percentage of ‘YES’ responses for the illusion and ungrammatical baseline conditions in all speeded acceptability experiments, ordered by effect size. Here we show only the embedded-quantifier, short-distance conditions. For experiments in which a manipulation did not matter to illusions (e.g. inserting a PP, as in Experiment 7), we collapse the data. Error bars indicate standard error of the mean across subjects.

6.2.1 The licenser effect

One of the key generalizations we have demonstrated here is the contrast between licensors like *no* and *very few*, which regularly yield clear illusions, and licensors like *not*, *didn’t*, and *haven’t*, which regularly do not yield clear illusions. We have understood this to be a contrast between quantificational and non-quantificational licensors, though in principle, there may be other ways to distinguish between these two types of licenser. As shown in Figure 6.2, sentences with these embedded licensors yield different acceptance rates (in the same direction) in 7 out of the 8 experiments that compared them (again, the only exception is Experiment 15, which had independent problems)

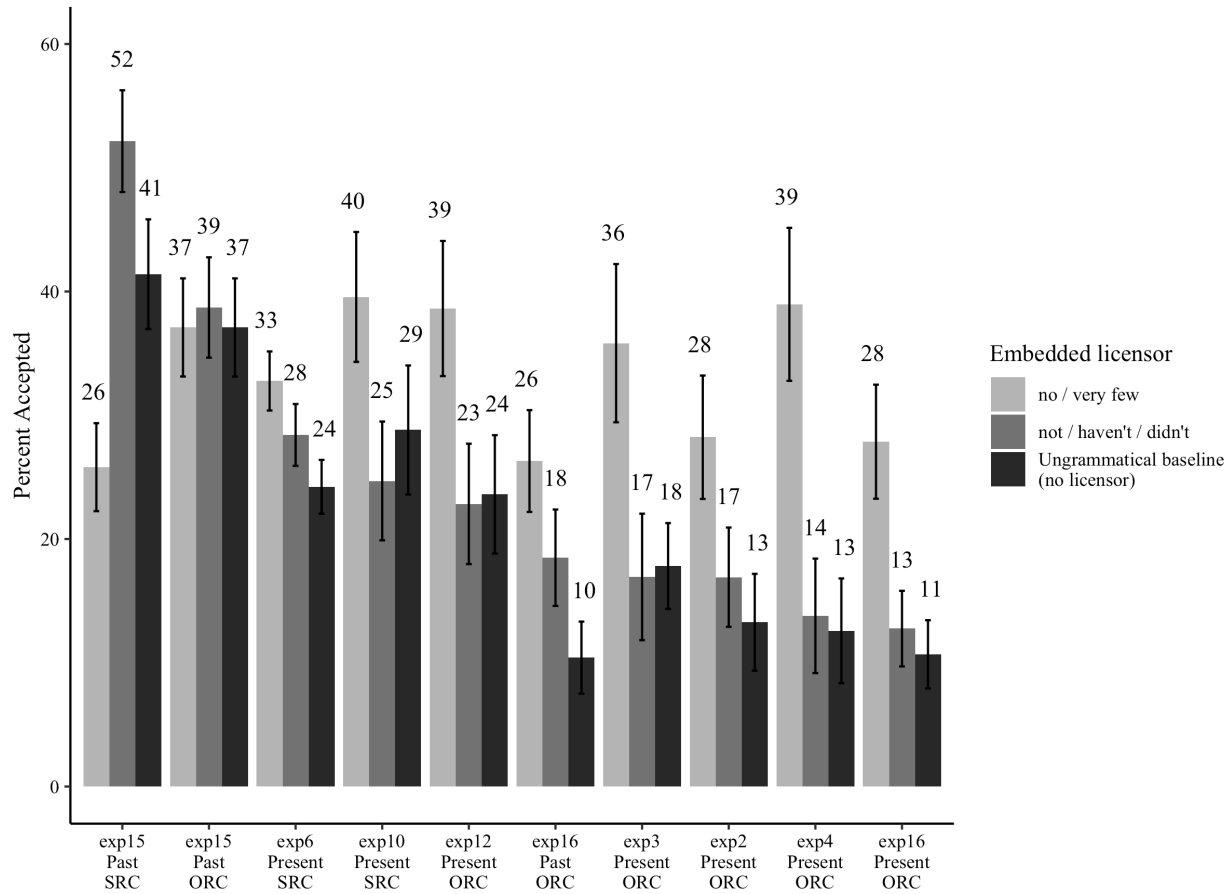


Figure 6.2: Mean percentage of ‘YES’ responses for the embedded *no* and embedded *not* conditions in all speeded acceptability experiments that compared them, ordered by effect size. For experiments in which we also tested non-quantificational licensors paired with embedded NPIs, we removed this condition. Error bars indicate standard error of the mean across subjects. MC tense and relative clause type are noted for each experiment.

This contrast does not map onto any distinction in these licensors’ ability to function as NPI licensors. For example, there is much work demonstrating that different NPIs (e.g. strong and weak NPIs) have different licensing requirements — both *no* and *at most 5* can license *anyone*, but only *no* can license *in weeks* (examples from Gajewski 2011, though the weak/strong distinction pre-dates this).

- (113) a. No doctor has seen anyone.
b. At most 5 doctors have seen anyone.
- (114) a. No doctor has seen Mary in weeks.

- b. * At most 5 doctors have seen Mary in weeks.

(Gajewski 2011:114)

Although differences between licensors clearly exist in the theoretical literature, the licensor effect observed for illusions is not obviously related to this distinction. All of the licensors tested here⁴⁵ can license both strong and weak NPIs, and, moreover all of the NPIs tested here are weak NPIs, which can be licensed by the broadest set of licensors. One distinction that has been made between *no/few* and *not/n't* in the literature on the licensing of different varieties of NPIs is in their ability to license so-called “superstrong” NPIs like *one bit*.

- (115)
- a. * Few people were one bit happy about these facts.
 - b. * No linguist was one bit happy about these facts.
 - c. The men weren't one bit happy about these facts.

(Zwarts 1998:190)

Note, however, that this contrast cuts in the opposite direction of the observed NPI illusion contrast — *n't* seems to license more NPIs than *no*, but *no* seems to cause more NPI illusions than *n't*. Thus it is not possible to reduce the licensor effect for NPI illusions to any contrast in these licensors' ability to function as licensors. There is no sense in which *not* is a “less good” NPI licensor than *no* — it is arguably the best NPI licensor there is, and yet it does not yield illusions.

A separate issue is that our “quantificational” and “non-quantificational” labels for (116a) and (116b), respectively, are not entirely uncontroversial.

- (116)
- a. * The authors [that no critics have recommended in their reviews] have ...
 - b. * The authors [that the critics haven't recommended in their reviews] have ...

Some analyses of *the* consider it to be quantificational. The argument for a quantificational treatment of definite descriptions is based primarily on observed interactions between definite descriptions and the

⁴⁵In fact, the exception to this is *(very) few*, which has a somewhat controversial status as a licensor of strong NPIs. Zwarts 1998 reports that *few* cannot license strong NPIs, but Hoeksema 2005 and Rullmann 2003 argue otherwise. This is somewhat beside the point, since the disagreement concerns NPIs like *in years* and *until*, which were not tested here. Moreover, even if the resolution of this disagreement were that *few* does not license all NPIs, this would not explain much regarding the licensor contrast for NPI illusions, which has been demonstrated many times for licensors besides *few*.

scope of negation as in the two readings of (117), an observation which is typically attributed to Russell 1905.⁴⁶

(117) The king of France is not bald.

If the two meanings of (117) are to be attributed to scope ambiguity and this is taken as evidence for a quantificational meaning for *the*⁴⁷, then both (116a) and (116b) involve quantifiers in the RC. However, this may not be a deep problem. Even if *the* is quantificational, it's clearly not a quantificational NPI licenser. Rather, the NPI licenser in (116b) is *haven't*, and we know of no argument that *haven't* is quantificational. Thus it may not matter, for our purposes, whether *the* is quantificational.

We have glossed the contrast between *no* and *not* as a contrast between quantificational licensers and non-quantificational licensers, but this of course does not tell us what it is about quantificational licensers that makes them so likely to cause illusions (or, what it is about non-quantificational licensers that makes them so immune to illusions). We have considered three possible explanatory factors: differences between quantificational and non-quantificational licensers in the kind of inferences (scalar or non-scalar) they invoke⁴⁸; differences between quantificational and non-quantificational licensers in their scope-taking properties⁴⁹; and differences between quantificational and non-quantificational licensers in the probability of an edit to the string that puts them in a position to take scope over an NPI. These characterizations of the effect are closely related to three mechanistic hypotheses we consider for the NPI, and so we discuss their plausibility in the sections dedicated to those hypotheses (sections 6.3.3, 6.3.4, and 6.3.5, respectively). Here we focus on the empirical generalization, and apparent exceptions to it.

⁴⁶Though Glanzberg 2009 clarifies that although Russell made the observation that the sentence has two meanings, he did not explicitly make the argument that the sentence is structurally ambiguous, nor that *the* must therefore be a quantifier.

⁴⁷This is not actually a very compelling argument for a quantificational analysis of *the*, as the two meanings of (117) can be explained in terms of the use of meta-linguistic negation (Glanzberg 2009).

⁴⁸Here it could in principle matter if *the* is quantificational. We have argued that negatively-ordered scalar alternatives must exist at the clause level, which does not necessarily mean that the alternatives have to have come from the negative word itself. If containing a quantifier leads to scalar inferences and the presence of negation orders those inferences, then we may end up with the same alternatives for (116a) and (116b). However, merely containing quantificational expression may not be the critical factor for whether scalar inferences arise, since, *the* is not typically assumed to be part of a Horn scale, as is assumed for *no* (Horn 1972). It is therefore not clear that quantificational analyses of *the* necessarily lead to the prediction of parallel alternative sets for (116a) and (116b) under the scalar alternatives hypothesis.

⁴⁹Here again the analysis of (117) becomes relevant. If the purported scope ambiguity is to be analyzed through covert movement of *not*, then the scope miscalculation hypothesis's assumption that quantificational negation can move and non-quantificational negation cannot move is undermined. However, as we have previously noted, a scope ambiguity analysis of the two meanings of (117) may not be the best analysis.

There are essentially two distinct claims we might make about non-quantificational licensors: that they yield fewer illusions than quantificational licensors, and that they yield effectively no illusions whatsoever. The evidence for the first claim is abundant. The only experiment in which we observe acceptance rates for non-quantificational licensors that are equal to or greater than those for quantificational licensors is Experiment 15. Recall that one possible cause of the surprising data patterns observed in this experiment was the use of embedding under neg-raising verbs. Specifically, it is possible that sentences containing *no* (e.g. *Mary thought that the authors that no critics recommended had...*) were judged less acceptable because of a preference for a *didn't+any* construction (e.g. *Mary didn't think that the authors that any critics recommended had...*). If this is the right analysis, it would explain why acceptance rates for embedded *no* sentences were pushed down, potentially to rates equal to or lower than acceptance rates for embedded *not*. We additionally speculated that general sentence complexity effects may have encouraged guessing, pushing acceptance rates for all ungrammatical sentences up. Follow up experiments are clearly needed here. One path which we think would be valuable would be to test acceptability rates for the Experiment 15 stimuli, but with the NPI removed. This would reveal if there is in fact a penalty for sentences with *no* in these contexts. It might additionally offer at least suggestive evidence for general sentence complexity effects. However, since removing the NPIs results in full grammaticality for all conditions, we would not be able to tease apart complexity-driven penalties in acceptability from complexity-driven guessing.

The evidence for the second claim, that illusions for non-quantificational licensors do not arise, is more mixed. It is of course difficult to make a statistical argument for true equivalence between embedded *not* and ungrammatical baseline conditions, given the nature of null hypothesis significance testing. Options like equivalence tests (Lakens, Scheel, & Isager 2018) and fully Bayesian analyses exist, but these require specification of an effect size range that is “practically” equivalent to zero, based on a Smallest Effect Size of Interest (SESOI) or Region of Practical Equivalence (ROPE). These are rather difficult to define. One might attempt to use the illusion effect size for embedded *no* as a starting point, perhaps defining the SESOI as an effect one half the size of typical illusions. But this effectively brings us back to our first claim about the licensor effect — that quantificational and non-quantificational licensors differ.

We may be better off simply quantifying this effect size, rather than arbitrarily choosing some fraction of the *no* illusion effect and checking if the *not* illusion is smaller.

Of course, the extent to which it matters whether illusions for embedded *not* are truly equal to zero depends on the hypothesis being considered. For example, under the scope miscalculation hypothesis, illusions should be categorically impossible for non-quantificational licensors, whereas for either the scalar alternatives or noisy channel accounts, small but non-zero illusion rates for non-quantificational licensors are not surprising. In fact, even under the scope miscalculation hypothesis it may be possible to explain away very small illusion rates as a consequence of the speeded acceptability task — surely participants in these experiments occasionally miss part of a sentence, due to looking away from the RSVP-presented sentence or not paying attention. But they must report a decision anyway. One might imagine that a comprehender who misses most of the sentence (and so has no real analysis) but caught both *didn't* and *ever* may be slightly more inclined to guess that the sentence was acceptable than a comprehender who similarly missed a large part of the sentence but caught only *ever*. Thus the question of whether illusion effects for non-quantificational licensors are truly zero may be theoretically inconsequential.

That said, if there is some factor that causes illusions to appear (with a larger effect size than the near-zero effect that is typically observed, but not necessarily as large as illusions for embedded *no*) for non-quantificational licensors, this may be worth exploring. The surprising findings from Experiment 16 point us in this direction. While we have observed a near-zero illusion rate for non-quantificational licensors many times, these experiments almost all used ORCs and a present perfect MC.⁵⁰ Experiments using SRCs yielded similar outcomes, though Experiment 6 demonstrated statistically reliable (but very small) illusions for non-quantificational licensors. The finding in Experiment 16 that past perfect MCs may yield illusions for embedded non-quantificational licensors is thus a significant development. We discuss this finding in detail in section 6.2.7.

⁵⁰Note that the repeated use of highly similar stimuli has some value. By having relatively little variability in the structures used from one experiment to the next, we are able to make comparisons across experiments, and we can conduct each experiment with reasonable confidence that the sanity check manipulations will work out. These are all useful things. However, this means that if the profile of the illusion is restricted in a way we did not anticipate, we might not discover this. The situation here is in many ways parallel to the discovery of the licensor effect itself — the use of embedded quantifiers in every previous NPI illusion study was a reasonable choice, but it led to the perception that illusions were a more general phenomenon than they turned out to be.

In sum, it is clear that not all NPI licensors are equally likely to trigger illusions. It has been demonstrated many times that ungrammatical NPI-containing sentences with embedded non-quantificational licensors are less likely to be accepted than those with embedded quantificational licensors. Whether non-quantificational licensors yield precisely zero illusions, and whether illusions for non-quantificational licensors can be induced with different tense configurations, are not currently clear.

6.2.2 The distance effect

The second major generalization we have highlighted is the effect of the position of the NPI. The effect was first reported by Parker & Phillips 2016. Here we replicate the basic effect and elaborate on its profile. The key question we pursued in Chapter 4 was whether the NPI needs to be close to the embedded licensor in order for illusions to occur (as Parker & Phillips claimed), or if instead the NPI needs to be close to the RC edge in order for illusions to occur. In two experiments, we showed that added distance from the embedded licensor to the NPI has no clear impact on illusion rates, as shown in Figure 6.3. We additionally replicated effects reported in Parker & Phillips 2016 showing that added distance from the RC to the NPI does reduce illusion rates, as shown in Figure 6.4. Experiment 11 further clarified that the very fine-grained distance effects that had previously been found, in which only one added intervening word can reduce illusion rates, appear specific to circumstances where the word that intervenes is a verb. Without an intervening verb, there is no clear difference between one and two intervening words (but note also that without a second intervening word, there is no clear difference between verb and no verb). Finally, we found a surprising effect in which sentence-final NPIs appear to be vulnerable to illusions as shown in Figure 6.5. We take these findings to suggest that NPI illusions involve interference from the entire RC representation, not just the individual negative word. The illusion for sentence-final NPIs makes this generalization somewhat more complicated, but it's possible that the RC comes back into the focus of attention during sentence-final wrap up processing. We note also that these effects were only investigated with prepositional phrases containing *any* like *at any time* or *in any way* and may be subject to interference from free choice readings.

One somewhat unexpected generalization that appears when these findings are combined is that the

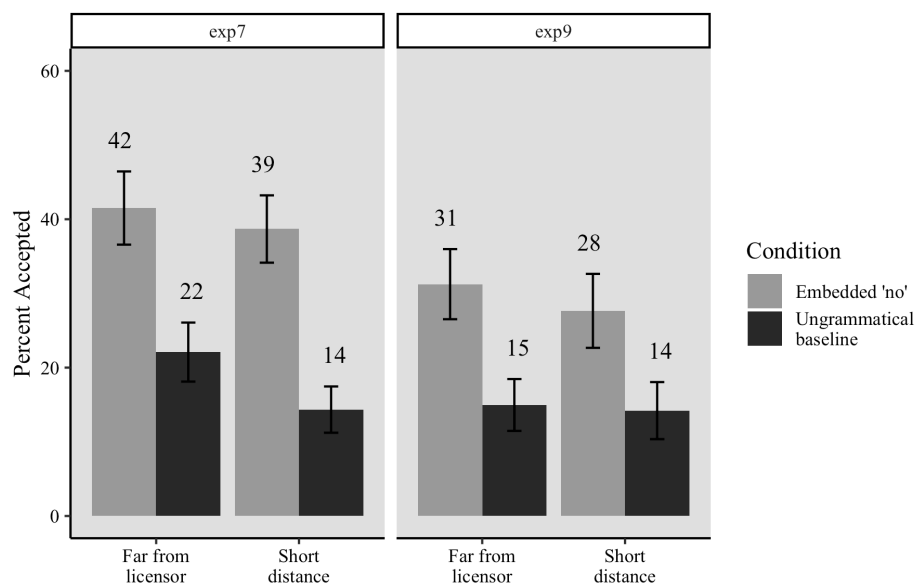


Figure 6.3: Mean percentage of ‘YES’ responses for short-distance and long-distance conditions in all speeded acceptability experiments that compared the impact of added distance from the embedded licenser. Error bars indicate standard error of the mean across subjects.

distance effect is only partly realized by a decrease in acceptance rates for the illusion condition when the NPI is positioned farther from the RC. Rather, a relatively large part of the effect is in the *increase* in acceptance rates for the ungrammatical baseline condition. In fact, this is true for Parker & Phillips’s data as well. They helpfully make their raw data available, allowing us to directly compare all known experiments investigating the distance effect for NPI illusions. In Figure 6.6 we plot the acceptance rates for short-distance and long-distance conditions, separating out the ungrammatical baseline, grammatical baseline, and illusion conditions. This visualization makes it clear that a number of distance manipulations impact the acceptance rate for the ungrammatical baseline condition.

The impact of distance on the ungrammatical baseline condition is rather unexpected under the accounts of the distance effect that we have considered here. Our discussion has focused entirely on the timecourse of declining accessibility of an interfering representation (either a negative word or a negative RC). In the ungrammatical baseline condition, there is no interfering representation, so the question of its accessibility is irrelevant. Perhaps this effect can be explained by a general uncertainty about the exact content of prior material — if the NPI is encountered late in the sentence, the participant is un-

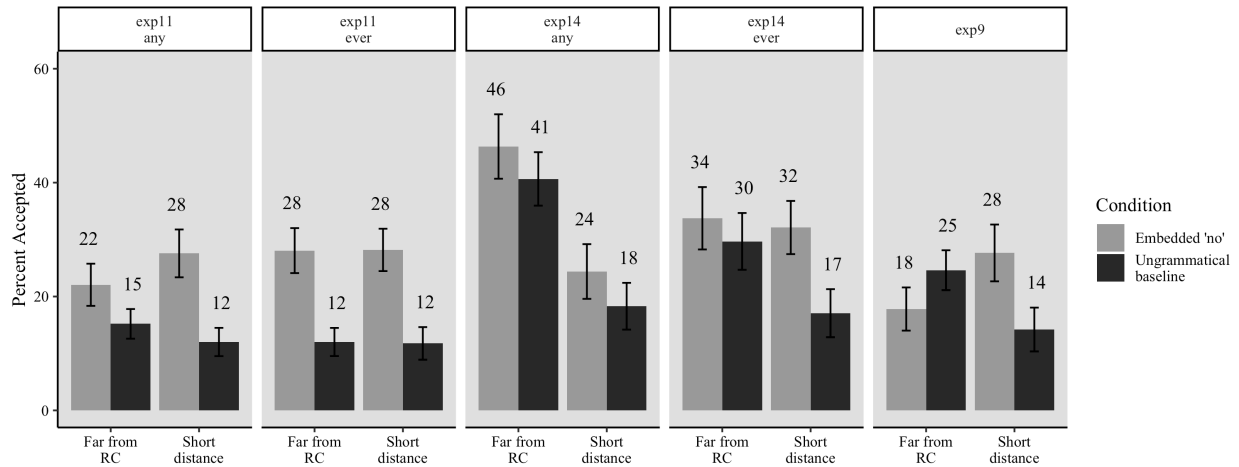


Figure 6.4: Mean percentage of ‘YES’ responses for short-distance and long-distance conditions in all speeded acceptability experiments that compared the impact of added distance from the relative clause. Error bars indicate standard error of the mean across subjects.

sure of what happened earlier in the sentence, and therefore is unsure of whether the NPI was licensed and resorts to guessing. However, it’s not clear why this wouldn’t also lead to a boost for the embedded negation condition.

Note also that in the grammatical baseline conditions (Figure 6.6, top panel) we observe no clear effects of any distance manipulations. This is somewhat mysterious if late-arriving NPIs lead to guessing. It is, however, very consistent with what is known independently about NPI licensing — namely, that NPIs can occur at an unbounded distance from the negative word, with no consequences for grammatical status or acceptability.

Setting this aside, the fact that there is a distance effect at all is somewhat more natural under environment-based licensing accounts than accounts that treat NPI licensing as an item-to-item dependency between the NPI and the negative word. Under environment-based accounts, licensing is an inherently local relation — the NPI must occur *within* the environment that has the appropriate properties. It is therefore reasonable that an NPI that is *nearly* within that environment would be more likely to be subject to illusions than an NPI that is very far from that environment. If licensing is an item-to-item dependency, it must be formulated in such a way that the two elements can be unboundedly far apart, making it somewhat unintuitive that NPI illusions should be dependent on the position of the NPI at all.

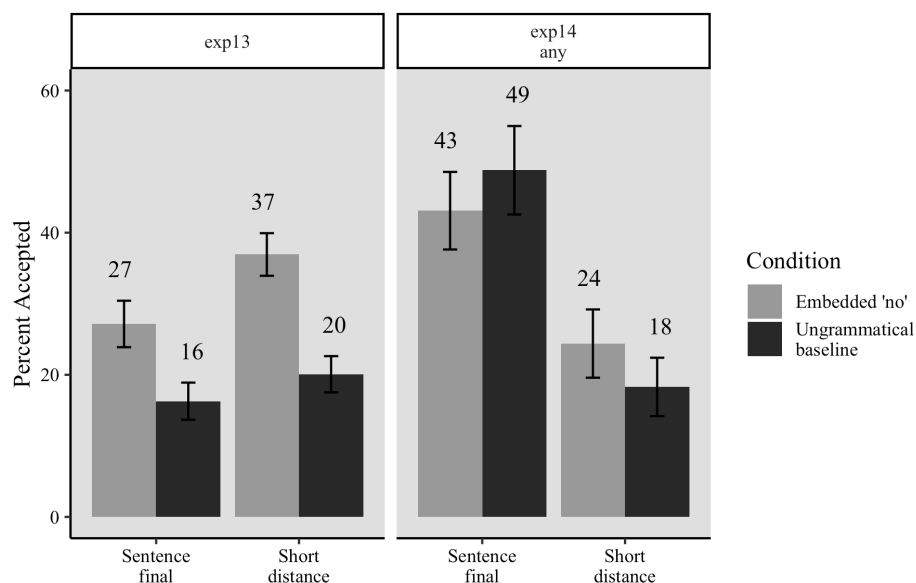


Figure 6.5: Mean percentage of ‘YES’ responses for short-distance and long-distance conditions in all speeded acceptability experiments that compared the impact of moving the NPI to the end of the sentence. Error bars indicate standard error of the mean across subjects.

However, as we noted in Chapter 4, there is a way for item-to-item licensing to account for the observed distance effects. One would have to say that the accessibility of the negative word is uniform throughout the scope of that negative word, and begins to decline once the scope of the word ends. Uniform accessibility throughout the licenser’s scope might be independently desirable to account for the non-effect of linear distance for truly licensed NPIs. However, it is unclear how such a mechanism could be implemented. Moreover, evidence from bound variable interpretations of pronouns (Kush, Lidz, & Phillips 2015) suggest that quantifiers do not remain accessible *at all* after their scope ends, which is inconsistent with this story.

6.2.3 NPI identity

The question of whether NPI illusions are specific to *ever* or general across all NPIs remains mostly unsettled. We have followed prior work in referring to the phenomenon as the “NPI illusion” (presupposing generality across NPIs), though the fact remains that it is unknown whether a wide variety of NPIs exhibit the same behavior. Certainly all hypotheses currently under consideration predict that they should.

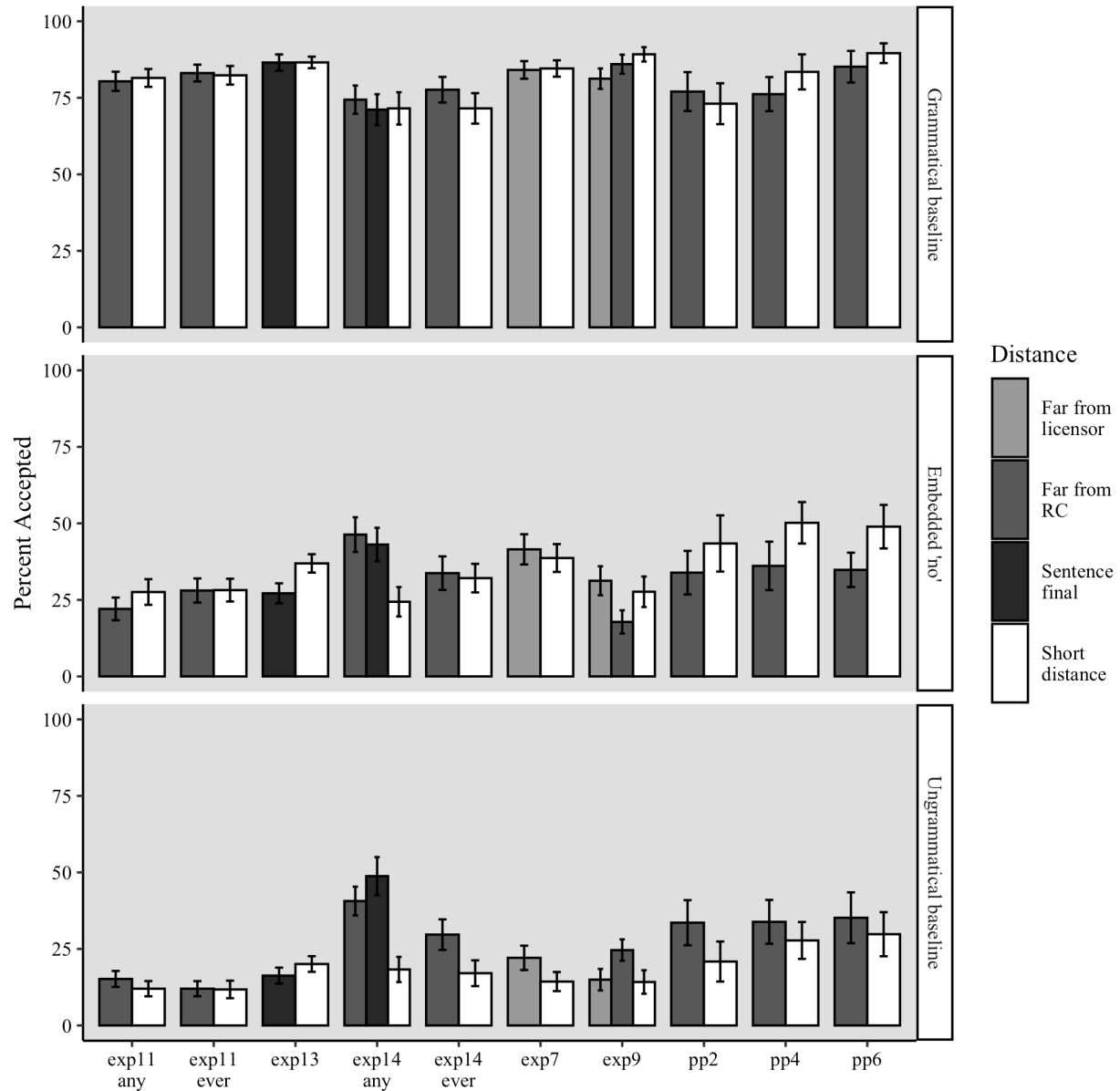


Figure 6.6: Mean percentage of ‘YES’ responses for short-distance and long-distance conditions in all speeded acceptability experiments that compared the impact of added distance. Error bars indicate standard error of the mean across subjects. Ungrammatical baseline sentences (bottom panel) are repeatedly judged acceptable more often in long distance (dark bar) than short distance (white bar) conditions.

One area where we now have greater certainty that the illusion is somewhat general concerns the contrast between *ever* and *any*. Parker & Phillips 2016 found illusions for *ever* but not *any* in one demonstration of the distance effect. They attributed the contrast to the distance, not the NPI identity, but illusions for *any* have been difficult to demonstrate (and, in general, not pursued) because of the impossibility of

positioning *any* pre-verbally. In Experiment 11 we found that *any* can yield illusions even post-verbally, if there is only one intervening word between the NPI and the RC. We additionally found illusions for prepositional phrases containing *any* such as *at any time* and *in any way* in Experiment 13. Surprisingly, we also found illusions for these items when they were positioned sentence-finally, suggesting that NPI illusions for *any* are quite robust. Thus, while it appears that NPI illusions are not specific to *ever*, our results raise the interesting possibility that the profile of the illusion may not be identical for all NPIs.

6.2.4 Relative clause type

Across experiments, we have investigated NPI illusions in both SRCs and ORCs. In general, illusions appear to be possible for both clause types. Moreover, the licenser effect is found in both, ruling out structural-distance accounts of that effect.

However, it appears that illusion rates are lower for SRCs than for ORCs. We have not compared these directly and so we cannot be confident that this contrast is reliable, but it is notable that in Figure 6.1, three of the four experiments showing the smallest illusion effects are the only three experiments that used SRCs.⁵¹ We have previously speculated that the illusion rate for SRCs may be reduced because of a penalty for sentences with *no* in the object position of the RC. However, presenting the data as we have in Figure 6.1, it's not clear that the illusion sentences suffer any penalty in SRCs — rather, the effect size is reduced because the acceptance rate for the ungrammatical baseline is boosted, relative to ORC experiments.⁵² Or, put differently, it appears that the acceptance rate for the ungrammatical baseline in ORC experiments is suppressed, relative to SRC experiments.

It's not clear why such an effect should arise, but answers might be found in the rich literature on the processing of SRCs and ORCs. It is well known that ORCs incur a processing cost, which manifests in many measures including both sentence-final comprehension and online measures such as reading times (King & Just 1991; Staub 2010; Staub, Dillon, & Clifton 2017; Lowder & Gordon 2021; among others). Some work suggests that manipulations that make the two noun phrases more discriminable

⁵¹Given the problems with Experiment 15, one might be inclined to simply disregard the illusion effect sizes in this experiment. In that case, the two smallest illusion effects were both in experiments that used SRCs.

⁵²This is empirically similar to what we observed for the distance effect, though we don't know of any reason to think the mechanism is the same.

ease some of the difficulty associated with ORCs. For example, changing the second noun to a name (*the banker that Ben praised...* versus *the banker that the barber praised...*) appears to make ORCs less difficult (Gordon, Hendrick, & Johnson 2001). Similarly, changing one of the nouns to *everyone* appears to help (e.g. *the salesman that everyone contacted ...* versus *the salesman that the accountant contacted ...*) (Gordon, Hendrick, & Johnson 2004). Extending this reasoning to NPI illusion stimuli, it's possible that ungrammatical baseline sentences for experiments using ORCs have an extra penalty, in addition to the unlicensed NPI, since the two noun phrases in the RC are the most similar (e.g. *the authors* and *the critics*, compared to pairs like *no authors* and *the critics* or *the authors* and *no critics* in other conditions).⁵³

This would predict that the sentences that function as ungrammatical baselines in ORC illusion experiments would be slightly less likely to be accepted than grammatical baseline and illusion sentences even if the NPIs were removed, since part of their unacceptability comes from the two definites in the subject. This prediction can be evaluated with existing data: in Experiment 4 we tested ORC sentences both with and without NPIs. We did not observe any trend toward decreased acceptability for the sentences with two definites, relative to the sentences where one of the nouns had a quantifier, making this explanation somewhat less plausible.

Another possibility that we considered for the trend toward an SRC/ORC asymmetry is in the noisy channel framework: since the lower noun in an ORC is also a subject (of the RC), the two subjects may be more confusable and therefore more susceptible to edits.⁵⁴ Because the two nouns in an ORC have more in common than the two nouns in an SRC, they may be more likely to be swapped in a mental representation. It is worth noting that these proposals predict differences in acceptance for illusion sentences in ORCs versus SRCs, not differences in acceptance for ungrammatical baseline sentences in ORCs versus SRCs. Further work on this issue is needed to identify with greater certainty whether clause type does in fact influence illusion rates.

⁵³This explanation is made somewhat less plausible by the finding that just making one of the nouns an indefinite does not appear to alleviate ORC processing costs (Gordon, Hendrick, & Johnson 2004). That is, just changing the determiner might not be enough to make a difference.

⁵⁴Note that such a proposal could also be translated into a reanalysis framework. In the noisy channel model, all candidate edits are computed and maintained in parallel with a probability distribution over them, but one might instead consider a model in which candidate edits are computed only after processing difficulty is encountered.

6.2.5 Interpretation

Experiment 4 investigated, for the first time, the interpretation that comprehenders construct for NPI illusion sentences. This was done with sentence-final comprehension questions targeting the polarity of the MC. We found that comprehenders believe NPI illusion sentences to express negative propositions in a large proportion of trials. We additionally found that illusions arise in both positively-interpreted and negatively-interpreted trials, suggesting that there is not a direct correspondence between a comprehender's assessment of an NPI illusion sentence's acceptability and their assessment of the polarity of the MC. This has some bearing on the question of whether NPI illusions reflect a processing error that occurs on a small proportion of trials, or a small boost in perceived acceptability that occurs on all trials, but which only crosses the threshold of an "acceptable" judgment on a small proportion of trials. Hypotheses that are committed to a stochastic processing error which has consequences for both interpretation and acceptance seem to be ruled out by this finding.

It should also be noted that we cannot be confident that the interpretations comprehenders reported in Experiment 4 reflect the interpretations they had before the question was asked. That is, by asking comprehenders questions like *Did the authors receive acknowledgement for a novel?*, we typically assume that comprehenders have an internal representation of the sentence meaning that specifies whether the authors that were described received acknowledgement for a novel. If comprehenders do not have such a representation (or if they have a representation that does not specify the answer but merely allows an inference about the answer to the question), the question itself may trigger interpretive processes that would not have occurred otherwise.

Further work is clearly needed to better understand the interpretation of NPI illusion sentences. Knowing whether the event described by the MC did or did not happen is obviously a critical component of knowing what a sentence means, but it is a fairly coarse-grained measure of interpretation. We are currently designing a repetition task in the style of Potter & Lombardi 1990 (as applied to linguistic illusions by Wellwood et al. 2018), in which comprehenders comprehend NPI illusion sentences, perform a distractor task, and then repeat the sentence. One advantage of this task is that it allows us to tease apart globally-negative interpretations that result from distorting the mental representation of the

string (i.e., moving *no* to the matrix subject, either as a result of noisy channel mechanisms or a reanalysis process) from those that result from covert syntactic operations on the sentence (i.e., quantifier raising). There may also be value in more implicit sentence comprehension measures, such as processing difficulty on a subsequent sentence that presupposes either a positive or negative global meaning for the illusion sentence.

6.2.6 Timing

The experiments presented here overwhelmingly use the speeded acceptability task with RSVP-presented sentences to probe illusions. Note, however, that even in untimed likert judgment tasks in which the sentence is presented all at once (Experiments 1, 5, and 8), we sometimes observe a trend toward illusions. This might be explained by experimental participants who, despite being told to consider each sentence carefully, approach an untimed acceptability experiment in essentially the same way as a speeded acceptability experiment. Splitting untimed data by response time might go some of the way toward determining whether this is the case, but it's possible for very quick judgments to have very slow RTs (if, for example, a participant is switching between tasks). Alternatively, the illusion could also simply be persistent. As we discussed at length in Chapter 1, illusions are not always fleeting, and a temporal profile is a poor diagnostic for illusion status.

One potentially interesting thing to consider about the speeded acceptability task is in what ways it is intended to probe different representations from an untimed acceptability task. There are three ways the tasks differ — sentence presentation (all at once or RSVP), response timing (unlimited or a few seconds), and response type (Likert or binary). Setting aside the third of these,⁵⁵ there are essentially two ways the speeded acceptability task could work: by not giving participants enough time to reanalyze their assessment of the sentence, or by not giving participants access to the string to verify that the analysis they've assigned is consistent with the input. In natural reading, reanalysis is often paired with regression

⁵⁵It's hard to make a good 1-7 judgment quickly, so researchers generally don't put participants in the position of having to do this. On the other hand 1-7 judgments provide greater detail about a comprehender's perception of a sentence, so when timing isn't an issue this is preferable. Thus the difference in this aspect of the task may not be very deep. However, speeded Likert judgments (if it can be done) could shed light on the question of whether NPI illusions reflect a stochastic processing error that occurs on only a few trials versus a boost in acceptability that occurs on every trial.

(e.g., Frazier & Rayner 1982; Meseguer, Carreiras, & Clifton 2002), suggesting that these two issues are related. However, a change in one's assessment of a string does not strictly require re-accessing the input: Parker 2019 presented participants with agreement attraction sentences in RSVP (making returning to the stimulus impossible) and compared acceptance rates for binary judgments of acceptability under time pressure versus binary judgments of acceptability with unlimited time. The agreement attraction effect was clearly reduced and possibly absent for untimed judgments. It appears that just having the time to think about the sentence more is enough for agreement attraction effects to go away.

This does not appear to be the case for NPI illusions. The one experiment that used RSVP but did not set a time limit on responses was Experiment 4. This was done because participants had two tasks (acceptance and comprehension questions) and we did not wish to make the experiment overly complicated. But the result of this choice is that the comparison between Experiment 4 and all of our other experiments is roughly analogous to Parker's untimed versus speeded tasks. Yet we find clear illusions in Experiment 4 (in fact, this was the largest illusion effect out of any experiment, based on the raw difference in means). One might be concerned that this comparison is not really parallel to Parker's because comprehenders had a second task in Experiment 4, so they might have felt rushed to complete the acceptability judgment before they forgot the sentence, so that they could answer both questions accurately. However, this is undermined by the fact that RTs for the acceptability judgments in Experiment 4 were in fact slower on average than RTs for other experiments. Thus it appears that what is needed to recover from an NPI illusion is not just time to think about the sentence but the ability to re-access to the sentence. Some corroborating evidence for this claim comes from Xiang, Grove, & Giannakidou 2013, which demonstrated clear illusion effects in binary untimed acceptability judgments following sentences presented with self-paced reading. Self-paced reading, like RSVP, does not allow the participant to return to previous parts of the sentence.

Of course, these findings should be replicated in an experiment directly comparing timed and untimed judgments as in Parker 2019. However, the pattern suggests that agreement attraction and NPI illusions may differ in their persistence.

6.2.7 Tense

Perhaps the most surprising individual finding in Chapters 3-5 is the apparent emergence of illusions for sentences with non-quantificational licensors when the MC is in the past perfect, as opposed to the present perfect. Note again that this pattern was not statistically significant and requires replication.

We might ask if the theoretical literature on NPI licensing has anything to say about this surprising finding. If, for example, the licensing conditions for NPIs are known to be different for clauses in different tenses, we might leverage such a fact to explain the emergence of the illusion. However, it is not clear that such differences exist. It's possible to license *ever* with either a negative quantifier or non-quantificational negation, in either a clause in the present perfect or the past perfect.

- (118) a. No critics have ever reviewed that novel.
b. The critics haven't ever reviewed that novel.
c. * The critics have ever reviewed that novel.
- (119) a. No critics had ever reviewed that novel.
b. The critics hadn't ever reviewed that novel.
c. * The critics had ever reviewed that novel.

The *haven't* and *hadn't* examples might be slightly odd, only because linearly adjacent *not+ever* is sometimes dispreferred because of the possibility of *never*. This can be remedied by putting the whole sentence under a neg-raising predicate: *I don't think the critics have ever reviewed that novel* or *I didn't think the critics had ever reviewed that novel*. Nonetheless, these facts do not help clarify the tense findings, since NPIs can clearly be licensed in clauses of any tense.⁵⁶

⁵⁶The one case that we know of where differences in NPI licensing arise as a consequence of tense changes is illustrated in (1), from Uribe-Echevarria 1994.

- (1) a. Mary didn't say that Ann would read any books tomorrow.
b. Mary didn't say that Ann had read any books last week.
c. */?* Mary didn't say that Ann will read any books tomorrow.
(Uribe-Echevarria 1994:95)

Uribe-Echevarria argues that the unacceptability of (1c) cannot be attributed to the mere use of the future in an embedded clause or even the combination of past and future tenses, since both (2) and (3) are acceptable.

A few comments about the role of tense in NPI illusions can be made. First, the tense of the RC does not appear to matter, at least when the MC is in the present perfect. Across experiments we have sometimes used the present perfect in the RC and sometimes the simple past, and clear illusions for embedded *no* (and a clear lack of illusions for embedded *not*) arise for both. However, this is not a very broad sample of tense-aspect configurations, so we cannot rule out the possibility that some impact of RC tense exists. Second, while we observe a trend toward illusions for embedded non-quantificational negation when the MC is in the past perfect, we cannot be confident that this is precisely the right characterization of the effect. That is, since the sentences in which we observe this effect happened to use the simple past in the RC, we cannot know whether the effect has to do with the tense of the MC or the sequence of tense between the two clauses.

As a first attempt at making sense of the finding, let us consider the possibility that the effect is in fact about the relationship between the tenses of the two clauses. If we assume that the past perfect and present perfect require a reference time (Reichenbach 1947; Comrie et al. 1985), then one possibility is that illusions arise for non-quantificational negation when the RC provides a possible reference time for the MC tense. Perhaps the mechanism underlying this has to do with the comprehender's attentional state — if the RC is relevant to MC processing because of its tense, it may be more difficult to disregard the RC in processing the NPI. This would predict that putting the RC in a different tense so that its event time is not a candidate reference time for the MC (e.g. *the authors that the critics won't recommend had ever written a best-selling novel*) would not yield illusions.

Another possibility is that the tense effect is related to the dispreference for *didn't* adjacent to *ever*, and the improvement in sentences where an NPI is licensed by sentential negation that is achieved by separating them with a neg-raising verb. This consideration is related to hypotheses in which the difference between *no* and *didn't* reduces to differences in the ways prior parts of the sentence might be edited

-
- (2) Mary will not say that Ann will read any books this fall.
(Uribe-Echevarria 1994:96)
 - (3) Mary didn't say that Ann will read those books tomorrow.
(Uribe-Echevarria 1994:100)

However, it's not clear that this pattern is directly related to the tense effect in illusion sentences.

to result in grammaticality. That is, perhaps the presence of *didn't* leads comprehenders to consider the possibility that the prior part of the sentence was like (120b).

- (120) a. * The authors [that the critics didn't recommend] had ever...
- b. Mary didn't think that the authors had ever...

This is obviously a major edit to the sentence, which should be assigned a very low probability under any noisy channel model. But it becomes a slightly less implausible hypothesis when one considers the fact that many of the fillers in Experiment 16 had structures just like (120b), since they were designed for Experiment 15, which used NPI illusion sentences embedded under neg-raising predicates. Perhaps the edit is considered because of interference from a prior trial. To explore whether this was possible, we looked at the data from Experiment 16, split by the previous trial type: embedded under a first-name+verb+*that* construction, or not. The Experiment 16 findings split by previous trial type are shown in Figure 6.7.

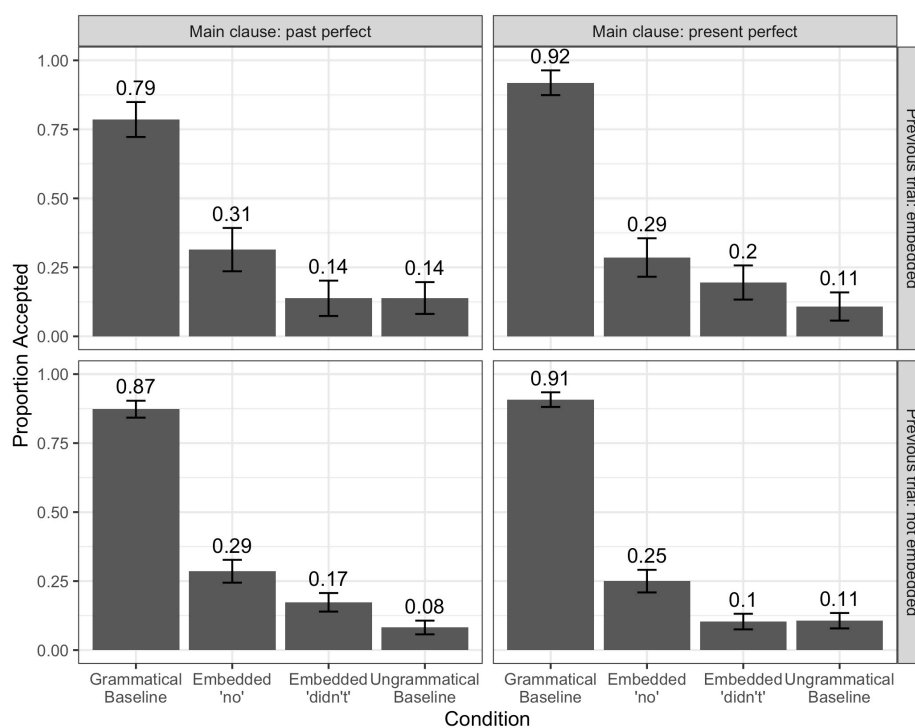


Figure 6.7: Mean percentage of ‘YES’ responses for the experimental conditions in Experiment 16, split by whether the immediate previous trial had an embedded sentence. Error bars indicate standard error of the proportion (i.e. trials are not grouped by subjects or by items).

This analysis does not clarify things. For the past perfect sentences (which is where we saw a trend toward illusions for non-quantificational licensors in the aggregate) we see the exact opposite pattern from what we expected under the hypothesis currently being explored: illusions arise only when the previous trial was *not* the embedding type that we speculated might cause interference. Even more surprisingly, the present perfect sentences show the opposite pattern (despite not showing illusions in the aggregate). Of course, we should not draw strong conclusions from these patterns, since each category consists of very few trials and the analysis is entirely exploratory. But it's safe to say the pattern is complicated.

A third consideration is that there seems to some expectation, when comprehending a sentence in the past perfect combined with *ever*, that the state of affairs being described changes at some point in time, maybe the reference time. That is, for a sentence like (119a), we might expect a continuation like *...until it became a best-seller*.⁵⁷ A hypothesis in which NPIs are related to contrastive inferences (such as the pragmatic rescuing account) might leverage this fact to explain the appearance of illusions when the MC is in the past perfect, though the details of such an account would need to be worked out.

One final note about the surprising tense effect is that even if it turns out that this trend is robust, it would not make the Experiment 15 findings make any more sense, since Experiment 15 did not reveal *any* illusions for ORCs (that is, neither illusions for embedded *no* nor illusions for embedded *didn't*), whereas the tense finding from Experiment 16 was, effectively, a case where we observed more illusions than we expected (illusions for embedded *didn't*, which are not typically observed).

6.2.8 Summary

The empirical contributions of the present dissertation with respect to NPI illusions are as follows. We have (1) documented a clear effect of the type of embedded licensor on the magnitude of the NPI illusion, (2) clarified the role of the NPI position for susceptibility to NPI illusions, (3) identified (some properties

⁵⁷We're not sure why this would be the case, but we can speculate a bit. It could be a consequence of the combination of the meanings of the specific word *ever* and the past perfect. Some analyses of *ever* treat it as a domain-widener (Chierchia 2006), such that the possible exceptions to the claim are minimized when *ever* is included. Importantly, the domain which *ever* widens is about possible event times. This is, to some extent, in conflict with the meaning of the past perfect, which imposes a boundary (the reference time) on the event times under consideration. So it seems that a past-perfect+*ever* sentence is making a claim about the event not occurring in the maximal span of possible event times *prior to* the reference time. This could, through Gricean mechanisms, lead to the implicature that the event *did* occur at some point after the reference time.

of) the sentence-final interpretation of NPI illusions, and (4) observed three surprising effects which future work might pursue: (a) illusions for sentence-final NPIs, (b) the possible impact of embedding NPI illusion sentences under neg-raising verbs, and (c) the possible role of MC tense. Although these findings may be fruitful areas of future research, we are, of course, not interested in characterizing the profile of the NPI illusion simply for the sake of characterizing the profile of the NPI illusion. That is to say, the form that future research of these phenomena takes should be guided by the hypothesized mechanisms that could explain the illusion, and, more importantly, explain how a comprehender uses their knowledge of language in real time to understand an incoming sentence.

6.3 Proposed mechanisms

Because the existing proposals as to the cause of NPI illusions have been explained in detail in other parts of this dissertation, we do not review their properties here. Rather we only highlight the issues that each hypothesis would need to account for in light of our findings from Chapters 3-5. That is, although each hypothesis is not obviously well-suited to account for the pattern of illusions, an explanatory hypothesis that fails to explain a single, not-yet-replicated data point should not be written off.

6.3.1 Cue-based retrieval

As we have previously noted, the cue-based retrieval explanation for NPI illusions proposed by Vasishth et al. 2008 is remarkably successful in accounting for a wide variety of illusion phenomena (agreement attraction, NPI illusions, illusions in anaphora and ellipsis processing). However, its generality is inconsistent with our empirical findings concerning the profile of the NPI illusion.

Specifically the licensor effect is a deep problem for such an account. There are essentially two options for making sense of the difference in illusion rates for quantificational licensors and non-quantificational licensors: changing the retrieval cues so that they distinguish quantificational and non-quantificational licensors, or adding a reanalysis process that captures the contrast. The first option is undesirable because it predicts differences between quantificational and non-quantificational forms of negation in their ability

to serve as true licensors. That is, if the licensors have different features (and those features are part of the NPI licensing search), we can capture the difference in illusion rates as a consequence of different retrieval probabilities, but this necessarily means that the retrieval probabilities will be different in general. Perhaps this can be avoided by weighting different cues differently, so that when the structural cue (e.g. [+c-command]) matches, retrieval is guaranteed even when other cues mismatch. But we suspect this would cause other problems, like predicting illusions whenever there's a word that c-commands the NPI (which is always).

The second option is somewhat more promising. The cue-based retrieval explanation might be maintained if it can be shown that illusions *do* occur at equal rates for both types of embedded licensor in the earliest stage of processing, but reanalysis processes that occur between the point when the NPI is encountered and the sentence-final judgment lead to differences in acceptance rates. Intuitively, such a reanalysis process might involve consideration of possible edits that would make the sentence grammatical, similar to the edits that are central to the noisy channel hypothesis. We explore the possible reanalysis processes that may follow an illusion in section 6.4. A related possibility is to adjust the linking hypotheses for how retrieval outcomes lead to acceptability judgments, so that the licensor effect might be attributed to these processes instead. As we noted in Chapter 2, there are many possible linking hypotheses one might consider, and the one assumed by Vasishth et al. 2008 is not particularly well-motivated. However, it's not clear whether a linking hypothesis that captures the licensor effect is possible, and a deeper exploration of this issue is needed.

There are other problems for the cue-based retrieval hypothesis, such as the fact that the distance effect exists, and, more worryingly, that it seems to be based on the RC boundary, not the distance from the licensor itself. These concerns are, to some extent, secondary. If the cue-based retrieval account is going to have a chance of making sense of the profile of the NPI illusion, it will need to be refined to make sense of the licensor effect.

6.3.2 Pragmatic rescuing

Like the cue-based retrieval hypothesis, the pragmatic rescuing hypothesis does not predict a contrast in illusion rates for different embedded licensors. In order to make sense of this effect the hypothesis would need some mechanism by which the contrastive implicature is inferred only for restrictive RCs that contain negative quantifiers.

However, we believe there is a deeper issue with the hypothesis, which is that it relies on a grammar of NPI licensing which is not plausible. In fact, it is the very NPI illusion data we seek to explain which reveals the problem with this version of the grammar. Recall that the RESCUING operation of Giannakidou 2006 is intended to account for the acceptability of sentences like (121) and (122), in which the NPIs are not in an non-veridical context (and so cannot be LICENSED, in Giannakidou's model) but the sentences are nonetheless fully acceptable. Giannakidou accounts for this acceptability through the RESCUING operation which is defined as in (123).

(121) Only Larry ate anything.

(122) Larry regrets that he said anything.

(Giannakidou 2006:577)

(123) A PI [polarity item] α can be RESCUED in the scope of a veridical expression β in a sentence S, if (a) the global context C of S makes a proposition S' available which contains a non-veridical expression β ; and (b) α can be associated with β in S'.

(Giannakidou 2006:596)

Given this definition, it is not obvious how the hypothesis can account for the unacceptability (in carefully considered judgments) of NPI illusion sentences like (124a) and (125a) or even ungrammatical baseline sentences like (126a). Each of these sentences makes the contrastive implicature in (b) available in virtue of the use of a restrictive RC, each of the implicatures in (b) contains the non-veridical element *not*, and each of the polarity items in (b) can be “associated with” (i.e., licensed by) *not*. Thus each of the (a) sentences satisfies the RESCUING conditions and is predicted to be just as acceptable as (121) and (122).

- (124) a. * The authors [that no critics have recommended in their reviews] have **ever** received acknowledgement for a best-selling novel.
- b. The authors [that some critics have recommended in their reviews] have **not ever** received acknowledgement for a best-selling novel.
- (125) a. * The authors [that the critics haven't recommended in their reviews] have **ever** received acknowledgement for a best-selling novel.
- b. The authors [that the critics have recommended in their reviews] have **not ever** received acknowledgement for a best-selling novel.
- (126) a. * The authors [that the critics have recommended in their reviews] have **ever** received acknowledgement for a best-selling novel.
- b. The authors [that some critics haven't recommended in their reviews] have **not ever** received acknowledgement for a best-selling novel.

It is of course possible to reign in the RESCUING operation so that it does not predict acceptability for NPIs in so many contexts in which they do not occur. Various proposals for a secondary mechanism through which NPI-containing sentences can be made acceptable in the absence of grammatical licensing exist. In our view, the viability of such hypotheses (as a grammatical theory, setting aside illusions) will come down to their ability to define this secondary mechanism in such a way that it captures both the full acceptability of (121) and (122) and the full unacceptability of (126a). The possibility for explaining illusions, then, comes down to whether that mechanism makes any distinction between (124a) on the one hand and (125a) and (126a) on the other. It remains to be seen whether this is the case.

6.3.3 Scalar alternatives

The scalar alternatives hypothesis accounts for the licensor effect in virtue of the differences in (scalar) inferences that quantificational and non-quantificational licensors are assumed to invoke, and accounts for the distance effect by treating licensing as an environment-based (not c-command-based or scope-based) phenomenon.

This hypothesis suffers from two significant failed predictions: comprehenders’ tendency to interpret the MCs of NPI illusion sentences as if they are negative (Experiment 4), and the lack of illusions for non-quantificational negation even when scalar alternatives are evoked through the use of NPIs (Experiment 12). One possibility for making sense of the first finding is to postulate some kind of re-analysis process that happens between the NPI and the sentence-final judgment would need to give rise to these effects. Perhaps the comprehender, finding themselves with a syntactic analysis which places the NPI in the MC and an impression of acceptability for the NPI, reconciles these representations in the only way they can, which is to assume that the MC was negative. Under such a hypothesis, comprehenders’ responses to the comprehension questions are not simply the result of “reading off” the meaning from the representation that was constructed at the NPI, but reflect an active inference procedure based on uncertain (and conflicting) representations.

Depending on the details of the reanalysis process (i.e. whether it involves a distortion of the string), it may be that the contrast between *no* and *not* is a consequence of this process as well, due to their different edit possibilities (as we have discussed in the context of the noisy channel model). This would additionally make sense of the lack of sentence-final illusions for *not* in scalar contexts, observed in Experiment 12. However, this results in some redundancy: the licenser effect is attributed to differences in scalar alternatives at the NPI, but attributed to differences in reanalysis possibilities after the NPI. In fact if we assume a reanalysis strategy based on possible edits, it’s not clear that the scalar alternatives hypothesis contributes much above and beyond a hypothesis that simply treats licensing as context-based (with or without scalar inferences). We return to this issue in section 6.4. But it appears that the value of the scalar alternatives hypothesis turns on whether the licenser effect is apparent even when the NPI is first encountered, and if this effect is mediated by the presence of scalar alternatives (as we manipulated them in Experiment 12). This suggests that follow up experiments to test whether the licenser effect arises in reading times at the NPI may be useful.

A separate option for explaining away the non-effect of scalar alternatives in Experiment 12 is to consider the particular alternatives evoked by the added NPIs in the RC. The details of the evoked alternatives have been mostly ignored in our discussions of the scalar alternatives hypothesis, because at least

three possibilities exist and we are unable to determine which is responsible, which we noted in Chapter 3. Recall that there is an apparent conflict with treating the alternatives evoked by the quantifier *no* as the NPI-licensing alternatives, because lexical alternatives to *no* and lexical alternatives to the NPI give rise to different sets of propositional alternatives. The three possible solutions to this are to say that the NPI *ever* is interpreted as if it quantifies over individuals (thus allowing it to make use of the quantifier’s alternatives), the licenser is interpreted as if it quantifies over events (thus allowing it to generate alternatives for the NPI *ever*), or the alternatives are not the propositions that arise from lexical alternatives, but rather some ad-hoc scale corresponding to an inferred function for the RC (i.e., *the authors that no critics recommended* is inferred to pick out the authors lowest on the recommendability scale). If it is the third option that is responsible, then surely any NPI in the RC could contribute to the inference of scalar alternatives. However, under the first two options, it’s less obvious that the inclusion of various NPIs in various RC positions would be expected to yield a representation that interferes with *ever*. Thus, it may be useful to identify with greater specificity the nature of the interfering alternatives.

In sum, it’s clear that the scalar alternatives is not as promising as it appeared in light of only the Chapter 3 data. However, it’s possible that further elaboration of the specific alternatives that are assumed and the reanalysis process will make sense of the surprising findings from Experiments 4 and 12.

6.3.4 Scope miscalculation

The scope miscalculation hypothesis treats the licenser effect as a consequence of the scope-taking possibilities of quantificational and non-quantificational terms. However, as we have noted previously, there is no difference in the scope-taking properties of quantificational and non-quantificational *negation*, and so the hypothesis must attribute the illusion to negative quantifier processing strategies that essentially treat the item as a quantifier first, and negation second. We have argued that the interpretation findings from Experiment 4 are inconsistent with this account. However, there are in fact a number of possible hypotheses within the category of scope miscalculation accounts, and it is worthwhile to consider whether they all make the same predictions. We’ll refer to constructed representations in which the negative quantifier takes scope over only the RC (i.e. the grammatical option) as “in situ” interpretations of

the quantifier and representations in which the negative quantifier takes scope over the entire sentence (contrary to the grammar) as “QR” interpretations, though we’re not necessarily committed to quantifier raising as the mechanism through which scope is achieved.⁵⁸

Before exploring the possibilities for a quantifier scope explanation for NPI illusions, it should be noted that a good deal is known independently about the processing of genuinely scopally ambiguous constructions. The question of the timecourse of the assignment of scope has been studied primarily using sentences with two quantifiers which have both a surface scope reading and an inverse scope reading. Many studies then use processing times on a follow-up sentence which in some way assumes a particular interpretation of the first sentence (usually in virtue of plural or singular marking on a noun phrase) to determine which scope configuration was assigned, or if both were (e.g., Tunstall 1998; Anderson 2004). Such studies generally find that surface scope is preferred and inverse scope is computed only when needed. Bott & Schlotterbeck 2015 investigated the assignment of quantifier scope *within* the timecourse of the unfolding of the sentence using bound variable manipulation, and again found that inverse scope is pursued only when other aspects of the sentence (e.g. a variable needing to be bound) make it clear that it is necessary. Depending on one’s analysis of how surface scope is represented⁵⁹, we might take these findings to suggest that *in situ* representations are preferred, and QR representations of quantifiers are pursued only when needed.

In Chapter 3 we considered “early” and “late” scope assignment accounts. The “early” version we considered was a hypothesis in which scope assignment occurs at the quantifier and is stochastic — on some proportion of trials, the comprehender represents the quantifier *in situ* and in the other trials the quantifier undergoes QR.⁶⁰ This version of the hypothesis is not obviously consistent with independent findings about the processing of quantifiers. It is also straightforwardly ruled out by the Experiment 4

⁵⁸We’re also not necessarily committed to the “*in situ*” representation being literally *in situ*. If some movement within the RC is needed to get the actual scope over the RC, that’s fine. These are just shorthands.

⁵⁹If surface scope and inverse scope are both the result of covert quantifier movement, but in different orders, this conclusion does not hold.

⁶⁰It is of course mysterious that the comprehender would randomly decide to construct a grammar-violating representation, especially when a grammatical option is available. We tend to think of grammars as things that determine which representations are buildable and which are not. Under a hypothesis like this one, there are representations that get built using the same tools that exist in the grammar (QR, or whatever gets you wide scope of quantifiers), but applied in ways that result in representations that the grammar itself cannot build. Similar concerns arise throughout our discussion of the quantifier scope hypothesis.

findings. If the grammar-violating QR representation is constructed on some proportion of trials, that proportion should be precisely the illusion rate (because illusions occur in all and only these trials) and precisely the rate of negative interpretations of the MC (because it is the QRed quantifier that causes these interpretations). These values do not align in Experiment 4.⁶¹ It should also be the case that, within illusion trials, negative interpretations occur whenever the sentence is accepted, and positive interpretations occur whenever the sentence is rejected, also contrary to Experiment 4. One might argue that negative interpretations for rejected trials can be accommodated if some comprehenders, some of the time, responded to the comprehension question based not on their internal representation of the sentence but based on what they reasoned the sentence was supposed to be. That is to say, perhaps on some trials no QR occurred, no illusion occurred, and the comprehender knew the sentence was unacceptable and ungrammatical, but, forced to answer a comprehension question about an ungrammatical sentence, they reasoned about what a speaker who produced such an utterance might have had in mind as their message, and responded based on that. This is entirely possible but it does not solve the problem of the low rate of negative interpretations (lower than the illusion rate) in the absence of an NPI or the existence of positively-interpreted illusion trials that are accepted. Thus the hypothesis does not, on the whole, make the right predictions.

Under the “late” version of the hypothesis, the comprehender does not commit, at the quantifier, to either scope. This could be spelled out as computing both the *in situ* and QR representations and maintaining both in parallel, or computing a single representation that does not (yet) specify the quantifier scope.⁶² The main-clause NPI, then, effectively “disambiguates”, favoring the QR option. In the absence of an NPI, the comprehender would presumably eventually disambiguate to favor the *in situ* option. As we have previously noted, this hypothesis does not predict PPI illusions, which may exist.

⁶¹This finding is actually a little bit different from the pattern we usually emphasize in discussions of Experiment 4. We typically highlight the fact that rates of negative interpretations are different for sentences with and without NPIs. Here we’re discussing the (relatively strong) prediction of the early, stochastic quantifier scope hypothesis that the negative-interpretation rate for embedded-quantifier sentences without NPIs (which was in fact 15%) should match the illusion rate for embedded-quantifier sentences with NPIs (which was in fact 39%). We have not previously emphasized this finding because the prediction is unique to this version of the hypothesis.

⁶²If the “*in situ*” option is truly achieved by interpreting the quantifier *in situ*, it’s somewhat hard to imagine an interim representation that truly doesn’t commit to scope. This is fine, we would just say that the *in situ* option is computed initially but the option that this will change is kept open.

More worryingly, it again predicts a correspondence between interpretation and acceptance that, as Experiment 4 showed, does not arise. This hypothesis does make another option available to account for the Experiment 4 data. If the comprehender never really *chooses* the QR representation in the NPI-containing sentence, but rather continues to maintain both representations in parallel indefinitely, we might say that the acceptability question is answered based on one representation and the comprehension question is answered based on another. If independent evidence of such behavior in the case of parallel representations can be established, this version of the scope miscalculation hypothesis may be possible.

The third option we discussed in Chapter 3, which is actually what Orth, Yoshida, & Sloggett 2021 propose, is that the QR representation is only attempted late (at the NPI), but not because it makes the NPI *licensed* (as our “late” scope assignment version claims). Rather, QR is attempted only because it *affects* the NPI. This is how the authors manage to account for PPI illusions while avoiding an early scope-assignment hypothesis. However, we think this is the least plausible account for PPI illusions. This amounts to hypothesizing that the comprehender has a single representation of the sentence which is consistent with their grammar, and then, upon encountering a PPI, they distort this representation, changing it into something that is *not* consistent with the grammar (QRing the negative quantifier) only to see what would happen, and what happens is that the PPI is made unacceptable. This kind of mechanism cannot possibly generalize. Not only do comprehenders not seem to like to deviate from a representation that is working well for them (as garden paths illustrate), but there is no reason to believe that comprehenders actively pursue analyses that are inconsistent with the grammar when a grammatical option is available, and many reasons *not* to believe that the comprehender is actively looking for ways to make the current sentence less acceptable. All of these assumptions together would mean that a comprehender might, at any point in the processing of a grammatical sentence, attempt to change their representation of the first part of the sentence just to see if it would make the later part of the sentence unacceptable, and then judge the sentence less acceptable because of it. We do not find this account plausible.

We finally turn to one further possibility (which we did not explicitly consider in Chapter 4). Because scope possibilities for quantifiers are in general rather broad, but narrower only for downward entailing quantifiers, we might treat the assignment of quantifier scope as a two-stage process. Essentially, the first

thing the comprehender notices about a negative quantifier is that it's quantifier, and the second thing the comprehender notices is that it's negative. Spelling this out in further detail, we would have to say that first, a QR representation is assigned for any and all quantifiers⁶³; then, after some delay, the QR representation is undone if the particular quantifier does not allow it. Thus, if the NPI is encountered before the QR representation has been reconsidered, the NPI will appear unacceptable. This version of the hypothesis not only predicts the licenser effect, but can also make sense of (a version of) a distance effect⁶⁴. One problem with this version of the hypothesis is that if the undoing of the QR representation is just part of the normal course of events for the processing of negative quantifiers, why does it not happen at all in trials where the illusion occurs? That is to say, such a hypothesis predicts that an NPI, if encountered between step one and step two of negative quantifier processing, will appear acceptable, but once step two has occurred, the unacceptability of the NPI should become apparent. Yet illusions persist until at least the end of the sentence (since this is when we measure them in the speeded acceptability task). One might conceptualize this situation again as a case in which the comprehender has conflicting cues (the presence of the NPI suggests that the QR representation of the quantifier should *not* be withdrawn, whereas the quantifier's own properties suggest that it *must* be) and must find a way to reconcile them. Perhaps on some trials the comprehender favors one source of information, and judges the sentence acceptable, and on other trials the comprehender favors the other source of information and judges the sentence unacceptable. Or perhaps both representations are maintained, resulting in the mixed judgments we see in Experiment 4.

In sum, the scope miscalculation hypothesis requires some assumptions that are not especially consistent with what is independently known about quantifier processing. It also requires stipulating that in cases where comprehenders have multiple representations in parallel at the end of a sentence, they may answer some questions based on one representation and some questions based on another. If these assumptions can be shown to be reasonable, the hypothesis may be able to account for the NPI illusion. However, it is worth repeating that the quantifier scope hypothesis is essentially just a re-allocation of

⁶³This is of course not entirely consistent with (one interpretation of) the literature on quantifier processing showing that surface scope is preferred.

⁶⁴The distance effect that the hypothesis predicts, however, would make reference to the time since the negative quantifier was encountered, not the edge of the RC.

blame. Rather than saying that the NPI illusion arises because the online comprehender is (mysteriously) unfaithful to the grammar of NPI licensing, we say that the NPI illusion arises because the online comprehender is (just as mysteriously) unfaithful to the grammar of quantifier scope. Some clarity on which aspects of the grammar are violable and which are not is needed.

6.3.5 Noisy channel

Experiment 15 was intended to test the predictions of the noisy channel explanation for NPI illusions, but this experiment was largely unsuccessful, in that the baseline sentences were not accepted at rates we expected and we saw no clear illusions for embedded quantificational negation. Thus the proposal remains on the table. Here we discuss a few reasons to be skeptical of it anyway.

One issue is that the version of edit distance that we must assume in order to make sense of the NPI illusion (which treats re-ordering as a more likely edit than deletions or additions) is directly at odds with versions that are assumed in the noisy channel literature (which don't tend to consider re-ordering at all). A clear definition of the hypothesized edit distance function, which accounts for both NPI illusions and other phenomena which have been attributed to the noisy channel hypothesis, is therefore needed. A related issue is that what needs to be said in order to make sense of the licensor effect (i.e., that edits to extremely recent material are unlikely) and what needs to be said in order to make sense of the distance effect (i.e., that edits to more distant material are unlikely) are at odds. It may be possible to define a u-shaped curve of edit probability to capture these effects, but one would also hope that this could be independently verified.

One final issue is that we are somewhat skeptical of the idea that all possible edits to the string are considered in parallel. Intuitively, this leads to a huge number of parses that must be maintained, and the processing cost for so many representations would presumably be high. Traxler 2014 makes essentially this argument in a review of noisy channel architectures and alternatives. A hypothesis in which misperceptions of the input are considered as part of a reanalysis strategy (i.e., only when the analysis being pursued starts to seem untenable) may be more plausible. This may actually be very close to the kind of re-analysis strategy we tend to assume but don't often make explicit. We return to the possibility of a

noisy-channel-related reanalysis mechanism in section 6.4.

6.3.6 Local coherence

Local coherence accounts treat the NPI illusion for sentences like (127a) as a consequence of the incrementally-generated locally-coherent parse of the substring in (127b), in which the negative quantifier c-commands the NPI (and *recommended* is a reduced RC). However, illusions also arise for sentences like (128a) (from our Experiment 3 stimuli) which does not have an equivalent locally-coherent substring (i.e., (128b) is ungrammatical).

- (127) a. * The authors [that **no** critics recommended] have **ever** received acknowledgement for a best-selling novel.
- b. **No** critics recommended have **ever** ...
- (128) a. * The authors [that **no** critics have recommended in their reviews] have **ever** received acknowledgement for a best-selling novel.
- b. * **No** critics have recommended in their reviews have **ever** ...

Thus we do not consider the local coherence hypothesis a plausible explanation for the illusion. It's not clear that there's any modification to the hypothesis that would make it account for the illusion for sentences like (128a).

6.4 Towards a plausible mechanism for NPI licensing

It is clear from the above discussion that each of the hypotheses we have considered is in some ways inadequate. Throughout this summary we have attempted to identify the modifications or additional stipulations that could supplement each hypothesis to make it account for the findings that are apparently inconsistent with it. Here we pursue a different approach to generating hypotheses about the NPI illusion. Starting from what has been proposed the linguistic knowledge underlying the distribution of NPIs, we ask what a reasonable theory for the successful processing of licensed NPIs and the successful

detection of unlicensed NPIs might look like, and where we might expect such an algorithm to fail. As our discussion of possible amendments to existing hypotheses has highlighted, it is desirable to have a theory of not only when the algorithm fails to immediately detect an unlicensed NPI, but also of the subsequent processing stages that lead to sentence-final judgments and eventual recovery from the illusion. Thus there are two components to our theories of the processing of NPI illusion sentences — what happens at the NPI and what happens after (though relatively little is known about the latter).

As we discussed in Chapter 2, there are roughly three categories of approaches to the grammar of NPIs: those that treat NPI licensing as a syntactic dependency between the NPI and another node in the tree (sometimes with a secondary mechanism to explain the acceptability of NPIs in other contexts); those that treat NPI licensing as a restriction on the kinds of environments NPIs can occur in, based on those environments' entailment patterns; and those that treat NPI licensing as a consequence of the meanings of NPI-containing sentences as compared to their alternatives. This means that there are effectively two kinds of representation (in terms of unit size) that might need to be accessed at the NPI in order to attempt to construct a grammatically-sanctioned representation of the sentence: the licensor or the environment. We discuss each of these in turn, followed by the possibilities for post-NPI re-analysis. We also discuss the related phenomenon of NPI illusions in Turkish and Korean, which may require fundamentally different processing strategies.

6.4.1 Processing algorithms for licensor-based grammars of NPIs

Although we noted in section 2.2 that there are some approaches where the feature that licenses the NPI dominates it (Herburger & Mauck 2013), the present discussion adopts the more conventional assumption that the critical relation is one of c-command, and the critical feature is encoded on the licensor itself⁶⁵. The comprehender's problem, upon encountering the NPI, can be conceptualized as a problem of memory retrieval, as Vasishth et al. 2008 suggested. The licensor is not the item currently being processed, and, assuming a very narrow focus of attention, this means the licensor can only be brought into a dependency with the NPI by retrieving it from memory. Framed this way, NPI licensing is strongly anal-

⁶⁵Here we are somewhat agnostic about what that feature is. We'll just assume that there is some feature that identifies all and only the NPI licensors.

ogous to anaphor binding. Accordingly, our discussion here borrows heavily from suggestions in Kush 2013 concerning the possibilities for executing such a retrieval. Vasishth et al. 2008 propose a particular way this retrieval operates, but we will step back from these assumptions and consider space of the logical possibilities. One important assumption we make is that in normal processing the comprehender knows that there *is* a licenser and it just needs to be found (in order to construct the dependency that the grammar requires). Processing hypotheses in which the comprehender’s goal is to determine whether there is a licenser or not, i.e., whether the sentence is grammatical (instead of assuming that it is grammatical and trying to determine what it meant) are not going to generalize to typical comprehension situations.

Given that there could be many nodes that c-command the NPI, it is not trivial to identify the particular node which contains the licenser. Serially checking each of these nodes for the target representation is in principle possible — the algorithm would simply need to trace the dominance relations through the tree to identify the c-commanding ones, as in Kush’s TREE-TRAVERSAL algorithms — but could be costly. Moreover, a serial procedure would predict that the time required to license an NPI that has many c-commanding nodes should be longer than the time required to license an NPI that has fewer c-commanding nodes, which does not appear to be the case⁶⁶.

This suggests that the evaluation of candidate nodes should be executed in parallel. However, even within parallel-evaluation models there are a few options that one could consider. Given how we have stated things so far, only c-commanding nodes are candidate locations where a licenser might be found and so only these nodes should be evaluated for whether they contain the target representation, identifiable by its NPI-licensing feature. One might instead assume that the only candidates are the ones with an NPI-licensing feature, and from among these the comprehender aims to find one in a c-command relation with the NPI. Or one might assume that all nodes are candidates, and they must be searched based on both the NPI-licensing feature and the c-command relation. As we have already noted, hypotheses in the third category such as Vasishth et al. 2008 do not make the right predictions.

However, given common assumptions about the memory architecture, there is no clear way to implement parallel search among only a small set of candidates, as in our first two options. Rather, the

⁶⁶For example, studies of the distance effect for NPI illusions do not find degraded acceptability for grammatical baseline sentences in which the NPI is farther from the licenser.

only way to do parallel search is to search everything. Thus, there is no pool of *c*-commanders to search among, and creating such a pool would require a tree-traversal algorithm to identify *c*-commanders serially, which we have already ruled out. However, identifying a pool of items with the NPI-licensing feature may be more straightforward. That is, an incremental comprehender could search among all nodes for just the NPI-licensing ones, and then serially check these for *c*-command. There is a risk of high runtimes, since there is no real upper bound on the number of NPI licensors that can occur in a sentence. However, the actual number of such words in a particular sentence is almost always going to be extremely small (i.e., most sentences do not have that many negations). Such a mechanism predicts illusion effects in a surprisingly straightforward manner — first, all NPI-licensing words are retrieved in parallel, which would include embedded negation in the relative clause, resulting in an initial impression that licensing is proceeding smoothly; then the retrieved items are checked for *c*-command, resulting in the delayed detection of unacceptability in precisely the cases where a licensor had been found but it turns out to be non-*c*-commanding. The hypothesis is rather elegant, except that it predicts the wrong profile for illusions — that is, the hypothesis fails to predict both the licensor effect and the distance effect.

Another kind of strategy for identifying a *c*-commanding node with the right features is delineated in Kush 2013. In order to account for the comprehender’s apparent success at accessing only structurally-relevant antecedents for bound variables, Kush proposes a combination of dynamically-updated features, *LOCAL* and *ACTIVE*, which jointly allow the comprehender to precisely target the candidate antecedents. Much of the work of ensuring access to structurally-appropriate antecedents is achieved through cue weighting, but a notable component of the theory is the need to re-write the *LOCAL* and *ACTIVE* features, specifically at clause boundaries. Kush argues that the hypothesis would allow interference from RC-embedded quantificational phrases in the processing of dependent elements (in his case pronouns) if they were encountered very soon after the relative clause boundary: “if it were possible to place a pronoun between the edge of a relative clause and the matrix verb of the higher clause, we might predict interference from the QP” (Kush 2013:296). That is, the hypothesis seems to predict a distance effect for pronoun binding of precisely the sort that is found for NPI illusions. Because much is made of QPs in this account, it’s possible that even the licensor effect would be predicted. However, substantial work is needed to

translate the mechanisms that Kush defined for pronoun processing into mechanisms that could capture the basic profile of NPI licensing. This may be a promising area for future research.

In sum, there are several imaginable algorithms for identifying a c-commanding NPI licensor that predict both efficient licensing in the truly-licensed case and a basic NPI illusion: (1) parallel cue-based activation of all items that match a set of cues such as [+negation] and [+c-command], with interference arising in partial match cases as in Vasishth et al. 2008; (2) parallel activation of all items matching a [+negation] cue, followed by serial evaluation of the retrieved items' structural relations to the NPI; and (3) possibly, something like Kush's (2013) dynamically-updated LOCAL and ACTIVE features, adapted for NPIs. However, both of the first two options clearly predict interference for any non-c-commanding licensor, contrary to the licensor effect. They also do not predict a distance effect. Thus, they are only plausible if it turns out that these effects do not arise immediately at the NPI, but are the result of reanalysis processes that occur between the NPI and the end-of-sentence judgment. We return to this possibility in section 6.4.3.

6.4.2 Processing algorithms for environment-based grammars of NPIs

As we noted in Chapter 2, the fact that a negative morpheme can be found in memory may not be the only way the comprehender's state differs at *in months* in (129a) versus (129b). Rather, there are interpretive consequences of the negative word, and if those consequences are tracked in real time, it could be this interpretation that the NPI is sensitive to, not the lexical item that caused it.

- (129) a. We haven't left the house in months.
 b. * We have left the house in months.

Thus, there is a key difference between (some versions of) environment-based approaches to real-time NPI licensing and licensor-based approaches — a licensor is necessarily remote and must be retrieved, whereas the properties of an environment that license NPIs could be already available when the NPI is encountered, because they were already constructed for other purposes (i.e. understanding the sentence). This is importantly separate from the question of the *size* of the unit that enters into a dependency with

the NPI (roughly, word or clause). A unit could in principle be quite large and still need to be retrieved from memory. It is also not exactly about the fact that the unit that carries the relevant property is *ongoing*. For example, one could imagine that a [+downward-entailing] feature is encoded on a verb phrase node, thereby licensing NPIs within that verb phrase. This could still require retrieval of that verb phrase node in order to license the NPI, because that node is not in the focus of attention. Rather, the key factor is whether the property that the NPI requires is something that is independently needed (in addition to the fact that it is ongoing). Thus, if one imagines that downward-entailment is not only a mathematical property that holds of some contexts and not others, but some critical part of what it means to interpret a negated clause, then the downward-entailment property will already be available and immediately accessible when the NPI is encountered.⁶⁷ The same argument can be made for other properties which have been argued to license NPIs at the environment level, if they might be properties that are central to sentence understanding.

If the relevant property is already constructed and immediately accessible, the real-time licensing of NPIs is trivial — the critical property is already there, and if it isn't there, there's no way to make it exist, so the sentence is obviously unacceptable. The one case where there could be vulnerability is in moments of change. At the RC boundary in illusion sentences, all of the interpretive consequences of the negation must be switched off. If this transition is at all delayed or costly, an NPI could slip in unnoticed before it's completed. This would straightforwardly predict distance effects but not licenser effects.

Environment-based approaches to licensing need not be committed to the computation and availability of the NPI-licensing property prior to, and independent of, the appearance of an NPI. It could be that this property is only ever computed after the NPI is encountered. In this case, we would need a theory of under what circumstances that property can and cannot be computed. Presumably this will have to come down to the detection of a licenser in a c-commanding position, and so all of our discussion of that process from section 6.4.1 would apply. One could instead hypothesize that the relevant property is *sometimes* pre-computed and sometimes not, based on the particulars the stimulus. In cases when it is

⁶⁷ Depending on how seriously one takes the claim that the only things that don't require memory retrieval are the things in the focus of attention, this might mean that the interpretation (or whatever aspect of it licenses NPIs) is in the focus of attention when the NPI is encountered.

pre-computed, licensing is rapid and effortless, whereas in cases where the property needs to be computed at the NPI there may be additional steps. The scalar alternatives hypothesis is in this family. The possibility that some but not all licensors (and some but not all trials) allow a comprehender to pre-compute the licensing property allows the hypothesis to account for the licensor effect.

6.4.3 Post-NPI reanalysis

We now turn our attention to the processes that occur after an illusion-prone NPI is encountered. It is useful to clarify at the outset that re-analysis procedures are only relevant under some conceptualizations of the nature of the illusion. As we have previously noted, it is not known whether NPI illusions involve an error in processing that occurs on a subset of trials and yields full acceptability (i.e. the absence of any error signal for the unlicensed NPI) on those trials, or if instead the illusion involves intermediate acceptability on a large number of trials. Put simply, does the sentence sometimes seem perfect and sometimes seem terrible or always seem mediocre? Under the stochastic version, there is little to be said about re-analysis. On the trials that seem perfect, there is no reason to re-analyse⁶⁸. On the trials that seem terrible, the sentence is just as bad as ungrammatical baseline sentences, and whatever reanalysis strategies are attempted, they will mostly fail. Thus, when we think about re-analysis processes that occur after the NPI, we are thinking about processes that occur under a hypothesis where the sentence-final representation may be one of intermediate acceptability or intermediate confidence in the judged acceptability.

Our consideration of possible reanalysis strategies is largely motivated by the observation that small edits assumed by the noisy channel model align rather reasonably with an intuitive strategy for figuring out what went wrong when something has gone wrong. That is to say, the noisy channel framework's commitment to the parallel maintenance of a very large number of parses, encompassing a large number of edits, of any given string might be too costly to be reasonable. But there are other directions one could

⁶⁸The fact that people often eventually recover from illusions seems to contradict this. If the sentence seems perfect and no re-analysis is attempted, how could it ever stop seeming perfect? However, we don't believe this is a serious barrier to a stochastic-error explanation for the illusion. It could simply be that comprehenders re-read the entire sentence to double check its acceptability status, run the same stochastic processes, but arrive at a different outcome at least some of the time, resulting in recovery from the illusion due to the second read-through. This of course predicts that illusion recovery requires access to the sentence, and recall from section 6.2.6 that we have some evidence that this is correct. It also predicts that in tasks where the comprehender does not have to determine the acceptability of the string, they would not necessarily re-read it and would not recover from the illusion (which may well be true, and we have no way of knowing).

go with the original idea motivating the framework — that comprehenders reading *The coach smiled at the player tossed a frisbee...* are willing to consider the possibility that the string was actually *The coach smiled as the player tossed a frisbee...* (Levy 2008). For us, the change from *at* to *as* is considered precisely because the reduced relative clause structure is infrequent enough to trigger re-analysis, whereas for Levy that edit was already constructed well before the reduced relative clause was encountered. That is to say, we assume that in the process of understanding a sentence, decisions must be made under conditions of uncertainty, and it is only when the comprehender faces evidence that they may have made the wrong decision that they re-evaluate it.

If the comprehender had complete certainty about every decision that was made in the course of processing a sentence, there would be no need for reanalysis — at the point when a sentence crashes, there is no hope to save it because the current analysis is the only analysis. The fact that reanalysis happens at all (assuming that it does) seems to suggest that comprehenders are open to the possibility that some decision made along the way was the wrong decision. Whether the wrong decision was a decision about a syntactic structure given multiple options consistent with the grammar (as in garden paths) or a decision about a particular word's identity given a brief exposure to the stimulus does not seem like the kind of thing a comprehender could know a priori. Thus reanalysis should consist of the reconsideration of all kinds of decisions. Relatedly, studies of regression paths for garden paths suggest that comprehenders relatively often reanalyze by simply “starting over” and reading the sentence from the beginning, not targeting the specific point of the ambiguity (Frazier & Rayner 1982; Von der Malsburg & Vasishth 2013). This, of course, doesn't guarantee that the kind of reanalysis strategy we're considering here is right, but it's consistent with it.

Turning to NPI illusions, we can speculate about the reanalysis processes that a comprehender might attempt in the time between the NPI and the judgment. We noted above that there are many hypotheses that predict NPI illusions as a general phenomenon, in which any non-c-commanding licenser causes interference. However, this is in conflict with the actual finding that non-quantificational licensors do not seem to interfere. In order to make sense of this pattern, one possibility is that non-quantificational licensors *do* interfere, in the same way that quantificational licensors interfere, but these illusions are “rean-

alyzed away”. Since we are considering reanalysis processes that consist of edits to the prior representation in the spirit of the noisy channel framework, the same kind of “shuffling” explanation that we discussed in section 6.3.5 could be considered. Given this possibility, it is important for future work to investigate whether the licenser effect exists even at the earliest stages of NPI processing — that is, is there a moment where both NPIs preceded by embedded-*no* and NPIs preceded by embedded-*didn’t* show reduced processing disruptions relative to the ungrammatical baseline? An eye-tracking while reading study could usefully bear on this question, and would have the added benefit of potentially revealing evidence of reanalysis through regression patterns.

6.4.4 NPI illusions in other languages

We have noted in passing that NPI illusions also arise in languages in which the NPI can come before the negative word, such as Turkish (Yanilmaz & Drury 2018b) and Korean (Yun, Lee, & Drury 2018), though we have not explored these phenomena in detail. The kind of structures in which illusions are found in these languages include sentences like (131), compared to grammatical baselines like (130) and ungrammatical baselines like (132).

- (130) **Kimse** [Ali’nin çalış-tığ-ı]-nı söyle-**me**-dı
 anybody [Ali-GEN work-FN-AGG]-ACC say-NEG-PST.3SG
 ‘Anybody didn’t say that Ali worked’

- (131) ***Kimse** [Ali’nin çalış-**ma**-dığ-ı]-nı söyle-dı
 anybody [Ali-GEN work-NEG-FN-AGG]-ACC say-PST.3SG
 ‘Anybody said that Ali did not work’

- (132) ***Kimse** [Ali’nin çalış-tığ-ı]-nı söyle-dı
 anybody [Ali-GEN work-FN-AGG]-ACC say-PST.3SG
 ‘Anybody said that Ali worked’

(Yanilmaz & Drury 2018b:116)

The comprehender’s task in processing a sentence one of these sentences is different from the English equivalent in meaningful ways. We’ve explored the possible ways a comprehender could quickly determine, at an NPI, whether the prior context had supplied the critical ingredients for the NPI to be

licensed. In a sentence like (130), at the NPI *kimse*, there is no prior licenser or already-constructed negative meaning. This is not a problem in the way it would be in English, since a licenser could (and should) still come⁶⁹. Thus, processing an NPI in Turkish intuitively involves predictive processes that are not relevant in English — the comprehender must generate a prediction, at the NPI, that a licenser must come at some later, syntactically appropriate, point.

We might therefore expect closer parallels between Turkish NPI illusions and other prediction-related phenomena like wh-movement or cataphor processing, rather than between Turkish NPI illusions and English NPI illusions. Given the difference in mechanisms, it is not so surprising that Turkish NPI illusions don't seem to have exactly the same profile as English NPI illusions — that is, they are not specific to quantificational licensers, as the illusion for (131) demonstrates. Further work is clearly needed to understand the key differences between Turkish and English NPI illusions, but we suspect that the word order differences are critical.

6.5 Broader implications for grammatical illusions

We began our discussion of NPI illusions in Chapter 2 with a brief overview of grammatical illusions, which seem to follow the same basic pattern: a dependent element is missing something that it needs, but an intervening element with the right properties but in the wrong position seems to alleviate the processing disruption. At this level of detail, NPI illusions and agreement attraction and ellipsis and anaphor processing are all just minor variants of the same problem.

What we have shown is that not only is a description at this level of detail unsatisfactory in that it predicts uniformity *across* illusions where none exists (e.g. the comparison between agreement attraction and NPI illusions) but it also predicts uniformity *within* illusions where none exists. That is to say, the

⁶⁹It's worth noting that NPIs that precede their licensers are dispreferred but not impossible in English. Clifton & Frazier 2010 tested sentences in which NPIs occur in relative clauses modifying subjects, preceding sentential negation in the main clause, and asked whether the use of a reduced relative clause (which is independently known to give rise to garden path effects, such that the comprehender believes themselves to be parsing a main clause, not a relative clause) impacted the NPI's acceptability. The main question was whether there would be a penalty for the reduced relative clause sentences with *ever*, because the current parse of the string suggests that it cannot be licensed, compared to full relative clauses with *ever*, in which, even though a licenser cannot currently be found, the current parse suggests that there's still a chance it will come. In untimed acceptability judgments they find the predicted trend. We might take these findings to suggest that the English comprehender is not totally unlike the Turkish comprehender — they, too, allow for the possibility that a licenser is coming later.

circumstances under which an NPI illusion arises are far narrower than this description would suggest and far narrower than was expected under all previous accounts. A pessimistic way to state this outcome is that there is little hope for what was once thought to be a highly parsimonious theory of illusions — that is, the treatment of all these dependencies as fundamentally memory retrieval operations that differ only in the features they use. However, there is a more hopeful angle as well. Although such a theory as the one just described is elegant in that it explains many phenomena in one fell swoop, it requires assuming online sentence comprehension algorithms that are rather unlike the grammar they are meant to implement. In other words, there is little parsimony to be found in a theory that treats the online resolution of NPI licensing as something fundamentally unlike the grammar of NPI licensing, and treats the online resolution of subject-verb agreement as something fundamentally unlike the grammar of subject-verb agreement, etc. Moving away from strong parallels across processing phenomena and toward strong parallels between processing phenomena and the grammatical knowledge that underlies them is not necessarily a bad move.

This is of course not to say that we expect a different explanation for every illusion which applies only to that illusion. Rather, we expect alignment in virtue of the mechanism not the superficial shape of the phenomenon. For example, if NPI illusions arise because of the difficulty of rapidly suppressing pragmatic alternatives, we might expect to find independent consequences of this difficulty in phenomena like the processing of focus. If agreement attraction effects arise because of partial matches that result from the parallel cue-based activation of prior chunks in memory, this hypothesized mechanism may simply make no predictions whatsoever about NPI illusions, if NPI licensing is not fundamentally a memory retrieval operation. While a satisfactory explanation for the NPI remains elusive, we have made progress in identifying the shape such an explanation would need to take and have ruled out some approaches that are unlikely to ultimately be satisfying.

Chapter 7 Substitution illusions: overview

What we refer to here as “substitution illusions” are most commonly referred to as “Moses illusions” throughout the literature, due to the early and influential example in (133), first demonstrated by Erickson & Mattson 1981. Importantly, similar effects arise for examples like (134).

(133) How many animals of each kind did Moses bring on the ark?

(134) What is the name of the holiday when children dress up in costumes and walk door to door giving out candy?

We adopt the terminology “substitution illusion” in order to avoid the suggestion that the use of religious trivia is in any way critical to the illusion. The key generalization, rather, is that when a word that is anomalous with respect to its context is inserted, its anomalous status is sometimes not detected by comprehenders — that is, for a question like (133), participants often say “two” and fail to detect the substitution of “Moses” in place of “Noah”, and for a question like (134), participants often say “Halloween” and fail to detect the substitution of “giving out” for “receiving”. Framed this way, the Moses illusion demonstrated by Erickson & Mattson 1981 is in fact not the first demonstration of a substitution illusion: both Hornby 1974 and L. Baker 1979 investigated the (non-)detection of word substitutions that gave rise to knowledge violations, though in the service of different research goals (understanding the nature of presupposition, and assessing students’ ability to monitor their own comprehension, respectively). Erickson & Mattson 1981, however, were the first to test the specific example in (133) and the first to investigate the illusion in a language-processing framework. Most research on substitution illusions therefore takes Erickson & Mattson 1981 as a starting point, and treats (133) as a defining example.

While it is tempting to explain Bible trivia examples like (133) as simply a case of fuzzy knowledge

of the distinction between the appropriate word, “Noah”, and the inappropriate word that has been inserted (which we will generally refer to as the “impostor”), “Moses”, this variety of explanation cannot be extended to (134), as we are confident that comprehenders are familiar with the distinction between “giving out” and “receiving”. The example in (134) thus usefully demonstrates that the failure to detect the anomaly cannot be purely due to a lack of relevant knowledge on the part of the participant. Experiments on substitution illusions typically control for knowledge of the domain using a post-test, which we will return to in section 7.3.2.

A second initially-appealing assessment of the illusion, which we do not pursue here, is that comprehenders do in fact detect the anomaly but respond by saying “two” or “Halloween” because this is the cooperative way to handle a speech error when one can infer the intended message. While such explanations are plausible for circumstances where the trivia question is asked out of the blue (or even in a trivia context), experiments on substitution illusions almost always inform participants that there will be anomalous questions in the experiment, and the participant’s task is to identify those anomalies, often by responding “can’t say” or “there’s an error”. Thus the illusion cannot be reduced to a reluctance to be rude by pointing out a conversational partner’s speech error.

Having ruled out these two fundamentally non-linguistic descriptions of the phenomenon, we may consider the possibility that the failed detection reflects some property of the mechanisms for understanding language, or how these mechanisms interact with stored world knowledge. If the illusion is ultimately a product of the sentence processing system, this would be highly informative, as it has the potential to shed light on aspects of incremental interpretation that are typically so successful that they become invisible. In principle, a number of language processing errors that could lead to the failed detection of the anomaly, including problems in lexical access, problems in semantic composition, and problems in the mapping from linguistic to non-linguistic representations. These are, of course, critical operations for sentence processing more generally, making the study of substitution illusions broadly relevant.

The remainder of this chapter is structured as follows. Section 7.1 identifies a few key generalizations that arise in the literature on substitution illusions. Accounting for the most robust of these generalizations is considered a minimal criterion for a successful account of the illusion. Section 7.2 then turns

to the set of processing stages that would need to occur successfully in order for the impostor to be detected, which constitutes the set of in principle possible error points. Then in section 7.3 we overview some alternative ways of thinking about the illusion, that aren't as committed to determining *where* in the sequence of operations something went wrong, but rather about *why* they went wrong (e.g., subconscious pragmatic processes, a lack of attention or motivation, generally “shallow” sentence processing strategies). Then in section 7.4 we discuss what can be learned from neural and reading measures, and from L2. Finally, we summarise the goals of the present work in section 7.5.

7.1 Key factors and their influence on illusions

Here we present four key manipulations that arise in the literature on substitution illusions: the similarity in meaning between the intended word and the impostor, as illustrated in (135); the similarity in form between the intended word and the impostor, as illustrated in (136); the positioning of the impostor in presuppositional versus asserted content, as in (137); and the use of questions as opposed to declaratives to be judged true or false, as in (138).

- (135) a. About which archipelago did Great Britain wage war in the eighties with Brazil (*Argentina*)?
 b. About which archipelago did Great Britain wage war in the eighties with Iceland⁷⁰ (*Argentina*)?

(van Jaarsveld, Dijkstra, & Hermans 1997)

- (136) a. How many animals of each kind did Moses (*Noah*) bring on the ark?
 b. How many animals of each kind did Abraham (*Noah*) bring on the ark?

(Erickson & Mattson 1981)

- (137) a. Hieroglyphics, which is usually associated with the Russians (*Egyptians*), is a kind of picture writing.

⁷⁰ Apparently, there was also a war with Iceland, but in the seventies. This makes this stimulus not an ideal example, but it's the one van Jaarsveld, Dijkstra, & Hermans 1997 used.

- b. Hieroglyphics, which is a kind of picture writing, is usually associated with the Russians (*Egyptians*).

(Baker & Wagner 1987)

- (138) a. Snow White (*Sleeping Beauty*) slept for how long after she pricked her finger?⁷¹

- b. Snow White (*Sleeping Beauty*) slept for 100 years after she pricked her finger.

(Büttner 2007)

We begin with these contrasts because they have each been investigated on multiple occasions, and often in the service of answering different questions about the nature of the substitution illusion. Thus, each of these generalizations, if it exists, has the potential to significantly influence the hypothesis space. It is additionally worth noting that some of these effects have not been as robust — the effect is found in some experiments, but not all — and so by identifying the most reliable generalizations we may determine some of the desiderata for an adequate theory.

Except where otherwise noted, these experiments took steps to rule out the potential speech-error-accommodation and lack-of-knowledge issues identified above. That is, we can be confident that comprehenders did not answer “two” to the Moses illusion question merely because they did not know the difference between Moses and Noah, because all included participants correctly answered a post-test question about this very fact (typically, “Who brought two animals of each kind on the ark?” or similar). Likewise, we can be confident that comprehenders did not answer “two” merely because they thought the experimenter had made a mistake and they did not wish to be rude, because the task instructions warned participants that there would be word substitutions making some questions unanswerable in their stated form, gave examples of such substitutions, and instructed participants to respond with “can’t say” or a similar remark when they identified such anomalies.

⁷¹This example demonstrates another important fact about substitution illusions, which we will not explore in much detail: the impostor can arrive very early in the sentence or very late.

7.1.1 Similarity in meaning

We begin with the effect of the similarity in meaning⁷² of the intended word and the impostor, as exemplified in (135) above. This effect was first investigated by Erickson & Mattson 1981, based on their intuition that not all substitutions would be equally likely to miss. They compared (139), as well as four other trivia questions with impostors designed to be more or less similar to the intended word. For the ark question in particular, they found that 19 out of 39 participants (49%)⁷³ experienced an illusion (that is, responded “two” instead of reporting that there was an error in the question) when the impostor was “Moses”, whereas 0 out of 42 participants (0%) experienced an illusion when the impostor was “Nixon”.

- (139) a. How many animals of each kind did Moses (*Noah*) bring on the ark?
 b. How many animals of each kind did Nixon (*Noah*) bring on the ark?

(Erickson & Mattson 1981)

As this example makes clear, there are two dimensions of similarity at stake: the similarity between “Moses” and “Noah” is intuitively greater than the similarity between “Nixon” and “Noah”, and at the same time the relatedness between “Moses” and the context “How many animals of each kind did ... bring on the ark?” is intuitively greater than the relatedness between “Nixon” and “How many animals of each kind did ... bring on the ark?” This second type of relatedness is importantly not the same as “goodness of fit” of the word itself to the sentential context, though it is sometimes described as such — that is, neither “Moses” nor “Nixon” is a good fit⁷⁴, as both result in a world knowledge violation (though not

⁷²In this section we use “meaning” in a relatively informal way, intending only to contrast the kind of similarity discussed here with “form”-based similarity, discussed in the next subsection. We are currently concerned with characterizing some previously-demonstrated empirical generalizations, not necessarily identifying the theoretical underpinnings of these generalizations. However, in section 7.2 we explore the distinction between versions of meaning that focus on what is encoded within the linguistic system compared to the conceptual / world knowledge system with which the linguistic system interacts.

⁷³As we will see, 49% is somewhat on the high end for illusion rates, but this is not particularly surprising since we’re looking at only one stimulus here. As will become clear throughout our review of the literature (and as is central to our exploration in Chapter 8), there is substantial variability in the illusion effect size across stimuli.

⁷⁴That is not to say that *nothing* about “Moses” is a good fit to the context. As our discussion of impostor-context relatedness is meant to highlight, there could be some component of the meaning of “Moses” (e.g. *figure from the Bible*) that fits well. Alternatively, we might say that “Moses” and “How many animals of each kind did ... bring on the ark?” are good fits to the same kind of (broader) context — conversations about Bible trivia. However, in the interest of clarity, we will use the term “goodness of fit” to refer to the extent to which the *word* “Moses” is a good fit to the *sentential context* “How many animals of each kind did ... bring on the ark?”, and we will use other labels for the other kinds of impostor-context relatedness just mentioned.

necessarily a detected one). Rather, the relationship between sentential context and the related vs unrelated impostors might be spelled out in terms of the conceptual features they evoke (e.g. *story from the old Testament*) or the broader topics in which they are appropriate (e.g. Bible trivia). As for which of these factors — impostor-intended similarity or impostor-context relatedness — is driving the effect observed for sentences like (139), we cannot tell based on these results alone, and we will return to the issue briefly.

This basic similarity effect has been replicated many times. van Jaarsveld, Dijkstra, & Hermans 1997 compared detection rates in a true/false judgment task for declarative versions (the exact materials are not provided) of sentences like (140), in which a “true” response indicates an illusion.

- (140) a. Om welke eilandengroep voerde Groot-Brittannië in de jaren tachtig oorlog met Brazilië?
- About which archipelago did Great Britain wage war in the eighties with Brazil (*Argentina*)?
- b. Om welke eilandengroep voerde Groot-Brittannië in de jaren tachtig oorlog met IJsland?
- About which archipelago did Great Britain wage war in the eighties with Iceland (*Argentina*)?

(van Jaarsveld, Dijkstra, & Hermans 1997)

They report 41% illusions for low-similarity impostors and 61% illusions for high-similarity impostors. The effect of similarity was statistically discernable. In a second experiment, they additionally manipulated whether the impostor occurred early or late in the sentence and found 49% and 46% illusions for low-similarity impostors in early and late positions, respectively, and 66% and 62% illusions for high-similarity impostors in early and late positions, respectively. The effect of similarity was statistically discernable at both positions and they observed no interaction with position.

van Oostendorp & de Mul 1990 compared highly-similar and less-similar impostors, such as “Moses” versus “Adam” in the ark sentence, the idea being that “Adam” and “Noah” share many features, though not as many as “Moses” and “Noah” (whereas “Nixon” and “Noah” have very little in common). Note that Erickson & Mattson 1981 also tested this particular substitution, though their investigation of the

degree of similarity was not as systematic. van Oostendorp & de Mul 1990 confirmed their intuition that their low-similarity impostors were in fact less similar to the intended word with a norming study. Participants generated as many attributes as they could think of for “Moses”, “Adam”, and “Noah” (and the equivalent words for other items), allowing confirmation that “Moses” and “Noah” in fact share more attributes than “Adam” and “Noah”. They additionally had participants generate attributes of “the omitted concept” for the sentential context “... took two animals of each kind on the ark”, again confirming greater overlap with “Moses” than with “Adam”. Note that while this comparison ostensibly disentangles impostor-intended similarity from impostor-context relatedness, the fact that the attribute generation task for contexts was about “the omitted concept” opens up the possibility that in this task, participants simply inferred that the omitted word was “Noah” and then generated attributes for Noah. Additionally, the ultimate experiment used stimuli that deliberately confounded impostor-intended similarity and impostor-context relatedness (since their norming study revealed a contrast between “Adam” and “Moses” on both dimensions), undermining any attempt to disentangle these. That said, the illusion experiment ultimately revealed illusions in 29% of trials for highly-similar impostors and 16% of trials for less-similar impostors, averaging across the 20 items tested. van Oostendorp & Kok 1990 replicated this effect, finding 30% and 17% illusions for highly-similar and less-similar substitutions, respectively, and additionally tested the effects of memorizing word pairs like “animals—Moses”, “ark—Moses”, “animals—Adam”, and “ark—Adam” prior to the illusion experiment. They found that having studied word pairs relating the impostor to words from the context (“animals” and “ark”) increased illusion rates by about 15 points, and this effect did not statistically discernably interact with the similarity effect. This suggests that the existence of associations between the impostor and the context can drive illusions up, though it does not tell us whether the contrast between “Moses” and “Adam” in illusion rates (or between “Moses” and “Nixon”, for that matter) is driven by such associations, *per se*.

Hannon & Daneman 2001 investigated why “susceptibility to semantic illusions is influenced by the semantic relatedness of both the impostor word and the surrounding context” (Hannon & Daneman 2001:449). Although this aim appears to mirror our distinction between impostor-intended similarity and impostor-context relatedness, their investigation of the properties of the context is importantly dis-

inct. That is, they are interested in the extent to which the context cues the answer to the trivia question, not the extent to which it is related to the impostor. This can be seen in (141), one of their stimuli which manipulates properties of the context and properties of the impostor orthogonally. The additional contextual cues “passenger liner” and “tragically sunk” are in not intended to be more related to the impostor “Pacific” than they are to the impostor “Indian”; rather these terms provide additional cues to the *answer*, “the Titanic”.

- (141) a. What passenger liner was tragically sunk by an iceberg in the Pacific (*Atlantic*) Ocean?
 (=*strong context, strong impostor*)
- b. What passenger liner was tragically sunk by an iceberg in the Indian (*Atlantic*) Ocean?
 (=*strong context, weak impostor*)
- c. What is the name of the ship that encountered an iceberg in the Pacific (*Atlantic*) Ocean?
 (=*weak context, strong impostor*)
- d. What is the name of the ship that encountered an iceberg in the Indian (*Atlantic*) Ocean?
 (=*weak context, weak impostor*)

(Hannon & Daneman 2001)

As in prior work, they report a boost in illusions for highly-related (“strong”) impostors relative to less-related impostors (34% and 46% illusions, respectively). They additionally report a small (4-point) but statistically discernable boost for contexts with more cues to the answer, relative to those with fewer cues. Their main investigation concerns the correlation of these effects with individual difference measures indexing participants’ working memory capacity and skill in accessing information in long-term memory. It is not clear from their analyses that there is sufficient variability across individuals in these measures to justify an individual differences approach (see Hedge, Powell, & Sumner 2018), so we do not explore these effects here.

Budiu & Anderson 2008 investigated accuracy and response times for highly-similar and less-similar

substitution illusions with and without sentential negation, as in (142). Note that this study, unlike the experiments discussed so far, used a true/false sentence judgment task, rather than a question-answering task. We will return to possible differences between these task types in section 7.1.4.

- (142) a. The ancient temple of Parthenon was built in Athens by the Romans (*Greek*⁷⁵).
 (=*highly-similar, affirmative*)
- b. The ancient temple of Parthenon was built in Athens by the Egyptians (*Greek*).
 (=*less-similar, affirmative*)
- c. The ancient temple of Parthenon was not built in Athens by the Romans (*Greek*).
 (=*highly-similar, negative*)
- d. The ancient temple of Parthenon was not built in Athens by the Egyptians (*Greek*).
 (=*less-similar, negative*)
- (Budiu & Anderson 2008)

Within the affirmative conditions, they found illusion rates of 37% and 31% for highly-similar and less-similar impostors, respectively. Note that this contrast is numerically smaller than what has been observed in other studies manipulating similarity, and the authors do not test whether the contrast is statistically discernable (as they are primarily interested in the processing of negation). For the negative sentences, note that the correct answer (for both the highly-similar and less-similar substitutions) is “true”, and so the illusion rate is the proportion of trials on which the participant selected “false”. For highly-similar negative sentences, the illusion rate was 41% and for less-similar negative sentences, it was 40%. These values were again not subjected to a statistical test.

Cook et al. 2018 also manipulated impostor similarity, though their primary aim was not the manipulation of illusion rates, but the pattern of eye-movements associated with detected and illusory impostors, which we return to in section 7.4.1. They also observe a contrast between highly-similar and less-similar impostors, with illusion rates of 37% and 11%, respectively. However, this contrast was only statistically discernable in the subjects analysis.

⁷⁵Presumably this should be the plural *Greeks* but they report it in the article as *Greek*.

A separate series of studies interested in the processing of full names in substitution illusions also manipulated similarity in meaning. Shafto & MacKay 2000 compared substitutions like (143). The critical comparison is between “Alan Shepard”, who, like the intended referent, Neil Armstrong, was an astronaut, and “Dizzy Gillespie”, who was a jazz musician (this will be relevant to their manipulation of phonological similarity in the same study, which we return to in section 7.1.2).

- (143) a. What was the famous line uttered by Alan Shepard (*Neil Armstrong*) when he first set foot on the moon?
- b. What was the famous line uttered by Dizzy Gillespie (*Neil Armstrong*) when he first set foot on the moon?

(Shafto & MacKay 2000)

Note that this experiment, unlike the others reported here, did not include a post-test to ensure that comprehenders had the relevant world knowledge that would enable them to detect the substitution. They did conduct several norming studies to ensure that the facts referenced by their stimuli were likely to be known by a large proportion of participants, but the possibility remains that some participants did not have the relevant knowledge for some trials. They also used a multiple choice answer format (rather than the more typical free response format), in which the four possible answers were “can’t say” (which is the correct response for questions with impostors), “don’t know”, the answer that would be correct if not for the impostor, and a distractor response. Trials receiving “don’t know” answers were discarded. This should somewhat alleviate concerns that participants did not have the relevant knowledge, but this is not a perfect control, since one could in principle know that the person who first stepped on the moon said “That’s one small step for man, one giant leap for mankind” without knowing the name of that person (thus producing data that looks like an illusion, but only due to a lack of world knowledge, not anything having to do with the processing of the sentence). They find illusion rates of 33% for the highly-similar impostors and 9% for the less-similar impostors.

Shafto & MacKay 2010 followed up on findings from Shafto & MacKay 2000 concerning the processing of full names with a manipulation intended to isolate the contribution of phonological similarity by reducing semantic similarity to zero. The basic idea is that although Neil Armstrong and Dizzy Gille-

spie are not particularly similar to each other, they were both men, both public figures, and both famous in the mid-twentieth century, and thus share some similarities. Shafto & MacKay 2010 therefore introduced made-up names, so that comprehenders would know nothing about the referent except the gender inferred on the basis of the first name, and so there is nothing that could potentially overlap. An example item is given in (144).

- (144) a. During which decade did Gerald Ford (*Henry Ford*) introduce his Model T to the world?
- b. During which decade did William Ford (*Henry Ford*) introduce his Model T to the world?
- c. During which decade did Laura Ford (*Henry Ford*) introduce his Model T to the world?
- (Shafto & MacKay 2010)

This manipulation is importantly different from other studies of semantic similarity, in that typically a highly-similar impostor is compared to a less-similar impostor, whereas here a less-similar impostor is compared to an impostor that is intended to be *even less* similar to the intended word, based on the assumption that unknown names have no semantic features associated with them. Thus (144b) overlaps only in gender and (144c) overlaps in no features at all. Confusingly, the authors do not seem to have modified the gendered pronoun “his” in (144c), even though the inference that the referent is a woman based on the name “Laura” is central to their investigation. They report illusions in 22% of trials for the less-similar impostors (Gerald Ford), 40% of trials for the unknown same-gender impostors (William Ford), and 27% of trials for the unknown different-gender impostors (Laura Ford). The contrast between low-similarity impostors and unknown same-gender impostors was statistically discernable.

This pattern is importantly different from the similarity effects reported in other studies, since (one of) the conditions that was intended to be *even less similar* to the intended word than the low-similarity impostor yielded more illusions. While this finding could be merely a Type S error (see Gelman & Carlin 2014), other explanations are possible. In principle there could simply be a non-linear trend in the effect of similarity, such that highly-similar impostors yield many illusions, less-similar impostors yield fewer

illusions, and the least-similar impostors again yield many illusions. This is somewhat unlikely, though. For example, the substitution “Taylor Swift” in the moon landing question discussed here would very likely be noticed very frequently, even though Taylor Swift is, intuitively, even less similar to Neil Armstrong than Dizzy Gillespie is. This contrast, assuming it exists, motivates an alternative way of thinking about the similarity effect — it is not that the impostor must have many encoded features in common with the intended word in order for illusions to occur, but rather the impostor must *not* have many encoded features that *distinguish* it from the intended word. One final possibility is that some of our assumptions about the way full names are processed are simply wrong — that is, comprehenders may identify referents based on only one piece of the name (in this case “Ford”) and effectively ignore the other, in a way that does not generalize to other linguistic units. This seems particularly plausible in the “Ford” case, since the present-day car company is called “Ford Motor Company”. Thus it may be that the high illusion rate for made-up names arises specifically because of the overlapping last name, and we should not expect such effects to arise when contrasting low-similarity and made-up names that do not share a surname with the intended referent (e.g., in the Model T example, “Herbert Hoover” versus “William Hoover” versus “Laura Hoover”).

One further study of the processing of full names compared the similarity of referents based on both visual similarity (i.e., Brad Pitt and Chris Hemsworth look alike, whereas Hugh Jackman looks somewhat less like either of them) and occupational similarity (i.e., Brad Pitt, Chris Hemsworth, and Hugh Jackman are all actors, whereas Rick Santorum is a politician) as in (145) (Davis & Abrams 2016). They additionally were interested in the impact of seeing pictures of the referents before presentation of the target sentences, which we do not explore in detail here.

- (145)
- a. Which movie features Chris Hemsworth (*Brad Pitt*) attempting to rob a casino?
 - b. Which movie features Hugh Jackman (*Brad Pitt*) attempting to rob a casino?
 - c. Which movie features Rick Santorum (*Brad Pitt*) attempting to rob a casino?
- (Davis & Abrams 2016)

For the conditions not preceded by a photo, they find illusions in 21% of trials for the visually-and-

occupationally similar impostors (Chris Hemsworth), 19% of trials for occupationally similar impostors (Hugh Jackman), and 18% of trials for dissimilar impostors (Rick Santorum). They do not subject these values to any statistical test.

Note that studies of substitution illusions using neural measures typically also include control comparisons intended to demonstrate the typical neural response to a (detected) anomaly, and they ensure frequent detection by using highly-dissimilar impostors in these control comparisons (Sanford et al. 2011; Bohan et al. 2012; Raposo & Marques 2013; Tune et al. 2014). However, these conditions are typically not closely matched to the illusion conditions, and so we do not discuss them here.

In sum, there appears to be strong evidence that substitution illusions are sensitive to the meaning of the impostor, such that low-similarity impostors yield fewer illusions than high-similarity impostors. Across the 13 experiments reviewed here, only one (Shafto & MacKay 2010) found an effect in the opposite direction, and only one (Budiu & Anderson 2008) found effectively no effect, as is shown in Figure 7.1. These anomalous findings might be independently explained by other factors distinguishing those studies — namely, the use of surname-overlapping full names, and the use of negation. Note, however, that both the size of the similarity effect (that is, the contrast between high-similarity and low-similarity impostors) and the illusion rate itself is highly variable across studies. This will become central to our investigation in Chapter 8. For the present purposes, it is clear both that similarity matters and that similarity is not the only thing that matters. Many studies report relatively high illusion rates even for low-similarity items, and many studies report relatively low illusion rates even for high-similarity items. Of course, much of this variability can likely be explained by differences in the participant samples (and sample sizes) and in the stimuli, task, and instructions. But we are confident that there is no single stimulus such that a high-similarity impostor would yield illusions in 100% of trials and a low-similarity impostor would yield illusions in 0% of trials. Thus, claims along the lines of “comprehenders do not notice the substitution because the impostor is highly similar” are unsatisfying not only because they are not mechanistic but also because they are empirically inadequate — a high degree of similarity is neither necessary nor sufficient for illusions to occur on some proportion of trials.

There is strong evidence that the similarity of the impostor matters for illusion rates, but the question

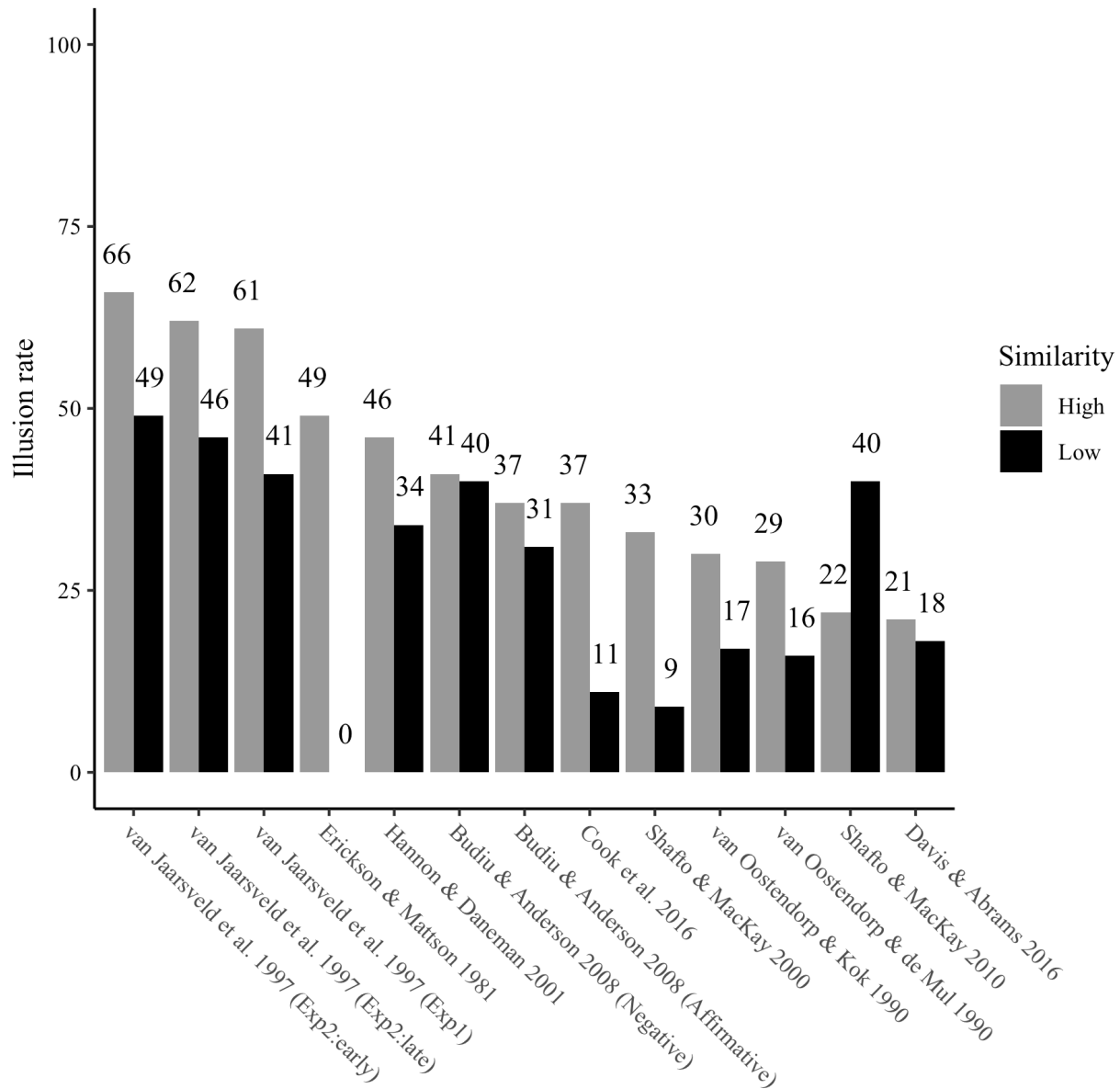


Figure 7.1: Substitution illusion rates for similar and dissimilar impostors, across studies

remains: similarity to *what*? Are illusion rates sensitive to the degree of overlap between the impostor and the intended word or the degree of overlap between the impostor and the context? None of the studies reviewed here successfully disentangle these factors. However, we might reasonably suspect that both factors matter. Consider the examples in (146) and (147).

- (146) a. Which British monarch formally opened the Olympic winter (*summer*) games in Lon-

don in 2012?

- b. Which British monarch formally opened the Olympic spring (*summer*) games in London in 2012?

- (147)
- a. What is the name of the holiday when children dress up in costumes and walk door to door giving out (*receiving*) candy?
 - b. What is the name of the holiday when children dress up in costumes and walk door to door scaring (*receiving*) candy?

Both examples in (146a) and (147a) are known to yield relatively high illusion rates. We have not collected data on the (b) examples, but we will assume, based on our own impressions of the sentences, that they yield fewer illusions, compared to their counterparts. First, consider (146a) and (146b). Without taking context into account, the meaning of “spring” is unlikely to be dramatically less similar to “summer” than “winter” is similar to “summer”. The intuition that there is likely a difference in the illusion rates of these stimuli suggests that it is not merely decontextualized impostor-intended similarity that matters — that is, the fact that “spring” seems to “jump out” as irrelevant, in virtue of the fact that it is not a season in which the Olympics are held suggests that impostor-context relatedness must be taken into account. Similarly, if we conceptualize impostor-context relatedness as simply shared topic-hood or the likelihood of evoking similar concepts, the intuition that (147b) would be less likely to yield illusions than (147a) becomes critical. “Scaring” is presumably highly related to the concepts evoked by “What is the name of the holiday when children dress up in costumes and walk door to door ... candy?”, in that both express Halloween-related ideas. So, a difference between (147b) and (147a) could be attributed to the lack of semantic similarity between “scaring” and “receiving”. One might object that “scaring” and “giving out” are simply not well matched, since “scaring candy” is an unlikely event whereas “giving out candy” is a more plausible one. Thus, the contextual fit of the two items is not identical. In other words, one could argue that although “scaring” is topically related to the sentential context, shared topic-hood is not precisely what is meant when we consider the effects of context-relatedness. This is a reasonable concern, but a precise definition of context relatedness remains elusive. As noted above, it cannot be reduced to goodness-of-fit of the word itself because, importantly, no impostor is a good fit to the con-

text — they all result in world knowledge violations, which is inherent to the definition of the illusion. We might attempt to incorporate a *local* goodness-of-fit component, such that the appropriateness of “giving out candy” as compared to “scaring candy” can be taken into account. But we suspect that this is not all there is to say on the matter, since impostors like “buying” might also yield reduced illusions, even though “buying candy” is a perfectly reasonable event. What this discussion highlights, however, is that it is quite difficult to independently manipulate impostor-intended similarity and impostor-context relatedness. We suspect that the difficulty of generating such items explains the lack of such an investigation in the literature, despite the fact that many authors have observed that both dimensions of similarity could in principle be relevant.

Of course, the findings discussed here do not tell us *why* similarity matters, only that it does. But given the robustness of the finding, any plausible mechanism proposed to explain the illusion should be able to accommodate this effect.

7.1.2 Similarity in form

While the impact of similarity between the meaning of the impostor word and the meaning of the intended word (and/or the meaning of the context) is well established, the importance of similarity in forms is less clear. The possibility that phonological similarity could matter is suggested by the initial “Moses” stimulus, repeated in (148). Both “Moses” and “Noah” have two syllables, those syllables feature similar vowels, and both begin with a nasal. They are also, to some extent, orthographically similar, being approximately equal in length. A second early example, also from Erickson & Mattson 1981, given in (149) also displayed some degree of phonological and orthographic overlap.

(148) How many animals of each kind did Moses (*Noah*) bring on the ark?

(149) In the biblical story, what was Joshua (*Jonah*) swallowed by?

Note that while these examples raise the possibility that phonological similarity may matter, it is clear from the examples discussed already that phonological similarity cannot be a critical ingredient for illusions. Examples like (150) have much less phonological overlap but still demonstrate robust illusions.

- (150) What is the name of the holiday when children dress up in costumes and walk door to door giving out (*receiving*) candy?

In order to determine if phonological overlap contributes to illusions, Erickson & Mattson 1981 tested both “Adam” and “Abraham” as impostors in the contexts from both (148) and (149). The idea was to use impostors that were still highly related (since they are also biblical figures) but with less phonological overlap, though note that later work found that at least “Adam” is less similar to “Noah” than “Moses” is, based on an attribute-generation task (van Oostendorp & de Mul 1990), so we should not consider these impostors to be perfectly well-matched for semantic similarity. Nonetheless, Erickson & Mattson 1981 report illusion rates of 70% for “Adam” and 44% for “Abraham” in the ark question, compared to a baseline of 49% for “Moses”. In the question frame from (149) they found illusion rates of 0% for “Adam” and 17% for “Abraham”, compared to a baseline of 39% for “Joshua”. Thus, the results are rather mixed, but the authors conclude that there is no clear impact of phonological similarity.

Later work has come to somewhat different conclusions. Shafto & MacKay 2000 and Shafto & MacKay 2010 both investigated this issue through the lens of full names with overlapping surnames. Shafto & MacKay 2000, as discussed above in Section 7.1.1, compare sentences like (151), finding 23% illusions for (151a) compared to only 9% for (151b). While both Louis Armstrong and Dizzy Gillespie were jazz musicians and therefore semantically dissimilar to astronaut Neil Armstrong, “Louis Armstrong” yields substantially more illusions as an impostor due to the shared surname. They label the phenomenon the “Armstrong illusion”. Shafto & MacKay 2010 followed up on this finding, demonstrating that such name-overlapping impostors can yield illusions even in the absence of any semantic overlap (see section 7.1.1).

- (151) a. What was the famous line uttered by Louis Armstrong (*Neil Armstrong*) when he first set foot on the moon?
- b. What was the famous line uttered by Dizzy Gillespie (*Neil Armstrong*) when he first set foot on the moon?

(Shafto & MacKay 2000)

These results suggest that phonological overlap may be relevant to substitution illusions, though these effects are far less extensively demonstrated than semantic similarity effects. It is important to bear in mind that all investigations of this effect have focused on names. Presumably this is because of the difficulty in generating stimuli that manipulate phonological similarity while holding semantic similarity constant. That said, there may be important was that the processing of full names differs from the processing of other linguistic units, and so these findings may not extend to the more general substitution illusion effect.

7.1.3 Presuppositional status

One key observation about the original “Moses illusion”, repeated in (152) is that the question-asker seems to presuppose that Moses was involved in the ark story. Presuppositions, largely by definition, are not “up for debate” and so it is not surprising that a comprehender hearing (152) does not instinctively consider the possibility that this is false.

(152) How many animals of each kind did Moses (*Noah*) bring on the ark?

In fact, the first demonstrations that substitutions in presupposed content are less detectable than substitutions in focused content predates the first demonstration of the Moses illusion. Hornby 1974 showed participants pictures of simple scenes, such as a boy petting a cat. Participants then heard a sentence like those in (153) and (154), which they had to judge to be true or false. Participants made about twice as many erroneous “true” judgments when the impostor was in the presupposition, as in (154) compared to when it was in the focused content, as in (153).

- (153) a. It is the girl (*boy*) that is petting the cat.
b. It is the dog (*cat*) that is being petted by the boy.
- (154) a. It is the cat that is being petted by the girl (*boy*).
b. It is the boy that is petting the dog (*cat*).

(Hornby 1974)

Baker & Wagner 1987 demonstrated a similar effect for trivia statements, rather than pictures being evaluated. They also usefully controlled for potential order effects using coordinated structures as in (156). Participants were less likely to identify the falsehood (that is, more likely to experience an illusion) for (155a), in which the falsehood is presupposed, than for the other three sentence types.

- (155) a. Hieroglyphics, which is usually associated with the Russians (*Egyptians*), is a kind of picture writing.
- b. Hieroglyphics, which is a kind of picture writing, is usually associated with the Russians (*Egyptians*).
- (156) a. Hieroglyphics is usually associated with the Russians (*Egyptians*) and is a kind of picture writing.
- b. Hieroglyphics is a kind of picture writing and is usually associated with the Russians (*Egyptians*).

(Baker & Wagner 1987)

Bredart & Modolo 1988 applied this manipulation to typical substitution illusion sentences like the Moses illusion, comparing illusion rates for sentences like (157) in a true/false judgment task. Again, illusion rates were substantially higher when the impostor is part of the presupposed content as in (157a). Notably, this trend arose for every single one of the 10 items they tested.

- (157) a. It was two animals of each kind that Moses (*Noah*) took on the ark.
- b. It was Moses (*Noah*) who took two animals of each kind on the ark.

(Bredart & Modolo 1988)

Sturt et al. 2004 found effects of presuppositional status in a change-detection paradigm. Participants read passages like (158a) and (158b), which differ with respect to whether “cider” is part of the presupposed content. After reading the passage, participants were presented with the same text, but with “cider” changed to “beer”. Their task was to report changes in the text when they occurred. Detection rates were lower when “cider” was part of the presupposed content as in (158a).

- (158) a. Everyone had a good time at the pub. A group of friends had met up there for a stag night. It was Jamie who really liked the cider, apparently.
- b. Everyone had a good time at the pub. A group of friends had met up there for a stag night. What Jamie really liked was the cider, apparently.

(Sturt et al. 2004)

Other work has shown that using capitalization or boldface font for the impostor can similarly reduce illusion rates (Bredart & Modolo 1988; Kamas, Reder, & Ayers 1996; Cantor & Marsh 2017). This is expected if these font changes are to be understood as indicating contrastive focus, thereby guaranteeing that the impostor is part of the asserted, not presupposed, content. However, other explanations are also possible — capitalization and boldface font may simply draw the comprehender’s attention.

In sum, it appears that illusions are in fact more likely to arise when the impostor is part of the presupposed content of the sentence compared to when it is part of the asserted content, though illusions are possible in both cases.

7.1.4 Questions versus statements

The effect of presuppositional status has some bearing on the issue of whether the use of a question format is central to the illusion, since questions, under some accounts, presuppose that an answer exists. Independently, questions may induce higher illusion rates because they typically involve a kind of “multi-tasking” in which participants have to recover the answer to the question *and* determine if the question contains an impostor. All else equal, we might expect doing both of these things simultaneously to be more taxing, and thus result in more errors, compared to the “mono-tasking” situation of simply evaluating whether a statement is true or false. That said, it is clear from the literature that questions are not a necessary ingredient for illusions to occur. Out of 31 studies investigating substitution illusions, 16 used primarily questions which participants answered and 15 used primarily sentences which participants verified. All of these found some degree of illusions.

In addition, two studies have directly compared questions and statements. Both Erickson & Mattson 1981 and Büttner 2007 found reduced but non-zero illusion rates for statements, as compared with

questions. For Erickson & Mattson 1981, the reduction was from approximately 52% illusions to approximately 27%, but note that they only tested four stimuli. For Büttner 2007, the reduction was from approximately 48% illusions to approximately 31%. Thus it appears that successful accounts of the substitution illusion should accommodate this generalization. However, any account that predicts that illusions will occur *exclusively* in questions is clearly inappropriate. Furthermore, as we will see in Chapter 9, the question-versus-sentence effect may not be as robust as prior studies suggested.

7.2 What could go wrong?

The above discussion of some key empirical generalizations places some constraints on the possible mechanisms that may explain the substitution illusion. In brief, there is robust evidence that impostor similarity in meaning (to the intended word and/or the context) and presuppositional status can influence illusion rates, though neither factor can eliminate them completely. There is additionally some evidence that impostor similarity in form (to the intended word), and the use of questions versus statements may matter, though these generalizations are less clear. With these empirical issues in mind, we now turn to the possible mechanisms underlying the illusion.

Recall that the basic fact we want to explain is the failed detection of world knowledge violations in substitution illusion sentences. This is in some ways a different issue from the question of the successful retrieval of the answer to the (intended) question — that is, “two” in the classic Moses example repeated in (159a) and “Halloween” in the giving-out-candy example (159b). This latter fact is likely explained by the mechanisms underlying the retrieval of facts from long term memory. That is, there is a relative abundance of cues to the queried fact, such that the questions can be answered based on the contextual information alone, as in versions like (160a) and (160b). The cues provided by “animals of each kind” and “ark” are sufficient to retrieve the answer. What we can determine about this process based on substitution illusions is that retrieval of the answer does not appear to be derailed by the existence of a single mis-matching cue in the search of long term memory (though in fact, even this may not need to be true, depending on how far the processing of the impostor goes). But this is not surprising given independently

motivated assumptions about the nature of memory retrievals — namely, that they involve parallel activation of cues, resulting in an increase in the activation of the target until it passes some threshold (e.g., McElree 2006, among others). Such models typically do not incorporate a mechanism whereby a cue that does not target the same representation as the other cues would decrease the activation of that representation, preventing successful retrieval, and so the successful retrieval of “two” and “Halloween” from memory are expected.

- (159) a. How many animals of each kind did Moses (*Noah*) bring on the ark?
 b. What is the name of the holiday when children dress up in costumes and walk door to door giving out (*receiving*) candy?
- (160) a. How many animals of each kind were brought on the ark?
 b. What is the name of the holiday when children dress up in costumes and walk door to door?

Our question, rather, is why comprehenders say “two” or “Halloween” instead of flagging the violation. This is presumably because they do not detect the violation⁷⁶, and so the question becomes why the violation goes undetected. Some brief discussion of why this non-detection is surprising is warranted. There are effectively two varieties of reasons to expect the violation to be detected: the empirical generalization (world knowledge violations typically are detected, and quite quickly) and the theoretical motivation (our assumptions about sentence processing predict that the violation should be detectable).

On the first point, much evidence suggests that a wide variety of anomalies in sentences are obvious to comprehenders. For example, the anomaly in the question in (161) is clear, and comprehenders likely have no trouble identifying this. Many similar examples can easily be constructed. In fact, generating anomalies that are unlikely to be detected (i.e., generating good substitution illusion stimuli) is much more challenging.

- (161) What do purple giraffes eat for breakfast?

⁷⁶In principle it could have been otherwise: comprehenders might have failed to report the violation because they determined that it was a speech error. This possibility is ruled out by the use of tasks in which comprehenders are informed of the existence of errors and asked to point them out, as discussed in the introduction to this chapter

Note that substitution illusions involve world knowledge violations, not linguistically ill-formed strings. That is, a sentence like (162) is perfectly linguistically sound, it just happens to be false. Similarly, a question like (163) is well-formed, but presupposes a fact that happens to be false (i.e. that there is such a holiday, in which children are the candy-givers).

(162) The name of the holiday when children dress up in costumes and walk door to door giving out (*receiving*) candy is Halloween.

(163) What is the name of the holiday when children dress up in costumes and walk door to door giving out (*receiving*) candy?

One reasonable way to construct a system that can represent sentence meanings and determine if they are true or false would be to treat these operations as two distinct processes, the latter of which is dependent upon the completion of the former. As we will see, under such assumptions, the existence of (some) substitution illusions is not particularly surprising; however, we will also see that these assumptions are not well-motivated.

Such a two-stage model is effectively the nature of the procedure Clark & Chase 1972 propose, though their model is primarily concerned with evaluating the truth of statements with respect to pictures, not general world knowledge, so it involves additional steps to encode the picture. In such a system, the evaluation procedure cannot begin until the sentence has been successfully encoded, raising the possibility that the substitution illusion in (163) reduces to a timing problem — the answer to the intended question is accessed (and produced) before the evaluation procedure is complete, maybe even before the sentence is over (and so before the evaluation procedure has begun). However, this is unlikely for two reasons. First, such a model cannot explain the existence of illusions for declarative sentences like (162), which also yield illusions (though possibly at lower rates; see section 7.1.4). In a true/false judgment task there is nothing else to do but evaluate the sentence, and so there's no process that could finish first. Second, more recent research has called into question the assumption in the Clark & Chase 1972 model that evaluation can only begin after the sentence encoding stage is complete. Intuitively, the world knowledge violation of “purple giraffes” in (161) is detectable before the conclusion of the sentence. Furthermore, Hagoort et al.

2004 demonstrated that the world knowledge violation in (164a) is reflected in the same ERP component (the N400) as the semantic violation in (164b).

- (164) a. The Dutch trains are white (*yellow*) and very crowded.
 b. The Dutch trains are sour (*yellow*) and very crowded.

(Hagoort et al. 2004)

Later work suggests that the immediate detection of the anomalies is not likely attributable to low-level associations between lexical items. Nieuwland & Van Berkum 2006 found that in the context of a fictional story about an anthropomorphized peanut who sings, dances, and falls in love, the N400 reflects the anomaly of (165a) relative to (165b), despite the low-level association between “peanut” and “salted”.

- (165) a. The peanut was salted...
 b. The peanut was in love...

(Nieuwland & Van Berkum 2006)

Thus, it is clear that many falsehoods are not only easily detected by comprehenders, but detected immediately upon encountering the anomalous word. Substitution illusions are therefore a surprising exception to this robust generalization.

We now turn to some standard assumptions about the nature of sentence processing which predict (correctly, for the most part, in light of the generalization just mentioned) that world knowledge violations should be quickly detected. This is a brief overview of all of the processing steps that would need to go right in order for the impostor to be detected, and thus constitutes a set of candidate operations that could be the locus of the error underlying failed detection.

To be brief but also as explicit as possible, in order to detect the impostor, a comprehender must complete all (or at least most) of the following steps, though not necessarily precisely in this order: (a) represent the linguistic material prior to the impostor at a variety of levels of representation, including syntactic and semantic levels, as well as potentially retrieving the world knowledge that is encoded in long term memory about both the individual entities mentioned in the context and the compositional meaning of the context, (b) fixate the impostor (or hear it in auditory tasks, which are in fact rarely used in

substitution illusion experiments), (c) identify the visual word form corresponding to the impostor, (d) identify the lemma corresponding to the impostor, (e) identify the semantic information corresponding to the impostor, (f) identify relevant world knowledge⁷⁷ about the impostor in long term memory, (g) incorporate the impostor lemma into the syntactic representation of the sentence in a way sanctioned by the grammar, (h) incorporate the impostor semantic representation into the semantic representation of the sentence compositionally, (i) transform the newly composed representation into a probe for world knowledge, and (j) detect, at a conscious level, the misalignment between the probe and what is already encoded in long term memory.

Note that step (a) glosses over many operations that must occur for the prior words in the context, which are made explicit with regard to the impostor in steps (b)-(i). Furthermore, these steps must all be executed for the contextual information that *follows* the impostor, as well, since the information generating the conflict between the context and the impostor is not always prior to the impostor in the string.

Having spelled out these operations, we can ask which ones could go wrong such that impostor detection fails. There are a few operations that we consider unlikely candidates. Importantly, the failure of these operations does not result in global failure in comprehension. So, for example, if comprehenders encountered difficulty in step (g), and struggle to identify a syntactic parse that is consistent with the input string, this would plausibly result in perceived difficulty in sentence comprehension. It is furthermore not clear how a parsing failure could result in failed detection of the world knowledge violation, since substitution illusions are always syntactically well-formed. Thus, we will not further consider step (g) as a candidate error point. In addition, we know from eye-tracking analyses that illusions are not simply a consequence of word skipping — they occur even when the impostor is directly fixated (Bohan & Sanford 2008; Cook et al. 2018). This makes (b) an implausible error point. We will also make the simplifying assumption that the context is processed successfully, thus ruling out step (a) as a candidate error point.

This leaves us with a series of plausible error points, which we group and evaluate as follows: problems in lexical access (steps (c)-(e)), problems in accessing world knowledge related to the impostor word

⁷⁷The ways in which this step is meaningfully distinct from the previous step will depend on our theory of word meanings, as we will see in the following sections.

(step (f)), problems in semantic composition (step (h)), and problems in accessing and evaluating world knowledge at the sentence level (steps (i)-(j)). We now spell out how exactly each of these operations could be fallible in such a way as to result in non-detection of the impostor.

7.2.1 Lexical access

Assuming lexical representations roughly analogous to those proposed in Levelt, Roelofs, & Meyer 1999 (that is, representations at the word form, lemma, and semantic level⁷⁸; see Figure 7.2), the possibility of problems in lexical access consists of three possible problems, corresponding to failure at each of those levels. Here we assume lexical access procedures at each of these levels in which candidate entries increase in activation as a function of the activation of the nodes they are connected to, until one entry reaches enough activation to cross some threshold and be selected (as in, for example, the TRACE model, McClelland & Elman 1986). Selection at a lower level feeds forward to increase the activation for the corresponding representation at the next highest level, and only the selected representations are used for computations that require such representations (for example, only the selected lemma enters into the syntactic representation of the sentence)⁷⁹.

For each of these, there are, in principle, two ways things could go wrong: either the entry that is selected for subsequent computations is the entry for the intended word⁸⁰ rather than the entry for the impostor, or nothing is selected, and subsequent operations that require, as input, a representation at that level instead use some kind of “dummy” representation that has the appropriate syntactic category but

⁷⁸In the discussion here we aim to be relatively agnostic about the precise nature of semantic-level lexical representations. Note that Levelt, Roelofs, & Meyer are committed to atomic “lexical concepts” at the semantic level. Lexical concepts are concepts that happen to have lemmas attached to them; in this sense, they are a part of the conceptual system, not a distinct linguistic representation. Some alternative ideas about what the semantic level could consist of include decomposable semantic representations (which Levelt, Roelofs, & Meyer 1999 argue against as the representation of the word meaning *per se*, but they do have conceptual features in the system, which can be accessed *through* the lexical concept), instructions to fetch a concept at an address (Pietroski 2018), or pointers to regions in conceptual space (Carston 2012). The last of these will become highly relevant to our discussion of polysemy in the following section.

⁷⁹This assumption is not adopted by all models of sentence comprehension. Notably, the noisy channel framework (Levy 2008) discussed in Chapter 5 and Chapter 6 assumes instead that multiple parses are constructed in parallel using multiple distinct lexical entries that might have been the seen/intended word. We cannot rule out such a model but we adopt the simplifying assumption here that only one word form / lemma / semantic representation is selected for participation in subsequent representations.

⁸⁰Technically, there is a third option, which is mis-selection of some other word (i.e. neither the impostor nor the intended word). But it’s not clear how this would lead to illusions so we don’t consider it further.

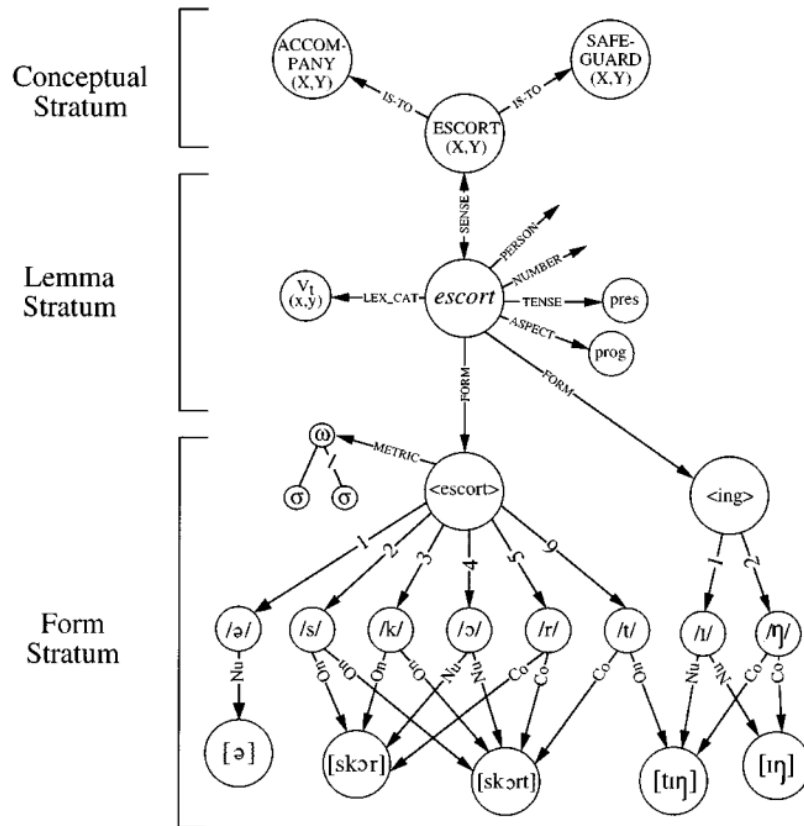


Figure 7.2: The lemma model of lexical representations (Levelt, Roelofs, & Meyer 1999:4)

very little else. Such dummy representations may seem counter-intuitive but something like this seems to be assumed for parsing Jabberwocky sentences, (i.e. there is no entry in the lexicon with that form, but comprehenders appear to be able to construct a lexical entry with the right syntactic category on the fly). The mis-selection version of this hypothesis has much in common with proposals from Shafto & MacKay 2000, Shafto & MacKay 2010, and Davis & Abrams 2016, though they are all specifically concerned with full name processing, so some of their assumptions are unique to names. Roughly, the idea is that some representation of the intended word (for them it is a “name phrase node”, but we might translate this to any of the three lexical representations) increases in activation due to top-down influence of the context. In the classic Moses illusion case, this would mean that the comprehender infers that the Bible story about an ark is being discussed, and this leads to increased activation of the “Noah” representation at the semantic and/or lemma and/or word form level of representation (i.e. “pre-activation” or “prediction” of a likely upcoming word based on context). The representation of the intended word additionally receives

bottom-up activation if it overlaps in form with the impostor (this activation comes directly from the activated orthographic or phonological features). Thus the convergence of these sources of activation is, on some trials, enough for the entry for the intended word to cross the threshold for selection.

Given these two possibilities for the nature of the lexical access problem, we can consider them at each lexical level. Both Erickson & Mattson 1981 and Shafto & MacKay 2000 attempt to rule out at least the first possible error points, failure in word form access. They use a read-aloud task and a shadowing task, respectively. The logic is similar, so we focus on the simpler read-aloud task. The basic idea is that the mapping from the orthographic representation of the impostor to the phonological representation of the impostor requires at least access to the (shared) word form. Thus, the existence of illusions even for read-aloud trials suggests that illusions are not a product of either mis-selection or failed selection at the word form level. The logic here is not airtight, since many standard models of read-aloud tasks, such as the Dual Route Cascaded model (Coltheart et al. 2001), allow for a path from orthography to phonology that does not require access to a shared word form node — that is, roughly, the ability to “sound it out”. But this is somewhat unlikely given that many impostor words are extremely common words that competent adult readers would know on sight.

A second consideration regarding the possibility of failures in word form access is the existence of robust effects of the similarity in meaning between the impostor and the intended word and/or relatedness in meaning between the impostor and the context. It is not immediately obvious how such effects can be explained if the processing of the impostor is cut off at the word form level⁸¹. We first consider the implications of this for mis-selection accounts (as opposed to failed selection). In the classic Moses illusion case, the claim is that the word form representation of “Moses” is not selected, and the “Noah” word form is selected instead. From this point on, “Noah” representations are the only ones that are accessed at all. Neither the semantic representation of “Moses” nor the world knowledge associated with

⁸¹Mis-selection accounts might fare better with respect to the semantic similarity effect if we assume a predictive coding mechanism (Rao & Ballard 1999). In such a framework, the current high-level representation allows predictions to be pushed down through the network, such that the only strong signal moving bottom-up is the *error signal* — that is, the extent to which the representation of the actual stimulus deviates from the predicted stimulus. This is a meaningfully different model than the convergent-activation explanation for illusions proposed by Shafto & MacKay 2000 and others. It is possible that such a model would straightforwardly account for the similarity effect by generating a bigger error signal for, e.g., “Nixon” than for “Moses”, but without an explicit model of how predictive coding could be applied to the substitution illusion, this is difficult to evaluate.

it is ever encountered at all. Thus, “Nixon” should be just as good an impostor. In brief, if we never reach any meaning-related representation of the impostor, there’s no way for impostors with different meanings to have different effects. The exact same issues arise for mis-selection at the lemma level, and even, potentially, at the semantic level (depending on whether the relevant type of “related meaning” is encoded as part of the semantics of the linguistic object or as part of world knowledge that is associated with the referent which lives outside the lexical representation altogether).

Importantly, top-down activation from the context, as Shafto & MacKay 2000, Shafto & MacKay 2010, and Davis & Abrams 2016 assume, does not help. This mechanism serves to increase the activation of “Noah” enough that it might sometimes reach threshold, but it would not necessarily have any bearing on the activation of “Moses”. In fact, if it were to influence the “Moses” representation, it would presumably be to increase its activation through spreading activation from the topic evoked by the context or from the conceptual features that “Moses” and “Noah” share. Top down influences would surely not have such an effect on “Nixon”. This ultimately predicts, counter-intuitively, that “Noah” representations should be mistakenly selected (resulting in illusions) *less often* when competing against “Moses”, which receives some top-down activation and is therefore strong competition, than when competing against “Nixon”, which receives no top-down activation.

One way out of the issue of the semantic similarity effect is to incorporate lateral connections between the impostor and the intended word, which are based on something like semantic relatedness, semantic similarity, or shared topic-hood. If such connections exist at the semantic level, and this allows activation to flow from “Moses” to “Noah”, then mis-selection at the semantic level would be more likely for a related word with strong connections to “Noah” than for an unrelated word with weaker ones. The same logic applies at the lemma level and word form level, though the possibility of topic- or meaning-based connections at these levels are somewhat less in the spirit of what the distinction in levels is intended to capture. Note also that lateral facilitation is not often assumed in models of word recognition, and in fact lateral inhibition is sometimes proposed (e.g. in the TRACE model of spoken word recognition, McClelland & Elman 1986).

Another option is to allow activation to flow through the entire system prior to a selection at any

individual level. Thus, at some early time point when the question of whether the current word form (or current lemma, or current semantic node) is “Moses” or “Noah” is not yet settled, activation continues to percolate up to the world knowledge representations associated with those lexical entries. This would allow the “Nixon” world knowledge to be accessed before mis-selection occurs⁸², thus allowing for some error signal to be generated quickly, possibly preventing the selection error from even taking place (or, even if it does take place, perhaps it does not matter because the comprehender already knows there’s a problem).

Turning now to failed-selection versions of the lexical access hypothesis (as opposed to mis-selection of the intended word), some of the same concerns about the semantic similarity effect apply. That is, if the meaning of the impostor is never accessed, there is no obvious way for the irrelevance of the meaning of “Nixon” to matter. The same varieties of explanations could be adopted to accommodate the effect — lateral connections between semantically-related nodes or rapid activation of world-knowledge based on candidate words (none of which has been selected). One interesting difference is that in failed selection accounts, there is no reason the lateral connections would need to be facilitatory — lateral inhibition between two strong candidates could be responsible for the lack of a winner.

In sum, problems in lexical access could in principle underlie the substitution illusion in either of two ways: the lexical access procedure may select the “Noah” lexical entry even though “Moses” is being seen, or the lexical access procedure may fail to select any lexical entry at all. There are two possible refinements to an intuitive theory of mis-selection or failed selection which allow such models to capture the frequently-observed semantic similarity effect: lateral connections between lexical representations, or a world-knowledge-access process that is not contingent on lexical selection. Each of these possibilities makes some predictions for both illusion processing and lexical access outcomes more generally. Some of these predictions are tested in Chapter 9.

⁸²Shafto & MacKay 2000, Shafto & MacKay 2010, and Davis & Abrams 2016 seem to have in mind roughly a hybrid of these two possibilities. In their model, “Noah” receives activation from “Moses” (as in our lateral-connections version) but only indirectly, via nodes that they both connect to. Those nodes are world knowledge representations (e.g. “saved people” and “in the Old Testament”), indicating that the authors assume that world knowledge of candidate words can be activated before selection occurs. Note that they use the term “activation” where we use “selection” and the term “priming” where we use “activation”.

7.2.2 Identifying concepts at the (sub-)word level

We now assume that lexical access is successful, contrary to the hypotheses discussed in 7.2.1. Here we consider the possibility that the problem lies instead in the identification of world knowledge features that are associated with the impostor. We assume that there are many things comprehenders know about the impostor — using the classic “Moses” illusion example, comprehenders have, stored in long term memory, the knowledge that Moses is an Old Testament figure, that he was involved in the story of the crossing of the Red Sea, and potentially much else. Comprehenders also, of course, know much about “Noah” and some of this knowledge will be very similar to what is known about “Moses” (e.g. Old Testament figure). These overlapping facts could be encoded as connections to shared nodes (as Shafto & MacKay 2000, Shafto & MacKay 2010, and Davis & Abrams 2016 assume) or simply as distinct representations that happen to share content.

The critical idea for this family of hypotheses is that not everything one knows about a referent needs to be accessed in order to successfully comprehend a sentence about that referent, or even to judge it true or false. For example, in order to verify the sentence “The official cat of the state of Maryland is the calico”, one does not need to access the world knowledge representation of the fact that Maryland borders Virginia. This knowledge may be encoded in long term memory and may even be encoded in a way that is closely connected to the lexical representation of “Maryland”, but it can be accessed or not accessed, with no consequences for the comprehension of the target sentence. Regularly accessing everything that is known about the referent may in fact cause problems, since, in some cases there is so much stored knowledge. Thus, an efficient system might regularly not access much of what is known about a referent in the course of normal sentence comprehension.

Thus, a comprehender processing the non-illusory sentence in (166) might reasonably not access everything that is known about Noah, and instead only retrieve a subset of that information like “Old Testament figure”, because that is what is needed for understanding the sentence⁸³. The process of selec-

⁸³It’s worth noting that the “Old Testament figure” representation may be all that’s needed for understanding the sentence, it’s clearly not all that’s needed for judging the sentence true or false. Thus a key part of the claim, which has been mostly implicit, is the assumption that the sentence comprehension system is optimized for the comprehension task, not the verification task, even though verification is something we can use the system to do.

tively targeting the “Old Testament figure” sub-part of what is known about the referent would yield an identical output for (167), since “Moses” and “Noah” have this knowledge in common.

(166) Noah brought two animals of each kind on the ark.

(167) Moses brought two animals of each kind on the ark.

This account successfully predicts the well-documented semantic similarity effect. Moreover, it predicts that not only impostor-intended similarity but also impostor-context relatedness matter, since the impostor and intended must not only share conceptual features, but must share precisely the features that the context is likely to pull out. Although we have so far articulated the error in terms of the retrieval of world knowledge about a referent based on a name, the account straightforwardly extends to other word substitutions. It merely requires that not everything that is known about the meaning of a word is retrieved on every encounter with that word, and retrieval emphasizes contextually-relevant features over contextually-irrelevant ones⁸⁴.

This hypothesis strongly resembles what Erickson & Mattson 1981 originally proposed. They similarly note that much is known about many referents, and a large proportion of that knowledge is not contextually relevant for any particular sentence about that referent. “Thus interrupting processing every time a semantic feature failed to fit would be disadvantageous ... A more likely mechanism would be to interrupt processing only if there were very few or no semantic features of a word which fit, making construction of a complete description of the meaning of the sentence difficult or impossible” (Erickson & Mattson 1981:550). There are a couple of potential differences between this account and what we propose here. First, in the Erickson & Mattson 1981 version, non-contextually relevant facts (which they call “semantic features”) are accessed, but the detection of these facts does not lead to a disruption in processing⁸⁵. A second, related difference is that, for them, it is only when there are few or no contextually-related

⁸⁴Note that we have so far framed the problem as one in which the retrieval of world knowledge is targeted to only contextually-relevant features. However, it is equally possible that retrieval itself is not targeted (i.e. everything that is known about the referent is accessed at first), but the retrieved knowledge is quickly and efficiently filtered, resulting in only the contextually-relevant features.

⁸⁵Why not? Their answer is, essentially, that a system that triggers a disruption every time it encounters a word that is associated with some non-contextually-related facts would be a system that spends a lot of its time dealing with disruptions. That is, most (content) words have a lot of different pieces of world knowledge attached to them, and most of that knowledge is not relevant most of the time. Thus, under their account, the reason there’s no error signal is because the existence of

facts associated with the impostor that sentence processing effectively fails and the anomaly is detected. This is a somewhat odd claim because comprehenders can, with relatively little difficulty, make sense of a sentence with a novel word, as long as there is sufficient contextual support to infer the meaning of the novel word. Thus, while both versions are in principle possible, we will assume a version in which non-contextually-related facts associated with the impostor are simply not accessed (or are quickly filtered out), resulting in the illusion, rather than a version where they are accessed but do not disrupt processing.

Somewhat mysteriously, subsequent work on substitution illusions did not pursue this early proposal. Rather, much of the research in this area investigated the related “partial matching”⁸⁶ hypothesis proposed in Reder & Cleeremans 1990, Reder & Kusbit 1991, and Kamas, Reder, & Ayers 1996. The hypothesis space was defined by Reder & Kusbit 1991 as three possible problems: encoding (i.e., the computation of a non-linguistic representation based on the sentence stimulus), retrieval (i.e., accessing a separate non-linguistic representation in long term memory), or match (i.e., determining whether the first representation and the second representation are the same)⁸⁷. These possibilities are initially stated at the sentence level: a partial match is a case where two representations match on all but one “word”⁸⁸. However, they note in the general discussion that “on reflection, it seems obvious that the match process involves concepts and features rather than words” (Reder & Kusbit 1991:402). Later work further suggested that “the matching process operates below the word level, at the level of distinctive features” (Kamas, Reder, & Ayers 1996:687). What this hypothesis shares with the hypothesis Erickson & Mattson’s proposal discussed above (and our related proposal) is an emphasis on the existence of a subset of features which are shared by the impostor and intended word. They differ, however, in the mechanism through which these features come to matter — in the partial matching hypothesis, a direct comparison between the concepts associated with the intended word and the concepts associated with the impostor

non-contextually-related world knowledge isn’t the kind of thing that *can* trigger an error signal.

⁸⁶We have seen the phrase “partial matching” in discussions of NPI illusions. This is not the same hypothesis.

⁸⁷Note that essentially all hypotheses being discussed in the present section could be considered “encoding” problems in the framework articulated by Reder & Kusbit 1991. They rule out encoding problems based on the failure to detect reliable reading time differences on impostor words that are detected and those that are not detected. However, we believe that not all encoding problems predict precisely the RT difference they assume, and moreover we are hesitant to draw strong conclusions from a null finding.

⁸⁸Scare quotes because we’re talking about non-linguistic representations, not actual sentences, so the units are not really words.

is carried out. In our version, the critical interaction is between the sentential context and the word — context narrows down all of the concepts associated with the word to only what is needed to understand the current sentence, and the impostor-intended overlap allows for the possibility that this narrowing will result in the same subset of concepts when applied to the impostor as when applied to the intended word.

A similar idea is discussed in relevance-theoretic approaches to substitution illusions, such as Allott & Rubio Fernandez 2002 and Maillat & Oswald 2009. They argue that “the addressee will only summon some contextual assumptions when interpreting [the Moses illusion], thereby shallow constructing an *ad hoc concept* in which only some (minimal) assumptions associated with Moses are activated” (Maillat & Oswald 2009:365-366). Their proposed “ad hoc concept” for “Moses” is “biblical figure,” as compared to the fuller conceptual representation of “Moses”, “person who led the people of Israel out of Egypt” (Maillat & Oswald 2009:366). While there are some obvious parallels, the ad hoc concepts proposal differs from the polysemy-resolution mechanism in that the authors suggest it is a mechanism that is specific to a shallow processing state. That is, in their model, the processor uses an impoverished representation of “Moses” even though it could, if the comprehender were more engaged, access the full representation. This has consequences for the expected role of motivation in the illusion, which we return to in section 7.3.2.

Key evidence supporting the idea we propose here comes from Kamas, Reder, & Ayers 1996. They compared illusion rates for substitution illusion questions that were preceded by one of the following: a question that highlighted the shared features of the impostor and intended word, a question that highlighted the features that distinguish them, or an unrelated question, . For example, for the standard Moses illusion question, they compared illusion rates when the question was preceded by “What religions study the story of Moses?” (shared features), by “What sea did Moses part?” or “How many sons did Noah have?” (distinguishing features), or by an unrelated question. They collapse the two distinguishing-feature questions into a single category. They found that questions that emphasize shared properties had no clear influence on illusion rates, relative to the unrelated-question baseline, but questions that emphasize distinctive properties reduced illusion rates (30% compared to 42% for the unrelated baseline).

Though it may also be possible to explain these findings under other accounts, they are straightforwardly predicted by a hypothesis in which the illusion arises due to the comprehender's selection of only the contextually-related conceptual features or facts associated with the impostor.

We have so far described the present hypothesis in a way that assumes that a mental representation of the referent of the impostor has been identified, and the filtering of conceptual features or facts is executed by subsequent computations which involve accessing other information stored in long term memory. An alternative description of the hypothesis, which assumes a very different kind of word meaning, is equally possible and would treat substitution illusions as a consequence of the exact same mechanism that resolves polysemy. A brief detour on possible models of polysemy is therefore warranted.

While many approaches to polysemy are in principle possible, McCourt 2021 argues that in fact there are essentially only two viable possibilities for the meanings of polysemous words: lists of concepts ("chimerical" representations) or pointers to regions of conceptual space with vague boundaries. If we adopt the first view, this would not change the present substitution illusion hypothesis in any meaningful way — some words are polysemous and their meanings are lists of concepts, but once a particular item on the list is identified based on context, the same selection of the other facts that are associated with that concept can proceed as we have already described it. If instead we adopt the second view, the distinction between what, in conceptual space, is a central piece of the meaning of the word and what is merely associated with the word is blurred (see McCourt 2021, section 4.4). There is no longer an atomic "lexical concept" associated with each individual lemma; rather, through the pointer, everything in a certain conceptual neighborhood becomes highlighted (with gradually "less light" on the concepts at the vague boundary) such that context can then pick out the concepts that are needed. Substitution illusions are then a case in which a concept which is common to the impostor and intended, and not necessarily central to the definition of either one, is picked out. This is more straightforwardly exhibited with an example in which the impostor is not a name, unlike the classic "Moses" illusion. Consider (168).

(168) The name of the raised bumps on paper that enable deaf (*blind*) people to read is Braille.

Under this hypothesis, the lexical meaning of "deaf" is a pointer to a region of conceptual space. Some concept having to do with *inability to hear* is surely in this region, but so is something like *sensory deficit*,

which is also in the region one arrives at via the pointer from “blind”. *Sensory deficit* is not a definition of either “deaf” or “blind” but it (or something like it) is within the conceptual neighborhoods that each of those words point you to. On some trials, context then selects this particular concept for participation in the inferred speaker meaning, resulting in the exact same representation for the “deaf” version and the “blind” version of the sentence. This hypothesis has some appeal due to its use of the exact same mechanism to account for polysemy resolution and substitution illusions. Moving forward, we will refer to the hypotheses discussed in this section as “polysemy-resolution” explanations for the illusion, though technically only this most recent version is strongly committed to a parallel with polysemy.

7.2.3 Semantic composition

Here we assume that lexical access and conceptual selection all proceed without error — that is, the lexical representation of “Moses” and all conceptual features associated with it, even the contextually-irrelevant ones which distinguish it from “Noah”, are successfully identified. We instead consider the possibility that the error lies in the comprehender’s strategy for combining the meanings of the parts of the sentence in order to construct the meaning of the whole. If there are sentence comprehension strategies that take in the meanings on “Moses” and “took two animals of each kind on the ark” and return a meaning identical to either the (compositional) meaning of “Noah took two animals of each kind on the ark” or the (compositional) meaning of “Someone took two animals of each kind on the ark”, or something similar, then we would expect failed detection of the anomaly on the trials where this strategy is used.

The only suggestion of non-compositional processes in the literature on substitution illusions is from Ferreira, Bailey, & Ferraro 2002. They state that “[the] assumption of compositionality seems eminently plausible, but results in the literature on the psychology of language call it into question. ... The Moses illusion ... is typically viewed as demonstrating the fallibility of memory processes, but it is also relevant to issues of language interpretation and compositionality” (Ferreira, Bailey, & Ferraro 2002:11). The precise mechanism they have in mind is not spelled out in detail, but they appear to be arguing that substitution illusions suggest that non-compositional strategies for determining sentence meanings exist. To be clear, the authors are not obviously committed to a mechanistic understanding of non-compositionality as an

explanation for the substitution illusion. They may have merely intended the claim descriptively — in the case of substitution illusions, the meaning that comprehenders ultimately entertain appears to not be identical to the compositional meaning of the sentence, and so the meaning is in that sense “non-compositional”. However, the claim that such illusions “call into question” the assumption of compositionality (which they previously define as “the assumption that interpretations of utterances are compositionally built up from words clustered into hierarchically organized constituents”) implies something stronger. If, for example, the actual failure underlying substitution illusions is the mis-selection of the “Noah” lexical item instead of the “Moses” lexical item, as we have previously discussed, it would still be perfectly reasonable to assume that the interpretation of the utterance is then built up from the words (as they’ve been identified), combined hierarchically. In other words, the lexical access hypothesis about the cause of the illusion does not at all call into question the assumption of compositionality. Thus, if the authors are committed to the claim that substitution illusions do call into question the assumption of compositionality, then they seem to be claiming that the actual mechanism underlying the illusion is the failure to combine the units compositionally.

It is not clear what exactly such a non-compositional approach to sentence comprehension would consist of (Ferreira, Bailey, & Ferraro 2002 do not define such a process in any more detail beyond stating that it seems to exist), but a key problem is the robust semantic similarity effect. If, for example, the comprehender can simply take a subject and a predicate and arrive at a meaning that combines “Noah” or an existential quantifier phrase with the predicate instead of the input subject, this process would yield exactly the same output when the input subject is the impostor “Moses” as when it is the impostor “Nixon”. One might consider incorporating an additional processing step, which simply determines whether there is any conceptual overlap in the world knowledge evoked by the meanings of individual words, so that it is this step that catches “Nixon”. This is roughly what van Oostendorp & Kok 1990 call a “conceptual cohesion” check. However, such a mechanism may be too coarse grained. Such a mechanism clearly predicts that any on-topic word would yield illusions, contrary to our intuitions about (169)⁸⁹.

⁸⁹It may be tempting to rule out illusions for (169) by other means, such as the selectional violation in *scaring candy*. If this selectional restriction truly lives in the syntax, then (169) can be consciously recognized as anomalous (because it is ungrammatical, not because it is a world knowledge violation) before the non-compositional semantic system can even begin to wreak its havoc. This is a quirk of the particular item and not a deep issue. One could instead consider *decorating* as the

- (169) What is the name of the holiday when children dress up in costumes and walk door to door scaring (*receiving*) candy?

Furthermore, the possibility non-compositional processes in sentence comprehension seems to predict rampant failed detection of conceptually cohesive falsehoods, and rampant misunderstandings of conceptually cohesive claims. That is, if a comprehender can simply insert a word other than the one that was encountered, and not even know that they’ve done so, how would they ever notice that a coherent claim is a lie? Here it becomes relevant that the claims about substitution illusions put forth by Ferreira, Bailey, & Ferraro 2002 are part of a larger argument that comprehenders have available to them a “shallow” or “good enough” processing mode, *in addition* to careful compositional processing. These shallow mechanisms are hypothesized to deliver an analysis of the sentence that is not necessarily perfectly veridical, but can be determined more quickly and with less effort than the alternative procedure, and the slightly-inaccurate output is often sufficient for the task at hand. We return to this issue in section 7.3.3.

7.2.4 Matching sentence meanings to non-linguistic concepts

In this section, we assume that lexical access, conceptual feature identification, and compositional processing all proceed without error. Here we evaluate the possibility that the problem is in converting a linguistic sentence-level representation into a probe for a search in long term memory, or matching this representation against the memory representation that is retrieved. In principle, there is also a third option in this family, which is that the retrieval of stored information is itself incomplete. That is, if a comprehender has the knowledge that “Noah took two animals of each kind on the ark” and searches their knowledge in the hopes of finding this information but only recovers the knowledge corresponding to “two animals of each kind on the ark”, there is little hope of detecting the mismatch between this retrieved representation and the impostor-containing probe — the information that misaligns just isn’t there. Early work on substitution illusions ruled out explanations of this variety by making the to-be-

impostor and make the same argument — *decorating* is on-topic in a discussion of Halloween (in fact, the word2vec semantic similarity measure we use Chapter 8 rates *Halloween* and *decorating* as even more similar than *Halloween* and *scaring*, though this is not my intuition) and so a “conceptual cohesion” check should fail to detect it, and *decorating candy* is not a selectional violation so there’s no way for the grammar to catch it either.

retrieved information easily accessible using a study phase before the experiment (Reder & Cleeremans 1990; van Oostendorp & Kok 1990; Reder & Kusbit 1991; Kamas, Reder, & Ayers 1996). Multiple experiments showed that the illusion persists even when the knowledge being probed has been made more accessible, contrary to the predictions of this explanation. Thus we do not consider it particularly likely that the information in long term memory is not retrieved in its entirety, and focus our attention instead on the conversion of the sentence representation into a non-linguistic representation and the matching process.

The key idea in this family of explanations is that not everything that is specified in a linguistic representation needs to also be specified in the non-linguistic thought that a comprehender works with after the sentence has been processed. Thus, although a substitution illusion sentence is successfully and veridically represented as a linguistic object, the non-linguistic representation that enters into subsequent computations (for evaluating truth or finding the answer) may not contain all of the information that the linguistic signal provides (i.e., it may not include the information corresponding to the impostor). In fact, on a long enough timescale it is obvious that not all information in a sentence is kept around — comprehenders can successfully recall the “gist” of a text long after reading it, though the details are quickly lost (e.g., Bransford & Franks 1971, among others). The discarding of details that are not perceived to be required for subsequent cognitive operations could in principle happen quite rapidly. Any non-restrictive or over-specified descriptor could in principle be discarded with no consequences for comprehension in the typical case. For example, a comprehender of a sentence like “The official cat of the state of Maryland, which borders Virginia, is the calico” could reasonably represent the sentence at a linguistic level in a way that incorporates “which borders Virginia” only to then encode it at a conceptual level in a way that discards this information, since the referent can easily be identified in memory without the added description. Note that such an account very straightforwardly predicts the contrasts observed by Baker & Wagner 1987, described in 7.1.3, since the presupposed content is placed in non-restrictive relative clause. Note, however that the fact that one could discard redundant information does not guarantee that comprehenders do this, and in fact some work shows that comprehenders benefit from over-specification, suggesting that redundant descriptors are not (always) discarded in the conversion from linguistic to

non-linguistic representations (e.g., Sonnenschein & Whitehurst 1982; Arts et al. 2011). Nonetheless, since substitution illusion sentences always involve multiple cues to the targeted trivia fact⁹⁰, this is line of inquiry remains worthy of consideration.

One piece of evidence that bears on this hypothesis comes from Hannon & Daneman 2001, which compared substitution illusion rates for questions with many cues to the intended answer (i.e. the answer that would be correct if the intended word had been used instead of the impostor) to questions with fewer cues. One might expect that with more cues the probability that the impostor is discarded should increase, and so the illusion rate should increase. However, Hannon & Daneman 2001 found only a 4-point difference in illusion rates (though note that this difference was statistically discernable). This is of course not evidence against such an account, but the effect size is somewhat smaller than one might have expected. A related finding comes from Barthel 2021, which compared not the amount of over-specifying information but the order. They found that when sufficient information to allow the question to be answered arrives before the impostor, illusion rates are higher than when the critical information arrives after the impostor (37% and 27%, respectively). This effect is expected under the current account if information becomes less likely to make it into the non-linguistic representation once a useful representation can be constructed. However, they do not find the same effect when comprehenders' task is only to detect anomalies, not to answer the questions, suggesting that this might not be a general fact about language processing but something that is specific to the task of finding question answers in long term memory. Thus an explanation that emphasizes a loss of information in the conversion from linguistic to non-linguistic representations remains plausible, though the evidence for or against such an account is inconclusive.

Finally we turn to the possibility of difficulty in comparing two representations (the thought expressed by the sentence and the fact retrieved from long term memory) to one another. Early work on substitution illusions tended to favor this account (Reder & Cleeremans 1990; Reder & Kusbit 1991; Kamas, Reder, & Ayers 1996), though typically this conclusion was primarily motivated by evidence against other accounts. This account places the error outside of the linguistic system altogether and so it predicts

⁹⁰This is a necessary characteristic of the stimuli; if the impostor or intended word was the *only* cue to the answer, the impostor-containing question wouldn't be anomalous it would just be a different question.

that a similar variety of illusion should occur when the form of the stimulus that encodes the world knowledge violation is not a language — for example, anomalies in pictures. This is possible but beyond the scope of the current work. Thus we consider it plausible that substitution illusions are caused by partial matching processes in evaluating the thought expressed by a sentence, but we do not aim to explicate or directly test this hypothesis here.

We additionally note that the explanations discussed in this section again do not straightforwardly predict effects of semantic similarity. One possibility is to again supplement the account with a conceptual coherence check to accommodate these facts.

7.2.5 Summary

Here we have considered four possible error points driving substitution illusions: lexical access, polysemy resolution, semantic composition, and matching sentence meanings to non-linguistic thoughts. Within the lexical access group, errors could occur at any of three levels (word form, lemma, or semantic), and could in principle involve either mis-selection (of the intended word instead of the impostor) or failed selection. The plausibility of these hypotheses turns on how they account for the semantic similarity effect: this could be either by stipulating lateral connections between world-knowledge-related lexical nodes or allowing access to world knowledge based on lexical candidates that have not (yet) been selected. The polysemy resolution explanation is partly motivated by the idea that non-contextually-relevant conceptual features associated with a word may not be regularly accessed, or, if accessed, may quickly filtered out. The semantic composition account, which is not often discussed in the literature on substitution illusions, amounts to a claim that the comprehender has multiple strategies available for determining sentence meanings, some of which respect the principle of compositionality and others of which do not. The account cannot explain the existence of semantic similarity effects unless a (fairly sophisticated) conceptual coherence check is additionally stipulated. And finally, there could be information loss in the conversion from linguistic to non-linguistic representations at the sentence level, or problems in matching processes between non-linguistic representations. These explanations also require additional mechanisms to account for semantic similarity effects.

7.3 Other accounts of the illusion

Given the above discussion, we believe that a plausible mechanistic account of the substitution illusion will involve one of the above sub-processes, as these are effectively the full list of operations that need to occur for the impostor to be detected. However, much work on substitution illusions has not focused on these mechanistic possibilities, and instead explores the question of why illusions occur at a somewhat higher level. These answers are not incompatible with the above possibilities, but in some sense orthogonal to them. For example, one might argue that substitution illusions occur because of problems in lexical access (one of the options from above) and at the same time argue that substitution illusions occur because speech errors are frequent and the comprehender's goal is to understand the speaker's meaning, making the "illusion" functionally useful in some circumstances. In this section we briefly review the literature on substitution illusions from the lens of three explanations that operate at this higher level.

7.3.1 Subconscious accommodation

An appealing initial analysis of the substitution illusion is that comprehenders simply fail to point out the impostor because they infer that the impostor was a speech error, and pointing out speech errors is generally weird. This could be articulated either in terms of purely social, non-linguistic factors (i.e., drawing attention to another person's mistake is rude) or specifically pragmatic factors (i.e., such behavior is not in line with the Cooperative Principle, Grice 1975). A distinct but related possibility is that the comprehender recognizes that some part of the sentence meaning misaligns with their own world knowledge, but, since the speaker is presupposing it, the comprehender adopts the presupposition for the purposes of the present conversation. Thus, under the first version, a comprehender hears (170) and thinks something like *the speaker said "Moses" but I know they know it was Noah on the ark, so I'll just ignore the speech error and answer the question*, whereas under the second version, a comprehender hearing the same question thinks something like *I think it was Noah on the ark, but the speaker seems to think it was Moses. They must be mis-remembering the story, or I am, but it doesn't matter for the question, so we'll just assume for now that it was Moses*. Note that the latter variety is intuitively somewhat less plausible for a question

like (171), since the question of who gives and who receives the candy on Halloween feels much less misrememberable than the question of the name of the ark character. We will refer to these two options as speech-error-accommodation and presupposition-accommodation hypotheses, respectively.

(170) How many animals of each kind did Moses (*Noah*) bring on the ark?

(171) What is the name of the holiday when children dress up in costumes and walk door to door giving out (*receiving*) candy?

We have argued against accounts along these lines since the illusion exists even in tasks where the comprehender is asked to point out anomalies. Early work on substitution illusions (e.g., Reder & Cleeremans 1990 Reder & Kusbit 1991) extensively compared tasks where comprehenders were told to answer the “gist” of the question to tasks where comprehenders were told to answer the literal question as stated and to point out any anomalies they identified (though note that these experiments did not always include post-tests to verify that participants had the relevant knowledge). While there may be differences in illusion rates between tasks⁹¹, it is clear that the illusion is not reducible to a deliberate attempt on the part of the comprehender to be cooperative. One way to make sense of these findings but maintain a speech-error-accommodation or presupposition-accommodation view is to infer that these processes occur at a subconscious level. One way to motivate this idea is that although the current task may have different goals, the Cooperative Principle may be so deeply ingrained that its effects can manifest even without conscious intervention.

Such a hypothesis is similar to the “reduced awareness” hypothesis considered (and ultimately rejected) by Sanford et al. 2011, in which “the anomaly is detected at a system level, but the listener is unaware of this detection” (Sanford et al. 2011:521). Note that they do not connect this hypothesis to Gricean principles but rather to non-detection phenomena in the visual processing literature (e.g., Simons & Levin 1997). One key piece of evidence that would bear on this hypothesis is the possible existence of processing disruptions in *implicit* measures of processing, such as reading times and ERPs. This is the approach Sanford et al. 2011 and several other studies pursue. This work is reviewed in section 7.4.2. In brief, the findings are inconsistent.

⁹¹This is actually debatable, since the “gist” task makes it impossible to diagnose whether an illusion has occurred.

Under the speech-error-accommodation version of this hypothesis, we might also expect that the likelihood of accommodation is tied to the likelihood of that particular speech error. Unrelated words like “Nixon” would be detected in part because speakers do not often make substitution errors by inserting a completely off-topic word. Work on substitution illusions would then likely benefit from insights from the literature on substitution errors in production (e.g., Fromkin 1973). Harley & MacAndrew 2001 explore effects of imageability and frequency on both semantic and phonological substitution errors in production, as in (172). Whether the same factors that influence speech error production also influence substitution illusion rates is not clear, but may be an interesting direction for future work.

- (172) a. I mean, you’ve put too much hot (*cold*) water in.
 b. You put that curtain (*cushion*) right in my eye.

(Harley & MacAndrew 2001)

Under the presupposition-accommodation version of the hypothesis, we do not necessarily expect any relationship between detection rates and speech error rates. We do, however, expect some consequences of the illusion for comprehenders’ belief in the presupposed (false) fact, at least in the short term. In other words, if the comprehender simply unconsciously adopts a belief in the presupposition for the purposes of the conversation, evidence of this new belief should be detectable.

The “illusory truth” effect was first demonstrated by Hasher, Goldstein, & Toppino 1977. The basic finding is that after being presented with a false claim, comprehenders become somewhat more likely to demonstrate belief that the claim is true in a subsequent task (e.g. trivia questions). Work in this area often describes such effects in terms of “availability heuristics” or “processing fluency”, but the basic phenomenon is also predicted by presupposition-accommodation mechanisms like the ones discussed here. Importantly, careful analysis of the types of responses provided in subsequent tasks suggest that the rate of *correct* answers is typically unaffected or only very slightly affected by the prior presentation of the false claim. While comprehenders are somewhat more likely to say the false thing they heard, these trials are mostly coming out of the “I don’t know” and miscellaneous incorrect categories.

Moreover, recent work suggests that the effect is reduced or eliminated when comprehenders adopt an “evaluative mindset” or “accuracy focus” during their first exposure to the false claim (Brashier, Eliseev,

& Marsh 2020; Salovich, Kirsch, & Rapp 2022). Framing this in terms of accommodation, it seems that comprehenders are much less inclined to adopt a belief in a false claim if they are asked to evaluate whether the claim is false. Since substitution illusions almost always involve an evaluation component, we would not expect them to adopt a belief in the claim at high rates, making this a somewhat less plausible explanation. It is additionally worth noting that the illusory truth effect persists on a longer timescale — potentially weeks — than is anticipated under a presupposition-accommodation account (Hasher, Goldstein, & Toppino 1977). Note that one related consequence of this work is that if the illusory truth effect is real and has a substantial effect on later judgments, the knowledge checks that are typically used in substitution illusion studies are not a pure reflection of participants knowledge, since they always follow presentation of (anomalous) illusion statements. Bottoms, Eslick, & Marsh 2010 investigated this issue and found some impact of substitution illusions on later accuracy, though again these effects are primarily borne out in a switch from “I don’t know” and miscellaneous incorrect responses to the presented falsehood, and much less of a switch from correct responses to the presented falsehood.

Thus, while (subconscious) pragmatic accommodation explanations for the illusion are plausible, there is little independent evidence for them, and more research is clearly needed.

7.3.2 Motivation and attention

Another variety of explanation for substitution illusions claims that comprehenders just aren’t trying very hard at the task, or just aren’t reading very carefully (or aren’t reading carefully because they aren’t trying hard).

First we consider the possibility that a lack of motivation is the key problem. One way of thinking about this is to imagine that experimental participants don’t think it’s that important to answer the questions correctly. One reason they might think this is because they assume, unless otherwise instructed, that providing quick responses is just as important as providing accurate responses. Another reason they might think this is that they’re going to get paid (or get class credit) either way.

van Jaarsveld, Dijkstra, & Hermans 1997 compared the effects of task instructions. One group of participants were told to be both as fast as possible and to catch as many errors as possible, while another

group was told only to catch as many errors as possible. The accuracy-focused group had fewer illusions than the speed-and-accuracy group, but they also more false alarms (that is, reporting an error when there wasn't one there). The authors conclude that focusing on error detection can improve sensitivity to errors somewhat, but it also simply makes participants more biased toward reporting errors. Thus, simply informing participants that their main goal is successful detection of anomalies doesn't make the illusion go away, and so it is unlikely that the illusion is purely a consequence of the extent to which participants are trying to detect anomalies.

Speckmann & Unkelbach 2021 manipulated monetary incentives in order to test the motivation question more directly. Participants were offered increased pay if they answered more questions correctly. Pooling the data from two experiments, they found that monetary incentives can reduce illusion rates, but the effect size is small in practical terms: “we paid participants about 4€ (about \$4.50) on average for them to give one less Moses response [out of 20 trials]” (Speckmann & Unkelbach 2021:854). Even in their highest compensation group, illusion rates were around 30 – 40%. It is also worth noting that there are limits on what we can learn from this kind of manipulation — surely if Mechanical Turk participants were paid an additional hundred dollars for each correct answer they would answer many more of them correctly, but this would likely arise because of strategies that we're not interested in, such as googling the question, asking a friend, or re-reading the sentence many many times.

We therefore conclude that motivation is not central to the substitution illusion. Of course, if participants try less hard we will see more errors, but, critically, even when one is doing one's best it seems that some proportion of illusions are unavoidable. We now consider the possible role of attention. Substitution illusions can be described as a case in which comprehenders “don't pay attention to the world knowledge violation”, meaning that they don't consciously note the world knowledge violation. This is in some sense not falsifiable in that it uses the label of “attention” purely descriptively. It is also unsatisfying, in the same way that it is unsatisfying to say that the Zöllner illusion in Figure 7.3 arises because people “don't pay attention to the parallel-ness of the lines”.

Another (unsatisfying) way to explore the role of attention in substitution illusions is to simply document the effects of increasing or decreasing the resources available for sentence processing by introduc-



Figure 7.3: The Zöllner illusion, in which parallel lines appear to be non-parallel (Zöllner 1860)

ing simultaneous tasks (thereby pulling attention away). Büttner 2012 found that illusion rates increase when comprehenders perform a simultaneous number repetition task or random number generation task, though this finding is not especially surprising. We might also conceive of the typical substitution illusion task for questions as a kind of multitasking, in which comprehenders must simultaneously perform both a trivia-question-answering task and an error-detection task. This would explain the contrast between questions and statements that is sometimes found (see section 7.1.4). Accordingly, Barthel 2021 found that when comprehenders read the same question stimuli but their only task is to detect errors, illusion rates go down (though note that Kamas, Reder, & Ayers 1996 investigated the same question and found similar illusion rates for the two tasks). It appears that pulling attention away from error-detection can make comprehenders worse at detecting errors, but we are hesitant to draw the conclusion that a lack of attention is the reason error-detection fails in the first place. Most things are harder to do if you pay attention to something else while you're doing them.

A potentially more interesting question is whether, even when there isn't a secondary task, comprehenders pay little attention to what they're reading. In principle, comprehenders could be paying so little attention to the stimulus that they skim it instead of reading it, and thus experience "illusions" purely because they skipped over the impostor. However, we know this to be incorrect because eye-tracking studies show that illusions arise even when the impostor is directly fixated (Bohan 2008; Cook et al. 2018). A more plausible version of this hypothesis is that comprehenders are engaging in "mind wandering during reading" in substitution illusion experiments. This is the phenomenon where one feels as though they are reading, and one's eyes continue to saccade through text, but once reaches the bottom of the page

and realizes nothing was actually understood. A few studies of this phenomenon have shown that during mindless reading many standard eye-tracking effects — like longer fixation times for less frequent or more complex words — are reduced or eliminated (Smallwood 2011; Franklin, Smallwood, & Schooler 2011; Schad, Nuthmann, & Engbert 2012; Foulsham, Farley, & Kingstone 2013). These findings suggest that eye-tracking methods may be well-suited to determining the role of attention in substitution illusions. In fact, Schad, Nuthmann, & Engbert 2012 used detection of anomalies, including lexical substitutions, as a diagnostic for whether mind wandering has occurred. Thus it is possible that mind-wandering is related to the substitution illusion. However, we are cautious about pursuing this variety of explanation, since mind-wandering seems to be at least partly subject to conscious control, and therefore should become less prevalent when comprehenders are highly motivated to perform well on the task. Speckmann & Unkelbach's (2021) findings indicating that even highly-motivated participants experience illusions cast some doubt on this claim. Furthermore, it is clear that comprehenders are not completely inattentive to the stimulus, since they successfully identify the intended answer.

An alternative way to think about attention is the possibility that the attentional state fluctuates throughout the sentence because of the demands of other processes. This is analogous to what we considered in Chapter 4 regarding the post-relative clause shift of attention in NPI illusions. We might imagine, for example, that if the current parsing operation is highly demanding, less resources are available for lexical access, making lexical access errors of the type discussed in section 7.2.1 more likely. This could be similar to the “attentional blink” (Shapiro, Raymond, & Arnell 1997). Under this type of hypothesis, manipulations of the text such as boldface font or capitalization pull attention back to lexical processing.

Attention-related hypotheses may also explain the presupposition effects, if one assumes that comprehenders are, in general, inclined to pay less attention to presuppositions. A more intuitive analysis of the presupposition effects, however, is that they are about presupposition *per se*. Not being “up for debate” is central to many ideas of what it means to be presuppositional, so it should not be surprising that comprehenders are disinclined to question the veracity of presupposed content. How this works mechanistically is of course up for debate but simply reducing the problem to “attention” is not necessarily useful. Similarly, it is not clear if font manipulations are in fact manipulations of attention or

manipulations of presuppositional status (i.e. capital letters are inferred to indicate contrastive focus). The existence of such effects is therefore not airtight evidence for the role of attention.

A final point on the topic of attention is that there is a sense in which attention is a central component of one of the mechanistic hypotheses we considered in section 7.2. That is, the polysemy-resolution mechanism can be thought of as a way in which context serves to direct a comprehender's attention towards a subset of the concepts associated with a word. This may be a useful way to frame the polysemy resolution account, if what is independently known about attention mechanisms can clarify the details of how the filtering of concepts works.

7.3.3 Shallow processing

We now turn to an idea that has been extremely influential in the literature on linguistic illusions, and especially substitution illusions: the “shallow processing” account. It is critical to note that the label “shallow” (and related terms like “good enough” and “underspecified”) are used in a wide variety of ways. Sometimes these terms appear to be merely descriptive — that is, when some researchers state that a phenomenon is due to “shallow processing”, they may merely mean that an error has occurred. For example, Sanford & Sturt 2002 describe the substitution illusion and then state “People are clearly using shallow processing here” (Sanford & Sturt 2002:384). Similarly, Ferreira & Patson 2007 state the effect, then state “thus, it appears that people’s comprehension of sentences can be quite shallow” (Ferreira, Bailey, & Ferraro 2002:73), and Bohan 2008 claims that “the best illustration of shallow processing is, in our view, when readers fail to notice semantically anomalous words or phrases in text.” (Bohan 2008:206). Either these authors mean simply that comprehenders sometimes make errors and “shallow processing” is a description of that fact⁹², or they consider the evidence for shallow processing mechanisms as a cause of the substitution illusion to be uncontroversial. However, it is not clear to us that the mere existence of illusions is sufficient evidence to determine the mechanism. In section 7.2 we explicated four possible error points that could underlie the illusion, including problems in lexical access, polysemy resolution,

⁹²At least some of the time, “shallow processing” really seems to just refer to the *outcome* (failed detection) and have nothing to do with the process. For example “shallow processing may be the *result* of inefficient processing at either one of these stages of language processing, that is, inefficient retrieval or integration of lexical information [emphasis ours]” (Bohan 2008:210). This use of terminology is, in our view, not helpful.

semantic composition, and world knowledge matching. In our view, there is currently not sufficient evidence to decide between these theories. Some researchers who argue for “shallow processing” do not appear to disagree: “future experiments must decide whether it is recovery of word meaning, or integration into a final interpretation that is the locus of shallow processing” (Sanford & Sturt 2002:384). This suggests that many — perhaps all — hypotheses about substitution illusions can be considered “shallow processing” hypotheses, making the label somewhat less useful.

Note also that not all research invoking “shallow” or “good enough” processing is quite so agnostic about the underlying source of the illusion: Ferreira, Bailey, & Ferraro 2002 claim that substitution illusions and other phenomena demonstrate that “language processing is not always compositional” (Ferreira, Bailey, & Ferraro 2002:12), which appears to be committed to a particular locus of the error (i.e., semantic composition). However, they include the misinterpretation of garden paths as an additional example of “good enough” processing. It is unlikely that non-compositionality is also responsible for garden path interpretation, and so we might infer that “good enough” is also intended as a cover term, encompassing many mechanistically distinct hypotheses. Later work uses different definitions of “good enough” processing: “Good Enough processing refers to situations in which the parser carries on interpreting new input without having completely pruned interpretations that are no longer compatible with this input” (Slattery et al. 2013:115). However, we do not believe the authors intended to claim that “good enough” processing refers *only* to that case (if they did, it is clear that substitution illusions have nothing to do with “good enough” processing, under the revised definition). Thus the updated version merely broadens the definition of “good enough” even further. In general, it appears that the only hypotheses are *not* in the category of “shallow”, “good enough”, or “underspecified” are those which predict that sentence processing is always error-free. It is not obvious to us that there are or ever have been hypotheses that make this prediction, since it has always been clear that, at a minimum, comprehenders can make mistakes when not paying attention to the stimulus.

Sanford et al. 2011, unlike other studies in this literature, do consider an alternative account of the illusion. They contrast shallow processing with a “reduced awareness hypothesis”: “As an alternative to the shallow processing hypothesis, it is conceivable that the comprehension system retrieves the meaning

of the anomalies and attempts to integrate the semantics of the word in question with the rest of the text ... the anomaly is detected at a system level, but the listener is unaware of this detection” (Sanford et al. 2011:515-521). This hypothesis has some features in common with the (subconscious) pragmatic accommodation hypothesis we considered in section 7.3.1, in that both assume that the impostor is detected, just not at a conscious level. They suggest that finding an ERP component that is sensitive to the impostor, even when it is not consciously detected “would support the reduced awareness interpretation and run counter to the shallow processing account” (Sanford et al. 2011:517). They do not find such an effect, and thus conclude that the shallow processing account is correct. However, we are hesitant to draw strong conclusions based on a null finding in an ERP experiment. Moreover, subsequent work sometimes does find such effects, weakening the argument (see section 7.4.2).

Cook et al. 2018 argue against the shallow processing hypothesis based on very similar logic. They find evidence of processing difficulty for substitution illusions in eye-tracking measures, even for illusory impostors. Like Sanford et al. 2011, they argue that a shallow processing account predicts that illusory impostors should result in no disruption at all. Note that eye-tracking findings have also been inconsistent (see section 7.4.1).

Davis & Abrams 2016 also disfavor shallow processing accounts, based on a different argument. The version of the shallow processing hypothesis that they consider is somewhat more concrete than what is sometimes proposed — specifically, they address a theory of shallow *semantic* processing, in which “not every word’s meaning is thoroughly checked for congruity” (Davis & Abrams 2016:75). Because the shallow processing hypothesis they consider involves a check for semantic congruity only, the existence of similarity effects in dimensions other than semantic fit (i.e. phonological similarity, as reported in Shafto & MacKay 2000 and Shafto & MacKay 2010, and visual similarity as Davis & Abrams 2016 find) constitute evidence against the hypothesis. However, we suspect that the authors of various shallow processing hypotheses would not agree that all hypotheses within the shallow processing framework have been addressed by this argument.

Christianson 2016 connects the shallow processing literature in language processing to the more general idea of fast, “heuristics”-based processing as compared to slow, “algorithmic” processing in cogni-

tion more broadly (e.g., Kahneman et al. 1982). Under Christianson’s proposal, *both* the fast and slow systems may be used on the same sentence, but the heuristics-derived representation can override the algorithmic one under certain circumstances. Other approaches treat heuristic processing as a “mode” that comprehenders are sometimes in and sometimes not in. The particular heuristics that comprehenders are thought to use are typically not spelled out, with the exception of work on the processing of passives (Ferreira 2003). Relatedly, it is not always clear how shallow processing theories account for the fact that comprehenders sometimes do notice anomalies. One might suppose that those trials on which the anomaly is detected are the trials on which the “algorithmic” system has been deployed, or might instead suppose that they are trials on which the heuristics yield a veridical representation (which they must sometimes do, otherwise they’re very bad heuristics). If the latter idea is intended, spelling out what the heuristics actually consist of is critical. Note that a “conceptual coherence” heuristic is not enough: the exact same stimuli, with for which this lone heuristic would result in the exact same output, result in illusions on some trials and detection on other trials.

One proposed heuristic that has received attention in work on substitution illusions is the “fluency heuristic”. The basic idea is that ease of processing functions as a signal to the comprehender for whether their *other* heuristics are doing a good job or if analytic processing needs to be engaged. A now-famous study from Alter et al. 2007 states this explicitly: “Metacognitive experiences of difficulty or disfluency appear to serve as an alarm that activates analytic forms of reasoning that assess and sometimes correct the output of more intuitive forms of reasoning” (Alter et al. 2007:569). They report that participants are more likely to answer a tricky math problem like (173) correctly if it is presented in a hard-to-read (“disfluent”) font.

(173) A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost?

However, a meta-analysis of 16 subsequent attempts to replicate this effect clearly demonstrated that the effect first reported does not exist (Meyer et al. 2015). Nonetheless, the fluency idea remains hugely influential. Deckert 2015 investigated the role of processing fluency on substitution illusions by manipu-

lating what they call the “degree of conventionalisation of language structure” — that is, they compared a more typical illusion question like in (174a) to an more convoluted version like in (174b).

- (174) a. What was the famous line uttered by Louis Armstrong (*Neil Armstrong*) when he first set foot on the moon?
- b. If possible to recall, quote the memorable phrase that has come to be conventionally ascribed to Louis Armstrong (*Neil Armstrong*) when he first set foot on the moon.

(Deckert 2015)

Note that these are not the most typical substitution illusion stimuli — the similarity between impostor and intended is primarily phonological (shared surname), not semantic. They report illusions in 88% of trials for the fluent version and 100% of trials for the disfluent version. Note, however, that it’s not clear that they ran a knowledge check, not clear that they warned participants that there would be substitutions and their task was to identify them, and not clear that they tested more than this one item. Contrary to these findings, Song & Schwarz 2008 found *fewer* illusions for “disfluent” stimuli, using font readability instead of the structure of the sentence to manipulate this. Janouskova et al. 2022 demonstrated that this effect does not replicate, and the effect size is statistically equivalent to zero, based on equivalence testing using two one-sided tests (Lakens, Scheel, & Isager 2018) and the “small telescopes” approach to estimating effect size (Simonsohn 2015). Given the statistical equivalence findings, we might conclude that a difficult font neither helps comprehenders catch illusions (as the processing fluency heuristic in Alter et al. 2007 predicts) nor harms their ability to catch them (as Song & Schwarz 2008 claimed), it simply does not matter. One final related study reported that substitution illusion rates decreased when participants were seated in a room that smelled like fish (with the idea that the “fishiness” metaphor for something being suspicious triggered more vigilant processing) (Lee, Kim, & Schwarz 2015). While we know of no attempts to replicate this finding, we are hesitant to take it at face value.

Summing up, there are some issues with the shallow processing literature, including both a lack of specificity with respect to what “shallow processing” consists of as a cognitive mechanism, and replication failures in the related “processing fluency” literature. While we cannot claim with certainty that the sub-

stitution illusion is not the result of shallow processing, we focus our attention on more specific, testable hypotheses.

7.4 Other methodological approaches

We now turn our attention to a handful of findings regarding the eye-movement patterns, neural responses, and second language processing for substitution illusions. While these kinds of investigations can be theoretically meaningful (we have already seen a few hypotheses that make predictions about eye-movements and ERPs), the findings are mixed and sometimes difficult to interpret. Moreover, research using implicit measures like eye-movements and ERPs must split the data based on a behavioral response indicating whether the impostor is detected, leading to potential issues with statistical power. Bearing these issues in mind, we turn to the reported findings.

7.4.1 Reading times

Reder & Kusbit 1991 were the first to investigate reading times for substitution illusion sentences, using a moving window self-paced reading paradigm. As we have previously noted, their hypothesis space consists of (a) problems with the encoding of the sentence meaning, (b) problems with the retrieval of the fact from memory, and (c) problems with the process matching these representations to one another. The reading time experiment was designed to address the “encoding” hypothesis. Their motivation is as follows: “if the illusion is caused by a failure to carefully encode, then one would expect that the reading time for a distorted target word [i.e., impostor] ... would be less when the subject failed to notice the distortion than when the distortion was noted (and responded ‘can’t say’)” (Reder & Kusbit 1991:396). While this is a reasonable expectation, we are not confident that the prediction of a contrast in reading times holds for all versions of “encoding” hypotheses, nor that the prediction of similar reading times holds for all hypotheses that are not about encoding.

Consider the hypothesis that the lexical item has been mis-selected, resulting in selection of the intended word instead of the impostor that was actually seen, as discussed in section 7.2.1. We suspect that

this hypothesis would fall under the “encoding” umbrella (it’s certainly not a problem of retrieving the fact from long term memory or of matching propositions). But it is not obvious that we should expect selection of one lexical entry to take any longer than selection of another, thus equivalent RTs for detected and illusory impostors would not be surprising under this account. On the other hand, under a “matching” hypothesis, if the long term memory representation has been successfully retrieved *before* the impostor is encountered, matching could be carried out incrementally, as the sentence unfolds. Then the trials on which the comprehender actually carries out the matching procedure for the impostor (resulting in detection) may have longer impostor reading times than the trials on which the comprehender skips the matching procedure for that word (resulting in the illusion). Note that this version of the “matching” hypotheses makes some assumptions about the format of facts retrieved from long term memory (i.e. that they are the same as the format of sentence meanings) which we do not necessarily endorse. For related ideas about how the illusion could arise as a consequence of the mapping between sentence representations and non-linguistic thoughts, see section 7.2.4. Nonetheless, it seems that the relevant hypotheses do not actually make clearly distinct predictions regarding reading times.

In their first reading time experiment, Reder & Kusbit 1991 observe a trend toward *faster* reading times on the impostor for trials where the impostor was detected compared to illusion trials (i.e. trials on which the impostor was not detected) (539ms and 628ms, respectively). They unfortunately do not report any measure of variability in the data, such as standard error, making it difficult to determine how precise these estimates are. An additional experiment found approximately the same pattern (476ms for detected impostors and 539ms for illusory impostors). An important note for interpreting these data is that neither experiment used a post-test to verify that participants knew the relevant facts. Thus a plausible explanation for the trend toward faster reading times (which the authors acknowledge in a footnote) for detected impostors is due to differences in knowledge — i.e., not knowing the relevant fact leads to both slower reading (because it’s unfamiliar) and failed detection of impostors (because the comprehender doesn’t know enough to know it’s an impostor). This analysis is further supported by their findings in a novel condition that was tested in the follow-up experiment. Participants memorised the (true version of the) targeted facts for some of the upcoming substitution illusion sentences. For studied facts,

the RT difference discussed above did not arise, and instead a small but not statistically discernable trend in the opposite direction was observed (556ms for detected impostors and 545ms for illusory impostors). In sum, the findings are not conclusive, though the misleading finding that RTs were faster for detected impostors than illusory ones demonstrates the importance of controlling for participants' knowledge in such investigations.

Later work investigated the question of reading times using eye-tracking during normal reading. Both Bohan & Sanford 2008 and Cook et al. 2018 investigate reading times for detected impostors, illusory impostors, and non-anomalous sentences, and both are primarily interested in reading time evidence as it pertains to the “shallow processing” hypothesis, though they arrive at different conclusions regarding the viability of the hypothesis. Under a shallow processing account, they argue, reading times for illusory impostors should be identical to reading times for non-anomalous trials⁹³.

As an alternative to the shallow processing hypothesis, Bohan & Sanford 2008 consider the possibility that “an anomaly might register in the comprehension system, but not be available for conscious report” (Bohan & Sanford 2008:237). This strongly resembles the “reduced awareness” hypothesis that Sanford et al. 2011 consider. Cook et al. 2018, in contrast, pit the shallow processing hypothesis against

⁹³As we have previously noted, the shallow processing hypothesis may in fact be a family of mechanistically-distinct hypotheses, and it is not clear that this prediction holds for all of them.

For one thing, if we assume a shallow processing hypothesis that is committed to the use of “heuristics”, this prediction very much depends on whether one assumes that detected impostors are the product of “algorithmic” (non-shallow) processes that are only sometimes deployed, or the result of the same heuristics that are used on illusory trials, but those heuristics yield a different output (obviously this will depend very much on what the heuristics are, which is never stated).

Under the first version, comprehenders stochastically decide, on each trial, to run either their heuristic or algorithmic processes. Successful detection of impostors occurs if and only if algorithmic processing has occurred — that is, illusion trials are quick, detection trials are slow. Non-anomalous trials consist of a mixture of slow algorithmic trials and quick heuristic trials (which have no consequences, since there's no anomaly to detect). Thus, this version of the hypothesis actually predicts that non-anomalous sentences should be read *slower* than illusory impostors, since all illusory trials used heuristic processing whereas non-anomalous trials are a mixture of heuristic and algorithmic trials.

Under the second version of the hypothesis, in which heuristics are always used and the heuristics themselves can catch the impostor some of the time, we see different predictions as a function of how the outputs of multiple heuristics are assumed to combine, resulting in a final binary judgment of the anomaly. One possibility is that if *any* heuristic indicates a problem, processing disruption occurs and the impostor is detected — in which case the non-anomalous trials and illusory impostor trials should truly be identical in reading times. Another possibility is that each failed heuristic contributes a little bit of disruption, and conscious detection occurs when the summed disruption crosses some threshold — in which case the illusory impostors may result in more processing difficulty and slower reading times than the non-anomalous trials.

Thus, the shallow processing hypothesis is not specific enough to make clear predictions about reading times. All three logically possible outcomes for the comparison of non-anomalous sentences and illusory impostors (equal reading times, slower for non-anomalous, slower for illusory impostors) are compatible with versions of the hypothesis. The versions of the hypothesis that we consider here are in fact only a subset of the hypotheses that are compatible with the shallow processing claim.

their “RI-Val” model (O’Brien & Cook 2016), which is primarily concerned with the processes that allow comprehenders to reach a “coherence threshold”, and the evaluation computations that continue passively even after a comprehender has saccaded past a particular word. A thorough evaluation of the RI-Val model is beyond the scope of the present work.

Bohan & Sanford 2008 report no differences in reading times for illusory impostors⁹⁴ and non-anomalous sentences for a number of eye-tracking measures (first pass, regression probability, total time, etc.) on the critical and post-critical region. Cook et al. 2018, in contrast, report slower reading times in per-character second pass durations for illusory impostors compared to non-anomalous sentences. Summing up, the findings regarding reading times for substitution illusions are mixed and we cannot definitively conclude whether such effects exist. Moreover, it is not obvious that the hypotheses that these experiments were intended to test make clear predictions about reading times.

7.4.2 Neural measures

The ERP literature for substitution illusions is similarly murky. Sanford et al. 2011, Bohan et al. 2012, Tune et al. 2014 all investigate differences in N400 amplitude for detected impostors, illusory impostors, and non-anomalous sentences. As in the eye-tracking literature, the primary goal is to determine if the processing of illusory impostors is more difficult than the processing of non-anomalous sentences, contrary to the predictions of (some versions of) the shallow processing hypothesis. Sanford et al. 2011 find no differences between the three conditions in the N400 time-window. Bohan et al. 2012 found that the non-anomalous condition had the smallest (least negative) N400, followed by the detected impostor condition, which had a larger (more negative) N400, followed by the illusory impostor condition, which had the largest (most negative) N400 of the three. Note that a larger N400 for illusory impostors than for detected impostors is not obviously predicted by any account. Tune et al. 2014, in two experiments using German stimuli, found a larger (more negative) N400 for the detected impostor condition and smaller (less negative) N400s for the undetected and non-anomalous conditions, which were not distinguishable from one another. In one experiment in English they found no differences between the three

⁹⁴It’s worth noting that their detection task was non-standard. They had participants read passages and “knock on the table” when they detected an anomaly.

conditions.⁹⁵ In sum, the ERP literature exhibits even greater variability in the pattern of findings than the eye-tracking literature, and further work is needed to clarify the empirical picture.

There are two other investigations of substitution illusions that approach the question from a neurophysiological perspective: Raposo & Marques 2013 collected fMRI data while participants read substitution illusion sentences and controls, and Izaute, Paire-Ficout, & Bacon 2004 collected behavioral judgments of substitution illusions from participants under the influence of the benzodiazepine lorazepam, a drug which is known to induce transient anterograde amnesia, among other cognitive effects. While these investigations are potentially informative, they are primarily concerned with using the substitution illusion to learn something about the brain, rather than using measurements / manipulations of the brain to learn something about substitution illusions. For this reason we do not go into the details of these studies.

7.4.3 Second language processing

A small number of studies have attempted to extend the research on substitution illusions to second language processing. Both Vaessen 2017 and Dhaene et al. 2021 explored the processing of substitution illusions in native speakers of Dutch who spoke English as a second language. Vaessen 2017 did not find statistically discernable differences in illusion rates when sentences were presented in the L1 or L2, whereas Dhaene et al. 2021 found a relatively small but statistically discernable increase in illusion rates in the L2 relative to the L1.

7.5 The current investigation

Having reviewed the existing work on substitution illusions, we turn to the contributions of the present study. Recall that we identified four candidate loci for the processing error that underlies the substitution illusion: lexical access, polysemy resolution, semantic composition, and the mapping of sentence

⁹⁵They attribute this to cross-linguistic differences in the relative weighting of top-down and bottom-up processing due to differences in word order. But importantly the equipment used for the German and English studies was different and the threshold for impedances was 10 times higher for the English experiment. The apparent cross-linguistic differences are therefore plausibly explained by noise.

meanings onto non-linguistic representations. We also considered the role of attention, which may interact with these processes such that errors become more likely in a low-attention state. The following experiments aim to identify the precise circumstances under which the illusion occurs, so that we may determine the hypothesis that best aligns with this pattern.

7.5.1 Item-wise variability

Whereas in our investigation of NPI illusions we were able to categorically eliminate the illusion with surprisingly subtle changes to the stimuli, substitution illusions present a somewhat different profile of variability. Because these illusions are critically dependent on content words and world knowledge, we have much less precise control over the relevant variables. For this reason, we begin with a correlational approach. Many first-pass accounts of substitution illusions (for example, the low-motivation explanation, mind-wandering, or pragmatic accommodation to speech errors) predict relative uniformity in the illusion — that is, any semantically-related substitution should go undetected on some proportion of trials. Experiment 17 demonstrates that this prediction for uniformity is not borne out. Across the 49 substitution illusion stimuli we tested, illusion rates varied wildly. Experiments 18-21 then evaluate the extent to which this variability can be explained by various factors: participants degree of familiarity with the targeted facts; the semantic similarity between the impostor and the intended word; and the degree of uncertainty about how the sentence will unfold at the point of the substitution, measured using both language model predictions and human data in a cloze task. We find that while some of these factors correlate with illusion rates, much of the variability remains unexplained.

7.5.2 Manipulating ease of access

We then turn to the question of where in the series of steps between fixating the impostor and responding to the question the comprehender seems to have gone wrong. Specifically, we ask whether the error seems to lie in the lexical access pipeline or in post-lexical processing steps. We address this by manipulating the ease of lexical access through priming, with the idea being that this manipulation should substantially reduce illusion rates if the problem lies in lexical access, but not if the problem lies elsewhere (and lexical

access was already proceeding smoothly). In order to be able to investigate this, Experiment 22 established that illusion rates are high even with a sentence judgment task. Experiment 23 then integrated this sentence judgment task into an experimental paradigm in which the impostor is primed using a lexical decision task. We find that priming the impostor does not dramatically reduce illusion rates, suggesting that the error lies in post-lexical-access processes.

Chapter 8 Substitution illusions: item-wise variability

As we noted in Chapter 7, the substitution illusion exhibits substantial variability in effect size across studies. Here we measure the extent to which that variability is driven by differences in the illusion rate across items, and to what extent the variability across items can be explained by independently-measured characteristics of the items.

The overall logic is similar to the strategy we employed in the study of NPI illusions. If we describe the NPI illusion as simply a case in which an unlicensed NPI is perceived to be acceptable when it co-occurs with a non-c-commanding negative word, the set of candidate hypotheses to explain the effect may be quite broad. If instead the NPI illusion is better described as a case in which an unlicensed NPI that occurs within two-words of a licensing environment created by a quantificational licenser (which, based on our findings in Chapters 3-5, seems to be a more accurate description), the requirements for a plausible explanation look rather different. Similarly, substitution illusions have typically been described as a case in which a word which is anomalous in its context (in that the sentence expresses a world knowledge violation) is not consciously noticed. We suspect this definition is much too broad. All of the stimuli we investigate here fit this definition, but as we will see, they are not equally likely to yield illusions.

This of course raises the question of how we might better define the profile of the substitution illusion. To that end, we measure various properties of the illusion stimuli and correlate these measures with illusion rates. If, for example, the variability across items were strongly related to imposter-intended similarity, we might refine our generalization to say that substitution illusions are cases in which a word which has been substituted for a word it is highly similar to and which is anomalous in its context is not consciously noted. This is not an explanation of the cause of the illusion, merely a refinement of its definition. Such refinements are not irrelevant to the explanation — some hypotheses predict that illusions will take a particular shape — but it is importantly not an explanation in itself. In section 7.2 we reviewed

mechanistic hypotheses that target the underlying cause of the illusion. We return to these hypotheses in Chapter 9.

8.1 Experiment 17: variability in the illusion

Experiment 17 aimed to replicate the basic illusion effect, and to determine the extent to which different substitution illusion stimuli yield different substitution illusion rates.

8.1.1 Participants

102 workers recruited through Amazon Mechanical Turk completed the task. 2 participants were excluded due to a high rate (greater than 90%) of "I don't know" answers. The full task was designed to last 15 to 25 minutes and participants were compensated \$5.

8.1.2 Materials

Stimuli were adapted primarily from two sources: Vaessen 2017 and Cook et al. 2018, though some of the stimuli in those studies were originally taken from Erickson & Mattson 1981, Reder & Kusbit 1991, Büttner 2007, and Sanford et al. 2011. Stimuli were copied exactly from their sources when possible, but occasionally altered for two reasons: to create questions from stimuli that were previously formed as statements, and to improve the naturalness of questions in American English. One item was included in the experiment but ultimately excluded from all analyses because we discovered that the impostor did not result in a world knowledge violation, but in a different (but still reasonable) question. After this exclusion there were 49 stimuli.

8.1.3 Procedure

Participants were told that the task involved answering trivia questions and that some trivia questions had errors, which they should report. Following instructions and practice trials, they were presented with each of our 50 stimuli, along with three options for a response: they could type an answer into a

text box, they could select a button labeled “I cannot answer” or they could select a button labeled “I don’t know”. Extensive instructions and examples prior to the onset of the task ensured that participants understood the function of these two buttons - “I don’t know” was to be used when the participant had insufficient knowledge of the domain to be able to come up with an answer, whereas “I cannot answer” was to be used when the participant noticed an error in the question. Each question was presented in full, on the same screen as the three response options, and participants were instructed to take as much time as they needed to answer each question, but to answer based solely on their own knowledge, and not based on searching the internet for the answer. Each participant saw 25 anomalous and 25 non-anomalous questions, which were counter-balanced across participants.

Following the trivia task, participants completed a “knowledge check” task. The purpose of this task is to ensure that data points from the trivia task can be excluded in cases where the participant does not have enough domain knowledge to be able to detect the error (that is, if an individual does not know the difference between Noah and Moses, their failure to report the anomaly is uninteresting). In this task, participants saw declarative versions of the same 50 stimuli from the trivia task. Each item was missing the word corresponding to the impostor position and participants needed to choose the word to fill in the blank from six options which included the correct answer (e.g. ‘Noah’), the impostor (e.g. ‘Moses’), two additional distractors (e.g. ‘Isaac’, ‘Joseph’), ‘none of these’ and ‘I don’t know’. Care was taken in the instructions to clarify that although the statements would be familiar (since they were the same trivia facts from the first part of the experiment) participants’ goal in this task was to select the word that makes the statement true, not to select the answer that matches what they saw in the first part of the experiment.

8.1.4 Analysis

We first removed data for which participants erroneously provided more than one answer (i.e. they typed an answer but also selected “I cannot answer” or “I don’t know”) or provided no answer at all, which resulted in the loss of less than 1% of data points. Second, we removed data points for which participants do not have the relevant knowledge that would allow them to detect the anomaly. Because all participants completed both the trivia section and knowledge check section of the task for all 50 items, we were able

to determine, for each trivia data point, whether the participant knew the relevant information to detect the world knowledge violation. We removed any trivia question trial for which the participant answered the corresponding knowledge check question incorrectly. This resulted in the loss of 20% of remaining data points. The amount of data lost varied substantially (2% - 64%) across items, because the items were intended to span a range of difficulties. Thus the precision of our estimates of the illusion rate varies across items. Finally, we removed data points for which participants selected “I don’t know” as their answer on the trivia question, resulting in the loss of 15% of remaining data points.⁹⁶ After these exclusions 3,321 usable trials remained, of which 1,603 were anomalous trials on which an illusion could occur.

We categorized any trial on which the participant typed in a response instead of selecting “can’t say” as an illusion, regardless of whether the response would have been correct for the intended question. This is for two reasons: first, it allows us to avoid hand-coding responses for accuracy⁹⁷, and second, to some extent it doesn’t matter if they knew the answer. For our purposes, the key thing is that they represented the sentence in such a way that they believed it was not anomalous. For example, typed answers to the anomalous and non-anomalous versions of the question in (175) included “Saturn” (the intended answer), “Earth”, “Mars”, and “Neptune”, among others. While these are not correct, they do suggest that the comprehender did not detect the impostor “star”. Of course not all incorrect typed answers are such clear cases, but we think this strategy is overall reasonable.

(175) What is the second largest star (*planet*) in our solar system, after Jupiter?

The data were analysed in two ways. First, for the purposes of visualizing the variability across items, we computed each item’s illusion rate as well as a 95-percent confidence interval based on the standard error of the proportion. We also simulated the variability across items that is expected to occur due to

⁹⁶There is variability in how “I don’t know” answers are handled in the literature on substitution illusions. Obviously we are only interested in substitution illusion trials for which the corresponding knowledge check was passed. But among these trials, there are effectively three possible outcomes: (A) the participant detects and reports the impostor, (B) the participant fails to detect the impostor and answers the question as if it is non-anomalous (i.e., the illusion occurs), or (C) the participant selects “I don’t know”. Some researchers discard the “I don’t know” data points, as we have done, so that the illusion rate is $\frac{B}{A+B}$. Alternatively, some researchers include the trials receiving an “I don’t know” response in the denominator, so that the illusion rate is $\frac{B}{A+B+C}$. Obviously, inclusion of “I don’t know” responses will result in numerically lower illusion rates. The extent to which the illusion rate is lowered based on the inclusion of “I don’t know” responses will be different for easier and more difficult items. For this reason, we exclude “I don’t know” responses.

⁹⁷Typos and spelling errors such as “Enstain” for “Albert Einstein” make it non-trivial to automate this.

chance, by randomly redistributing trials across 49 pseudo-items, each of which constituted a random sample of the full dataset. We similarly computed illusion rates and confidence intervals for these pseudo-items. Second, we ran a linear regression predicting, across items, illusion rates for a random half trials based on illusion rates for the other half of trials, which corresponds to the approach of measuring split-half reliability in individual differences research. We used a bootstrapping procedure to derive a median reliability measure and 95-percent confidence interval.

8.1.5 Results

Per-item and overall illusion rates are shown in Figure 8.1. The overall illusion rate, collapsing across items, was 31%, but per-item illusion rates varied from 0% to 84%. In order to determine if this variability was systematic or could be explained by random noise, we computed illusion rates for random pseudo-items, as described above. These are shown in Figure 8.2. As is clear from the comparison of these figures, the variability across items is substantially more than would be expected if each item was a random sample from the same population. This is reinforced by split-half correlation analyses which found that the median reliability was .76 (95CI: .65, .85).

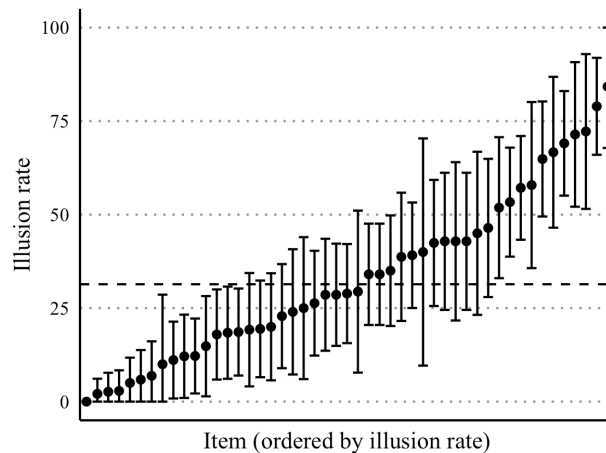


Figure 8.1: Substitution illusion rates for each of 49 question stimuli tested in Experiment 17, shown with 95-percent confidence intervals. The average illusion rate across items is indicated by the black dashed line.

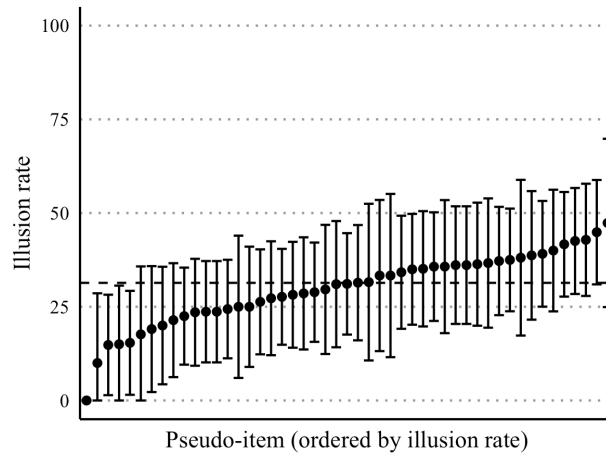


Figure 8.2: Substitution illusion rates for each of 49 pseudo-stimuli generated by randomly partitioning the data, shown with 95-percent confidence intervals. The average illusion rate across pseudo-items is indicated by the black dashed line.

8.1.6 Discussion

Two findings from Experiment 17 are of note. First, we replicate previously-reported findings of substitution illusions, with an overall illusion rate of 31%, which is typical of substitution illusion experiments (see Figure 7.1). Second, we find that variability in illusion rates across items is greater than would be expected by chance, and highly reliable — that is, items with high illusion rates consistently demonstrate high illusion rates, and items with low illusion rates consistently demonstrate low illusion rates. Some of the items we tested revealed quite low illusion rates, even though all stimuli used in this experiment were drawn from prior substitution illusion studies.

These findings cast doubt on any account of substitution illusions which treats them as a highly general or random phenomenon. For example, if the explanation for the illusion is random mind-wandering during reading or low motivation, it is surprising that the illusion occurs so frequently for some stimuli and so infrequently for others. However, it is important to determine the extent to which this variability can be explained by factors that are already known to influence illusion rates. Experiments 18-21 therefore tested the degree to which variability in illusion rates correlates with variability across items with respect to comprehenders' familiarity with the target facts, semantic similarity between the impostor and the intended word, and the comprehender's ability to infer the rest of the question and its answer prior

to the impostor.

8.2 Experiment 18: familiarity

A common initial reaction to reports of substitution illusions, especially when the classic Moses illusion is used, is suspicion that comprehenders' knowledge of the distinction between the impostor and the intended word is fuzzy. That is, comprehenders might just not be very sure of the difference between "Moses" and "Noah". Such an explanation is intuitively less appealing for substitution illusions where the impostor is not a proper noun but an highly familiar content word (such as the impostor "giving out" in place of "receiving" in the Halloween example discussed at length in Chapter 7). If there is in fact variability across items in the extent to which comprehenders are familiar with the information being queried, this could explain some of the variability in illusion rates.

Relatedly, early work on substitution illusions considered the possibility that difficulty in retrieving of the relevant facts from memory drives the illusion, and, in order to address this, compared illusion rates for facts that were studied ahead of time versus not studied (Reder & Cleeremans 1990; van Oostendorp & Kok 1990; Reder & Kusbit 1991; Kamas, Reder, & Ayers 1996). While their findings showed that the effect of studying was not as large as would be predicted under the memory-retrieval account (that is, the illusion did not disappear when the participant knew the targeted fact extremely well), some studies did find small effects of studying. Regarding our findings from Experiment 17, we might hope that the knowledge check conducted after the main experiment would control for such effects, but since these questions were presented in a multiple-choice format, it is possibility that although the facts were known well enough to make the correct selection, some facts were known better than others. Cantor & Marsh 2017 gave substitution illusion sentences about biology and history to graduate students in biology and history, and while participants were not immune to illusions in their own area of expertise, they did perform better. This effect of expertise arose even though Cantor & Marsh only looked at trials where participants passed the knowledge check, suggesting that a deep familiarity with the target fact may help participants resist illusions.

It's worth clarifying that our question is not whether the Experiment 17 participants had different areas of expertise from one another and whether this influenced their illusion susceptibility. Rather, we assume that this is true — some participants may know a lot about space and were more likely to get (175) right whereas others knew a lot about Bible trivia and were more likely to get the Moses/Noah question right. This variability would not explain why some items had higher illusion rates than others. Instead, we are interested in whether some items were about trivia that virtually all of our participants knew a good deal about whereas others were about trivia that virtually none of our participants were experts in. That is, does item-wise variability correlate with how well-known each item was, in general. Experiment 18 collected familiarity ratings for the (non-anomalous versions of the) trivia facts tested in Experiment 17 from a separate set of participants.

8.2.1 Participants

50 workers recruited through Amazon Mechanical Turk completed the task. It was designed to last 5 to 15 minutes and participants were compensated \$3.

8.2.2 Materials

The non-anomalous versions of the 50 trivia questions from the previous experiment were turned into declarative statements.

8.2.3 Procedure

Following instructions and practice trials, participants were presented with the 50 items one at a time and asked to rate on a 7-point scale the extent to which they agreed that “most people are familiar with this statement.” Instructions clarified that when thinking about “most people” they should focus on adults living in the United States.

8.2.4 Analysis

For each item we computed the mean familiarity rating. We verified that familiarity ratings were reliable enough to be correlated with other measures using the same split-half reliability procedure used for Experiment 17. We then determined the extent to which familiarity predicts illusion rates with a linear regression, treating an item's mean familiarity score as the independent variable and each item's illusion rate from Experiment 17 as the dependent variable.

8.2.5 Results

Average ratings ranged from 1.76 to 6.58 across items (see Figure 8.3), suggesting that participants used the full 1-7 scale and there was substantial variability across items. Split-half correlation analyses found that the median reliability was .96 (95CI: .94, .97). A linear regression did not identify a statistically discernable correlation between familiarity ratings and illusion rates ($\beta=0.001$, $SE=0.03$, $z=0.02$, $p=0.98$, adjusted $R^2=-0.02$), as shown in Figure 8.4.

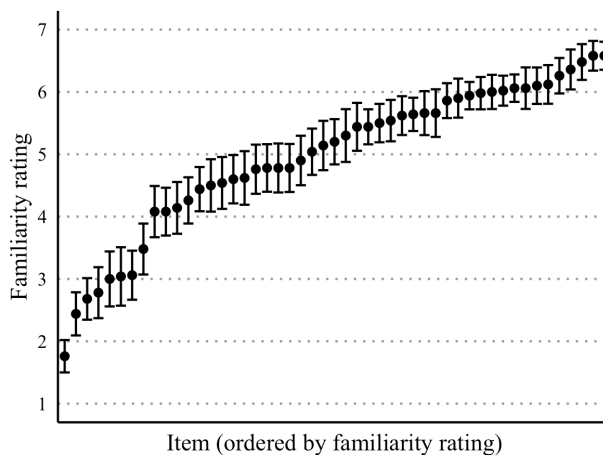


Figure 8.3: Familiarity ratings for each of 49 stimuli, shown with 95-percent confidence intervals.

8.2.6 Discussion

In order to investigate whether differences in illusion rates across items could be partly due to differences in comprehenders' degree of familiarity with the target facts, we collected familiarity judgments in a sepa-

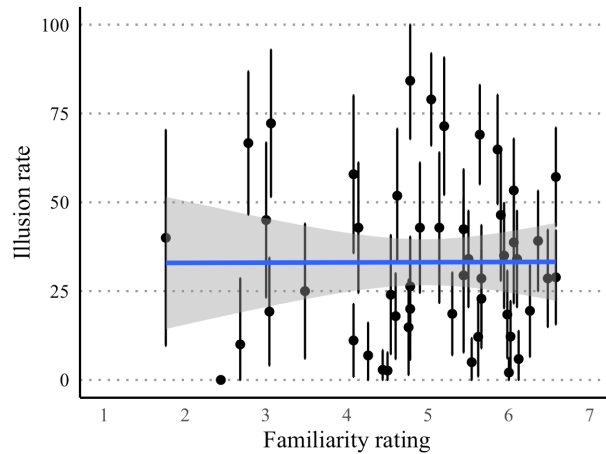


Figure 8.4: Correlation across 49 items between familiarity ratings (Experiment 18) and illusion rates (Experiment 17)

rate task. Although these ratings were highly reliable, they did not reliably correlate with illusion rates. It therefore appears unlikely that a substantial amount of the variability in substitution illusion rates across stimuli can be attributed to differences in comprehenders' familiarity with the target facts.

It is important to note that Experiment 18 collected familiarity ratings from a different sample of participants from those who participated in Experiment 17. Moreover, the task was not to judge one's *own* familiarity with the statement, but to judge the extent to which they agreed that "most people are familiar with" the statement. It would therefore be inappropriate to conclude anything about whether any particular participant in Experiment 17 was very familiar with the target fact for any particular stimulus. Rather, as previously noted, we assume that there was variability across individuals in their areas of expertise, and it may very well be that they performed better (i.e., had fewer illusions) on the trials that were more familiar to them individually. The key claim, rather, is that differences in familiarity are not responsible for the systematic variability in illusion rates across items. That is, the items that had the highest illusion rates did not have the highest illusion rates purely because they targeted facts that the fewest participants were familiar with.

8.3 Experiment 19: lexical similarity

As we discussed in section 7.1.1, much prior work shows that some aspect of the semantic properties of the impostor word influences illusion rates (Erickson & Mattson 1981; van Oostendorp & Kok 1990; van Oostendorp & de Mul 1990; van Jaarsveld, Dijkstra, & Hermans 1997; Shafto & MacKay 2000; Hannon & Daneman 2001; Budiu & Anderson 2008; Shafto & MacKay 2010; Davis & Abrams 2016; Cook et al. 2018). This may have to do with similarity to the intended word, relatedness to the context, or both. Here we focus on impostor-intended similarity. Manipulations of the meaning of the impostor typically compare only two categories, such as the high-similarity and low-similarity pair in (176), though both semantic similarity and semantic relatedness are presumably gradient properties.

- (176) a. How many animals of each kind did Moses (*Noah*) bring on the ark?
 b. How many animals of each kind did Nixon (*Noah*) bring on the ark?
 (Erickson & Mattson 1981)

Our Experiment 17 stimuli are all taken from the “high similarity” conditions of previous studies, when similarity was manipulated, but given this underlying gradience, we don’t expect that all substitutions were equal in semantic similarity. Thus it’s possible that some of the variability in illusion rates that we observe is a consequence of variability in similarity.

Testing this requires quantifying the similarity between word pairs. While much previous research has manipulated similarity, hardly any studies explicitly define or attempt to quantify the difference in similarity between conditions. van Oostendorp & de Mul 1990 is the exception to this. They had a group of participants (separate from the participants in the main illusion study) list attributes of the intended words and impostor words, presented in isolation, and quantified similarity as the number of attributes a pair shared. Other studies base classification into “high similarity” and “low similarity” categories on the authors’ intuitions. This is a reasonable choice given that norming stimuli in the way van Oostendorp & de Mul 1990 did takes time and may not be worthwhile if the similarity pair is uncontroversial⁹⁸. It does, however, risk some circularity in the identification of high-similarity (high illusion rate) and low-similarity

⁹⁸For example, no one is going to dispute the claim that “Moses” is more related to “Noah” than “Nixon” is.

(low illusion rate) stimuli. For example, Cook et al. 2018 categorize (177a) as a “high-related” item and (177b) as a “low-related” item. It is not obvious to us that the sun is more similar to the moon than the earth is. Based on the physical properties of the referents (size, location, temperature, composition), the moon seems to be much more similar to the earth, whereas based on perceptual experience (glowy thing in the sky), the moon may be more similar to the sun⁹⁹.

- (177) a. Neil Armstrong was the first man to walk on the sun (*moon*) in outer space.
 b. Neil Armstrong was the first man to walk on the earth (*moon*) in outer space.
 (Cook et al. 2018)

This is not a deep problem for Cook et al. 2018, since they use the relatedness manipulation only to create a baseline in which detection rates are high (regardless of the reason) against which to compare eye-tracking measures. For our purposes, however, precise measurements of word pair similarity are necessary. We therefore used computational measures of lexical similarity based on word embeddings to estimate similarity for each item, and computed the correlation between this similarity measure and illusion rates.

8.3.1 Computational measures of word similarity

Representing words is a critical component of many natural language processing (NLP) tasks. By far the dominant approach is to represent each word as a fixed-length vector, an idea which originated in the information retrieval literature (Salton, Wong, & Yang 1975). Because semantic similarity is relevant to many NLP tasks, one might hope for similar words to be encoded with similar vectors. As a way to generate such vectors, many models leverage the “distributional hypothesis” — that is, the idea that distributional similarity (i.e. the extent to which two words occur in the same linguistic contexts) is strongly correlated with meaning similarity (Harris 1954). Word vectors based on co-occurrence are sometimes

⁹⁹Cook et al. 2018 are not especially clear about whether they think impostor-intended similarity or impostor-context relatedness is the main factor, so one might argue that while *earth* and *moon* could be a more similar pair than *sun* and *moon*, the key claim could actually be that *sun* and *Neil Armstrong was the first man to walk on the ... in outer space* are more related than *earth* and *Neil Armstrong was the first man to walk on the ... in outer space*. This latter claim is actually hard to judge. The sun fits better in that it is in outer space, but the earth fits better in that it is walk-on-able. This is all to say that the judgment of similarity is not always as crisp as in the Nixon-Moses case.

called word embeddings (for a more refined definition of word embeddings and an overview of their history, see Almeida & Xexéo 2019).

Vector representations of words based on the distributional hypothesis are of course not the only option for representing words in a way that encodes similarity in meaning. WordNet (Miller et al. 1990) is an alternative approach, which uses a hand-constructed network based on human judgment which represents similarity in meaning through the connections in the network.

Word embeddings have been shown to exhibit a number of properties that may be desirable for representations of lexical meaning. Famously, embeddings in the word2vec model have been shown to combine in intuitively reasonable ways: the vector for “queen” added to the vector for “man” results in the vector for “king” (Mikolov, Yih, & Zweig 2013). Moreover, the kind of similarity that these models encode has long been suspected (or at least hoped) to relate to the kind of word similarity that is encoded in the human mind (see, e.g., Lund & Burgess 1996, Rohde, Gonnerman, & Plaut 2006). Some research shows that the cosine similarity between two word vectors correlates with priming effects in a lexical decision task (Auguste, Rey, & Favre 2017).

Applying models of this type to the study of substitution illusions, we expect to find that impostor-intended word pairs that are close together in a vector space representation of word meanings yield more illusions. We verified this assumption using data from Cook et al. 2018. Because they compare high-similarity and low-similarity impostors, and they report the full list of stimuli used, we can compute word similarity for each of their items. This comparison is illustrated in Figure 8.5. Note that even the dissimilar items are on the positive end of the scale of cosine similarity as measured by word2vec (which ranges from -1 to 1). This is unsurprising, since an impostor must be somewhat similar to the intended (in syntactic category, animacy, etc.) to even be usable as an impostor. Critically, we find that word2vec similarity ratings are higher for Cook and colleagues’ high-similarity items than for their low-similarity items, indicating that the version of word similarity targeted by word2vec is similar to the kind that is of interest in substitution illusion research.

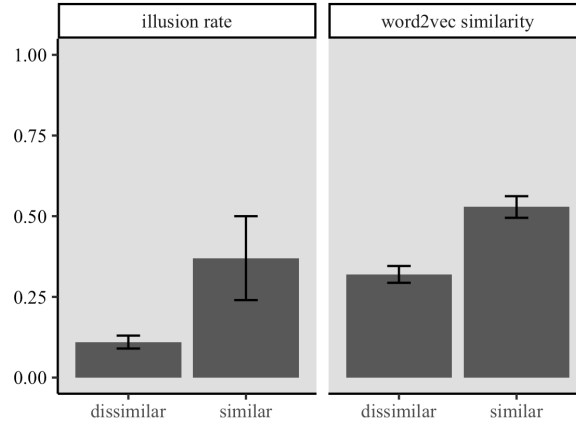


Figure 8.5: Substitution illusion rates reported by Cook et al. 2018 for their high-similarity and low-similarity impostors, and impostor-intended similarity for these conditions, as measured by word2vec.

8.3.2 Correlation with illusion rates

In order to determine whether variation in impostor-intended word similarity is responsible for the variability in illusion rates that was observed in Experiment 17, we first measured the similarity between the two words for each of the 49 items tested in Experiment 17. Similarity measures could not be obtained for four of the items, because the impostor string is not an identified token in word2vec (e.g. “Sleeping Beauty”). We exclude these four items from our correlation analysis. We then measured the correlation between these values and illusion rates. A linear regression did not identify a statistically discernable correlation between familiarity ratings and illusion rates ($\beta=0.30$, $SE=0.18$, $z=1.72$, $p=0.09$, adjusted $R^2=0.04$), as shown in Figure 8.6. Note that while there is a trend toward higher illusion rates for more similar word pairs, there is substantial variability in illusion rates that is *not* explained by impostor-intended similarity.

8.3.3 Discussion

Here we investigated whether the item-wise variability in illusion rates observed in Experiment 17 is due to variability in impostor-intended semantic similarity. While there is a weak (and not statistically discernable) correlation between these measures, substantial variability remains unexplained.

There are a few issues concerning semantic similarity to be considered. First, the cosine similarity of

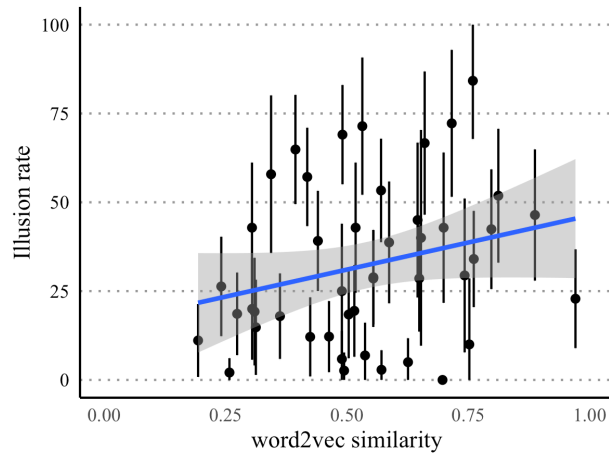


Figure 8.6: Correlation across 49 items between similarity ratings as measured by word2vec (Experiment 19) and illusion rates (Experiment 17).

word2vec embeddings is a measure of word similarity in the absence of context. Recall from section 7.1.1 that the question of whether the impostor must be similar to the intended word (which we measure here) or related to the context (which we have not measured) is unresolved. A related issue is that word2vec does not disambiguate word senses — thus, an ambiguous word like “bank” will have only one embedding which attempts to capture all of its uses (e.g., financial institution and edge of river). This effectively makes word2vec similarity a noisier measure of similarity than we might want.

Finally, it is worth clarifying that the present findings should not be taken as evidence that semantic similarity does not matter for substitution illusion rates. We do not dispute the claim that “Nixon” yields fewer illusions than “Moses” in the classic ark example, nor can we demonstrate that this effect is or is not a consequence of impostor-intended similarity per se. Recall that all of the stimuli we test here are on the high end of the similarity scale. Thus our findings are perfectly compatible with the existence of a strong effect of impostor-intended similarity. We only claim that this effect is not the same effect that drives variability in illusion rates across the 49 items tested here. Another way to put it is that relatively high levels of similarity are clearly necessary for robust illusions to occur, as Moses-vs-Nixon demonstrates, but we find that high levels of similarity are not sufficient.

8.4 Experiment 20: language model predictions

As we have previously discussed, substitution illusions could be exacerbated by fluctuations in the comprehender's attentional state as a sentence is being processed. Specifically, when attention is pulled away to another task, making lexical processing and integration secondary, illusions may become more likely. Tentative support for this idea comes from the observation that performing a simultaneous task increases illusion rates (Büttner 2012), and removing the question-answering component of the typical substitution illusion task sometimes reduces illusion rates (Barthel 2021, cf. Kamas, Reder, & Ayers 1996). However, illusion rates remain relatively high even in a single-task situation, suggesting that distractions from other tasks are not critical to the illusion.

The particular influence of the question-answering task on the attentional state is not well understood, but one plausible idea is that at the point when the comprehender has enough information to answer the question, resources are re-allocated to answer generation, and away from processing the sentence. This can be understood as a claim about attention as a function of the information state.

We find some support for the idea that some of the variability in illusion rates across stimuli is due to differences in the attentional state at the point of the substitution (which may in turn be due to differences in whether the subject can infer the answer to the question) in an analysis of position effects in our stimuli. Since there is substantial variability in sentence length, we normalize our position metric by dividing the impostor's position (in number of words since sentence onset) by the total number of words. A linear regression identified a statistically discernable correlation between sentence position and illusion rates ($\beta=0.34$, $SE=0.14$, $z=2.37$, $p=0.02$, adjusted $R^2=0.09$), as shown in Figure 8.7.

This correlation is fairly indirect evidence of an attention mechanism, since there are ways a position effect could arise that are not about changes in the information state or shifts in attention. It would therefore be valuable to determine if these position effects are in fact due to developments in what is known about the question over the course of the sentence. We therefore tried to quantify the information state at the substitution point for each of our stimuli using language models.

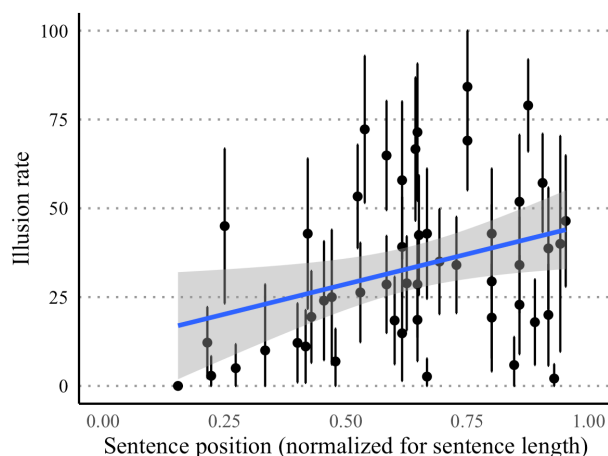


Figure 8.7: Correlation across 49 items between the impostor’s position in the sentence and illusion rates (Experiment 17).

8.4.1 Computational language models

Language models are statistical models of language use. They assign probabilities to strings based on the frequency with which similar strings occur in corpora. Language models can be extremely simple, as in an n -gram model which simply treats the probability of a string n words in length as the count of times that string has occurred in the corpus divided by the count of times the first $n - 1$ words occurred in the corpus (in other words, the conditional probability of word n given the prior context of words 1 through $n - 1$). More elaborate — and more successful, from the perspective of mirroring human language use and achieving various NLP tasks — language models use neural networks that are trained to predict a missing word based on its local context (drawn from a corpus). Two varieties of neural network language model that have been shown to function reasonably well at least for some applications are LSTM (long short-term memory) and transformer models. For our purposes, the distinction between these types of models is not critical; both succeed at least in part because of their ability to use relatively large windows of context.

Because language models assign probabilities to possible upcoming words given a prior string, they can be used to compute various information theoretic properties of a partial sentence, such as surprisal and entropy. Our primary goal in this project is to determine whether the predictions of two language models for upcoming content in the substitution illusion contexts of our stimuli align with our hypoth-

esis the information state as a driver of attentional shifts in substitution illusions. As a measure of the information state at the substitution point, we compute the entropy of the distribution of predictions of next words. Under our hypothesis, we expect that stimuli with lower entropy at this point will have a higher probability of illusions.

8.4.2 Knowledge check

Before we can evaluate the successes of our two models — an LSTM language model and a transformer model, GPT-2 — we need to establish that these models have some similar world knowledge¹⁰⁰ to the humans we are trying to model. In other words, the language models should pass a “knowledge check” equivalent to what the humans do. We implemented this in both models by computing, for each stimulus, the probability of the entire error-free trivia question (i.e. the version of the question with the intended word) and the probability of the entire anomalous trivia question (i.e. the version of the question with the impostor), and determining whether the model assigns a higher probability to the intended version than to the impostor version. 27 items passed the knowledge check for the LSTM language model and 39 items for GPT-2. We computed correlational analyses for both the full sets of items and filtered sets, using only the items that the models “knew”.

8.4.3 LSTM language model predictions

This model was trained to predict the following word and evaluated on its ability to capture long-distance dependencies in language Gulordava et al. 2018. The model is much smaller than GPT-2, but it is a sophisticated language model in its own right. It was trained on text from English Wikipedia and had 650 hidden units. One hot encoding was used, and infrequent tokens were replaced with the token <unk>.

Using this model, for each of our 49 stimuli, we measure the entropy of the probability distribution just before the impostor. We fit a linear model comparing each of this measure to the proportion of illusions for each item. We first focused on only the 27 items that passed the knowledge check. We did

¹⁰⁰Of course, the models don’t actually “know” anything about the world. But if, for example, they never saw anything in a corpus about Moses or Noah, it would be unreasonable to expect their predictions in the Moses illusion stimulus to have much to do with the information state of a human who knows the target fact.

not find a reliable correlation between entropy and illusion rate ($R^2 = .002$, $p = .31$), as shown in Figure 8.8. We additionally tested the correlation between entropy at the substitution point and illusion rate using the full set of 49 items, as shown in Figure 8.9. We again did not find a reliable correlation between entropy and illusion rate ($R^2 = .03$, $p = .13$)

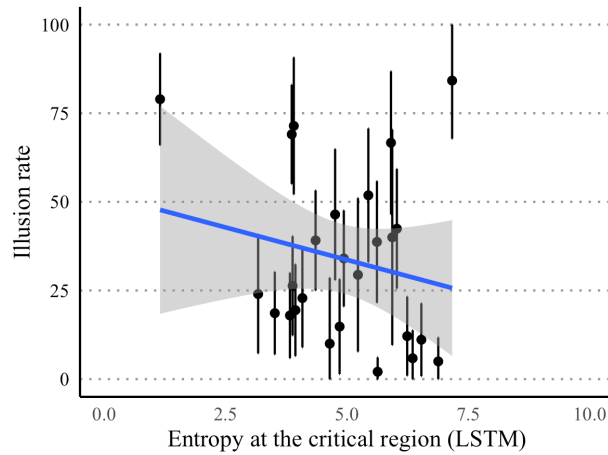


Figure 8.8: Correlation across 49 items between entropy at the critical region according to LSTM predictions (Gulordava et al. 2018) and illusion rates (Experiment 17), filtered to only the items that pass the knowledge check.

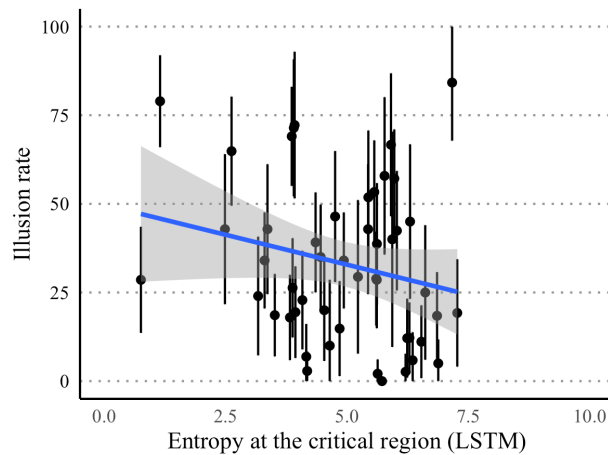


Figure 8.9: Correlation across 49 items between entropy at the critical region according to LSTM predictions (Gulordava et al. 2018) and illusion rates (Experiment 17).

8.4.4 GPT-2 language model predictions

GPT-2 was trained on a crawl of all outbound links from Reddit with karma ≥ 3 . This resulted in 8 million documents, and 40 GB of text. A Byte Pair Encoding (BPE) representation (Sennrich, Haddow, & Birch 2016) was learned for this data, and then a one hot encoding of this representation was used in training an LSTM with attention that had 1.5 billion parameters (Radford et al. 2019).

Using this model, for each of our 49 stimuli, we measure the entropy of the probability distribution just before the impostor. We fit a linear model comparing each of this measure to the proportion of illusions for each item. We first focused on only the 39 items that passed the knowledge check. We did not find a reliable correlation between entropy and illusion rate ($R^2 = .02$, $p = .19$), as shown in Figure 8.10. We additionally tested the correlation between entropy at the substitution point and illusion rate using the full set of 49 items, as shown in Figure 8.11. We again did not find a reliable correlation between entropy and illusion rate ($R^2 = -.006$,¹⁰¹ $p = .40$)

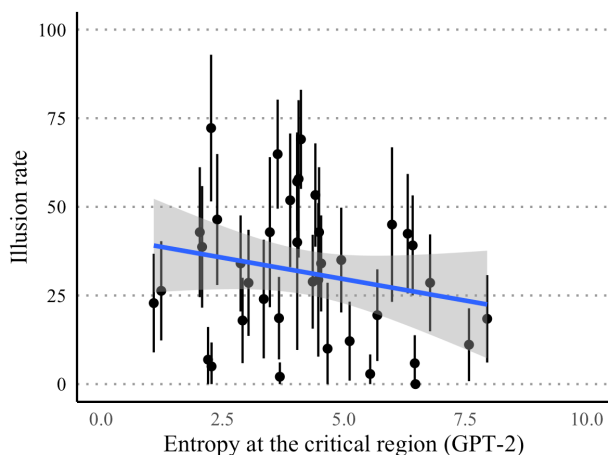


Figure 8.10: Correlation across 49 items between entropy at the critical region according to GPT-2 predictions (Radford et al. 2019) and illusion rates (Experiment 17), filtered to only the items that pass the knowledge check.

¹⁰¹Note that we report adjusted R^2 values, which can be negative.

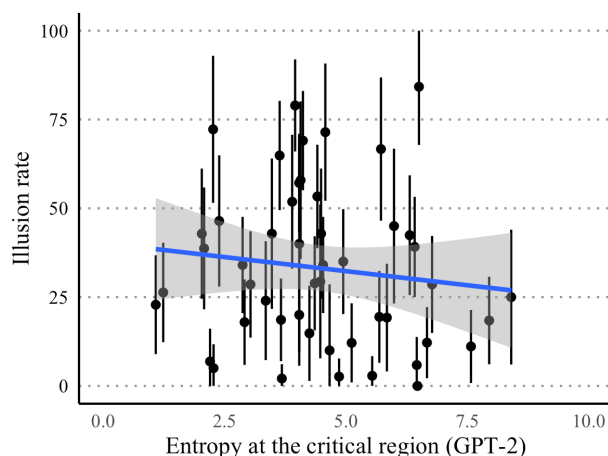


Figure 8.11: Correlation across 49 items between entropy at the critical region according to GPT-2 predictions (Radford et al. 2019) and illusion rates (Experiment 17).

8.4.5 Discussion

Here we investigated whether a substantial part of the variability in illusions across items is explained by differences in the information state at the point of the substitution. We quantified the information state as the entropy over possible next words at the substitution point, using two state-of-the-art language models. We found no reliable correlations between entropy and illusion rates.

It’s likely that the kind of knowledge the human has that enables a shift in attention is not well approximated by these metrics. The language models allow us to measure certainty about how exactly the *question* will unfold, but our hypothesis focuses on shifts in attention that are triggered by knowing the *answer* to the question. It is possible that there are situations in which a human is confident that the question is querying a particular fact (i.e. the answer is predictable) but there are many possible ways to ask the question (i.e. the literal question is not predictable). The opposite may also occur: a human is not yet certain of the fact being queried, but a few upcoming words can be predicted because of local co-occurrence regularities (e.g. if the question is in the middle of an idiom, the remainder of the idiom might be predicted). If either of these are true in our stimuli, the language model metrics simply misalign with the hypothesis of interest. We also suspect that these findings reflect the more general fact that the human sentence processing algorithm is fundamentally unlike the algorithms implemented by language models like GPT-2 and LSTM models.

8.5 Experiment 21: cloze

In light of the lack of correlations observed between illusion rates and language model entropy, we turned to human measures of the information state at the point of the substitution to try to obtain a more precise measure of the underlying source of hypothesized attention shifts. This was done using a version of the cloze task, in which participants are asked to continue a fragment of a sentence. In this case, we asked participants to provide both the full continuation of the question (whereas typical cloze tasks often ask for only the next word) and the answer. We are interested in two aspects of the provided continuations: whether question continuations include the intended word, and whether provided answers converge to the same answer across participants. These two metrics allow us to disentangle expectations about the particular form the question will take from expectations about the overall message of the question.

8.5.1 Participants

52 workers recruited through Amazon Mechanical completed the task. 2 participants were excluded for failing to follow instructions. The task was designed to last 10 to 20 minutes and participants were compensated \$4.

8.5.2 Materials

Each of the 50 stimuli from our Moses illusion experiment was trimmed at the point of the impostor to create a question fragment. That is, participants saw questions only up to but excluding the impostor.

8.5.3 Procedure

Following instructions and practice trials, participants were presented with the 50 question fragments one at a time. Their task was to type in a completion for the fragment such that it created a trivia question, and to additionally type in an answer to the trivia question they generated.

8.5.4 Analysis

There are at least two dimensions of this dataset that are potentially useful. First, the degree to which comprehenders have predicted the intended word prior to the impostor is reflected in the proportion of question completions that contain the intended word. Second, the amount of information contained in the context prior to the impostor that allows comprehenders to identify the answer before the impostor has even occurred is reflected in the proportion of answers that overlap (suggesting the same inferred question).

For the first of these metrics, we coded each question completion for whether it contains the intended / non-anomalous word. This was done first automatically by checking for string overlap, and then hand-checked to ensure that misspellings and minor variants of the intended word were included. For each item we then compute the proportion of question completions that contain the intended word.

For the second metric, we computed the proportion of answers that overlapped across participants. This was done first automatically by grouping identical strings and then checked by hand to ensure that misspellings and semantically equivalent completions were counted as overlapping. For each item we then take the maximum overlap proportion (i.e. for the answer that the largest number of participants gave, what proportion of participants gave it) as an index of the predictability of the answer prior to the impostor.

We then measured the correlation between each of these properties and illusion rates across items.

8.5.5 Results

First we consider the proportion of question completions that contain the intended word, for each item. Per-item proportions ranged from 0% to 100% (see Figure 8.12). Split-half correlation analyses found that the median reliability was .96 (95CI: .94, .98). A linear regression did not identify a statistically discernable correlation between the proportion cloze completions containing the intended word and illusion rates ($\beta=-0.0005$, $SE=0.001$, $z=-0.54$, $p=0.59$, adjusted $R^2=-0.02$), as shown in Figure 8.13.

We now consider the proportion of cloze completions for which participants generated questions that had the same answer, for each item. Per-item proportions ranged from 8% to 100% (see Figure 8.14).

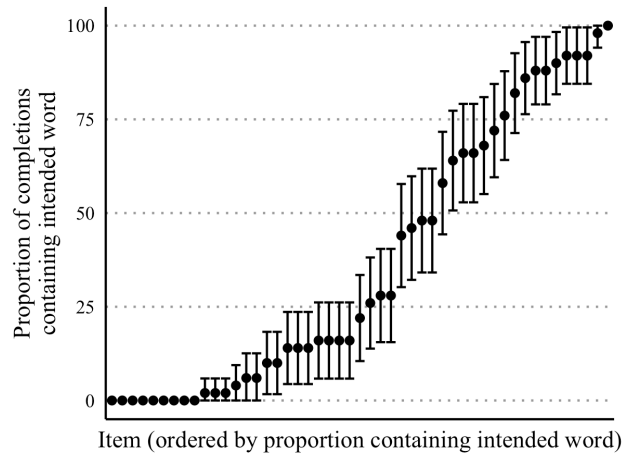


Figure 8.12: Proportion of cloze completions in which the participant-supplied question contained the intended word, for each of 49 stimuli, shown with 95-percent confidence intervals.

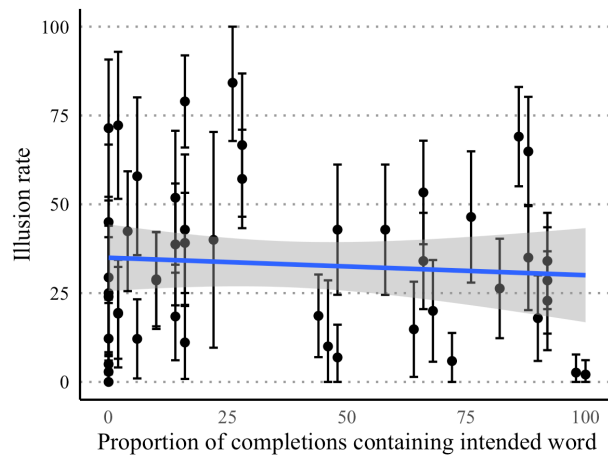


Figure 8.13: Correlation across 49 items between the proportion of cloze completions in which the participant-supplied question contained the intended word (Experiment 21) and illusion rates (Experiment 17)

Split-half correlation analyses found that the median reliability was .94 (95CI: .90, .96). A linear regression identified a small but statistically discernable correlation between the proportion cloze completions containing the intended word and illusion rates ($\beta=0.002$, $SE=0.001$, $z=2.25$, $p=0.03$, adjusted $R^2=0.08$), as shown in Figure 8.15.

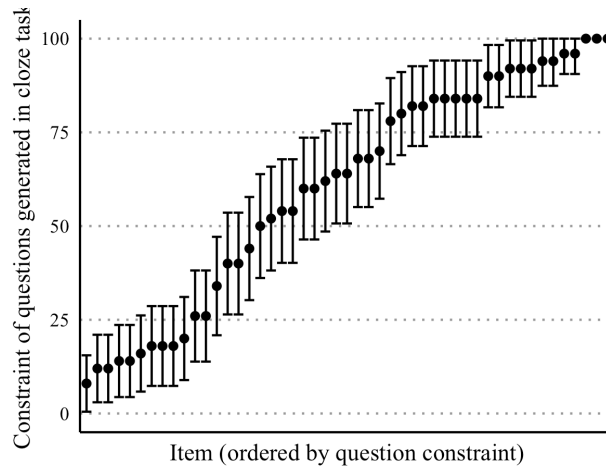


Figure 8.14: Proportion of cloze completions in which the participant-supplied question had the same answer, for each of 49 stimuli, shown with 95-percent confidence intervals.

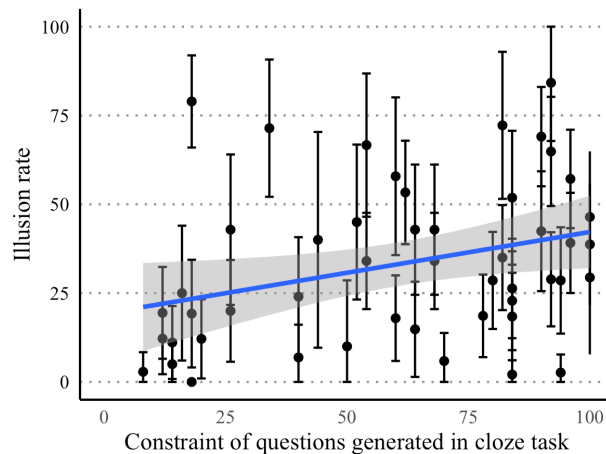


Figure 8.15: Correlation across 49 items between the proportion of cloze completions in which the participant-supplied question had the same answer (Experiment 21) and illusion rates (Experiment 17)

8.5.6 Discussion

We observed a reliable correlation between illusion rate and answer convergence, but not between illusion rate and the proportion of question completions that contained the intended word. This suggests that our concern that the lack of correlations with language model entropy may have been due in part to the entropy measure's focus on possible next words, rather than overall certainty about the message, was appropriate. More importantly, it suggests that some of the variability in illusion rates across items is due to differences in whether comprehenders were able to infer the answer to the question prior to

the impostor. However, as with our discussion of similarity measures, it is important to note that this correlation is weak and much of the variability in illusion rates remains unexplained.

8.6 Conclusion

In Experiments 18-21 we demonstrated that there is reliable itemwise variability in substitution illusion rates, and identified a few variables that correlate with this variability. However, as we have noted several times, much of the variability is not explained by these variables. in Figure 8.16 we show the results of an analysis in which all predictors we considered are combined. We still find that much of the variability in illusion rates is not explained.

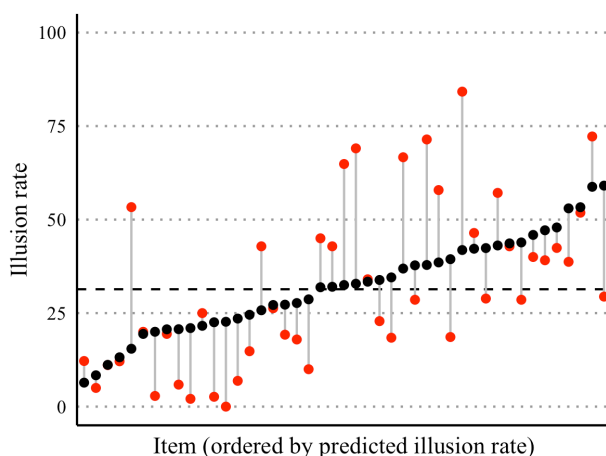


Figure 8.16: Predicted illusion rates (black dots) for each of 49 substitution illusion stimuli, based on a model using all predictors from Experiments 18-21, compared to actual illusion rates (red dots) from Experiment 17.

This variability is important partly in that it reinforces the value of investigating substitution illusions as a window into the sentence comprehension system. There are, of course, many errors in comprehension (and production) that happen to language users every day. Much of this is, from a psycholinguistic perspective, only noise. Our findings suggest that the substitution illusion is not noise — illusion errors are not random occurrences that arise when a participant happens to be thinking about what they’re going to have for lunch instead of the sentence in front of them. The systematicity of the illusion makes this clear. The unexplained systematicity (in the sense that the variables that we already know influence

illusion rates are not enough to predict illusion rates) makes it clear that our current understanding of the illusion is not complete.

One important note about these investigations is that the internal split-half reliability of each of our measures was quite high. This is important because there is relatively little hope of correlating two measures if each measure has low internal reliability (as has been shown to be the case for, e.g., attempts to explain itemwise variability in semantic priming effects; see Heyman et al. 2018). Since split-half reliability was consistently high for both illusion rates and the experimental measures we attempted to correlate with illusion rates (familiarity, cloze), these concerns should be alleviated.

There are a few ways that the measures we have attempted to correlate with illusion rates could be improved. For example, one might improve on similarity measures by collecting human judgments, or by measuring impostor-context relatedness in addition to impostor-intended similarity. Importantly, though, the ultimate goal of our exploration of substitution illusions is not to identify the variables that best predict illusion rates. A list of properties that correlate with illusion rates is not a theory. We therefore turn to a different strategy for investigating illusions in the next chapter.

Chapter 9 Substitution illusions: levels of representation

While it is clear from our findings in Chapter 8 that the substitution illusion is not well explained by hypotheses that treat the illusion as a highly general phenomenon, a more precise definition of the circumstances under which the illusion arises remains elusive. Here we turn to the specific mechanistic hypotheses that were considered in section 7.2, and consider possible strategies to test them. Note that this requires turning to different experimental strategies — rather than identifying factors that correlate with the naturally occurring variability across items, we aim to introduce variability by deliberately manipulating factors that are predicted to matter or not matter under various accounts. While explaining item-wise variability remains a viable strategy, it cannot be the end goal of an investigation of substitution illusions. Any variable with does or does not correlate with illusion rates can likely be accommodated by most candidate hypotheses by simply appealing to unmeasured characteristics of sentences with which the variable may or may not be confounded. Thus, while correlational analyses are a useful exploratory tool, other approaches are also necessary.

In section 7.2 we considered four possible cognitive operations that underlie successful sentence comprehension, and ways that each of these operations could be fallible such that they yield substitution illusions. We briefly review these hypotheses here before turning to a series of experiments designed to test them.

First, we considered the possibility that lexical access is fallible, such that the impostor word is never actually selected from the mental lexicon, and, as a result, there is no way for the world knowledge violation to be detected. This is in fact a family of hypotheses, since lexical access could go awry at any of the three levels discussed by Levelt, Roelofs, & Meyer 1999, and could err either by selecting the wrong lexical entry (presumably selecting the intended word instead of the actually-seen impostor) or by failing to select an entry from the mental lexicon at all. This latter possibility requires assuming some mechanisms

whereby syntactic structure-building can proceed despite the lack of an identified lexical entry in some position of the tree, but presumably such assumptions are independently required, since comprehenders are able to parse sentences with novel words, and can even parse sentences with *many* novel words as is the case with Jabberwocky. One difficulty for hypotheses in this category is accounting for the effect of the semantic similarity of the impostor (to either the intended word or the context; recall that we do not know which matters). If the impostor's lexical entry is never identified at all, it is not obvious how impostors whose lexical entries are tied to related world knowledge can have any different impact than impostors whose lexical entries are tied to unrelated world knowledge. This can be solved in a few ways. One possibility is to assume connections between world-knowledge-related words within the mental lexicon, such that mis-selection becomes more likely when many of these connections exist. Another possibility is to allow access to the world knowledge that is connected to a candidate lexical entry before that lexical entry has been "selected" as the current word.

If the problem does not lie in the lexical access pipeline, it could lie in a subsequent step. A second possibility is that it is the activation and selection of world knowledge that is related to the impostor that leads to the error. Not all of the world knowledge that is associated with any given lexical item needs to be retrieved for every instance of that lexical item — rather the context should spotlight those components of the meaning that are relevant to the present message. Thus, in cases where the impostor and intended word share components of their non-linguistic meaning, and those components are the ones emphasized by the sentential context, the comprehender will fail to detect the anomaly. A third possibility is that the comprehender uses a non-compositional strategy for generating a sentence meaning, which presumably fails to incorporate the meaning of the impostor into the meaning of the sentence. This again faces difficulty in accommodating the well-documented similarity effect, and so an initial "coherence check" would have to be stipulated. And finally, the mapping from linguistic representations of sentence meaning to non-linguistic thoughts (or the comparison between two thoughts — one coming from the sentence meaning and the other coming from long term memory) could involve some loss of information.

Notably, each of these hypotheses could be supplemented by independent explanations regarding attention and presuppositional status. That is, for any operation that is hypothesized to lead to the sub-

stitution illusion, we might wish to additionally determine whether this operation proceeds differently when the comprehender is attending to something else, or when the linguistic signal which these operations interact with is clearly expressing a thought that is presupposed by the speaker.

In Experiments 22 and 23 we are specifically concerned with determining whether the substitution illusion is a consequence of failed lexical access or some later computation. The basic idea is to increase the probability that the impostor will be successfully selected through priming. Identity priming effects are robust in lexical decision tasks (e.g., Gaston et al. 2021, among others) and typically understood to reflect the increased activation of the target word on its second presentation. Under the hypothesis where the illusion arises due to mis-selection of the intended word instead of the impostor, we would expect that priming the impostor prior to the presentation of the illusion sentence would increase the probability of accurate lexical retrieval, and reduce illusion rates. Similarly, under this hypothesis, priming the intended word prior to the illusion sentence should increase the probability of *mis*-selection, and increase illusion rates. If lexical access fails not due to mis-selection but due to failed selection, we expect a reduction in illusions when the impostor is primed, but no boost in illusions when the intended word is primed. Under the three alternative hypotheses, in which lexical access proceeds smoothly and errors arise in a subsequent stage, priming effects are not clearly predicted.

The experiments reported here address these predictions as follows. First, Experiment 22 tests the basic illusion effect using a truth value judgment task with declarative sentences, instead of the trivia-question format used in Experiment 17. The use of declaratives is useful because it allows us to reduce the influence of simultaneous tasks which may pull attention away and “artificially” inflate illusion rates. The true/false task also simplifies data analysis because it does not require participants to freely type in an answer. Thus Experiment 22 aimed to verify that such a task was feasible, and identify the best stimuli (defined as those that yield a clear and consistent illusion and which a large number of participants know) for the subsequent priming experiments. Experiment 22 also served the added goal of attempting to replicate findings from Büttner 2007, suggesting that illusion rates are systematically lower for sentences than for questions. Experiment 23 tests the predicted priming effect using a task that interleaves lexical decision trials and sentence judgment trials.

9.1 Experiment 22: declaratives

Prior to conducting the priming experiments of interest, we aimed to validate the use of the true/false judgment task for declaratives as a way to measure illusions. A variant of this task was used in subsequent experiments. We also tested a wide variety of stimuli so that the best ones could be selected for use in subsequent experiments.

9.1.1 Participants

100 workers recruited through Amazon Mechanical Turk completed the task. It was designed to last 45 to 60 minutes and participants were compensated \$12.

9.1.2 Materials

We used the 50 items¹⁰² from Experiment 17 as a starting point. These questions were turned into declarative statements that could be judged true or false. We added to these 26 items adapted from prior studies (Cook et al. 2018; Reder & Kusbit 1991; Umanath, Dolan, & Marsh 2014), and 44 newly-written substitution illusion stimuli. This resulted in 120 total stimuli.

9.1.3 Procedure

Participants were told that the task involved judging trivia statements as true or false. Following instructions and practice trials, they were presented with each of our 120 stimuli, along with three options for a response: true, false or “I don’t know”. Extensive instructions and examples prior to the onset of the task ensured that participants understood the kind of falsehood they were to identify. That is, we were worried that comprehenders might think they were supposed to judge all trivia statements referring to

¹⁰² Although only 49 of the 50 items used in Experiment 17 were analyzed for that experiment, all 50 items were used here. Recall that one item was problematic because the introduction of the impostor did not make the question anomalous, it simply made it a different, still reasonable question. This issue does not arise for declaratives.

fictional texts as false¹⁰³, or judge statements that don't tell the "whole truth" as false¹⁰⁴, so participants also completed an "instructions quiz" to ensure that they understood these aspects of the task. Each participant saw 25 anomalous and 25 non-anomalous questions, which were counter-balanced across participants.

Following the trivia task, these participants completed a "knowledge check" task. This was identical to the knowledge check used in Experiment 17.

9.1.4 Analysis

As with Experiment 17, we discarded any trial for which the participant answered "I don't know" or answered the corresponding knowledge check question incorrectly. 8,271 trials remained following these exclusions, of which 3,882 trials were anomalous sentences (i.e. potential illusions). We measured illusion rates for each item as the proportion of trials that received a "true" judgment.

9.1.5 Results

Per-item and overall illusion rates are shown in Figure 9.1. The overall illusion rate, collapsing across items, was 37%, but per-item illusion rates varied from 0% to 100%. In order to determine if this variability was systematic or could be explained by random noise, we again computed split-half reliability measures, which found that the median reliability was .77 (95CI: .70, .83).

In addition to measuring variability across items, we were interested in whether declaratives have lower illusion rates than questions. At first glance, this does not appear to be the case, since the present experiment revealed an overall illusion rate of 37%, whereas Experiment 17, which used questions, re-

¹⁰³The instructions included the following text to address this issue: *Some statements will be about fictional events. In these cases, we ask that you evaluate the statement with respect to the story. For example "Harry Potter is a wizard" should be judged true and "Harry Potter is a vampire" should be judged false. Even though there is no real wizard named Harry Potter, it is clear that the statements are about the fictional book / film series, in which the character named "Harry Potter" is a wizard, not a vampire.*

¹⁰⁴The instructions included the following text to address this issue: *Some (but not all) of the statements will have errors that make the statement false. For example, "In the United States, the king is elected by the citizens" is false, since the United States does not have a king. This is the type of falsehood you will encounter in this experiment. (You do not need to worry about identifying statements that aren't the "whole truth". For example, we consider "In the United States, the president is elected by the citizens" to be true. You may be concerned that, for example, this statement glosses over the fact that citizens do not directly elect the president, due to the electoral college, or the fact that some citizens are not able to vote. For our purposes, however, you may judge such a sentence to be true.)*

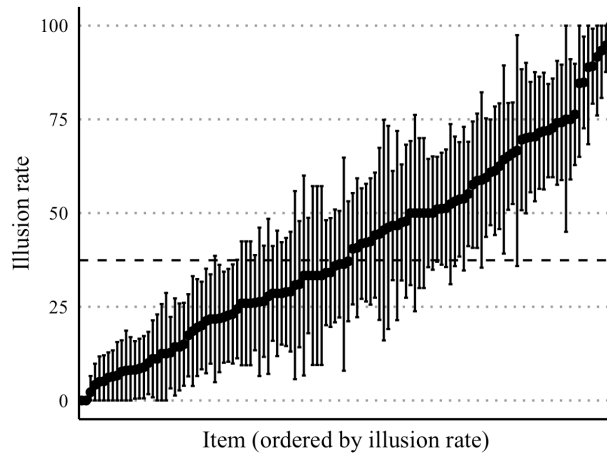


Figure 9.1: Substitution illusion rates for each of 120 declarative stimuli tested in Experiment 22, shown with 95-percent confidence intervals. The average illusion rate across items is indicated by the black dashed line.

vealed an overall illusion rate of 31%. But of course the present experiment used many stimuli that were not tested in Experiment 17, and since we know there’s substantial variation in illusion rates for different stimuli, this comparison is not appropriate. Rather, we can compare illusion rates for just the 49 stimuli that were tested in both experiments. Here we again do not find the previously-reported trend of an increase in illusion rates for questions (31% for questions, 34% for declaratives). These findings are presented in Figure 9.2, alongside reported data from Büttner 2007.

9.1.6 Discussion

In Experiment 22 we investigated whether robust illusions can be observed with a declarative true/false judgment task. We found similar overall illusion rates to previous studies and similar variability across items to what was observed in Experiment 17. Moreover, our findings do not replicate the previously reported effect of sentence type on illusion rates — that is, Büttner 2007 reported that illusion rates are lower for declaratives than for questions, but we find a small numerical trend in the opposite direction. One likely explanation for this discrepancy in findings is that we tested 49 stimuli with 100 participants in each task, whereas Büttner 2007 tested 8 stimuli with 40 participants total (20 in the questions-first task and 20 in the declaratives-first task). While the design used by Büttner 2007 benefits from the use of

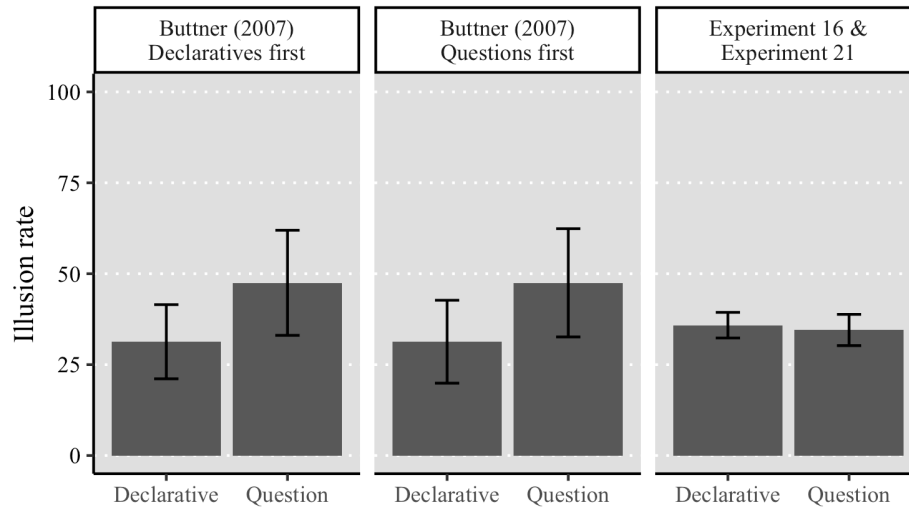


Figure 9.2: Average substitution illusion rates for questions and declaratives in our data (right panel) and data reported by Büttner 2007 (left panel, center panel). Error bars indicate 95-percent confidence intervals, based on standard deviations calculated across participants.

a within-subjects comparison (which has, all else equal, more statistical power than our between-subjects comparison), the small numbers of participants and items make their estimates less precise and more prone to type M errors (see Vasishth et al. 2018). We thus conclude that there may be a small effect of questions versus statements, but it is likely not as large as that reported by Büttner 2007. There may also be no effect.

The results of Experiment 22 provide a valuable test set for future correlational analyses of the type we explored in Chapter 8. Because the set of stimuli is larger (120 items instead of 49), there is greater statistical power to identify potential correlations. These results also suggest that future explorations of substitution illusions that aim to experimentally manipulate the illusion rate may be better off using a subset of stimuli with reliably high (and similar) illusion rates, since the variability across items, which allows for greater statistical power in correlational analyses, undermines power in experimental manipulations.

9.2 Experiment 23: priming through lexical decision

Having established that true/false judgments of declaratives are a valid way to induce and measure substitution illusions, we turn to our primary question, namely the influence of lexical priming on illusion rates. Recall that the hypothesis that illusions are due to errors in lexical access predicts that facilitating correct lexical access by priming the impostor should substantially reduce illusion rates. Additionally, the version of this hypothesis that specifically attributes the illusion to *mis*-selection of the intended word instead of the impostor predicts that increasing the likelihood of incorrect lexical access by priming the intended word should substantially boost illusion rates. Other hypotheses do not predict large differences in illusion rates as a function of priming (though they can accommodate small differences, as we discuss below).

In Experiment 23, participants completed two tasks, which were interleaved with one another: a lexical decision task, and a true/false sentence judgment task, similar to what was done in Experiment 22. The key manipulation was the identity of the word in the lexical decision task. There were four conditions: the impostor (e.g. “Moses”), the intended word (e.g. “Noah”), an unrelated word (e.g. “snarl”), or a word that occurs in the target sentence but is not the locus of the anomaly, which we label the “related” word (e.g. “kind”).

The key comparisons are (a) illusion rates when the sentence follows the impostor compared to illusion rates when it follows the unrelated word, and (b) illusion rates when the sentence follows the intended word compared to illusion rates when it follows the unrelated word. The fourth condition (the related word) was included to discourage a particular strategy that we worried participants might adopt in this kind of task. If participants pick up on the fact that the words they are judging in the lexical decision task are often in the next sentence they see, they might begin to treat the lexical decision word as a clue, pointing them toward the locus of the error. The inclusion of lexical decision words like “kind” prior to “Moses brought two animals of each kind on the ark” makes this a bad strategy. Thus, participants should be less likely to adopt it.

These same four lexical decision words were also paired with the non-anomalous version of the sentence. We do not expect any effect of the identity of the lexical decision word on accuracy rates for non-

anomalous sentence trials. The inclusion of these trials allows us to identify unintended strategies (for example, if participants simply always judge a sentence “false” if it contains a word they just saw, this would be detectable as a decreased accuracy for non-anomalous sentences preceded by the intended word and the related word).

Because the task was a lexical decision task, some non-word trials are needed. This was accomplished in two ways. First, each sentence was preceded by not one but four lexical decision trials (so that the structure of the experiment was WORD, WORD, WORD, WORD, SENTENCE, WORD, WORD, WORD, WORD, SENTENCE, etc.). For the experimental trials, the critical lexical decision was one of these four lexical decision judgments (counter-balanced for where in the sequence of four it appeared), and the other three trials were always two non-words and one unrelated word. Second, we included filler trials where participants saw four lexical decision trials paired with true/false sentence judgments, but we did not manipulate the relation between these two tasks. For the filler trials, the balance of words and non-words among the four lexical decision trials was varied, ranging from all-words to all-non-words, so that within a block of four lexical decision trials, the outcome of upcoming trials could not be predicted based on what had been seen (e.g., if we had made all blocks of four lexical decision trials balanced, with two words and two non-words, then a participant having seen two words and one non-word would know that the next stimulus was a non-word, without even having to see it). The details of how these stimuli were generated are described below.

9.2.1 Participants

128 workers recruited through Amazon Mechanical Turk completed the task. However, two participants’ data was not successfully recorded, due to problems with PCIBex. The task was designed to last 30 to 45 minutes and participants were compensated \$9.

9.2.2 Materials

The experimental materials for the present experiment were subject to a number of restrictions that did not arise for previous experiments. Here we describe the process of generating stimuli that fit these re-

quirements. Critically, while variability across items was a useful signal in previous experiments, here we are interested in comparisons between *conditions*, not between items, so item-wise variability undermines power and should be minimized.

We began with the 120 items for which we collected data in Experiment 22. The main constraints on sentence items for the present experiment were that all experimental items have a high yield (most participants pass the knowledge check), a high illusion rate, and both the impostor and intended word is a single orthographic word (so that it can be presented in a lexical decision task). In addition, all experimental items and fillers must be natural and have a clear truth value (i.e., not a matter of opinion).

We began by filtering the items based on this last criterion, removing any item that was not clearly a canonical substitution illusion.¹⁰⁵ Then, for each item, we computed the proportion of participants who correctly answered the knowledge check (the “yield”) in Experiment 22. We additionally computed each item’s illusion rate. Candidate experimental items were those with a yield of 50% or greater and an illusion rate of 25% or greater. The remaining items were classified as candidate fillers. Items with multi-word impostors or multi-word intended strings were also classified as candidate fillers. In order to achieve equal numbers of fillers and items (32 each) we removed the fillers and experimental items with the most words (since, all else equal, a shorter experiment is preferable to a longer experiment).

Among the selected experimental items, the average yield was 78%, ranging from 51% (e.g., “Lines of constant longitude (*latitude*), sometimes called ‘parallels’, run east–west as circles parallel to the equator.”) to 99% (e.g., “Santa Claus wears a red suit and gives out birthday (*Christmas*) presents from his sleigh.”). The average illusion rate was 49%, ranging from 26% (e.g., “Honolulu is the capital of the state of Hawaii, which consists of a chain of islands in the Atlantic (*Pacific*) Ocean.”) to 83% (e.g., “The famous speech that begins ‘Four score and twenty (*seven*) years ago’ was given by Abraham Lincoln.”). The 32 fillers fall into two general categories: “hard fillers” (those that were disqualified from being experimental items because less than 50% of participants passed the knowledge check) and “easy fillers” (those

¹⁰⁵As an example, consider the item *The Boston Tea Party, in which chests of tea were thrown into the Plymouth Harbor as a protest against Parliament’s Tea Act*. The sentence is anomalous because of the impostor “Plymouth Harbor” in place of Boston Harbor. However, it is reasonable to suspect that some proportion of participants genuinely did not know this, but were able to pass the knowledge check because it’s multiple choice and if we’re talking about the *Boston* tea party then probably *Boston* Harbor is the best option.

that were disqualified from being experimental items because less than 25% of participants who passed the knowledge check demonstrated illusions).¹⁰⁶

Having selected the target sentences, we generated the lexical decision strings using data from the English Lexicon Project (Balota et al. 2007) and the semantic similarity metric in SpaCy (Honnibal & Montani 2017)¹⁰⁷. There are six types of lexical decision strings that we needed: impostor primes (e.g. “Moses”), intended primes (e.g. “Noah”), related primes (e.g. “kind”), unrelated primes (e.g. “snarl”), filler words, and filler non-words. Impostor primes and intended primes are determined by the choice of sentence stimuli. Related primes were selected from the other words in the sentence contexts, with the constraints that no prime was used multiple times throughout the experiment and no related prime was substantially more related to the impostor than to the intended word or vice versa (operationalized as a difference in similarity measures greater than .1 on the -1 to 1 scale).

Each unrelated prime needed to be both unrelated to the sentence with which it was paired and a fairly good control for the impostor prime, since the comparison between these two conditions is critical. They were therefore length-matched and had similar RTs in a lexical decision task, per the English Lexicon Project norms¹⁰⁸. No unrelated prime had a similarity score greater than .2 for the comparison with the impostor, the intended word, or the sentence context.

Each filler word had to be similar in length to the primes (between 4 and 14 characters), similar in lexical decision RT to the primes, relatively low in similarity (less than .5) to all 32 experimental items, and very low in similarity (less than .1) to the experimental or filler sentence with which it was paired. Each filler non-word had to be similar in length and lexical decision RT to the selected words, and no filler non-word was an orthographic neighbor to any impostor prime or intended prime (e.g. “joses” is a nonword neighbor of “Moses”).

Example experimental stimuli and fillers are shown in Table 9.1

¹⁰⁶Note that this isn’t a perfect division of filler items because in principle an item can be both “too hard” (low knowledge check yield) and “too easy” (low illusion rate among those who pass the knowledge check).

¹⁰⁷This is qualitatively similar to word2vec, which was introduced in section 8.3

¹⁰⁸For three items, lexical decision RTs were not available so we matched on frequency instead

Sentence Condition	Prime Condition	lexical decision 1	lexical decision 2	lexical decision 3	lexical decision 4	sentence
anomalous	unrelated	leview	swamping	arboreel	cults	Santa Claus wears a red suit and gives out <i>birthday</i> presents from his sleigh.
anomalous	impostor	leview	birthday	arboreel	cults	Santa Claus wears a red suit and gives out <i>birthday</i> presents from his sleigh.
anomalous	intended	leview	Christmas	arboreel	cults	Santa Claus wears a red suit and gives out <i>birthday</i> presents from his sleigh.
anomalous	related	leview	suit	arboreel	cults	Santa Claus wears a red suit and gives out <i>birthday</i> presents from his sleigh.
non-anomalous	unrelated	leview	swamping	arboreel	cults	Santa Claus wears a red suit and gives out Christmas presents from his sleigh.
non-anomalous	impostor	leview	birthday	arboreel	cults	Santa Claus wears a red suit and gives out Christmas presents from his sleigh.
non-anomalous	intended	leview	Christmas	arboreel	cults	Santa Claus wears a red suit and gives out Christmas presents from his sleigh.
non-anomalous	related	leview	suit	arboreel	cults	Santa Claus wears a red suit and gives out Christmas presents from his sleigh.
anomalous hard filler	N/A	whitan	subs	Moobe	enchunting	The name of the country that is located between France and Spain in the <i>Alps</i> is Andorra.
non-anomalous hard filler	N/A	whitan	subs	Moobe	enchunting	The name of the country that is located between France and Spain in the Pyrenees is Andorra.
anomalous easy filler	N/A	sweetly	acrasive	angrily	teething	One can visit the Colosseum in the <i>Spanish</i> city of Rome.
non-anomalous easy filler	N/A	sweetly	acrasive	angrily	teething	One can visit the Colosseum in the Italian city of Rome.

Table 9.1: Example stimuli for Experiment 23. Non-word lexical decision trials are shown in boldface for clarity, though they were not bolded in the experiment. Impostors in anomalous sentences are shown in italics for clarity, though they were not in the experiment.

9.2.3 Procedure

Participants were told that the task involved judging trivia statements as true or false and judging character sequences as words or non-words. Following instructions and practice trials, they were presented with each of our 32 stimuli and 32 fillers. Each lexical decision trial was presented with three options for a response: word, non-word, or skip. Each sentence trial was presented with three options for a response: true, false, or skip. Choosing a response triggered the presentation of the next trial. Extensive instructions and examples prior to the onset of the task ensured that participants understood the kind of falsehood they were to identify and the relevant notion of “word” (for example, unlike Scrabble rules, we consider proper nouns to be words). We additionally asked participants to respond to each trial as quickly and accurately as possible, and interspersed random opportunities to take a break throughout the experiment. Although our prior substitution illusion experiments did not emphasize speedy responses, and had participants take as much time as they needed for each trial, we decided to encourage a faster pace

in the present experiment in order to make sure participants proceeded directly from the lexical decision trials to the corresponding sentence trial. Due to the nature of the priming effect under investigation, it would be disadvantageous if participants chose the fourth lexical decision trial of a trial sequence as an opportunity to take a break — by the time they come back, they would have forgotten the prime. Thus participants were encouraged to move through the task quickly without sacrificing accuracy, and to only take breaks at the designated times. Participants also completed an “instructions quiz” to ensure that they understood these aspects of the task. Each participant saw 64 trial sequences, each consisting of four lexical decision trials and one sentence trial. 32 of these were fillers (of which 16 used true trivia statements and 16 used false trivia statements), and 32 were experimental items, distributed across the eight experimental conditions depicted in Table 9.1 in a Latin Square design.

Following the trivia task, these participants completed a “knowledge check” task, which was the same as what was used in Experiments 17 and 22.

9.2.4 Analysis

We again removed any trial for which the participant did not answer the corresponding knowledge check correctly, and any trial for which the participant chose “skip” (the equivalent of the “I don’t know” option used in previous experiments) for the sentence judgment. Because we are interested in comparisons between conditions with a binary outcome in an experiment where both items and participants are intended to be random samples¹⁰⁹, a logistic mixed effects model is in principle appropriate. However, the removal of data points based on the knowledge check and the “skip” option has the potential to result in imbalances in the data, which, as Eager & Roy 2017 documented, can result in severe convergence problems, especially for binary data. Following their recommendation, we use Bayesian mixed effects models using the *brms* package (Bürkner 2017).

¹⁰⁹Of course, items are very much not “random” here in the sense that we deliberately selected items with high yields and high illusion rates

9.2.5 Results

As can be seen in Figure 9.3, the experiment revealed numerically small trends toward an increase in illusions for the intended-prime condition (“Christmas”, in the example item) and a decrease in illusions for the impostor-prime condition (“birthday”, in the example item). A Bayesian mixed effects model with default priors did not identify a reliable effect of either the intended prime ($\beta=0.20$, 95-percent credible interval = $[-0.25, 0.64]$), or the impostor prime ($\beta=-0.25$, 95-percent credible interval = $[-0.64, 0.14]$).

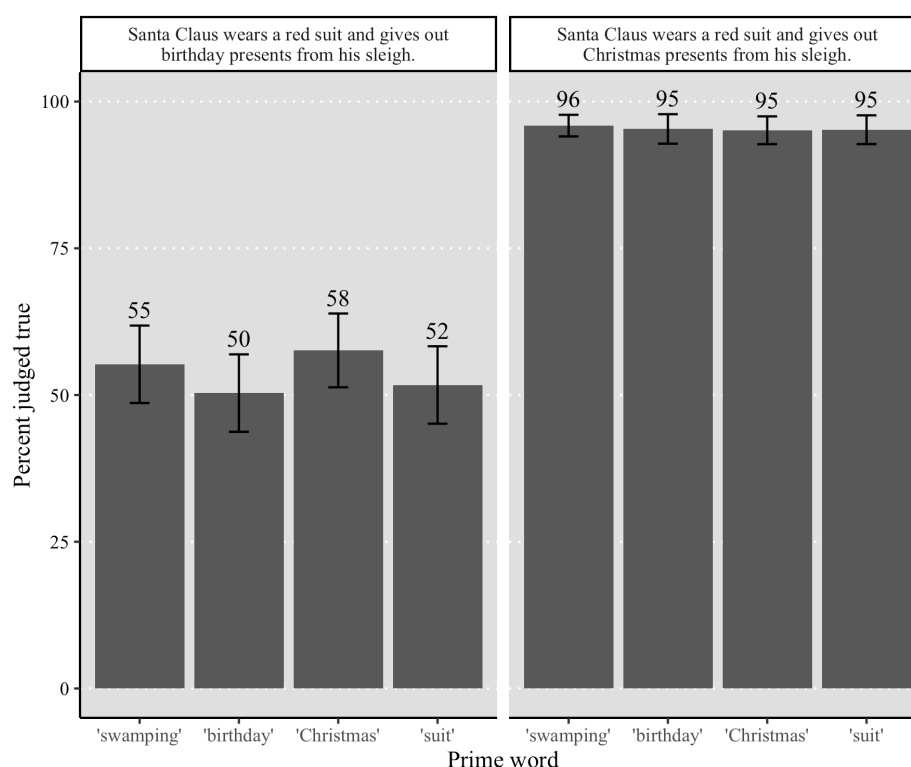


Figure 9.3: Proportion of sentence trials receiving a “true” judgment for each of the eight critical conditions for Experiment 23

9.2.6 Discussion

Experiment 23 tested a prediction made by the family of hypotheses that attribute substitution illusions to problems in lexical access processes, in contrast with hypotheses that attribute illusions to post-lexical processes, including polysemy resolution, semantic composition, and world knowledge access. In brief, we do not find conclusive evidence that the illusion rate changes predicted by lexical access accounts arise,

though there is of course difficulty in reasoning from a lack of statistically reliable effects. A related issue is that the lexical access hypothesis does not make quantitatively precise predictions, in part due to the qualitative nature of the hypothesis, and in part due to the complexity of the linking assumptions.

Recall that there are essentially two dimensions along which hypotheses in the lexical access group can vary: the particular level of lexical representation at which access goes awry, and the nature of the error (selection of the intended word, or no selection at all). First we focus on the prediction that the mis-selection and failed selection accounts share: a decrease in illusion rates when the impostor is primed. Considering the question of the level of lexical access, the extent to which the present experiment bears on the hypothesis depends on the levels of lexical representation we believe a comprehender accesses in a lexical decision task. A comprehender could in principle base their decision on having identified a word form that matches the stimulus in their mental lexicon, or having identified a lemma, or a semantic representation. Accordingly, identity priming effects could in principle be because access to the word form representation was facilitated, access to the lemma was facilitated, or access to a semantic representation was facilitated (or all three) on the second exposure. If we assume for a moment that lexical decisions are based only on the word form representation, and identity priming operates only at the word form level, the present findings merely suggest that facilitation in accessing the word form level is not sufficient to eliminate the illusion. If this is true, the present findings suggest only that word form access is not the locus of the error, but tell us nothing about lemmas or semantic representation. Similarly, if we assume that lexical decisions are based on lemmas and identity priming operates at the lemma level, the present findings suggest that neither word form access nor lemma access is the locus of the error, but leaves open the possibility of problems at the semantic level.

However, there is some evidence that a participant in a lexical decision task does access even the semantic representation of the word. This evidence comes from facilitation in lexical decision as a result of overlap in the meanings of adjacent words (often labeled “semantic priming” effects). Comprehenders respond more quickly and more accurately to *nurse* when it is preceded by *doctor* than when it is preceded by *table* (e.g., Meyer & Schvaneveldt 1971, among many others). This would seem to suggest that some level of representation at which *nurse* and *doctor* are related and *nurse* and *table* are not is accessed in the

lexical decision task¹¹⁰. If this is the right interpretation of such effects, then we should be able to assume that access to the word form, lemma, and semantic representations of the impostor are all facilitated by exposure to the impostor. Thus the non-effect in Experiment 23 would suggest that none of these is the locus of the error.

The evidence from semantic priming is not quite so straightforward, though. Many studies find effects of both relatedness proportion (that is, over the course of the experiment or block, how often do adjacent trials use meaning-related words) and stimulus onset asynchrony (SOA) on the size of the semantic priming effect (Neely 1977; Neely, Keefe, & Ross 1989; Hutchison, Neely, & Johnson 2001; Bodner & Masson 2003; among others). These effects are sometimes taken as evidence against automaticity in the semantic priming effect, and specifically against a “spreading activation” explanation for the effect. Importantly, for our purposes, the relatedness proportion effect seems to suggest that access to meaning-level representations of a word may not always occur in a lexical decision task, especially when the overall experimental conditions do not make it beneficial. Our experiment had no related word pairs, except to the extent that such pairs arose by chance, and so we should not necessarily assume that facilitation at the lexical semantic level occurred.

Thus, the extent to which our findings challenge lexical-access based explanations for the illusion depends on the level of lexical access that is assumed to be accessed in the lexical decision task. We can be relatively confident that failed access at the word form level is not the right analysis, because we can be confident that the lexical decision task involves access to and facilitation of the word form level. As we move up to the lemma and semantic level, we are less certain. This suggests that a useful follow up to the present experiment would be to either deliberately include a high proportion of related word pairs among the lexical decision fillers or to change the task for the isolated words from a lexical decision task to something that requires access to a representation of the meaning of the word (e.g. “is it an animal?”)

A second concern is that the priming manipulation we used could have simply not facilitated access to the relevant representation enough to influence illusion rates. For example, one might argue that a single exposure to the impostor word does not result in a sufficiently large boost to its activation at the

¹¹⁰Presumably the level at which they are related is some meaning-related level. See Lucas 2000 for a review concerning whether such effects are better accounted for by word co-occurrence effects, i.e. “association”.

point when the sentence is read to result in large, reliable changes in the outcome of the later lexical access operation. Relatedly, it could be that the problem isn't the single exposure but the timespan — perhaps by the time the word is encountered in the sentence, its activation for the primed word has decayed back to baseline. One critical piece of evidence against the latter concern is that identity priming effects are quite persistent. Gaston et al. 2021 found robust identity priming effects in the lexical decision task for both short-lag (i.e., adjacent) and long-lag (i.e. between 1 and 33 intervening trials) experiments. Though the effect size is numerically somewhat smaller for the long-lag comparison, it is clear that a return to baseline is not plausible for our timespan (0 to 3 intervening trials). We can also look at our data focusing only on the trials in which the critical word was fourth in the list of lexical decision trials, and therefore immediately prior to the illusion sentence (Figure 9.4). While the effect size for the reduction in illusions when the impostor is primed is numerically larger for this subset than in the aggregated data — a 14-point reduction, where before we saw a 5-point reduction — we can reasonably infer that this is mostly due to noise in the smaller sample, since it is primarily the unprimed baseline (i.e. *swamping*, in the figure) whose illusion rate is different (62% illusions for fourth-trial primes, compared to 55% illusions in the aggregate), and there is no reason for this condition to differ at all as a function of presentation order.

Thus we do not have reason to believe that the relative timing of the lexical decision prime and sentence target was too long for the prime to influence the target. We cannot rule out the possibility that a single exposure was simply not enough of a boost to the impostor's activation to influence illusion rates. However, given that illusion rates are around 50% overall, it seems that, under a lexical access story, the stimuli put the comprehender right on the knife's edge in terms of accessing the correct word. Thus we might expect that any boost in activation at all should lead to a relatively large reduction in the illusion rate. Of course, making these intuitions precise would require a much more explicit model of the lexical access processes that go wrong. A possible follow up experiment to mitigate these concerns might use a different task in order to guarantee that the activation level for the impostor remains high throughout the sentence judgment trial. For example, we might consider a task in which, rather than completing four lexical decisions before each sentence, the comprehender memorizes a list of four words, and then after the sentence, must decide whether a probe word was in the list.

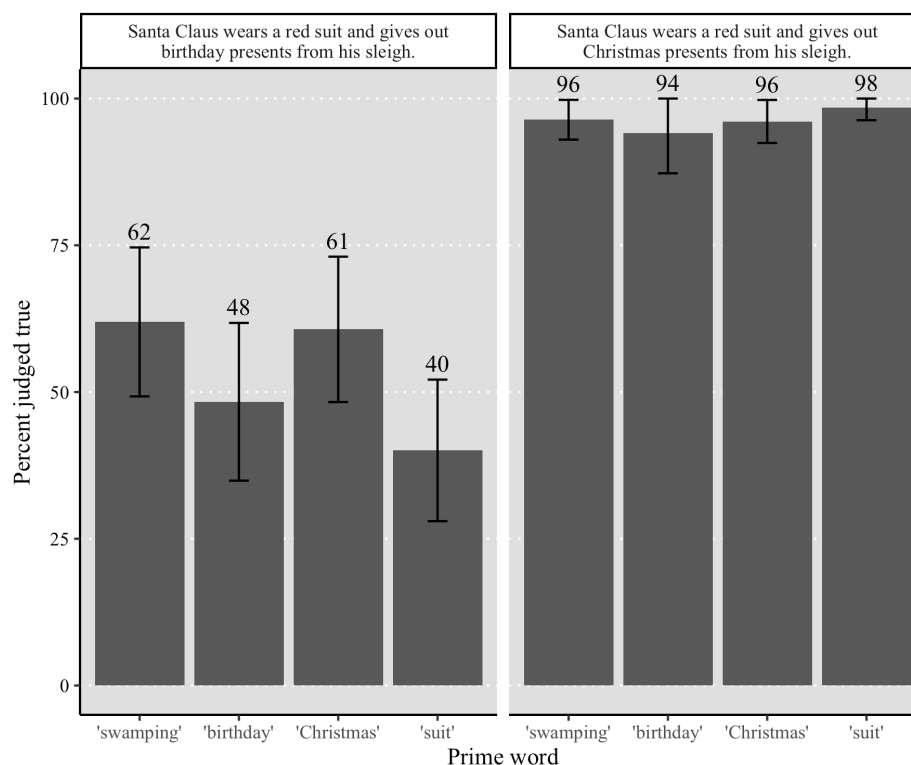


Figure 9.4: Proportion of sentence trials receiving a “true” judgment for each of the eight critical conditions for Experiment 23, filtered to only the trials in which the critical lexical decision trial was fourth out of four lexical decisions, and therefore immediately prior to the sentence judgment.

Another important dimension to explore, concerning the findings from this experiment, is the question of whether there is a boost in illusion rates for the condition where the intended word is primed (i.e. *Christmas* in the example in Figure 9.3). As we have previously noted, such an effect is predicted by models in which the lexical access error is one of mis-selection — as a consequence of boosting the activation of the intended word, the probability that it is selected and the error arises should increase. This prediction is not made by accounts in which the lexical access problem is a failure to select any lexical entry at all¹¹¹. We do not see clear evidence of such a boost, which may be taken as evidence against mis-selection accounts. But all of the concerns above apply, regarding how confident we can be that priming through a lexical decision trial causes access to all levels of representation, and whether this results in a big enough

¹¹¹The details of such an account might be elaborated in such a way that this prediction is actually made. One might imagine that the reason no entry is selected is because two candidates (the impostor and the intended) were both strongly competing, and because of lateral inhibition effects, each prevented the other from reaching a high enough level of activation to be selected. Whether this ultimately predicts that a boost in the activation for the intended word would lead to more illusions or less will depend on the details of the implementation.

boost in activation to influence illusion rates.

One final issue in the interpretation of this experiment is the question of whether the priming manipulation functions exclusively as a priming manipulation or if there are additional effects, such as directing attention to the part of the sentence that is repeated. Through manipulations of the fillers we have made it so that the conditional probability that a sentence has an anomaly, given that a word in the sentence is identical to one of the words in the immediately preceding lexical decision trials, is not different from the overall probability that a sentence has an anomaly (50%). This seems to have worked — looking at the trials with non-anomalous sentences, the existence of a repeated word does not seem to have biased participants toward rejecting the sentence.

We have also made it so that repeated words do not function well as a “clue” to the location of the anomaly when anomalies exist. One could imagine a situation where a comprehender figures out that a repeated word does not necessarily mean that there’s an impostor, but if there is an impostor, it’s going to be the repeated word.¹¹² Such inferences are discouraged by the inclusion of repeated-word trials in which the sentence is anomalous but the repeated word is not the location of the anomaly (i.e. the *suit* condition for the Santa Claus stimulus). One might still worry that comprehenders adopt a different strategy for repeated-word trials (even if the repeated word isn’t *always* the location of the anomaly, it could be a good place to start looking). We don’t see a boost in illusion rates for the repeated-word condition where the word is not the impostor (*suit* in Figure 9.3), making this somewhat less likely.

Such a strategy would also lead to more careful processing of the repeated word compared to other words in the sentence (and compared to the processing of that word if it hadn’t been repeated). Looking at the true sentences that were correctly judged true, we don’t see any trend in RTs toward slower processing for conditions where a word was repeated (the *suit* and *Christmas* conditions) compared to conditions in which nothing was repeated (the *birthday* and *swamping* conditions), which would be expected for such a strategy. But of course, total response time for the judgment of a relatively long sentence is not a very fine-

¹¹²This sounds like a somewhat convoluted inference, but it’s more reasonable when applied to an actual stimulus. One could infer that, since they saw *birthday* in a lexical decision task and now they’re seeing *Santa Claus wears a red suit and gives out birthday presents from his sleigh*, there are effectively two options: either *birthday* shouldn’t be there, or the sentence is fine. In contrast, on trials where the comprehender saw only unrelated lexical decisions, there are many more options to think through: *birthday* could be an impostor, or *sleigh* could be an impostor, or *Santa Claus* could be an impostor, etc., or the sentence could be fine.

grained measure, so we can't rule out such strategies on the basis of these data. A useful follow-up might be to use a self-paced reading task for the sentence to determine if repeated words are processed more slowly than not-repeated words (due to more careful processing) or more quickly (due to facilitation).

9.3 Conclusion

In this chapter, we present the results of two experiments that aim to narrow the hypothesis space concerning the possible cause of the substitution illusion. Experiment 22 demonstrated that illusion effects are robust for sentence judgment tasks, and not reliably less frequent compared to question-answering tasks. Experiment 23 used a priming manipulation with the lexical decision task to assess the predictions of lexical-access explanations for the illusion. Given that Experiment 23 did not reveal clear reductions in the illusion as a result of boosting the activation of the impostor, nor clear increases in the illusion as a result of boosting the activation of the intended word, we consider it somewhat less likely that the illusion arises because the lexical entry corresponding to the intended word is selected instead of the lexical entry corresponding to the (actually viewed) impostor. However, as we noted previously, there are ways for such an account to remain plausible despite these findings.

Chapter 10 Substitution illusions: general discussion

Here we review the theoretical and empirical contributions of this dissertation with respect to the study of substitution illusions. In brief, Chapter 7 provided a new way of deconstructing the problem into four possible mechanistic hypotheses, whereas previous work on substitution illusions has focussed on a more abstract level of explanation. Chapters 8 and 9 presented a series of experiments exploring variability in substitution illusion rates, first through the lens of naturally-arising variability between items, and then by attempting to deliberately modulate illusion rates in theoretically-driven ways. In section 10.1, we review these findings and in section 10.2 we return to the question of what could go wrong that would cause substitution illusions, and how the hypothesis space can be updated in light of our findings. We additionally consider the implications of the substitution illusion phenomenon for other domains of inquiry in sentence processing research in section 10.3.

10.1 Summary of findings

10.1.1 Experiment 17

Experiment 17 tested illusion rates for 49 substitution illusion stimuli that had been tested in prior research, using a question-answering task. Following prior work, we provided a clear option to identify an error on each trial, and used a multiple-choice post-test to verify that the only participants whose data we included in our measure of illusion rates were those who knew the relevant facts. We found overall illusion rates similar to previous findings (31%) as well as substantial itemwise variability that had not been previously noted. Per-item illusion rates ranged from 0% to 84% across the 49 stimuli we tested. Split-half reliability was shown to be high (.76) indicating that the itemwise variability we observed was not due to random sampling, but due to differences between items.

10.1.2 Experiment 18

Experiment 18 tested whether the itemwise variability in illusion rates observed in Experiment 17 was due to differences in the familiarity of the queried facts. A separate group of participants rated non-anomalous declarative versions of the Experiment 17 stimuli on a 1 to 7 scale indicating the extent to which they believed that “most people” are familiar with the stated trivia fact. Internal reliability was again high (.96) indicating that investigations of correlations between measures are possible. However, we found no reliable correlation between items’ familiarity ratings in Experiment 18 and their illusion rates in Experiment 17.

10.1.3 Experiment 19

Experiment 19 investigated whether gradient differences in the similarity between the meaning of the intended word and the meaning of the impostor underlied items’ differing illusion rates. Similarity between word pairs was measured as cosine similarity using word2vec, a popular word embedding tool from the NLP literature. Pre-testing based on data from Cook et al. 2018 indicated that the type of similarity measured by word2vec is related to the type of similarity that has been shown to influence illusion rates. However, we found only a weak and not statistically significant relationship between impostor-intended similarity as measured by word2vec and illusion rate as measured in Experiment 17.

10.1.4 Experiment 20

Experiment 20 measured whether the 49 items used in Experiment 17 differed in the information state at the point of the substitution. The motivating idea was that items for which the substitution occurs after the comprehender already has a clear expectation for how the sentence will unfold will be more likely to be missed because the clear expectation allows for attention to be shifted to question-answering. While tentative support for this idea was found in an analysis of position effects (impostors that occurred later in the sentence were more likely to cause illusions), analyses of entropy were not consistent with this idea. We used two language models, which assign probabilities to every word in the lexicon at the point of the

substitution, and which therefore allowed us to compute entropy at the substitution point for each of the 49 items. We found no reliable correlations between entropy and illusion rate. We additionally tested the correlation for only the subset of items that passed a version of the “knowledge check” for each of the models, and again found no correlations.

10.1.5 Experiment 21

Experiment 21 tested the same basic idea as Experiment 20 but with human data, which allowed us to target different aspects of comprehenders’ knowledge about how the sentence would unfold at the point of the substitution. In Experiment 21, participants completed a version of the cloze task for each of the 49 stimuli, in which they were presented with a fragment of the question, up to but excluding the impostor, and were asked to complete it in a way that created a trivia question, and provide the answer to the question. We computed two properties for each item based on their completions: the proportion of question completions that contained the intended word (as a measure of certainty about how the string would unfold) and the proportion of question answers that converged on the same answer (as a measure of certainty about the message). Internal reliability for each of these measures was high (.96 and .94, respectively), indicating that it is reasonable to assess their correlation with other measures. We found no statistically discernable relationship between the proportion of completions that contained the intended word and illusion rates, but we found a reliable (though weak) relationship between convergence in the answers and illusion rates.

10.1.6 Experiment 22

Experiment 22 tested illusion rates for a true/false judgment task for declarative versions of the 49 items tested in Experiment 17, as well as 71 additional items. We again used a multiple choice post-test to guarantee that participants had the relevant world knowledge. We again found robust illusions (37%) and substantial itemwise variability (split-half correlation: .77). A comparison of Experiment 17 and Experiment 22 found no evidence that illusion rates are reliably lower for declaratives than for questions.

10.1.7 Experiment 23

Experiment 23 directly tested a prediction of an account of substitution illusions in which lexical access is error-prone such that the lexical entry for the impostor is not successfully identified (and thus not a component of subsequent computations of sentence meaning), despite fixating it. The probability of successful lexical access was modulated through a priming task. Comprehenders completed four lexical decision trials prior to each sentence judgment trial. We manipulated the content of one of the lexical decision trials (the prime) across conditions, so that it was either an unrelated word, a word in the illusion sentence that is not the impostor, the intended word, or the impostor. We found no statistically reliable effects of prime condition on illusion rates, though we observed small numeric trends in the predicted direction.

10.2 Updated theoretical landscape

Recall from section 7.2 that we identified four possible loci of the error underlying the substitution illusion. These options were motivated by an exploration of the full set of operations that would need to occur in order to correctly detect the impostor. The plausible error points include (1) problems in lexical access, such that either the intended word is selected or no lexical entry is selected at all; (2) problems in the selection of concepts associated with the impostor lexical entry, such that the context biases the comprehender to selecting some sub-part that is common to the impostor and the intended word; (3) problems with semantic composition, such that the output of combining the meaning of the impostor with the meaning of the rest of the sentence is something that is not actually in any way related to the meaning of the impostor; or (4) problems in the mapping from a linguistic representation of the meaning of the sentence to a non-linguistic representation, such that information is lost. The correlations (and lack of correlations) observed in Chapter 8 do not directly bear on these hypotheses.

We also identified several explanations for the illusion that are not quite so mechanistic, but which may still be related to illusion phenomena. These include (1) subconscious accommodation, either of speech errors or of presuppositions; (2) a lack of attention being allocated to the impostor-detection

task, or, relatedly, low motivation to identify impostors; and (3) the use of “shallow” sentence processing strategies.

An explanation along these lines could be true in combination with one of the mechanistic hypotheses described above — for example, it could be that the lexical access procedure fails, and the system is designed to fail in precisely this way because it is optimized for inferring a likely intended word under circumstances where speech errors regularly occur. An attention mechanism could be combined with any of the mechanistic failures described above, if one assumes that attention is the kind of thing that makes everything run smoother. For example, under a lexical access story, increased attention could act as an amplifier on the bottom-up signal, making correct lexical identification more likely. If instead the hypothesized mechanism is about the mapping of the sentence to a non-linguistic representation, increased attention could boost the activation of all pieces of the representation, making information loss less likely.

Finally, a shallow processing explanation is compatible with all hypotheses stated here. That is, we might say that the comprehender mis-identifies the lexical item because they are using shallow processing¹¹³. Or we might say that the comprehender accesses the subset of the concepts that are associated with a word based on context because they are using shallow processing. And so on. Stating the claim in terms of shallow processing seems to imply that the comprehender could have done better if only they had used their “deep processing” strategies instead. This might reduce to the same claim as the attention/motivation explanation described above — if only the comprehender had tried harder, they could have succeeded¹¹⁴. But it’s also not the case that all hypotheses under the “shallow processing” umbrella are strongly committed to there being two distinct modes that a comprehender can switch between at will.

Setting these issues aside for the moment, it’s clear that mechanistic hypotheses can be combined

¹¹³One interesting difference between the shallow hypotheses considered here and those considered in our discussion of grammatical illusions like NPI illusions is that the question of whether the incremental comprehension system even makes use of linguistic knowledge does not seem to arise here. If the key claim in the grammatical illusion case is that even though the comprehender has detailed knowledge of the grammar, they might not always use it, then we might translate this into a claim that even though the comprehender has detailed knowledge of the lexicon, they might not always use it. While it’s clear that the path through the lexicon could be fallible, as we have discussed, it’s somewhat hard to imagine a hypothesis where it just isn’t used at all — what could be a shortcut from a visual stimulus to a meaning that does not go through the lexicon?

¹¹⁴Based on findings from Speckmann & Unkelbach 2021, this seems not to be true — even when you pay participants more based on the number of correct answers they give, illusions still happen.

with other kinds of explanation for the illusion in a useful way. However, without a mechanism, we find hypotheses in the second category unsatisfying and, in light of our findings from Chapter 8, mostly empirically inadequate. That is, if the explanation is *just* that comprehenders sometimes don't pay attention to the sentence, or *just* that comprehenders sometimes use shallow processing strategies, without a clear claim about the process that is different as a result, we might expect substitution illusions to arise randomly — illusions as a function of the comprehender's internal state on any given trial, not the content of the stimulus. The systematic itemwise variability that we find in Experiment 17 shows that this is the wrong prediction.

The more interesting question, for our purposes, is whether we can narrow the space of mechanistic hypotheses. Experiment 22 attempted to directly address a prediction of the lexical access hypothesis, and did not find clear evidence that this prediction was borne out. However, as we discussed in section 9.2.6, we are hesitant to rule out this hypothesis altogether on the basis of these findings, given the complexity of the linking assumptions. We therefore intend to follow up on these results with some modifications to the task, in order to obtain converging evidence for whether the substitution illusion is due to problems in lexical access or post-lexical processes. Of course, answering this question is only a first step toward identifying the cause of the illusion, since we have considered multiple possible hypotheses within the lexical access category, and multiple possible hypotheses outside it.

10.3 Broader implications

The mechanistic hypotheses we considered for the cause of the substitution illusion all concern processes that we believe occur as a necessary component of the comprehension of any sentence, not just substitution illusions. Thus the investigation of the illusion has significant potential consequences for sentence processing hypotheses more broadly. For example, the ideas we have considered regarding lexical access errors as an explanation for the illusion are not just hypotheses about how lexical access proceeds in illusion sentences, but about how lexical access proceeds in general. This is related to the goal of developing a theory that predicts not only that illusions will sometimes arise but also predicts the overwhelming success

of the comprehension system. However, given the currently uncertain status of each of these hypotheses, we are unable to draw strong conclusions about these mechanisms more generally.

In addition to these theoretical consequences, our investigation has some practical implications. Given our attempt to leverage the not-previously-noted itemwise variability in the illusion, one might wonder if other sentence processing phenomena exhibit the same systematic variability. In addition, we believe there is a valuable lesson to be learned from substitution illusions for how we interpret findings in which a word appears to fit well in a sentence. We discuss each of these implications in turn.

10.3.1 Item-wise variability

Although substitution illusions have been documented in the psycholinguistics literature for four decades, the observation that some of the stimuli that are used to demonstrate the phenomenon yield much larger effect sizes than others seems to be new. This variability is potentially informative, as we discussed at length in Chapter 8. It also raises the question of whether unnoticed itemwise variability arises in other phenomena. For example, we might ask if the size of the NPI illusion also varies systematically across items. There is a potential issue of statistical power, since our NPI illusion experiments typically use 30 to 40 items, and these are distributed in a Latin Square design across four to fifteen conditions, such that half to two thirds of those conditions are baselines (making the data from those trials less relevant to investigations of itemwise variability).

However, many of our NPI illusion experiments in Chapters 3, 4, and 5 use variants of the same stimuli, making it possible to merge datasets in order to obtain a greater number of illusion judgments per item. We combined the data sets from Experiment 3 (from Chapter 3) and Experiment 7 (from Chapter 4). Recall that Experiment 3 investigated the contrast in illusion rates between sentences with embedded *no* and sentences with embedded *haven't*, and Experiment 7 compared illusion rates for sentences with embedded *no*, with and without a prepositional phrase inside the relative clause. For 33 items, the embedded-*no* condition from Experiment 3 and the with-PP condition from Experiment 7 were identical strings. We therefore merged these two datasets and computed per-item acceptance rates for just the identical illusion condition. In Figure 10.1, we show the item-wise variability in acceptance rates, alongside the equivalent

substitution illusion data.

Note that after merging the Experiment 3 and Experiment 7 data, we had between 20 and 23 illusion judgments per item. This is fairly typical of the per-item dataset size from the substitution illusion experiment, Experiment 17, though there is much greater variability in Experiment 17 because of data loss due to the post-test. It's important to have similar amounts of data for the two illusion types because one could have the impression of greater itemwise variability only because one experiment has more data per item than the other, and therefore smaller confidence intervals. It's also important to compare datasets with similar numbers of items, since one could have the impression of greater itemwise variability only because one experiment tested more items than the other. We therefore excluded the 15 substitution illusion items that had the largest datasets. This means that for NPI illusions we're looking at 33 items with between 20 and 23 judgments per item, and for substitution illusions we're looking at 34 items with between 10 and 38 judgments per item¹¹⁵.

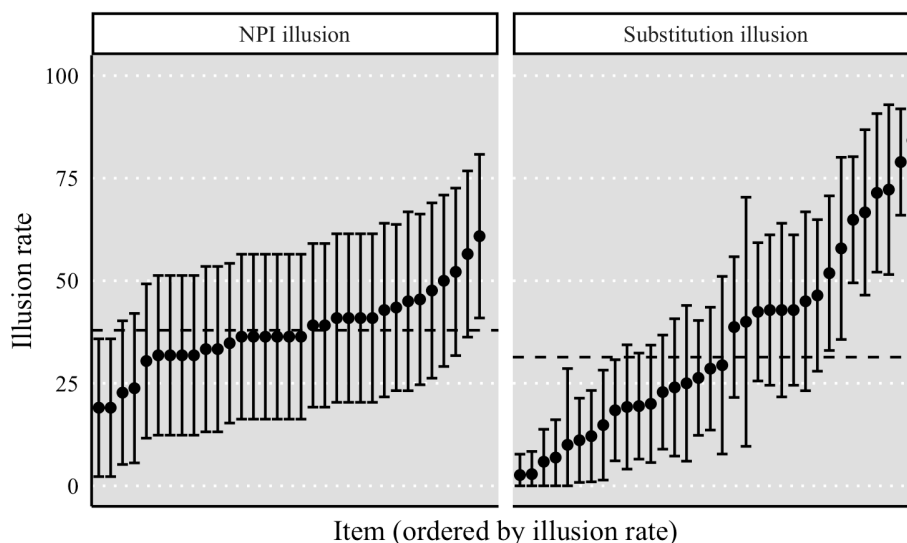


Figure 10.1: Itemwise variability in illusion rates for 33 NPI illusion stimuli (Experiment 3 and Experiment 7 data) and 34 substitution illusion stimuli (Experiment 17 data)

As the figure illustrates, NPI illusion rates do not systematically vary across the 33 items shown. Split half correlation reliability was .16, which is much lower than what we observed for substitution illusions.

¹¹⁵If we filtered to just the substitution illusion stimuli with 20 to 23 judgments, we would only have five stimuli to look at. The comparison between illusion phenomena is not perfect, but it's the best we can do.

However, some caution is warranted in interpreting these findings in light of how the stimuli for these experiments are typically constructed. First consider the substitution illusion. The phenomenon is typically described as failed detection of the world knowledge violation caused by an anomalous word which is similar to the intended word and/or context. Two items from Experiment 17 with some of the highest illusion rates and two items with some of the lowest illusion rates are shown in Table 10.1. All of these (and all 49 of the items in Experiment 17) fit the description — they are anomalous, and the impostor is similar to the intended word and/or context — but they have little else in common. In contrast, consider the NPI illusion. The phenomenon is typically described as failed detection of an unlicensed NPI when it is preceded by a non-c-commanding licenser. Again, two of the items with the highest illusion rates and two of the items with the lowest illusion rates are shown in Table 10.1.

Substitution illusions	
How did painter Vincent van Gogh lose his eye (<i>ear</i>) during his life?	2% illusions
What is the second largest star (<i>planet</i>) in our solar system, after Jupiter?	11% illusions
What is the name of the English rock band of the sixties and seventies that had Ringo Starr as their guitarist (<i>drummer</i>)?	46% illusions
Which British monarch formally opened the Olympic winter (<i>summer</i>) games in London in 2012?	72% illusions
NPI illusions	
The professors that no students have visited during office hours have ever experienced tiredness after a long lecture.	19% illusions
The lawyers that no businessmen have hired for legal advice have ever received criticism for lost fraud trials.	30% illusions
The nurses that no doctors have requested for the surgery have ever shown clumsiness in the operating room.	52% illusions
The politicians that no journalists have endorsed in the media have ever earned trust from the rural communities.	61% illusions

Table 10.1: Example stimuli for NPI illusions and substitution illusions

All of the items fit this description but they are far more similar than that — they all use the NPI *ever*, they all use the licenser *no*, they all place the embedded licenser inside an object relative clause, they all use topically-related content words throughout the sentence, they all position the NPI one word after the relative clause, they all use the same tense and aspect, etc. NPI illusion stimuli are designed to be highly similar in this way because homogeneity in the stimuli improves statistical power — we minimize

superfluous variability in acceptability judgments by making the items are very similar. In comparison, substitution illusion stimuli are much more diverse. This also makes practical sense — given that the illusion has to do with world knowledge and requires the use of trivia that is familiar to many participants, it’s difficult to construct highly homogeneous items in the way we have done for NPI illusions. The comparison is informative. If we had gone about creating NPI illusion stimuli in the same way we create substitution illusion stimuli, and thrown together anything that fits the description “an unlicensed NPI is preceded by a non-c-commanding licenser”, there would be much more variability in the items themselves, and, surely, much more variability in the illusion rate. In fact, we can be certain that items generated in this way would have more variability in the illusion rate because we know that many stimuli that fit that description — such as those in which the licenser happens to be non-quantificational or those in which the NPI happens to be later in the sentence — do not yield reliable NPI illusions.

From this perspective, the empirical progress we have made on NPI illusions and substitution illusions has a very similar flavor. We knew previously that people sometimes make errors, but the research presented here shows that our assumed generalizations about when those errors occur are too broad — there are a lot of cases where people are good at detecting unlicensed NPIs, but we hadn’t previously tested them, and there are a lot of cases where people are good at detecting impostors, but we hadn’t previously noticed that they were mixed in. Although this dissertation is about linguistic illusions, it is in fact mostly about non-illusions.

Thus, it appears that there isn’t much to be learned from item-wise variability in the NPI illusion stimuli that we have tested. However, this approach may be valuable for other phenomena. For example, Huang et al. 2022 investigated itemwise variability in the magnitude of the reading time penalty for garden path sentences, and further reported that itemwise variability in surprisal does not correlate with this variability. Measuring itemwise variability may be a useful path forward in sentence processing research, but a few words of caution are warranted. The first we have already noted several times, but it bears repeating: identifying variables that correlate with some phenomenon is not the same thing as building a theory. Even if the investigation in Chapter 8 had led to a list of variables that collectively predict illusion rates perfectly, this alone would not constitute better understanding of the sentence processing systems that

underlie the illusion. This is no different from any other empirical generalization — for example, knowing that quantificational and non-quantificational licensors do not yield the same rate of NPI illusions is valuable, but it is not a theory of the NPI illusion.

The second cautionary note concerns reliability. We have been careful to note in every case where we tested a correlation between two experimental measures that each measure had high internal reliability, measured by split-half correlations. This is critical. Without internal reliability, there is little hope of correlating a measure with anything else. Hedge, Powell, & Sumner 2018 demonstrate that this is a real risk in studies of individual differences across participants; the exact same concerns apply to differences across stimuli. Similarly Heyman et al. 2018 report that investigations of itemwise variability in semantic priming effects are unlikely to succeed because internal reliability is low. One way to think about this issue is to note that just because the mean value (of the illusion rate or any dependent measure) is different for different items does not mean that those differences reflect anything at all about the items themselves. In NPI illusions, the per-item illusion rate across 33 items ranges from 19% to 61% — while this may sound like a big enough difference that it could be correlated with something, 33 samples from the same population will always have numerically different means from one another.

10.3.2 Prediction

Substitution illusions can be described as cases in which a word is processed as if it's a better fit to its context than it truly is. That is, the impostor should cause some processing disruption (leading to conscious detection), but it does not. The impostor's non-disruptiveness may even extend to very early and implicit measures like first fixation times and the N400 ERP component. We do not generally take this finding to mean that comprehenders predict or pre-activate the impostor. This would seem like a rather odd claim for stimuli like (178) — why would a comprehender predict the word *deaf* in the context *What is the name of the raised bumps on paper that enable...?* It is of course not impossible that this happens — we might say that *blind* is predicted, and then through shared features (e.g., *sensory deficit*) some amount of activation spreads to *deaf* — but does not seem to be a logically necessary conclusion based on the finding that *deaf* is not disruptive.

(178) What is the name of the raised bumps on paper that enable deaf (*blind*) people to read?

In other contexts, however, very similar findings (a lack of disruption for an anomalous word) often lead to the conclusion that there was significant pre-activation. For example, in the role reversal literature, the lack of an increased (more negative) N400 amplitude for (179b) compared to (179a) is taken to mean that *served* is pre-activated in the context in (179b) (e.g., Chow et al. 2016, among others).

- (179) a. The restaurant owner forgot which customer the waitress had served ...
 b. The restaurant owner forgot which waitress the customer had served ...

The substitution illusion does not, in any way, tell us that this is the wrong analysis of (179b). But it does suggest that there may be a broader range of possibilities for how to think about the non-disruptiveness of an anomalous word.

Chapter 11 Conclusion

11.1 Summary of this dissertation

This dissertation has focused on clarifying the profile of two well-known linguistic illusions: the NPI illusion and the substitution illusion (sometimes called the “Moses illusion”). In **Chapter 1** we considered possible approaches, both theoretical and methodological, to studying linguistic illusions, and the potential for illusions to reveal the nature of the mechanisms through which a comprehender deploys their knowledge of language.

Chapter 2 reviewed the literature on grammatical illusions (of which NPI illusions are a type) and the grammar of NPI licensing, specifically considering possible implementations of such grammars for an online processor. One important takeaway from this discussion is that although memory-based explanations for grammatical illusions have the advantage of capturing numerous illusion phenomena under a single mechanism, they also assume uniformity in the online implementation of importantly different aspects of grammatical knowledge (e.g. subject-verb agreement and NPI licensing). **Chapter 3** presented six experiments that demonstrate that NPI illusions are substantially more restricted than was previously thought. We also investigated, for the first time, the sentence-final interpretation that comprehenders of illusion sentences construct. Chapter 3 additionally introduced the *scalar alternatives hypothesis*, which attempts to explain the NPI illusion as a consequence of an online licensing operation that is closely tied to (one version of) the grammar of NPIs. In **Chapter 4**, we saw five additional experiments further narrowing the scope of the NPI illusion. These experiments collectively suggest that illusions arise when NPIs are close to NPI-licensing environments, not when NPIs are close to negative words (“licensors”). This pattern suggests (but does not guarantee) that the online processing of licensed NPIs involves relating an NPI to the properties of its environment, rather than relating an NPI to an individual negative

lexical item in memory. **Chapter 5** presented some notable complications of the NPI illusion profile, including evidence against the scalar alternatives hypothesis and surprising (in that they were predicted by none of the hypotheses under consideration) effects of the NPI's position in the sentence and the tense of the clause containing the NPI. In **Chapter 6** we reviewed the findings from the 16 experiments presented in Chapters 3, 4, and 5, and considered each of the candidate explanations for the illusion in light of these findings. Although none of the proposals perfectly account for the data, possible strategies for refining these hypotheses were considered, and the properties of a potential successful explanation were identified.

Chapter 7 presented the basic profile of the substitution illusion and reviewed prior investigations of this effect. A critical contribution of this chapter was the clarification of the possible mechanistic explanations for the illusion, as distinct from other factors that may play a role in exacerbating illusions but are not a cause per se (e.g., the role of attention and motivation). **Chapter 8** demonstrated that illusion effects are not equally likely for all stimuli that fit the standard description of substitution illusions, and that this variability across stimuli is reliable. We further observed, across four additional experiments, that some computational and behavioral measures that also vary across our stimuli do not fully explain the variability in illusion rates. **Chapter 9** presented the results of two experiments that attempt to more directly target the predictions of the mechanistic hypotheses presented in Chapter 7. We found that illusion rates are not substantially reduced for truth value judgment tasks with declarative sentences compared to a question-answering task, and that boosting the activation level of the impostor word through a priming manipulation did not substantially reduce illusion effects. The implications of these findings were discussed in **Chapter 10**.

11.2 Conclusions

As we discussed in chapters 6 and 10, the findings we have presented here have significant consequences for the landscape of plausible explanations for the NPI illusion and substitution illusion respectively. In the interest of space, we do not reprise these arguments here but rather focus on the implications of these

findings for studies of linguistic illusions and sentence processing more broadly.

There is the question of whether the occasional failure of the incremental comprehension system to deliver a representation that is consistent with both the stimulus and the comprehender's stored knowledge (that is, the very existence of linguistic illusions) reveals the comprehender's general disregard for such knowledge in the comprehension task. We have argued that it does not. Recall our long division analogy from Chapter 1 — it is possible to give the wrong answer to a difficult division problem like $2052 \div 27$ not because you used a shortcut, but because a mistake was made along the way. It should also be noted (although this point is not necessarily strong evidence against a heuristic view of sentence processing) that humans are, in general, very good at understanding sentences. Mistakes like NPI illusions and substitution illusions are notable in part because they are rare — the work presented here suggests that they are even rarer than previously thought. It is worth marvelling at the fact that a comprehender faced with a sentence like (180) or a question like (181) is able to detect, essentially immediately, that (180) is unacceptable and (181) assumes a falsehood. Through our exploration of illusion phenomena we have attempted to show that this success is not trivial.

(180) * The authors that the critics didn't recommend have ever written a best-selling novel.

(181) In baseball, where does a player run after hitting the ball with a racket?

Turning to the specific phenomena investigated here, we do not yet have clear evidence favoring a particular mechanistic explanation for either NPI illusions or substitution illusions. However, we have made progress in sketching the landscape of possible explanations for both domains, and our findings have important theoretical consequences.

In the case of NPI illusions, it now appears that the illusion is much more specific than previously thought, suggesting that superficial parallels with other illusion phenomena will not necessarily yield insights. For example, although at a high level of abstraction, NPI illusions and agreement attraction appear to show the same pattern, that level of abstraction does not allow us to make distinctions between the cases where NPI illusions do and do not arise. We have argued that a sentence processing algorithm that hews closely to the hypothesized grammatical knowledge may be more parsimonious than an account

that treats diverse grammatical phenomena as underlyingly the same operation (i.e., a memory retrieval).

Though this has been said before, it bears repeating: the findings presented here do not suggest that memory is irrelevant to sentence processing, nor that the particular account of the memory retrieval operation that is proposed to explain illusions is the wrong account of the memory architecture. Rather, it is possible that NPI licensing simply does not involve the kind of memory operation it was thought to involve. One path that may be worth pursuing in light of this work is a deeper consideration of the extent to which the assumed implementation of grammatical knowledge is actually faithful to hypothesized grammars in other cases.

One further implication of this work is that more might be learned about NPI illusions by exploring phenomena that appear to rely on the same mechanism, rather than phenomena that appear to result in the same type of error. For example, if the NPI illusion is to be attributed to the inference and suppression of pragmatic alternatives, we might expect to find consequences of the same mechanisms in the processing of focus or contrastive implicatures, or other areas where pragmatic alternatives are thought to be in play. This doesn't mean that we should expect to find "focus illusions", whatever that would mean, but rather that given a concrete hypothesis about what is difficult about processing alternatives, we might find that the predictions are borne out in other ways. (Of course, our discussion of scalar alternatives is merely an illustrative example; we have seen some evidence that this may not be the right account of NPI illusions.)

The explanation of the substitution illusion also remains unresolved, but we are optimistic that further exploration will yield insights regarding the interaction of word-level and sentence-level meaning inference. One possibility currently on the table which we find promising is the suggestion that substitution illusion is related to the same mechanisms for focusing in on contextually-relevant components of the meaning of a word that serve a comprehender well in cases of polysemy resolution. Part of the appeal of this hypothesis is its reliance on operations that have to happen anyway. It is reasonable to assume that comprehenders cannot carry all of the concepts that are associated with every individual word with them throughout the processing of a sentence — the irrelevant parts must be pruned away sooner or later. The idea that it is this pruning that leads the comprehender to lose the *Red-Sea-guy* part of the meaning of "Moses", and carry on with only the *Bible-guy* part allows us to capture substitution illusions with the

same tools as normal sentence processing. Of course, it remains somewhat mysterious that the comprehender doesn't simply carry on with a *sports-equipment-for-hitting-a-ball* part of the meaning of "racket" in (181). More work is clearly needed to refine the hypothesis.

11.3 Next steps

While progress has been made in our understanding of both NPI illusions and substitution illusions, mysteries remain. Here we summarise some of the suggestions that have been made for future experimental investigations of these phenomena.

One critical finding from our exploration of NPI illusions is the licensor effect. However, many reasonable theories make the wrong prediction here, and we have considered some possible reanalysis-driven explanations for the lack of illusions with non-quantificational licensors. It would therefore be valuable to demonstrate the contrast in earlier and more implicit measures such as reading times. If it turns out that illusions actually do arise for non-quantificational licensors, but these illusions do not survive until sentence-final judgments, this would have significant implications for the hypotheses we consider. We have advocated for using simple measures whenever possible, both due to practical concerns concerning data collection and issues of statistical power. However, there is also value in subtler measurements and this may be a case where they are needed.

One might similarly pursue the question of illusion recovery in the post-NPI region through manipulations of the content that follows the NPI. If, for example, re-analysis is thought to involve fully swapping two noun phrases, manipulations of the plausibility of the embedded noun as a subject for the main clause predicate should modulate illusion rates. If instead only the determiners are swapped, we do not expect such effects. Related to the question of post-NPI re-analysis is the observation that NPI illusions, unlike agreement attraction, do not seem to go away with time. Rather, a comparison across experiments suggests that sentences presented with RSVP yield robust illusions even when there is no time limit on comprehenders' judgments. It may be worthwhile to demonstrate this more directly.

There have also been some deeply puzzling experimental findings which require follow-up investiga-

tions. For one thing, a replication of the surprising trend toward illusions for sentences with embedded non-quantificational negation only when the main clause is in the past perfect is required. Relatedly, our investigation of illusion sentences embedded under neg-raising verbs (which led to the surprising tense finding) yielded uninterpretable results. As a first step toward making sense of this pattern, one might measure the acceptability of these same sentences, but with the NPIs removed. One explanation we considered for the observed acceptance patterns was a penalty for sentences containing *no* where a *didn't+any* construction could have been used instead. If this is correct, we should find evidence of the penalty even when the later NPI *ever* is removed. This would not, of course, tell us much about the cause of the NPI illusion. But if it remains desirable to test the predictions of the noisy channel account using something like the stimuli tested in that experiment, a follow up such as this would guide us toward identifying the properties of more usable stimuli.

Finally, we have only begun to scratch the surface of the interpretation that NPI illusion sentences receive, using binary comprehension question responses. As we have previously discussed, some hypotheses make clear predictions about the sentence-final interpretation of illusion sentences, so it is valuable to seek a richer understanding of this. We are developing a sentence-repetition task in which we expect that comprehenders' errors in recalling illusion sentences will reveal the interpretations they assigned.

Turning our attention to substitution illusions, many questions remain. As a first step, we may wish to solidify our findings concerning the non-influence of priming (of the impostor or the intended word) on illusion rates. We have suggested a few ways this could be done. One is to guarantee that the activation of the prime remains high throughout the sentence by transitioning from a lexical decision task to a word-list memorization task — comprehenders would memorize four words, then judge the illusion sentence, then judge whether a probe word was in the word list. Another aspect of this investigation to pursue is the question of the level of representation that is accessed for the words presented in isolation. We might try to guarantee access to a meaning-related level by using a task in which comprehenders do not judge whether the string is a word, but something meaning-related like whether the word is an animal. Relatedly, lexical decision tasks show greater semantic priming effects (suggesting access to semantic levels of representation) when the proportion of related items is high, so a manipulation of the relatedness of

fillers might achieve the same goal. Then, if any of the priming experiments just discussed were to yield substantial changes in illusion rates, it would be worthwhile to follow up on the finding with a self paced reading version of the task, allowing us to measure whether processing time for the impostor is decreased (suggesting that the change in the illusion rate is in fact a facilitation effect due to priming) or increased (suggesting that the change is due to increased attention).

One hypothesis which we did not explore in great detail here was the idea that there is a functional explanation for the substitution illusion rooted in the probability of lexical substitutions in speech. Given that there is a fairly rich literature on the factors that influence the probability of speech errors, future work could test whether the same factors influence illusion rates, as a way to test the viability of this explanation.

We have also noted some interestingly unresolved issues in the basic profile of the illusion. It remains unclear whether the impostor must be similar in meaning to the intended word or related in meaning to the context, or both. One possible strategy for resolving this issue might be to generate new substitution illusion stimuli by taking existing sentence contexts and intended words and using computational measures like word embeddings or language models to generate impostors that dissociate these two factors. That is, if we can identify words that have relatively high similarity to the intended word but low probability in the context, as well as words that have relatively low similarity to the intended word but high probability in the context, we may disentangle these issues.

In conclusion, neither NPI illusions nor substitution illusions have received a fully satisfactory analysis here. However, the present experiments and discussion of the hypothesis space have advanced our understanding in substantive ways and identified potentially fruitful paths forward, and it is clear that an understanding of these illusions will have broad consequences.

Appendix A Experimental materials

A.1 Experiment 1

1	a	No authors that the critics recommended have ever received acknowledgment for a best-selling novel.
1	b	The authors that no critics recommended have ever received acknowledgment for a best-selling novel.
1	c	The authors that the critics did not recommend have ever received acknowledgment for a best-selling novel.
1	d	The authors that the critics recommended have ever received acknowledgment for a best-selling novel.
2	a	No soldiers that the diplomats supported have ever shown bravery in the controversial war.
2	b	The soldiers that no diplomats supported have ever shown bravery in the controversial war.
2	c	The soldiers that the diplomats did not support have ever shown bravery in the controversial war.
2	d	The soldiers that the diplomats supported have ever shown bravery in the controversial war.
3	a	No ambassadors that the diplomats consulted have ever seen brutality in the foreign war.
3	b	The ambassadors that no diplomats consulted have ever seen brutality in the foreign war.
3	c	The ambassadors that the diplomats did not consult have ever seen brutality in the foreign war.
3	d	The ambassadors that the diplomats consulted have ever seen brutality in the foreign war.
4	a	No professors that the students respected have ever wanted negativity in a class debate.
4	b	The professors that no students respected have ever wanted negativity in a class debate.
4	c	The professors that the students did not respect have ever wanted negativity in a class debate.
4	d	The professors that the students respected have ever wanted negativity in a class debate.
5	a	No customers that the salesmen assisted have ever expressed optimism for a full refund.
5	b	The customers that no salesmen assisted have ever expressed optimism for a full refund.
5	c	The customers that the salesmen did not assist have ever expressed optimism for a full refund.
5	d	The customers that the salesmen assisted have ever expressed optimism for a full refund.
6	a	No comments that the politicians ignored have ever caused bitterness toward the liberal newspapers.
6	b	The comments that no politicians ignored have ever caused bitterness toward the liberal newspapers.
6	c	The comments that the politicians did not ignore have ever caused bitterness toward the liberal newspapers.

6	d	The comments that the politicians ignored have ever caused bitterness toward the liberal newspapers.
7	a	No detergents that the housewives used have ever caused damage to the delicate clothing.
7	b	The detergents that no housewives used have ever caused damage to the delicate clothing.
7	c	The detergents that the housewives did not use have ever caused damage to the delicate clothing.
7	d	The detergents that the housewives used have ever caused damage to the delicate clothing.
8	a	No lawyers that the businessmen respected have ever received criticism for a bad trial.
8	b	The lawyers that no businessmen respected have ever received criticism for a bad trial.
8	c	The lawyers that the businessmen did not respect have ever received criticism for a bad trial.
8	d	The lawyers that the businessmen respected have ever received criticism for a bad trial.
9	a	No students that the teachers punished have ever expected friendliness from the strict principal.
9	b	The students that no teachers punished have ever expected friendliness from the strict principal.
9	c	The students that the teachers did not punish have ever expected friendliness from the strict principal.
9	d	The students that the teachers punished have ever expected friendliness from the strict principal.
10	a	No babysitters that the children obeyed have ever expected gratitude from the disappointed parents.
10	b	The babysitters that no children obeyed have ever expected gratitude from the disappointed parents.
10	c	The babysitters that the children did not obey have ever expected gratitude from the disappointed parents.
10	d	The babysitters that the children obeyed have ever expected gratitude from the disappointed parents.
11	a	No actors that the fans recognized have ever experienced soreness after a dangerous stunt.
11	b	The actors that no fans recognized have ever experienced soreness after a dangerous stunt.
11	c	The actors that the fans did not recognize have ever experienced soreness after a dangerous stunt.
11	d	The actors that the fans recognized have ever experienced soreness after a dangerous stunt.
12	a	No teachers that the parents recommended have ever caused problems for the new students.
12	b	The teachers that no parents recommended have ever caused problems for the new students.
12	c	The teachers that the parents did not recommend have ever caused problems for the new students.
12	d	The teachers that the parents recommended have ever caused problems for the new students.
13	a	No students that the librarians helped have ever made progress on the difficult assignment.
13	b	The students that no librarians helped have ever made progress on the difficult assignment.
13	c	The students that the librarians did not help have ever made progress on the difficult assignment.
13	d	The students that the librarians helped have ever made progress on the difficult assignment.
14	a	No nurses that the doctors appreciated have ever shown patience in the operating room.
14	b	The nurses that no doctors appreciated have ever shown patience in the operating room.
14	c	The nurses that the doctors did not appreciate have ever shown patience in the operating room.

14	d	The nurses that the doctors appreciated have ever shown patience in the operating room.
15	a	No criminals that the policemen caught have ever felt satisfaction from a petty crime.
15	b	The criminals that no policemen caught have ever felt satisfaction from a petty crime.
15	c	The criminals that the policemen did not catch have ever felt satisfaction from a petty crime.
15	d	The criminals that the policemen caught have ever felt satisfaction from a petty crime.
16	a	No employees that the managers recommended have ever wanted kindness from the rude customers.
16	b	The employees that no managers recommended have ever wanted kindness from the rude customers.
16	c	The employees that the managers did not recommend have ever wanted kindness from the rude customers.
16	d	The employees that the managers recommended have ever wanted kindness from the rude customers.
17	a	No accountants that the managers trusted have ever seen rises in the quarterly profits.
17	b	The accountants that no managers trusted have ever seen rises in the quarterly profits.
17	c	The accountants that the managers did not trust have ever seen rises in the quarterly profits.
17	d	The accountants that the managers trusted have ever seen rises in the quarterly profits.
18	a	No candidates that the voters supported have ever shown friendliness to the rude journalists.
18	b	The candidates that no voters supported have ever shown friendliness to the rude journalists.
18	c	The candidates that the voters did not support have ever shown friendliness to the rude journalists.
18	d	The candidates that the voters supported have ever shown friendliness to the rude journalists.
19	a	No surgeons that the patients trusted have ever seen appreciation from the hospital staff.
19	b	The surgeons that no patients trusted have ever seen appreciation from the hospital staff.
19	c	The surgeons that the patients did not trust have ever seen appreciation from the hospital staff.
19	d	The surgeons that the patients trusted have ever seen appreciation from the hospital staff.
20	a	No suspects that the witnesses identified have ever shown nervousness in the court room.
20	b	The suspects that no witnesses identified have ever shown nervousness in the court room.
20	c	The suspects that the witnesses did not identify have ever shown nervousness in the court room.
20	d	The suspects that the witnesses identified have ever shown nervousness in the court room.
21	a	No actresses that the housewives admired have ever caused excitement at a film festival.
21	b	The actresses that no housewives admired have ever caused excitement at a film festival.
21	c	The actresses that the housewives did not admire have ever caused excitement at a film festival.
21	d	The actresses that the housewives admired have ever caused excitement at a film festival.
22	a	No ambassadors that the government consulted have ever received hostility from the liberal media.
22	b	The ambassadors that no government consulted have ever received hostility from the liberal media.
22	c	The ambassadors that the government did not consult have ever received hostility from the liberal media.

22	d	The ambassadors that the government consulted have ever received hostility from the liberal media.
23	a	No politicians that the journalists endorsed have ever earned trust from the rural communities.
23	b	The politicians that no journalists endorsed have ever earned trust from the rural communities.
23	c	The politicians that the journalists did not endorse have ever earned trust from the rural communities.
23	d	The politicians that the journalists endorsed have ever earned trust from the rural communities.
24	a	No teenagers that the parents trusted have ever expressed kindness to a younger sibling.
24	b	The teenagers that no parents trusted have ever expressed kindness to a younger sibling.
24	c	The teenagers that the parents did not trust have ever expressed kindness to a younger sibling.
24	d	The teenagers that the parents trusted have ever expressed kindness to a younger sibling.
25	a	No survivors that the medics treated have ever felt courage during the extreme emergency.
25	b	The survivors that no medics treated have ever felt courage during the extreme emergency.
25	c	The survivors that the medics did not treat have ever felt courage during the extreme emergency.
25	d	The survivors that the medics treated have ever felt courage during the extreme emergency.
26	a	No players that the coaches drafted have ever felt nervousness before a championship game.
26	b	The players that no coaches drafted have ever felt nervousness before a championship game.
26	c	The players that the coaches did not draft have ever felt nervousness before a championship game.
26	d	The players that the coaches drafted have ever felt nervousness before a championship game.
27	a	No dictators that the citizens trusted have ever caused chaos at a public event.
27	b	The dictators that no citizens trusted have ever caused chaos at a public event.
27	c	The dictators that the citizens did not trust have ever caused chaos at a public event.
27	d	The dictators that the citizens trusted have ever caused chaos at a public event.
28	a	No professors that the students understood have ever experienced tiredness after a long lecture.
28	b	The professors that no students understood have ever experienced tiredness after a long lecture.
28	c	The professors that the students did not understand have ever experienced tiredness after a long lecture.
28	d	The professors that the students understood have ever experienced tiredness after a long lecture.
29	a	No actors that the judges nominated have ever had luck at the award ceremonies.
29	b	The actors that no judges nominated have ever had luck at the award ceremonies.
29	c	The actors that the judges did not nominate have ever had luck at the award ceremonies.
29	d	The actors that the judges nominated have ever had luck at the award ceremonies.
30	a	No actresses that the directors auditioned have ever shown elegance on a large stage.
30	b	The actresses that no directors auditioned have ever shown elegance on a large stage.
30	c	The actresses that the directors did not audition have ever shown elegance on a large stage.

30	d	The actresses that the directors auditioned have ever shown elegance on a large stage.
31	a	No champions that the competitors defeated have ever shown humility after a big game.
31	b	The champions that no competitors defeated have ever shown humility after a big game.
31	c	The champions that the competitors did not defeat have ever shown humility after a big game.
31	d	The champions that the competitors defeated have ever shown humility after a big game.
32	a	No paintings that the collectors liked have ever depicted tranquility with very bright colors.
32	b	The paintings that no collectors liked have ever depicted tranquility with very bright colors.
32	c	The paintings that the collectors did not like have ever depicted tranquility with very bright colors.
32	d	The paintings that the collectors liked have ever depicted tranquility with very bright colors.
33	a	No editors that the journalists respected have ever had patience for a missed deadline.
33	b	The editors that no journalists respected have ever had patience for a missed deadline.
33	c	The editors that the journalists did not respect have ever had patience for a missed deadline.
33	d	The editors that the journalists respected have ever had patience for a missed deadline.
34	a	No teenagers that the teachers motivated have ever experienced loneliness in the large class.
34	b	The teenagers that no teachers motivated have ever experienced loneliness in the large class.
34	c	The teenagers that the teachers did not motivate have ever experienced loneliness in the large class.
34	d	The teenagers that the teachers motivated have ever experienced loneliness in the large class.
35	a	No students that the professors tutored have ever had trouble in a math class.
35	b	The students that no professors tutored have ever had trouble in a math class.
35	c	The students that the professors did not tutor have ever had trouble in a math class.
35	d	The students that the professors tutored have ever had trouble in a math class.
36	a	No movies that the children watched have ever depicted gore during a fight scene.
36	b	The movies that no children watched have ever depicted gore during a fight scene.
36	c	The movies that the children did not watch have ever depicted gore during a fight scene.
36	d	The movies that the children watched have ever depicted gore during a fight scene.

Table A.1: Full experimental stimuli for Experiment 1

A.2 Experiment 2

1	a	No authors that the critics recommended have ever received acknowledgment for a best-selling novel.
1	b	The authors that no critics recommended have ever received acknowledgment for a best-selling novel.
1	c	The authors that the critics did not recommend have ever received acknowledgment for a best-selling novel.

1	d	The authors that the critics recommended have ever received acknowledgment for a best-selling novel.
2	a	No soldiers that the diplomats supported have ever shown bravery in the controversial war.
2	b	The soldiers that no diplomats supported have ever shown bravery in the controversial war.
2	c	The soldiers that the diplomats did not support have ever shown bravery in the controversial war.
2	d	The soldiers that the diplomats supported have ever shown bravery in the controversial war.
3	a	No ambassadors that the diplomats consulted have ever seen brutality in the foreign war.
3	b	The ambassadors that no diplomats consulted have ever seen brutality in the foreign war.
3	c	The ambassadors that the diplomats did not consult have ever seen brutality in the foreign war.
3	d	The ambassadors that the diplomats consulted have ever seen brutality in the foreign war.
4	a	No professors that the students respected have ever wanted negativity in a class debate.
4	b	The professors that no students respected have ever wanted negativity in a class debate.
4	c	The professors that the students did not respect have ever wanted negativity in a class debate.
4	d	The professors that the students respected have ever wanted negativity in a class debate.
5	a	No customers that the salesmen assisted have ever expressed optimism for a full refund.
5	b	The customers that no salesmen assisted have ever expressed optimism for a full refund.
5	c	The customers that the salesmen did not assist have ever expressed optimism for a full refund.
5	d	The customers that the salesmen assisted have ever expressed optimism for a full refund.
6	a	No comments that the politicians ignored have ever caused bitterness toward the liberal newspapers.
6	b	The comments that no politicians ignored have ever caused bitterness toward the liberal newspapers.
6	c	The comments that the politicians did not ignore have ever caused bitterness toward the liberal newspapers.
6	d	The comments that the politicians ignored have ever caused bitterness toward the liberal newspapers.
7	a	No detergents that the housewives used have ever caused damage to the delicate clothing.
7	b	The detergents that no housewives used have ever caused damage to the delicate clothing.
7	c	The detergents that the housewives did not use have ever caused damage to the delicate clothing.
7	d	The detergents that the housewives used have ever caused damage to the delicate clothing.
8	a	No lawyers that the businessmen respected have ever received criticism for a bad trial.
8	b	The lawyers that no businessmen respected have ever received criticism for a bad trial.
8	c	The lawyers that the businessmen did not respect have ever received criticism for a bad trial.
8	d	The lawyers that the businessmen respected have ever received criticism for a bad trial.
9	a	No students that the teachers punished have ever expected friendliness from the strict principal.
9	b	The students that no teachers punished have ever expected friendliness from the strict principal.
9	c	The students that the teachers did not punish have ever expected friendliness from the strict principal.

9	d	The students that the teachers punished have ever expected friendliness from the strict principal.
10	a	No babysitters that the children obeyed have ever expected gratitude from the disappointed parents.
10	b	The babysitters that no children obeyed have ever expected gratitude from the disappointed parents.
10	c	The babysitters that the children did not obey have ever expected gratitude from the disappointed parents.
10	d	The babysitters that the children obeyed have ever expected gratitude from the disappointed parents.
11	a	No actors that the fans recognized have ever experienced soreness after a dangerous stunt.
11	b	The actors that no fans recognized have ever experienced soreness after a dangerous stunt.
11	c	The actors that the fans did not recognize have ever experienced soreness after a dangerous stunt.
11	d	The actors that the fans recognized have ever experienced soreness after a dangerous stunt.
12	a	No teachers that the parents recommended have ever caused problems for the new students.
12	b	The teachers that no parents recommended have ever caused problems for the new students.
12	c	The teachers that the parents did not recommend have ever caused problems for the new students.
12	d	The teachers that the parents recommended have ever caused problems for the new students.
13	a	No students that the librarians helped have ever made progress on the difficult assignment.
13	b	The students that no librarians helped have ever made progress on the difficult assignment.
13	c	The students that the librarians did not help have ever made progress on the difficult assignment.
13	d	The students that the librarians helped have ever made progress on the difficult assignment.
14	a	No nurses that the doctors appreciated have ever shown patience in the operating room.
14	b	The nurses that no doctors appreciated have ever shown patience in the operating room.
14	c	The nurses that the doctors did not appreciate have ever shown patience in the operating room.
14	d	The nurses that the doctors appreciated have ever shown patience in the operating room.
15	a	No criminals that the policemen caught have ever felt satisfaction from a petty crime.
15	b	The criminals that no policemen caught have ever felt satisfaction from a petty crime.
15	c	The criminals that the policemen did not catch have ever felt satisfaction from a petty crime.
15	d	The criminals that the policemen caught have ever felt satisfaction from a petty crime.
16	a	No employees that the managers recommended have ever wanted kindness from the rude customers.
16	b	The employees that no managers recommended have ever wanted kindness from the rude customers.
16	c	The employees that the managers did not recommend have ever wanted kindness from the rude customers.
16	d	The employees that the managers recommended have ever wanted kindness from the rude customers.
17	a	No accountants that the managers trusted have ever seen rises in the quarterly profits.
17	b	The accountants that no managers trusted have ever seen rises in the quarterly profits.
17	c	The accountants that the managers did not trust have ever seen rises in the quarterly profits.

17	d	The accountants that the managers trusted have ever seen rises in the quarterly profits.
18	a	No candidates that the voters supported have ever shown friendliness to the rude journalists.
18	b	The candidates that no voters supported have ever shown friendliness to the rude journalists.
18	c	The candidates that the voters did not support have ever shown friendliness to the rude journalists.
18	d	The candidates that the voters supported have ever shown friendliness to the rude journalists.
19	a	No surgeons that the patients trusted have ever seen appreciation from the hospital staff.
19	b	The surgeons that no patients trusted have ever seen appreciation from the hospital staff.
19	c	The surgeons that the patients did not trust have ever seen appreciation from the hospital staff.
19	d	The surgeons that the patients trusted have ever seen appreciation from the hospital staff.
20	a	No suspects that the witnesses identified have ever shown nervousness in the court room.
20	b	The suspects that no witnesses identified have ever shown nervousness in the court room.
20	c	The suspects that the witnesses did not identify have ever shown nervousness in the court room.
20	d	The suspects that the witnesses identified have ever shown nervousness in the court room.
21	a	No actresses that the housewives admired have ever caused excitement at a film festival.
21	b	The actresses that no housewives admired have ever caused excitement at a film festival.
21	c	The actresses that the housewives did not admire have ever caused excitement at a film festival.
21	d	The actresses that the housewives admired have ever caused excitement at a film festival.
22	a	No ambassadors that the government consulted have ever received hostility from the liberal media.
22	b	The ambassadors that no government consulted have ever received hostility from the liberal media.
22	c	The ambassadors that the government did not consult have ever received hostility from the liberal media.
22	d	The ambassadors that the government consulted have ever received hostility from the liberal media.
23	a	No politicians that the journalists endorsed have ever earned trust from the rural communities.
23	b	The politicians that no journalists endorsed have ever earned trust from the rural communities.
23	c	The politicians that the journalists did not endorse have ever earned trust from the rural communities.
23	d	The politicians that the journalists endorsed have ever earned trust from the rural communities.
24	a	No teenagers that the parents trusted have ever expressed kindness to a younger sibling.
24	b	The teenagers that no parents trusted have ever expressed kindness to a younger sibling.
24	c	The teenagers that the parents did not trust have ever expressed kindness to a younger sibling.
24	d	The teenagers that the parents trusted have ever expressed kindness to a younger sibling.
25	a	No survivors that the medics treated have ever felt courage during the extreme emergency.
25	b	The survivors that no medics treated have ever felt courage during the extreme emergency.
25	c	The survivors that the medics did not treat have ever felt courage during the extreme emergency.

25	d	The survivors that the medics treated have ever felt courage during the extreme emergency.
26	a	No players that the coaches drafted have ever felt nervousness before a championship game.
26	b	The players that no coaches drafted have ever felt nervousness before a championship game.
26	c	The players that the coaches did not draft have ever felt nervousness before a championship game.
26	d	The players that the coaches drafted have ever felt nervousness before a championship game.
27	a	No dictators that the citizens trusted have ever caused chaos at a public event.
27	b	The dictators that no citizens trusted have ever caused chaos at a public event.
27	c	The dictators that the citizens did not trust have ever caused chaos at a public event.
27	d	The dictators that the citizens trusted have ever caused chaos at a public event.
28	a	No professors that the students understood have ever experienced tiredness after a long lecture.
28	b	The professors that no students understood have ever experienced tiredness after a long lecture.
28	c	The professors that the students did not understand have ever experienced tiredness after a long lecture.
28	d	The professors that the students understood have ever experienced tiredness after a long lecture.
29	a	No actors that the judges nominated have ever had luck at the award ceremonies.
29	b	The actors that no judges nominated have ever had luck at the award ceremonies.
29	c	The actors that the judges did not nominate have ever had luck at the award ceremonies.
29	d	The actors that the judges nominated have ever had luck at the award ceremonies.
30	a	No actresses that the directors auditioned have ever shown elegance on a large stage.
30	b	The actresses that no directors auditioned have ever shown elegance on a large stage.
30	c	The actresses that the directors did not audition have ever shown elegance on a large stage.
30	d	The actresses that the directors auditioned have ever shown elegance on a large stage.
31	a	No champions that the competitors defeated have ever shown humility after a big game.
31	b	The champions that no competitors defeated have ever shown humility after a big game.
31	c	The champions that the competitors did not defeat have ever shown humility after a big game.
31	d	The champions that the competitors defeated have ever shown humility after a big game.
32	a	No paintings that the collectors liked have ever depicted tranquility with very bright colors.
32	b	The paintings that no collectors liked have ever depicted tranquility with very bright colors.
32	c	The paintings that the collectors did not like have ever depicted tranquility with very bright colors.
32	d	The paintings that the collectors liked have ever depicted tranquility with very bright colors.
33	a	No editors that the journalists respected have ever had patience for a missed deadline.
33	b	The editors that no journalists respected have ever had patience for a missed deadline.
33	c	The editors that the journalists did not respect have ever had patience for a missed deadline.

33	d	The editors that the journalists respected have ever had patience for a missed deadline.
34	a	No teenagers that the teachers motivated have ever experienced loneliness in the large class.
34	b	The teenagers that no teachers motivated have ever experienced loneliness in the large class.
34	c	The teenagers that the teachers did not motivate have ever experienced loneliness in the large class.
34	d	The teenagers that the teachers motivated have ever experienced loneliness in the large class.
35	a	No students that the professors tutored have ever had trouble in a math class.
35	b	The students that no professors tutored have ever had trouble in a math class.
35	c	The students that the professors did not tutor have ever had trouble in a math class.
35	d	The students that the professors tutored have ever had trouble in a math class.
36	a	No movies that the children watched have ever depicted gore during a fight scene.
36	b	The movies that no children watched have ever depicted gore during a fight scene.
36	c	The movies that the children did not watch have ever depicted gore during a fight scene.
36	d	The movies that the children watched have ever depicted gore during a fight scene.

Table A.2: Full experimental stimuli for Experiment 2

A.3 Experiment 3

1	a	No authors that the critics have recommended in their reviews have ever received acknowledgment for a best-selling novel.
1	b	The authors that no critics have recommended in their reviews have ever received acknowledgment for a best-selling novel.
1	c	The authors that the critics haven't recommended in their reviews have ever received acknowledgment for a best-selling novel.
1	d	The authors that the critics have recommended in their reviews have ever received acknowledgment for a best-selling novel.
2	a	No soldiers that the diplomats have supported in the trial have ever shown respect to the war victims.
2	b	The soldiers that no diplomats have supported in the trial have ever shown respect to the war victims.
2	c	The soldiers that the diplomats haven't supported in the trial have ever shown respect to the war victims.
2	d	The soldiers that the diplomats have supported in the trial have ever shown respect to the war victims.
3	a	No ambassadors that the diplomats have consulted about the treaty have ever seen brutality in the foreign war.

3	b	The ambassadors that no diplomats have consulted about the treaty have ever seen brutality in the foreign war.
3	c	The ambassadors that the diplomats haven't consulted about the treaty have ever seen brutality in the foreign war.
3	d	The ambassadors that the diplomats have consulted about the treaty have ever seen brutality in the foreign war.
4	a	No professors that the students have challenged over low grades have ever wanted negativity in a class debate.
4	b	The professors that no students have challenged over low grades have ever wanted negativity in a class debate.
4	c	The professors that the students haven't challenged over low grades have ever wanted negativity in a class debate.
4	d	The professors that the students have challenged over low grades have ever wanted negativity in a class debate.
5	a	No customers that the salesmen have assisted in the outlet have ever expressed optimism for a full refund.
5	b	The customers that no salesmen have assisted in the outlet have ever expressed optimism for a full refund.
5	c	The customers that the salesmen haven't assisted in the outlet have ever expressed optimism for a full refund.
5	d	The customers that the salesmen have assisted in the outlet have ever expressed optimism for a full refund.
6	a	No diplomats that the politicians have informed of recent policies have ever caused controversy in the liberal newspapers.
6	b	The diplomats that no politicians have informed of recent policies have ever caused controversy in the liberal newspapers.
6	c	The diplomats that the politicians haven't informed of recent policies have ever caused controversy in the liberal newspapers.
6	d	The diplomats that the politicians have informed of recent policies have ever caused controversy in the liberal newspapers.
7	a	No maids that the housewives have thanked for their work have ever caused damage to the delicate clothing.
7	b	The maids that no housewives have thanked for their work have ever caused damage to the delicate clothing.
7	c	The maids that the housewives haven't thanked for their work have ever caused damage to the delicate clothing.
7	d	The maids that the housewives have thanked for their work have ever caused damage to the delicate clothing.

8	a	No lawyers that the businessmen have hired for legal advice have ever received criticism for lost fraud trials.
8	b	The lawyers that no businessmen have hired for legal advice have ever received criticism for lost fraud trials.
8	c	The lawyers that the businessmen haven't hired for legal advice have ever received criticism for lost fraud trials.
8	d	The lawyers that the businessmen have hired for legal advice have ever received criticism for lost fraud trials.
9	a	No students that the teachers have punished for bad behavior have ever expected friendliness from the strict principal.
9	b	The students that no teachers have punished for bad behavior have ever expected friendliness from the strict principal.
9	c	The students that the teachers haven't punished for bad behavior have ever expected friendliness from the strict principal.
9	d	The students that the teachers have punished for bad behavior have ever expected friendliness from the strict principal.
10	a	No babysitters that the children have disobeyed during an outing have ever expected gratitude from the disappointed parents.
10	b	The babysitters that no children have disobeyed during an outing have ever expected gratitude from the disappointed parents.
10	c	The babysitters that the children haven't disobeyed during an outing have ever expected gratitude from the disappointed parents.
10	d	The babysitters that the children have disobeyed during an outing have ever expected gratitude from the disappointed parents.
11	a	No actors that the producers have solicited for their films have ever experienced injuries from a dangerous stunt.
11	b	The actors that no producers have solicited for their films have ever experienced injuries from a dangerous stunt.
11	c	The actors that the producers haven't solicited for their films have ever experienced injury from a dangerous stunt.
11	d	The actors that the producers have solicited for their films have ever experienced injuries from a dangerous stunt.
12	a	No teachers that the parents have commended in PTA meetings have ever caused problems with the school administrators.

12	b	The teachers that no parents have commended in PTA meetings have ever caused problems with the school administrators.
12	c	The teachers that the parents haven't commended in PTA meetings have ever caused problems with the school administrators.
12	d	The teachers that the parents have commended in PTA meetings have ever caused problems with the school administrators.
13	a	No students that the librarians have helped with book reports have ever made progress on the difficult assignment.
13	b	The students that no librarians have helped with book reports have ever made progress on the difficult assignment.
13	c	The students that the librarians haven't helped with book reports have ever made progress on the difficult assignment.
13	d	The students that the librarians have helped with book reports have ever made progress on the difficult assignment.
14	a	No nurses that the doctors have requested for the surgery have ever shown clumsiness in the operating room.
14	b	The nurses that no doctors have requested for the surgery have ever shown clumsiness in the operating room.
14	c	The nurses that the doctors haven't requested for the surgery have ever shown clumsiness in the operating room.
14	d	The nurses that the doctors have requested for the surgery have ever shown clumsiness in the operating room.
15	a	No criminals that the policemen have caught in drug raids have ever felt satisfaction from a petty crime.
15	b	The criminals that no policemen have caught in drug raids have ever felt satisfaction from a petty crime.
15	c	The criminals that the policemen haven't caught in drug raids have ever felt satisfaction from a petty crime.
15	d	The criminals that the policemen have caught in drug raids have ever felt satisfaction from a petty crime.
16	a	No employees that the managers have recommended for a raise have ever expressed frustration with the rude customers.
16	b	The employees that no managers have recommended for a raise have ever expressed frustration with the rude customers.
16	c	The employees that the managers haven't recommended for a raise have ever expressed frustration with the rude customers.

16	d	The employees that the managers have recommended for a raise have ever expressed frustration with the rude customers.
17	a	No accountants that the managers have blamed for company losses have ever seen rises in the quarterly profits.
17	b	The accountants that no managers have blamed for company losses have ever seen rises in the quarterly profits.
17	c	The accountants that the managers haven't blamed for company losses have ever seen rises in the quarterly profits.
17	d	The accountants that the managers have blamed for company losses have ever seen rises in the quarterly profits.
18	a	No candidates that the voters have supported during the election have ever shown friendliness to the rude journalists.
18	b	The candidates that no voters have supported during the election have ever shown friendliness to the rude journalists.
18	c	The candidates that the voters haven't supported during the election have ever shown friendliness to the rude journalists.
18	d	The candidates that the voters have supported during the election have ever shown friendliness to the rude journalists.
19	a	No surgeons that the patients have consulted about the operation have ever expressed dissatisfaction with the hospital staff.
19	b	The surgeons that no patients have consulted about the operation have ever expressed dissatisfaction with the hospital staff.
19	c	The surgeons that the patients haven't consulted about the operation have ever expressed dissatisfaction with the hospital staff.
19	d	The surgeons that the patients have consulted about the operation have ever expressed dissatisfaction with the hospital staff.
20	a	No suspects that the witnesses have identified in photo line-ups have ever shown nervousness in the court room.
20	b	The suspects that no witnesses have identified in photo line-ups have ever shown nervousness in the court room.
20	c	The suspects that the witnesses haven't identified in photo line-ups have ever shown nervousness in the court room.

20	d	The suspects that the witnesses have identified in photo line-ups have ever shown nervousness in the court room.
21	a	No actresses that the moviegoers have praised for their performance have ever caused excitement at a film festival.
21	b	The actresses that no moviegoers have praised for their performance have ever caused excitement at a film festival.
21	c	The actresses that the moviegoers haven't praised for their performance have ever caused excitement at a film festival.
21	d	The actresses that the moviegoers have praised for their performance have ever caused excitement at a film festival.
22	a	No senators that the billionaires have supported with campaign donations have ever received hostility from online news media.
22	b	The senators that no billionaires have supported with campaign donations have ever received hostility from online news media.
22	c	The senators that the billionaires haven't supported with campaign donations have ever received hostility from online news media.
22	d	The senators that the billionaires have supported with campaign donations have ever received hostility from online news media.
23	a	No politicians that the journalists have endorsed in the media have ever earned trust from the rural communities.
23	b	The politicians that no journalists have endorsed in the media have ever earned trust from the rural communities.
23	c	The politicians that the journalists haven't endorsed in the media have ever earned trust from the rural communities.
23	d	The politicians that the journalists haven endorsed in the media have ever earned trust from the rural communities.
24	a	No teenagers that the parents have trusted with a car have ever expressed impatience with their rambunctious siblings.
24	b	The teenagers that no parents have trusted with a car have ever expressed impatience with their rambunctious siblings.
24	c	The teenagers that the parents haven't trusted with a car have ever expressed impatience with their rambunctious siblings.

24	d	The teenagers that the parents have trusted with a car have ever expressed impatience with their rambunctious siblings.
25	a	No survivors that the medics have cured of their injuries have ever felt regret for their military service
25	b	The survivors that no medics have cured of their injuries have ever felt regret for their military service
25	c	The survivors that the medics haven't cured of their injuries have ever felt regret for their military service
25	d	The survivors that the medics have cured of their injuries have ever felt regret for their military service
26	a	No players that the coaches have drafted for the team have ever felt nervousness before a championship game.
26	b	The players that no coaches have drafted for the team have ever felt nervousness before a championship game.
26	c	The players that the coaches haven't drafted for the team have ever felt nervousness before a championship game.
26	d	The players that the coaches have drafted for the team have ever felt nervousness before a championship game.
27	a	No voters that the senators have courted at campaign rallies have ever caused controversy in a major election
27	b	The voters that no senators have courted at campaign rallies have ever caused controversy in a major election
27	c	The voters that the senators haven't courted at campaign rallies have ever caused controversy in a major election.
27	d	The voters that the senators have courted at campaign rallies have ever caused controversy in a major election
28	a	No professors that the students have visited during office hours have ever experienced tiredness after a long lecture.
28	b	The professors that no students have visited during office hours have ever experienced tiredness after a long lecture.
28	c	The professors that the students haven't visited during office hours have ever experienced tiredness after a long lecture.
28	d	The professors that the students have visited during office hours have ever experienced tiredness after a long lecture.
29	a	No actors that the judges have nominated for an award have ever experienced derision from the tabloid gossip.
29	b	The actors that no judges have nominated for an award have ever experienced derision from the tabloid gossip.

29	c	The actors that the judges haven't nominated for an award have ever experienced derision from the tabloid gossip.
29	d	The actors that the judges have nominated for an award have ever experienced derision from the tabloid gossip.
30	a	No actresses that the directors have auditioned for the role have ever shown nervousness on a large stage.
30	b	The actresses that no directors have auditioned for the role have ever shown nervousness on a large stage.
30	c	The actresses that the directors haven't auditioned for the role have ever shown nervousness on a large stage.
30	d	The actresses that the directors have auditioned for the role have ever shown nervousness on a large stage.
31	a	No champions that the competitors have defeated in important races have ever shown humility after a big win.
31	b	The champions that no competitors have defeated in important races have ever shown humility after a big win.
31	c	The champions that the competitors haven't defeated in important races have ever shown humility after a big win.
31	d	The champions that the competitors have defeated in important races have ever shown humility after a big win.
32	a	No painters that the collectors have favored at prestigious exhibitions have ever depicted tranquility with very bright colors.
32	b	The painters that no collectors have favored at prestigious exhibitions have ever depicted tranquility with very bright colors.
32	c	The painters that the collectors haven't favored at prestigious exhibitions have ever depicted tranquility with very bright colors.
32	d	The painters that the collectors have favored at prestigious exhibitions have ever depicted tranquility with very bright colors.
33	a	No editors that the journalists have insulted in stressful meetings have ever had patience for a missed deadline.
33	b	The editors that no journalists have insulted in stressful meetings have ever had patience for a missed deadline.
33	c	The editors that the journalists haven't insulted in stressful meetings have ever had patience for a missed deadline.
33	d	The editors that the journalists have insulted in stressful meetings have ever had patience for a missed deadline.

34	a	No teenagers that the teachers have scolded for their chattiness have ever experienced loneliness in the large class.
34	b	The teenagers that no teachers have scolded for their chattiness have ever experienced loneliness in the large class.
34	c	The teenagers that the teachers haven't scolded for their chattiness have ever experienced loneliness in the large class.
34	d	The teenagers that the teachers have scolded for their chattiness have ever experienced loneliness in the large class.
35	a	No students that the professors have tutored on the weekends have ever had trouble in a math class.
35	b	The students that no professors have tutored on the weekends have ever had trouble in a math class.
35	c	The students that the professors haven't tutored on the weekends have ever had trouble in a math class.
35	d	The students that the professors have tutored on the weekends have ever had trouble in a math class.
36	a	No actors that the children have seen in family films have ever shown gore during a fight scene.
36	b	The actors that no children have seen in family films have ever shown gore during a fight scene.
36	c	The actors that the children haven't seen in family films have ever shown gore during a fight scene.
36	d	The actors that the children have seen in family films have ever shown gore during a fight scene.

Table A.3: Full experimental stimuli for Experiment 3

A.4 Experiment 4

1	a	No authors that the critics recommended have ever received acknowledgment for a best-selling novel.	Have the authors received acknowledgment for a novel?
1	b	The authors that no critics recommended have ever received acknowledgment for a best-selling novel.	Have the authors received acknowledgment for a novel?
1	c	The authors that the critics didn't recommend have ever received acknowledgment for a best-selling novel.	Have the authors received acknowledgment for a novel?
1	d	The authors that the critics recommended have ever received acknowledgment for a best-selling novel.	Have the authors received acknowledgment for a novel?
1	w	No authors that the critics recommended have received acknowledgment for a best-selling novel.	Have the authors received acknowledgment for a novel?

1	x	The authors that no critics recommended have received acknowledgment for a best-selling novel.	Have the authors received acknowledgment for a novel?
1	y	The authors that the critics didn't recommend have received acknowledgment for a best-selling novel.	Have the authors received acknowledgment for a novel?
1	z	The authors that the critics recommended have received acknowledgment for a best-selling novel.	Have the authors received acknowledgment for a novel?
2	a	No soldiers that the diplomats supported have ever shown bravery in the controversial war.	Have the soldiers shown bravery in the controversial war?
2	b	The soldiers that no diplomats supported have ever shown bravery in the controversial war.	Have the soldiers shown bravery in the controversial war?
2	c	The soldiers that the diplomats didn't support have ever shown bravery in the controversial war.	Have the soldiers shown bravery in the controversial war?
2	d	The soldiers that the diplomats supported have ever shown bravery in the controversial war.	Have the soldiers shown bravery in the controversial war?
2	w	No soldiers that the diplomats supported have shown bravery in the controversial war.	Have the soldiers shown bravery in the controversial war?
2	x	The soldiers that no diplomats supported have shown bravery in the controversial war.	Have the soldiers shown bravery in the controversial war?
2	y	The soldiers that the diplomats didn't support have shown bravery in the controversial war.	Have the soldiers shown bravery in the controversial war?
2	z	The soldiers that the diplomats supported have shown bravery in the controversial war.	Have the soldiers shown bravery in the controversial war?
3	a	No ambassadors that the diplomats consulted have ever seen brutality in the foreign war.	Have the ambassadors seen brutality in the war?
3	b	The ambassadors that no diplomats consulted have ever seen brutality in the foreign war.	Have the ambassadors seen brutality in the war?
3	c	The ambassadors that the diplomats didn't consult have ever seen brutality in the foreign war.	Have the ambassadors seen brutality in the war?
3	d	The ambassadors that the diplomats consulted have ever seen brutality in the foreign war.	Have the ambassadors seen brutality in the war?
3	w	No ambassadors that the diplomats consulted have seen brutality in the foreign war.	Have the ambassadors seen brutality in the war?

3	x	The ambassadors that no diplomats consulted have seen brutality in the foreign war.	Have the ambassadors seen brutality in the war?
3	y	The ambassadors that the diplomats didn't consult have seen brutality in the foreign war.	Have the ambassadors seen brutality in the war?
3	z	The ambassadors that the diplomats consulted have seen brutality in the foreign war.	Have the ambassadors seen brutality in the war?
4	a	No professors that the students respected have ever wanted negativity in a class debate.	Have the professors wanted negativity in the debate?
4	b	The professors that no students respected have ever wanted negativity in a class debate.	Have the professors wanted negativity in the debate?
4	c	The professors that the students didn't respect have ever wanted negativity in a class debate.	Have the professors wanted negativity in the debate?
4	d	The professors that the students respected have ever wanted negativity in a class debate.	Have the professors wanted negativity in the debate?
4	w	No professors that the students respected have wanted negativity in a class debate.	Have the professors wanted negativity in the debate?
4	x	The professors that no students respected have wanted negativity in a class debate.	Have the professors wanted negativity in the debate?
4	y	The professors that the students didn't respect have wanted negativity in a class debate.	Have the professors wanted negativity in the debate?
4	z	The professors that the students respected have wanted negativity in a class debate.	Have the professors wanted negativity in the debate?
5	a	No customers that the salesmen assisted have ever expressed optimism for a full refund.	Were the customers optimistic about receiving a refund?
5	b	The customers that no salesmen assisted have ever expressed optimism for a full refund.	Were the customers optimistic about receiving a refund?
5	c	The customers that the salesmen didn't assist have ever expressed optimism for a full refund.	Were the customers optimistic about receiving a refund?
5	d	The customers that the salesmen assisted have ever expressed optimism for a full refund.	Were the customers optimistic about receiving a refund?
5	w	No customers that the salesmen assisted have expressed optimism for a full refund.	Were the customers optimistic about receiving a refund?

5	x	The customers that no salesmen assisted have expressed optimism for a full refund.	Were the customers optimistic about receiving a refund?
5	y	The customers that the salesmen didn't assist have expressed optimism for a full refund.	Were the customers optimistic about receiving a refund?
5	z	The customers that the salesmen assisted have expressed optimism for a full refund.	Were the customers optimistic about receiving a refund?
6	a	No comments that the politicians ignored have ever caused bitterness toward the liberal newspapers.	Have the comments caused bitterness?
6	b	The comments that no politicians ignored have ever caused bitterness toward the liberal newspapers.	Have the comments caused bitterness?
6	c	The comments that the politicians didn't ignore have ever caused bitterness toward the liberal newspapers.	Have the comments caused bitterness?
6	d	The comments that the politicians ignored have ever caused bitterness toward the liberal newspapers.	Have the comments caused bitterness?
6	w	No comments that the politicians ignored have caused bitterness toward the liberal newspapers.	Have the comments caused bitterness?
6	x	The comments that no politicians ignored have caused bitterness toward the liberal newspapers.	Have the comments caused bitterness?
6	y	The comments that the politicians didn't ignore have caused bitterness toward the liberal newspapers.	Have the comments caused bitterness?
6	z	The comments that the politicians ignored have caused bitterness toward the liberal newspapers.	Have the comments caused bitterness?
7	a	No detergents that the housewives used have ever caused damage to the delicate clothing.	Have the detergents damaged the clothing?
7	b	The detergents that no housewives used have ever caused damage to the delicate clothing.	Have the detergents damaged the clothing?
7	c	The detergents that the housewives didn't use have ever caused damage to the delicate clothing.	Have the detergents damaged the clothing?
7	d	The detergents that the housewives used have ever caused damage to the delicate clothing.	Have the detergents damaged the clothing?
7	w	No detergents that the housewives used have caused damage to the delicate clothing.	Have the detergents damaged the clothing?

7	x	The detergents that no housewives used have caused damage to the delicate clothing.	Have the detergents damaged the clothing?
7	y	The detergents that the housewives didn't use have caused damage to the delicate clothing.	Have the detergents damaged the clothing?
7	z	The detergents that the housewives used have caused damage to the delicate clothing.	Have the detergents damaged the clothing?
8	a	No lawyers that the businessmen respected have ever received criticism for a bad trial.	Have the lawyers been criticized?
8	b	The lawyers that no businessmen respected have ever received criticism for a bad trial.	Have the lawyers been criticized?
8	c	The lawyers that the businessmen didn't respect have ever received criticism for a bad trial.	Have the lawyers been criticized?
8	d	The lawyers that the businessmen respected have ever received criticism for a bad trial.	Have the lawyers been criticized?
8	w	No lawyers that the businessmen respected have received criticism for a bad trial.	Have the lawyers been criticized?
8	x	The lawyers that no businessmen respected have received criticism for a bad trial.	Have the lawyers been criticized?
8	y	The lawyers that the businessmen didn't respect have received criticism for a bad trial.	Have the lawyers been criticized?
8	z	The lawyers that the businessmen respected have received criticism for a bad trial.	Have the lawyers been criticized?
9	a	No students that the teachers punished have ever expected friendliness from the strict principal.	Have the students expected friendliness from the principal?
9	b	The students that no teachers punished have ever expected friendliness from the strict principal.	Have the students expected friendliness from the principal?
9	c	The students that the teachers didn't punish have ever expected friendliness from the strict principal.	Have the students expected friendliness from the principal?
9	d	The students that the teachers punished have ever expected friendliness from the strict principal.	Have the students expected friendliness from the principal?
9	w	No students that the teachers punished have expected friendliness from the strict principal.	Have the students expected friendliness from the principal?

9	x	The students that no teachers punished have expected friendliness from the strict principal.	Have the students expected friendliness from the principal?
9	y	The students that the teachers didn't punish have expected friendliness from the strict principal.	Have the students expected friendliness from the principal?
9	z	The students that the teachers punished have expected friendliness from the strict principal.	Have the students expected friendliness from the principal?
10	a	No babysitters that the children obeyed have ever expected gratitude from the disappointed parents.	Have the babysitters expected the parents' gratitude?
10	b	The babysitters that no children obeyed have ever expected gratitude from the disappointed parents.	Have the babysitters expected the parents' gratitude?
10	c	The babysitters that the children didn't obey have ever expected gratitude from the disappointed parents.	Have the babysitters expected the parents' gratitude?
10	d	The babysitters that the children obeyed have ever expected gratitude from the disappointed parents.	Have the babysitters expected the parents' gratitude?
10	w	No babysitters that the children obeyed have expected gratitude from the disappointed parents.	Have the babysitters expected the parents' gratitude?
10	x	The babysitters that no children obeyed have expected gratitude from the disappointed parents.	Have the babysitters expected the parents' gratitude?
10	y	The babysitters that the children didn't obey have expected gratitude from the disappointed parents.	Have the babysitters expected the parents' gratitude?
10	z	The babysitters that the children obeyed have expected gratitude from the disappointed parents.	Have the babysitters expected the parents' gratitude?
11	a	No actors that the fans recognized have ever experienced soreness after a dangerous stunt.	Have the actors been sore from performing stunts?
11	b	The actors that no fans recognized have ever experienced soreness after a dangerous stunt.	Have the actors been sore from performing stunts?
11	c	The actors that the fans didn't recognize have ever experienced soreness after a dangerous stunt.	Have the actors been sore from performing stunts?
11	d	The actors that the fans recognized have ever experienced soreness after a dangerous stunt.	Have the actors been sore from performing stunts?
11	w	No actors that the fans recognized have experienced soreness after a dangerous stunt.	Have the actors been sore from performing stunts?

11	x	The actors that no fans recognized have experienced soreness after a dangerous stunt.	Have the actors been sore from performing stunts?
11	y	The actors that the fans didn't recognize have experienced soreness after a dangerous stunt.	Have the actors been sore from performing stunts?
11	z	The actors that the fans recognized have experienced soreness after a dangerous stunt.	Have the actors been sore from performing stunts?
12	a	No teachers that the parents recommended have ever caused problems for the new students.	Have the teachers caused problems for the new students?
12	b	The teachers that no parents recommended have ever caused problems for the new students.	Have the teachers caused problems for the new students?
12	c	The teachers that the parents didn't recommend have ever caused problems for the new students.	Have the teachers caused problems for the new students?
12	d	The teachers that the parents recommended have ever caused problems for the new students.	Have the teachers caused problems for the new students?
12	w	No teachers that the parents recommended have caused problems for the new students.	Have the teachers caused problems for the new students?
12	x	The teachers that no parents recommended have caused problems for the new students.	Have the teachers caused problems for the new students?
12	y	The teachers that the parents didn't recommend have caused problems for the new students.	Have the teachers caused problems for the new students?
12	z	The teachers that the parents recommended have caused problems for the new students.	Have the teachers caused problems for the new students?
13	a	No students that the librarians helped have ever made progress on the difficult assignment.	Have the students made progress on the assignment?
13	b	The students that no librarians helped have ever made progress on the difficult assignment.	Have the students made progress on the assignment?
13	c	The students that the librarians didn't help have ever made progress on the difficult assignment.	Have the students made progress on the assignment?
13	d	The students that the librarians helped have ever made progress on the difficult assignment.	Have the students made progress on the assignment?
13	w	No students that the librarians helped have made progress on the difficult assignment.	Have the students made progress on the assignment?

13	x	The students that no librarians helped have made progress on the difficult assignment.	Have the students made progress on the assignment?
13	y	The students that the librarians didn't help have made progress on the difficult assignment.	Have the students made progress on the assignment?
13	z	The students that the librarians helped have made progress on the difficult assignment.	Have the students made progress on the assignment?
14	a	No nurses that the doctors appreciated have ever shown patience in the operating room.	Have the nurses been patient?
14	b	The nurses that no doctors appreciated have ever shown patience in the operating room.	Have the nurses been patient?
14	c	The nurses that the doctors didn't appreciate have ever shown patience in the operating room.	Have the nurses been patient?
14	d	The nurses that the doctors appreciated have ever shown patience in the operating room.	Have the nurses been patient?
14	w	No nurses that the doctors appreciated have shown patience in the operating room.	Have the nurses been patient?
14	x	The nurses that no doctors appreciated have shown patience in the operating room.	Have the nurses been patient?
14	y	The nurses that the doctors didn't appreciate have shown patience in the operating room.	Have the nurses been patient?
14	z	The nurses that the doctors appreciated have shown patience in the operating room.	Have the nurses been patient?
15	a	No criminals that the policemen caught have ever felt satisfaction from a petty crime.	Have the criminals felt satisfaction?
15	b	The criminals that no policemen caught have ever felt satisfaction from a petty crime.	Have the criminals felt satisfaction?
15	c	The criminals that the policemen didn't catch have ever felt satisfaction from a petty crime.	Have the criminals felt satisfaction?
15	d	The criminals that the policemen caught have ever felt satisfaction from a petty crime.	Have the criminals felt satisfaction?
15	w	No criminals that the policemen caught have felt satisfaction from a petty crime.	Have the criminals felt satisfaction?

15	x	The criminals that no policemen caught have felt satisfaction from a petty crime.	Have the criminals felt satisfaction?
15	y	The criminals that the policemen didn't catch have felt satisfaction from a petty crime.	Have the criminals felt satisfaction?
15	z	The criminals that the policemen caught have felt satisfaction from a petty crime.	Have the criminals felt satisfaction?
16	a	No employees that the managers recommended have ever wanted kindness from the rude customers.	Have the employees wanted the customers to be kind?
16	b	The employees that no managers recommended have ever wanted kindness from the rude customers.	Have the employees wanted the customers to be kind?
16	c	The employees that the managers didn't recommend have ever wanted kindness from the rude customers.	Have the employees wanted the customers to be kind?
16	d	The employees that the managers recommended have ever wanted kindness from the rude customers.	Have the employees wanted the customers to be kind?
16	w	No employees that the managers recommended have wanted kindness from the rude customers.	Have the employees wanted the customers to be kind?
16	x	The employees that no managers recommended have wanted kindness from the rude customers.	Have the employees wanted the customers to be kind?
16	y	The employees that the managers didn't recommend have wanted kindness from the rude customers.	Have the employees wanted the customers to be kind?
16	z	The employees that the managers recommended have wanted kindness from the rude customers.	Have the employees wanted the customers to be kind?
17	a	No accountants that the managers trusted have ever seen rises in the quarterly profits.	Have the accountants seen rises in profits?
17	b	The accountants that no managers trusted have ever seen rises in the quarterly profits.	Have the accountants seen rises in profits?
17	c	The accountants that the managers didn't trust have ever seen rises in the quarterly profits.	Have the accountants seen rises in profits?
17	d	The accountants that the managers trusted have ever seen rises in the quarterly profits.	Have the accountants seen rises in profits?
17	w	No accountants that the managers trusted have seen rises in the quarterly profits.	Have the accountants seen rises in profits?

17	x	The accountants that no managers trusted have seen rises in the quarterly profits.	Have the accountants seen rises in profits?
17	y	The accountants that the managers didn't trust have seen rises in the quarterly profits.	Have the accountants seen rises in profits?
17	z	The accountants that the managers trusted have seen rises in the quarterly profits.	Have the accountants seen rises in profits?
18	a	No candidates that the voters supported have ever shown friendliness to the rude journalists.	Have the candidates been friendly?
18	b	The candidates that no voters supported have ever shown friendliness to the rude journalists.	Have the candidates been friendly?
18	c	The candidates that the voters didn't support have ever shown friendliness to the rude journalists.	Have the candidates been friendly?
18	d	The candidates that the voters supported have ever shown friendliness to the rude journalists.	Have the candidates been friendly?
18	w	No candidates that the voters supported have shown friendliness to the rude journalists.	Have the candidates been friendly?
18	x	The candidates that no voters supported have shown friendliness to the rude journalists.	Have the candidates been friendly?
18	y	The candidates that the voters didn't support have shown friendliness to the rude journalists.	Have the candidates been friendly?
18	z	The candidates that the voters supported have shown friendliness to the rude journalists.	Have the candidates been friendly?
19	a	No surgeons that the patients trusted have ever seen appreciation from the hospital staff.	Have the surgeons been appreciated by the hospital staff?
19	b	The surgeons that no patients trusted have ever seen appreciation from the hospital staff.	Have the surgeons been appreciated by the hospital staff?
19	c	The surgeons that the patients didn't trust have ever seen appreciation from the hospital staff.	Have the surgeons been appreciated by the hospital staff?
19	d	The surgeons that the patients trusted have ever seen appreciation from the hospital staff.	Have the surgeons been appreciated by the hospital staff?
19	w	No surgeons that the patients trusted have seen appreciation from the hospital staff.	Have the surgeons been appreciated by the hospital staff?

19	x	The surgeons that no patients trusted have seen appreciation from the hospital staff.	Have the surgeons been appreciated by the hospital staff?
19	y	The surgeons that the patients didn't trust have seen appreciation from the hospital staff.	Have the surgeons been appreciated by the hospital staff?
19	z	The surgeons that the patients trusted have seen appreciation from the hospital staff.	Have the surgeons been appreciated by the hospital staff?
20	a	No suspects that the witnesses identified have ever shown nervousness in the court room.	Have the suspects shown nervousness?
20	b	The suspects that no witnesses identified have ever shown nervousness in the court room.	Have the suspects shown nervousness?
20	c	The suspects that the witnesses didn't identify have ever shown nervousness in the court room.	Have the suspects shown nervousness?
20	d	The suspects that the witnesses identified have ever shown nervousness in the court room.	Have the suspects shown nervousness?
20	w	No suspects that the witnesses identified have shown nervousness in the court room.	Have the suspects shown nervousness?
20	x	The suspects that no witnesses identified have shown nervousness in the court room.	Have the suspects shown nervousness?
20	y	The suspects that the witnesses didn't identify have shown nervousness in the court room.	Have the suspects shown nervousness?
20	z	The suspects that the witnesses identified have shown nervousness in the court room.	Have the suspects shown nervousness?
21	a	No actresses that the housewives admired have ever caused excitement at a film festival.	Have the actresses caused excitement at a film festival?
21	b	The actresses that no housewives admired have ever caused excitement at a film festival.	Have the actresses caused excitement at a film festival?
21	c	The actresses that the housewives didn't admire have ever caused excitement at a film festival.	Have the actresses caused excitement at a film festival?
21	d	The actresses that the housewives admired have ever caused excitement at a film festival.	Have the actresses caused excitement at a film festival?
21	w	No actresses that the housewives admired have caused excitement at a film festival.	Have the actresses caused excitement at a film festival?

21	x	The actresses that no housewives admired have caused excitement at a film festival.	Have the actresses caused excitement at a film festival?
21	y	The actresses that the housewives didn't admire have caused excitement at a film festival.	Have the actresses caused excitement at a film festival?
21	z	The actresses that the housewives admired have caused excitement at a film festival.	Have the actresses caused excitement at a film festival?
22	a	No ambassadors that the government consulted have ever received hostility from the liberal media.	Have the ambassadors received hostility?
22	b	The ambassadors that no government consulted have ever received hostility from the liberal media.	Have the ambassadors received hostility?
22	c	The ambassadors that the government didn't consult have ever received hostility from the liberal media.	Have the ambassadors received hostility?
22	d	The ambassadors that the government consulted have ever received hostility from the liberal media.	Have the ambassadors received hostility?
22	w	No ambassadors that the government consulted have received hostility from the liberal media.	Have the ambassadors received hostility?
22	x	The ambassadors that no government consulted have received hostility from the liberal media.	Have the ambassadors received hostility?
22	y	The ambassadors that the government didn't consult have received hostility from the liberal media.	Have the ambassadors received hostility?
22	z	The ambassadors that the government consulted have received hostility from the liberal media.	Have the ambassadors received hostility?
23	a	No politicians that the journalists endorsed have ever earned trust from the rural communities.	Have the politicians earned the communities' trust?
23	b	The politicians that no journalists endorsed have ever earned trust from the rural communities.	Have the politicians earned the communities' trust?
23	c	The politicians that the journalists didn't endorse have ever earned trust from the rural communities.	Have the politicians earned the communities' trust?
23	d	The politicians that the journalists endorsed have ever earned trust from the rural communities.	Have the politicians earned the communities' trust?
23	w	No politicians that the journalists endorsed have earned trust from the rural communities.	Have the politicians earned the communities' trust?

23	x	The politicians that no journalists endorsed have earned trust from the rural communities.	Have the politicians earned the communities' trust?
23	y	The politicians that the journalists didn't endorse have earned trust from the rural communities.	Have the politicians earned the communities' trust?
23	z	The politicians that the journalists endorsed have earned trust from the rural communities.	Have the politicians earned the communities' trust?
24	a	No teenagers that the parents trusted have ever expressed kindness to a younger sibling.	Have the teenagers been kind to their siblings?
24	b	The teenagers that no parents trusted have ever expressed kindness to a younger sibling.	Have the teenagers been kind to their siblings?
24	c	The teenagers that the parents didn't trust have ever expressed kindness to a younger sibling.	Have the teenagers been kind to their siblings?
24	d	The teenagers that the parents trusted have ever expressed kindness to a younger sibling.	Have the teenagers been kind to their siblings?
24	w	No teenagers that the parents trusted have expressed kindness to a younger sibling.	Have the teenagers been kind to their siblings?
24	x	The teenagers that no parents trusted have expressed kindness to a younger sibling.	Have the teenagers been kind to their siblings?
24	y	The teenagers that the parents didn't trust have expressed kindness to a younger sibling.	Have the teenagers been kind to their siblings?
24	z	The teenagers that the parents trusted have expressed kindness to a younger sibling.	Have the teenagers been kind to their siblings?
25	a	No survivors that the medics treated have ever felt courage during the extreme emergency.	Have the survivors felt courage?
25	b	The survivors that no medics treated have ever felt courage during the extreme emergency.	Have the survivors felt courage?
25	c	The survivors that the medics didn't treat have ever felt courage during the extreme emergency.	Have the survivors felt courage?
25	d	The survivors that the medics treated have ever felt courage during the extreme emergency.	Have the survivors felt courage?
25	w	No survivors that the medics treated have felt courage during the extreme emergency.	Have the survivors felt courage?

25	x	The survivors that no medics treated have felt courage during the extreme emergency.	Have the survivors felt courage?
25	y	The survivors that the medics didn't treat have felt courage during the extreme emergency.	Have the survivors felt courage?
25	z	The survivors that the medics treated have felt courage during the extreme emergency.	Have the survivors felt courage?
26	a	No players that the coaches drafted have ever felt nervousness before a championship game.	Have the players felt nervous before the game?
26	b	The players that no coaches drafted have ever felt nervousness before a championship game.	Have the players felt nervous before the game?
26	c	The players that the coaches didn't draft have ever felt nervousness before a championship game.	Have the players felt nervous before the game?
26	d	The players that the coaches drafted have ever felt nervousness before a championship game.	Have the players felt nervous before the game?
26	w	No players that the coaches drafted have felt nervousness before a championship game.	Have the players felt nervous before the game?
26	x	The players that no coaches drafted have felt nervousness before a championship game.	Have the players felt nervous before the game?
26	y	The players that the coaches didn't draft have felt nervousness before a championship game.	Have the players felt nervous before the game?
26	z	The players that the coaches drafted have felt nervousness before a championship game.	Have the players felt nervous before the game?
27	a	No dictators that the citizens trusted have ever caused chaos at a public event.	Have the dictators caused chaos in public?
27	b	The dictators that no citizens trusted have ever caused chaos at a public event.	Have the dictators caused chaos in public?
27	c	The dictators that the citizens didn't trust have ever caused chaos at a public event.	Have the dictators caused chaos in public?
27	d	The dictators that the citizens trusted have ever caused chaos at a public event.	Have the dictators caused chaos in public?
27	w	No dictators that the citizens trusted have caused chaos at a public event.	Have the dictators caused chaos in public?

27	x	The dictators that no citizens trusted have caused chaos at a public event.	Have the dictators caused chaos in public?
27	y	The dictators that the citizens didn't trust have caused chaos at a public event.	Have the dictators caused chaos in public?
27	z	The dictators that the citizens trusted have caused chaos at a public event.	Have the dictators caused chaos in public?
28	a	No professors that the students understood have ever experienced tiredness after a long lecture.	Have the professors experienced tiredness?
28	b	The professors that no students understood have ever experienced tiredness after a long lecture.	Have the professors experienced tiredness?
28	c	The professors that the students didn't understand have ever experienced tiredness after a long lecture.	Have the professors experienced tiredness?
28	d	The professors that the students understood have ever experienced tiredness after a long lecture.	Have the professors experienced tiredness?
28	w	No professors that the students understood have experienced tiredness after a long lecture.	Have the professors experienced tiredness?
28	x	The professors that no students understood have experienced tiredness after a long lecture.	Have the professors experienced tiredness?
28	y	The professors that the students didn't understand have experienced tiredness after a long lecture.	Have the professors experienced tiredness?
28	z	The professors that the students understood have experienced tiredness after a long lecture.	Have the professors experienced tiredness?
29	a	No actors that the judges nominated have ever had luck at the award ceremonies.	Have the actors had luck at awards ceremonies?
29	b	The actors that no judges nominated have ever had luck at the award ceremonies.	Have the actors had luck at awards ceremonies?
29	c	The actors that the judges didn't nominate have ever had luck at the award ceremonies.	Have the actors had luck at awards ceremonies?
29	d	The actors that the judges nominated have ever had luck at the award ceremonies.	Have the actors had luck at awards ceremonies?
29	w	No actors that the judges nominated have had luck at the award ceremonies.	Have the actors had luck at awards ceremonies?

29	x	The actors that no judges nominated have had luck at the award ceremonies.	Have the actors had luck at awards ceremonies?
29	y	The actors that the judges didn't nominate have had luck at the award ceremonies.	Have the actors had luck at awards ceremonies?
29	z	The actors that the judges nominated have had luck at the award ceremonies.	Have the actors had luck at awards ceremonies?
30	a	No actresses that the directors auditioned have ever shown elegance on a large stage.	Were the actresses elegant on stage?
30	b	The actresses that no directors auditioned have ever shown elegance on a large stage.	Were the actresses elegant on stage?
30	c	The actresses that the directors didn't audition have ever shown elegance on a large stage.	Were the actresses elegant on stage?
30	d	The actresses that the directors auditioned have ever shown elegance on a large stage.	Were the actresses elegant on stage?
30	w	No actresses that the directors auditioned have shown elegance on a large stage.	Were the actresses elegant on stage?
30	x	The actresses that no directors auditioned have shown elegance on a large stage.	Were the actresses elegant on stage?
30	y	The actresses that the directors didn't audition have shown elegance on a large stage.	Were the actresses elegant on stage?
30	z	The actresses that the directors auditioned have shown elegance on a large stage.	Were the actresses elegant on stage?
31	a	No champions that the competitors defeated have ever shown humility after a big game.	Have the champions shown humility?
31	b	The champions that no competitors defeated have ever shown humility after a big game.	Have the champions shown humility?
31	c	The champions that the competitors didn't defeat have ever shown humility after a big game.	Have the champions shown humility?
31	d	The champions that the competitors defeated have ever shown humility after a big game.	Have the champions shown humility?
31	w	No champions that the competitors defeated have shown humility after a big game.	Have the champions shown humility?

31	x	The champions that no competitors defeated have shown humility after a big game.	Have the champions shown humility?
31	y	The champions that the competitors didn't defeat have shown humility after a big game.	Have the champions shown humility?
31	z	The champions that the competitors defeated have shown humility after a big game.	Have the champions shown humility?
32	a	No paintings that the collectors liked have ever depicted tranquility with very bright colors.	Have the paintings depicted tranquility with bright colors?
32	b	The paintings that no collectors liked have ever depicted tranquility with very bright colors.	Have the paintings depicted tranquility with bright colors?
32	c	The paintings that the collectors didn't like have ever depicted tranquility with very bright colors.	Have the paintings depicted tranquility with bright colors?
32	d	The paintings that the collectors liked have ever depicted tranquility with very bright colors.	Have the paintings depicted tranquility with bright colors?
32	w	No paintings that the collectors liked have depicted tranquility with very bright colors.	Have the paintings depicted tranquility with bright colors?
32	x	The paintings that no collectors liked have depicted tranquility with very bright colors.	Have the paintings depicted tranquility with bright colors?
32	y	The paintings that the collectors didn't like have depicted tranquility with very bright colors.	Have the paintings depicted tranquility with bright colors?
32	z	The paintings that the collectors liked have depicted tranquility with very bright colors.	Have the paintings depicted tranquility with bright colors?
33	a	No editors that the journalists respected have ever had patience for a missed deadline.	Have the editors had patience for a missed deadline?
33	b	The editors that no journalists respected have ever had patience for a missed deadline.	Have the editors had patience for a missed deadline?
33	c	The editors that the journalists didn't respect have ever had patience for a missed deadline.	Have the editors had patience for a missed deadline?
33	d	The editors that the journalists respected have ever had patience for a missed deadline.	Have the editors had patience for a missed deadline?
33	w	No editors that the journalists respected have had patience for a missed deadline.	Have the editors had patience for a missed deadline?

33	x	The editors that no journalists respected have had patience for a missed deadline.	Have the editors had patience for a missed deadline?
33	y	The editors that the journalists didn't respect have had patience for a missed deadline.	Have the editors had patience for a missed deadline?
33	z	The editors that the journalists respected have had patience for a missed deadline.	Have the editors had patience for a missed deadline?
34	a	No teenagers that the teachers motivated have ever experienced loneliness in the large class.	Have the teenagers experienced loneliness in large classes?
34	b	The teenagers that no teachers motivated have ever experienced loneliness in the large class.	Have the teenagers experienced loneliness in large classes?
34	c	The teenagers that the teachers didn't motivate have ever experienced loneliness in the large class.	Have the teenagers experienced loneliness in large classes?
34	d	The teenagers that the teachers motivated have ever experienced loneliness in the large class.	Have the teenagers experienced loneliness in large classes?
34	w	No teenagers that the teachers motivated have experienced loneliness in the large class.	Have the teenagers experienced loneliness in large classes?
34	x	The teenagers that no teachers motivated have experienced loneliness in the large class.	Have the teenagers experienced loneliness in large classes?
34	y	The teenagers that the teachers didn't motivate have experienced loneliness in the large class.	Have the teenagers experienced loneliness in large classes?
34	z	The teenagers that the teachers motivated have experienced loneliness in the large class.	Have the teenagers experienced loneliness in large classes?
35	a	No students that the professors tutored have ever had trouble in a math class.	Have the students had trouble in math class?
35	b	The students that no professors tutored have ever had trouble in a math class.	Have the students had trouble in math class?
35	c	The students that the professors didn't tutor have ever had trouble in a math class.	Have the students had trouble in math class?
35	d	The students that the professors tutored have ever had trouble in a math class.	Have the students had trouble in math class?
35	w	No students that the professors tutored have had trouble in a math class.	Have the students had trouble in math class?

35	x	The students that no professors tutored have had trouble in a math class.	Have the students had trouble in math class?
35	y	The students that the professors didn't tutor have had trouble in a math class.	Have the students had trouble in math class?
35	z	The students that the professors tutored have had trouble in a math class.	Have the students had trouble in math class?
36	a	No movies that the children watched have ever depicted gore during a fight scene.	Have the movies shown gory fight scenes?
36	b	The movies that no children watched have ever depicted gore during a fight scene.	Have the movies shown gory fight scenes?
36	c	The movies that the children didn't watch have ever depicted gore during a fight scene.	Have the movies shown gory fight scenes?
36	d	The movies that the children watched have ever depicted gore during a fight scene.	Have the movies shown gory fight scenes?
36	w	No movies that the children watched have depicted gore during a fight scene.	Have the movies shown gory fight scenes?
36	x	The movies that no children watched have depicted gore during a fight scene.	Have the movies shown gory fight scenes?
36	y	The movies that the children didn't watch have depicted gore during a fight scene.	Have the movies shown gory fight scenes?
36	z	The movies that the children watched have depicted gore during a fight scene.	Have the movies shown gory fight scenes?

Table A.4: Full experimental stimuli for Experiment 4

A.5 Experiment 5

1	a	No critics that have recommended the authors of alternative genres have ever objected to mainstream literary trends.
1	b	The critics that have recommended no authors of alternative genres have ever objected to mainstream literary trends.
1	c	The critics that haven't recommended the authors of alternative genres have ever objected to mainstream literary trends.

1	d	The critics that haven't recommended any authors of alternative genres have ever objected to mainstream literary trends.
1	e	The critics that have recommended the authors of alternative genres have ever objected to mainstream literary trends.
2	a	No diplomats that have supported the refugees from war-torn countries have ever gained the respect of far-right journalists.
2	b	The diplomats that have supported no refugees from war-torn countries have ever gained the respect of far-right journalists.
2	c	The diplomats that haven't supported the refugees from war-torn countries have ever gained the respect of far-right journalists.
2	d	The diplomats that haven't supported any refugees from war-torn countries have ever gained the respect of far-right journalists.
2	e	The diplomats that have supported the refugees from war-torn countries have ever gained the respect of far-right journalists.
3	a	No paramedics that have treated the victims of heart attacks have ever underestimated the importance of a healthy lifestyle.
3	b	The paramedics that have treated no victims of heart attacks have ever underestimated the importance of a healthy lifestyle.
3	c	The paramedics that haven't treated the victims of heart attacks have ever underestimated the importance of a healthy lifestyle.
3	d	The paramedics that haven't treated any victims of heart attacks have ever underestimated the importance of a healthy lifestyle.
3	e	The paramedics that have treated the victims of heart attacks have ever underestimated the importance of a healthy lifestyle.
4	a	No professors that have taught the children of needy families have ever supported the plans for privatizing the school system.
4	b	The professors that have taught no children of needy families have ever supported the plans for privatizing the school system.
4	c	The professors that haven't taught the children of needy families have ever supported the plans for privatizing the school system.
4	d	The professors that haven't taught any children of needy families have ever supported the plans for privatizing the school system.

4	e	The professors that have taught the children of needy families have ever supported the plans for privatizing the school system.
5	a	No salesmen that have assisted the customers with bad attitudes have ever picked up extra shifts during the holiday season.
5	b	The salesmen that have assisted no customers with bad attitudes have ever picked up extra shifts during the holiday season.
5	c	The salesmen that haven't assisted the customers with bad attitudes have ever picked up extra shifts during the holiday season.
5	d	The salesmen that haven't assisted any customers with bad attitudes have ever picked up extra shifts during the holiday season.
5	e	The salesmen that have assisted the customers with bad attitudes have ever picked up extra shifts during the holiday season.
6	a	No politicians that have addressed the reporters from socialist websites have ever caused controversy over a democratic bill.
6	b	The politicians that have addressed no reporters from socialist websites have ever caused controversy over a democratic bill.
6	c	The politicians that haven't addressed the reporters from socialist websites have ever caused controversy over a democratic bill.
6	d	The politicians that haven't addressed any reporters from socialist websites have ever caused controversy over a democratic bill.
6	e	The politicians that have addressed the reporters from socialist websites have ever caused controversy over a democratic bill.
7	a	No designers that have hired the models from top-tier agencies have ever expressed concern about eating disorders.
7	b	The designers that have hired no models from top-tier agencies have ever expressed concern about eating disorders.
7	c	The designers that haven't hired the models from top-tier agencies have ever expressed concern about eating disorders.
7	d	The designers that haven't hired any models from top-tier agencies have ever expressed concern about eating disorders.
7	e	The designers that have hired the models from top-tier agencies have ever expressed concern about eating disorders.

8	a	No businessmen that have fired the salesmen with poor performance have ever received criticism from other powerful CEOs.
8	b	The businessmen that have fired no salesmen with poor performance have ever received criticism from other powerful CEOs.
8	c	The businessmen that haven't fired the salesmen with poor performance have ever received criticism from other powerful CEOs.
8	d	The businessmen that haven't fired any salesmen with poor performance have ever received criticism from other powerful CEOs.
8	e	The businessmen that have fired the salesmen with poor performance have ever received criticism from other powerful CEOs.
9	a	No teachers that have punished the organizers of school pranks have ever expected support from the liberal principal.
9	b	The teachers that have punished no organizers of school pranks have ever expected support from the liberal principal.
9	c	The teachers that haven't punished the organizers of school pranks have ever expected support from the liberal principal.
9	d	The teachers that haven't punished any organizers of school pranks have ever expected support from the liberal principal.
9	e	The teachers that have punished the organizers of school pranks have ever expected support from the liberal principal.
10	a	No judges that have dismissed the jurors with clear biases have ever been challenged on their rulings.
10	b	The judges that have dismissed no jurors with clear biases have ever been challenged on their rulings.
10	c	The judges that haven't dismissed the jurors with clear biases have ever been challenged on their rulings.
10	d	The judges that haven't dismissed any jurors with clear biases have ever been challenged on their rulings.
10	e	The judges that have dismissed the jurors with clear biases have ever been challenged on their rulings.
11	a	No parents that have criticized the teachers of difficult classes have ever expressed concerns about students' work ethic.
11	b	The parents that have criticized no teachers of difficult classes have ever expressed concerns about students' work ethic.
11	c	The parents that haven't criticized the teachers of difficult classes have ever expressed concerns about students' work ethic.

11	d	The parents that haven't criticized any teachers of difficult classes have ever expressed concerns about students' work ethic.
11	e	The parents that have criticized the teachers of difficult classes have ever expressed concerns about students' work ethic.
12	a	No coaches that have commended the athletes with bad sportsmanship have ever valued building camaraderie among the teammates.
12	b	The coaches that have commended no athletes with bad sportsmanship have ever valued building camaraderie among the teammates.
12	c	The coaches that haven't commended the athletes with bad sportsmanship have ever valued building camaraderie among the teammates.
12	d	The coaches that haven't commended any athletes with bad sportsmanship have ever valued building camaraderie among the teammates.
12	e	The coaches that have commended the athletes with bad sportsmanship have ever valued building camaraderie among the teammates.
13	a	No moms that have consulted the doctors with great credentials have ever placed absolute faith in medical science.
13	b	The moms that have consulted no doctors with great credentials have ever placed absolute faith in medical science.
13	c	The moms that haven't consulted the doctors with great credentials have ever placed absolute faith in medical science.
13	d	The moms that haven't consulted any doctors with great credentials have ever placed absolute faith in medical science.
13	e	The moms that have consulted the doctors with great credentials have ever placed absolute faith in medical science.
14	a	No policemen that have injured the suspects of drug crimes have ever received criticism for implicit racial bias.
14	b	The policemen that have injured no suspects of drug crimes have ever received criticism for implicit racial bias.
14	c	The policemen that haven't injured the suspects of drug crimes have ever received criticism for implicit racial bias.
14	d	The policemen that haven't injured any suspects of drug crimes have ever received criticism for implicit racial bias.

14	e	The policemen that have injured the suspects of drug crimes have ever received criticism for implicit racial bias.
15	a	No supervisors that have praised the workers for strong performance have ever defended the employees in managerial meetings.
15	b	The supervisors that have praised no workers for strong performance have ever defended the employees in managerial meetings.
15	c	The supervisors that haven't praised the workers for strong performance have ever defended the employees in managerial meetings.
15	d	The supervisors that haven't praised any workers for strong performance have ever defended the employees in managerial meetings.
15	e	The supervisors that have praised the workers for strong performance have ever defended the employees in managerial meetings.
16	a	No producers that have signed the singers with sexist attitudes have ever been criticized by the fans.
16	b	The producers that have signed no singers with sexist attitudes have ever been criticized by the fans.
16	c	The producers that haven't signed the singers with sexist attitudes have ever been criticized by the fans.
16	d	The producers that haven't signed any singers with sexist attitudes have ever been criticized by the fans.
16	e	The producers that have signed the singers with sexist attitudes have ever been criticized by the fans.
17	a	No moviegoers that have favored the celebrities with Oscar nominations have ever shown interest in an independent film.
17	b	The moviegoers that have favored no celebrities with Oscar nominations have ever shown interest in an independent film.
17	c	The moviegoers that haven't favored the celebrities with Oscar nominations have ever shown interest in an independent film.
17	d	The moviegoers that haven't favored any celebrities with Oscar nominations have ever shown interest in an independent film.
17	e	The moviegoers that have favored the celebrities with Oscar nominations have ever shown interest in an independent film.
18	a	No nurses that have treated the patients with infectious diseases have ever expressed anxiety about the required vaccinations
18	b	The nurses that have treated no patients with infectious diseases have ever expressed anxiety about the required vaccinations

18	c	The nurses that haven't treated the patients with infectious diseases have ever expressed anxiety about the required vaccinations
18	d	The nurses that haven't treated any patients with infectious diseases have ever expressed anxiety about the required vaccinations
18	e	The nurses that have treated the patients with infectious diseases have ever expressed anxiety about the required vaccinations
19	a	No detectives that have intimidated the witnesses of horrific crimes have ever received training in conflict de-escalation.
19	b	The detectives that have intimidated no witnesses of horrific crimes have ever received training in conflict de-escalation.
19	c	The detectives that haven't intimidated the witnesses of horrific crimes have ever received training in conflict de-escalation.
19	d	The detectives that haven't intimidated any witnesses of horrific crimes have ever received training in conflict de-escalation.
19	e	The detectives that have intimidated the witnesses of horrific crimes have ever received training in conflict de-escalation.
20	a	No mechanics that have overcharged the customers with minimal expertise have ever received positive online reviews.
20	b	The mechanics that have overcharged no customers with minimal expertise have ever received positive online reviews.
20	c	The mechanics that haven't overcharged the customers with minimal expertise have ever received positive online reviews.
20	d	The mechanics that haven't overcharged any customers with minimal expertise have ever received positive online reviews.
20	e	The mechanics that have overcharged the customers with minimal expertise have ever received positive online reviews.
21	a	No billionaires that have bankrolled the candidates with Republican support have ever expressed dissatisfaction with conservative fiscal policies.
21	b	The billionaires that have bankrolled no candidates with Republican support have ever expressed dissatisfaction with conservative fiscal policies.
21	c	The billionaires that haven't bankrolled the candidates with Republican support have ever expressed dissatisfaction with conservative fiscal policies.

21	d	The billionaires that haven't bankrolled any candidates with Republican support have ever expressed dissatisfaction with conservative fiscal policies.
21	e	The billionaires that have bankrolled the candidates with Republican support have ever expressed dissatisfaction with conservative fiscal policies.
22	a	No lawyers that have assisted the clients with hopeless cases have ever developed impressive track records.
22	b	The lawyers that have assisted no clients with hopeless cases have ever developed impressive track records.
22	c	The lawyers that haven't assisted the clients with hopeless cases have ever developed impressive track records.
22	d	The lawyers that haven't assisted any clients with hopeless cases have ever developed impressive track records.
22	e	The lawyers that haven assisted the clients with hopeless cases have ever developed impressive track records.
23	a	No landlords that have approved the tenants with large dogs have ever worried about noise levels in the apartments.
23	b	The landlords that have approved no tenants with large dogs have ever worried about noise levels in the apartments.
23	c	The landlords that haven't approved the tenants with large dogs have ever worried about noise levels in the apartments.
23	d	The landlords that haven't approved any tenants with large dogs have ever worried about noise levels in the apartments.
23	e	The landlords that have approved the tenants with large dogs have ever worried about noise levels in the apartments.
24	a	No authors that have engaged the readers with refined taste have ever gained a broad audience.
24	b	The authors that have engaged no readers with refined taste have ever gained a broad audience.
24	c	The authors that haven't engaged the readers with refined taste have ever gained a broad audience.
24	d	The authors that haven't engaged any readers with refined taste have ever gained a broad audience.
24	e	The authors that have engaged the readers with refined taste have ever gained a broad audience.
25	a	No coaches that have drafted the players from top-tier schools have ever felt nervousness before a championship game.
25	b	The coaches that have drafted no players from top-tier schools have ever felt nervousness before a championship game.
25	c	The coaches that haven't drafted the players from top-tier schools have ever felt nervousness before a championship game.

25	d	The coaches that haven't drafted any players from top-tier schools have ever felt nervousness before a championship game.
25	e	The coaches that have drafted the players from top-tier schools have ever felt nervousness before a championship game.
26	a	No students that have visited the professors during office hours have ever been stressed about exams.
26	b	The students that have visited no professors during office hours have ever been stressed about exams.
26	c	The students that haven't visited the professors during office hours have ever been stressed about exams.
26	d	The students that haven't visited any professors during office hours have ever been stressed about exams.
26	e	The students that have visited the professors during office hours have ever been stressed about exams.
27	a	No researchers that have published the papers with stunning findings have ever been frustrated with the public's apathy.
27	b	The researchers that have published no papers with stunning findings have ever been frustrated with the public's apathy.
27	c	The researchers that haven't published the papers with stunning findings have ever been frustrated with the public's apathy.
27	d	The researchers that haven't published any papers with stunning findings have ever been frustrated with the public's apathy.
27	e	The researchers that have published the papers with stunning findings have ever been frustrated with the public's apathy.
28	a	No actors that have impressed the directors of blockbuster films have ever been nervous before an important audition.
28	b	The actors that have impressed no directors of blockbuster films have ever been nervous before an important audition.
28	c	The actors that haven't impressed the directors of blockbuster films have ever been nervous before an important audition.
28	d	The actors that haven't impressed any directors of blockbuster films have ever been nervous before an important audition.
28	e	The actors that have impressed the directors of blockbuster films have ever been nervous before an important audition.
29	a	No sprinters that have faced the competitors from top teams have ever been over-confident before an important race.

29	b	The sprinters that have faced no competitors from top teams have ever been over-confident before an important race.
29	c	The sprinters that haven't faced the competitors from top teams have ever been over-confident before an important race.
29	d	The sprinters that haven't faced any competitors from top teams have ever been over-confident before an important race.
29	e	The sprinters that have faced the competitors from top teams have ever been over-confident before an important race.
30	a	No collectors that have endorsed the artists with controversial themes have ever wanted the art world to be more political.
30	b	The collectors that have endorsed no artists with controversial themes have ever wanted the art world to be more political.
30	c	The collectors that haven't endorsed the artists with controversial themes have ever wanted the art world to be more political.
30	d	The collectors that haven't endorsed any artists with controversial themes have ever wanted the art world to be more political.
30	e	The collectors that have endorsed the artists with controversial themes have ever wanted the art world to be more political.
31	a	No journalists that have challenged the editors of prestigious newspapers have ever expected difficulty on the job market.
31	b	The journalists that have challenged no editors of prestigious newspapers have ever expected difficulty on the job market.
31	c	The journalists that haven't challenged the editors of prestigious newspapers have ever expected difficulty on the job market.
31	d	The journalists that haven't challenged any editors of prestigious newspapers have ever expected difficulty on the job market.
31	e	The journalists that have challenged the editors of prestigious newspapers have ever expected difficulty on the job market.
32	a	No teachers that have suspended the teenagers with behavioral problems have ever expected cooperation from the parents.
32	b	The teachers that have suspended no teenagers with behavioral problems have ever expected cooperation from the parents.

32	c	The teachers that haven't suspended the teenagers with behavioral problems have ever expected cooperation from the parents.
32	d	The teachers that haven't suspended any teenagers with behavioral problems have ever expected cooperation from the parents.
32	e	The teachers that have suspended the teenagers with behavioral problems have ever expected cooperation from the parents.
33	a	No tutors that have helped the students in challenging classes have ever suggested improvements in the student to teacher ratio.
33	b	The tutors that have helped no students in challenging classes have ever suggested improvements in the student to teacher ratio.
33	c	The tutors that haven't helped the students in challenging classes have ever suggested improvements in the student to teacher ratio.
33	d	The tutors that haven't helped any students in challenging classes have ever suggested improvements in the student to teacher ratio.
33	e	The tutors that have helped the students in challenging classes have ever suggested improvements in the student to teacher ratio.
34	a	No children that have respected the rules at community pools have ever caused trouble for the lifeguards.
34	b	The children that have respected no rules at community pools have ever caused trouble for the lifeguards.
34	c	The children that haven't respected the rules at community pools have ever caused trouble for the lifeguards.
34	d	The children that haven't respected any rules at community pools have ever caused trouble for the lifeguards.
34	e	The children that have respected the rules at community pools have ever caused trouble for the lifeguards.
35	a	No students that have completed the assignments during school hours have ever turned in their homework late.
35	b	The students that have completed no assignments during school hours have ever turned in their homework late.
35	c	The students that haven't completed the assignments during school hours have ever turned in their homework late.
35	d	The students that haven't completed any assignments during school hours have ever turned in their homework late.
35	e	The students that have completed the assignments during school hours have ever turned in their homework late.

36	a	No families that have visited the resorts during summer break have ever worried about the weather.
36	b	The families that have toured no resorts during summer break have ever worried about the weather.
36	c	The families that haven't toured the resorts during summer break have ever worried about the weather.
36	d	The families that haven't toured any resorts during summer break have ever worried about the weather.
36	e	The families that have toured the resorts during summer break have ever worried about the weather.
37	a	No swimmers that have done the warmups before morning practice have ever worried about strained muscles.
37	b	The swimmers that have done no warmups before morning practice have ever worried about strained muscles.
37	c	The swimmers that haven't done the warmups before morning practice have ever worried about strained muscles.
37	d	The swimmers that haven't done any warmups before morning practice have ever worried about strained muscles.
37	e	The swimmers that have done the warmups before morning practice have ever worried about strained muscles.
38	a	No actors that have memorized the lines before final auditions have ever given a compelling performance.
38	b	The actors that have memorized no lines before final auditions have ever given a compelling performance.
38	c	The actors that haven't memorized the lines before final auditions have ever given a compelling performance.
38	d	The actors that haven't memorized any lines before final auditions have ever given a compelling performance.
38	e	The actors that have memorized the lines before final auditions have ever given a compelling performance.
39	a	No graduates that have attended the rehearsals before the ceremony have ever been confused about the procession.
39	b	The graduates that have attended no rehearsals before the ceremony have ever been confused about the procession.
39	c	The graduates that haven't attended the rehearsals before the ceremony have ever been confused about the procession.
39	d	The graduates that haven't attended any rehearsals before the ceremony have ever been confused about the procession.
39	e	The graduates that have attended the rehearsals before the ceremony have ever been confused about the procession.

40	a	No teachers that have graded the exams before spring break have ever had students complain about an incomplete semester grade.
40	b	The teachers that have graded no exams before spring break have ever had students complain about an incomplete semester grade.
40	c	The teachers that haven't graded the exams before spring break have ever had students complain about an incomplete semester grade.
40	d	The teachers that haven't graded any exams before spring break have ever had students complain about an incomplete semester grade.
40	e	The teachers that have graded the exams before spring break have ever had students complain about an incomplete semester grade.

Table A.5: Full experimental stimuli for Experiment 5

A.6 Experiment 6

1	a	No critics that have recommended the authors of alternative genres have ever objected to mainstream literary trends.
1	b	The critics that have recommended no authors of alternative genres have ever objected to mainstream literary trends.
1	c	The critics that haven't recommended the authors of alternative genres have ever objected to mainstream literary trends.
1	d	The critics that haven't recommended any authors of alternative genres have ever objected to mainstream literary trends.
1	e	The critics that have recommended the authors of alternative genres have ever objected to mainstream literary trends.
2	a	No diplomats that have supported the refugees from war-torn countries have ever gained the respect of far-right journalists.
2	b	The diplomats that have supported no refugees from war-torn countries have ever gained the respect of far-right journalists.
2	c	The diplomats that haven't supported the refugees from war-torn countries have ever gained the respect of far-right journalists.

2	d	The diplomats that haven't supported any refugees from war-torn countries have ever gained the respect of far-right journalists.
2	e	The diplomats that have supported the refugees from war-torn countries have ever gained the respect of far-right journalists.
3	a	No paramedics that have treated the victims of heart attacks have ever underestimated the importance of a healthy lifestyle.
3	b	The paramedics that have treated no victims of heart attacks have ever underestimated the importance of a healthy lifestyle.
3	c	The paramedics that haven't treated the victims of heart attacks have ever underestimated the importance of a healthy lifestyle.
3	d	The paramedics that haven't treated any victims of heart attacks have ever underestimated the importance of a healthy lifestyle.
3	e	The paramedics that have treated the victims of heart attacks have ever underestimated the importance of a healthy lifestyle.
4	a	No professors that have taught the children of needy families have ever supported the plans for privatizing the school system.
4	b	The professors that have taught no children of needy families have ever supported the plans for privatizing the school system.
4	c	The professors that haven't taught the children of needy families have ever supported the plans for privatizing the school system.
4	d	The professors that haven't taught any children of needy families have ever supported the plans for privatizing the school system.
4	e	The professors that have taught the children of needy families have ever supported the plans for privatizing the school system.
5	a	No salesmen that have assisted the customers with bad attitudes have ever picked up extra shifts during the holiday season.
5	b	The salesmen that have assisted no customers with bad attitudes have ever picked up extra shifts during the holiday season.
5	c	The salesmen that haven't assisted the customers with bad attitudes have ever picked up extra shifts during the holiday season.
5	d	The salesmen that haven't assisted any customers with bad attitudes have ever picked up extra shifts during the holiday season.

5	e	The salesmen that have assisted the customers with bad attitudes have ever picked up extra shifts during the holiday season.
6	a	No politicians that have addressed the reporters from socialist websites have ever caused controversy over a democratic bill.
6	b	The politicians that have addressed no reporters from socialist websites have ever caused controversy over a democratic bill.
6	c	The politicians that haven't addressed the reporters from socialist websites have ever caused controversy over a democratic bill.
6	d	The politicians that haven't addressed any reporters from socialist websites have ever caused controversy over a democratic bill.
6	e	The politicians that have addressed the reporters from socialist websites have ever caused controversy over a democratic bill.
7	a	No designers that have hired the models from top-tier agencies have ever expressed concern about eating disorders.
7	b	The designers that have hired no models from top-tier agencies have ever expressed concern about eating disorders.
7	c	The designers that haven't hired the models from top-tier agencies have ever expressed concern about eating disorders.
7	d	The designers that haven't hired any models from top-tier agencies have ever expressed concern about eating disorders.
7	e	The designers that have hired the models from top-tier agencies have ever expressed concern about eating disorders.
8	a	No businessmen that have fired the salesmen with poor performance have ever received criticism from other powerful CEOs.
8	b	The businessmen that have fired no salesmen with poor performance have ever received criticism from other powerful CEOs.
8	c	The businessmen that haven't fired the salesmen with poor performance have ever received criticism from other powerful CEOs.
8	d	The businessmen that haven't fired any salesmen with poor performance have ever received criticism from other powerful CEOs.
8	e	The businessmen that have fired the salesmen with poor performance have ever received criticism from other powerful CEOs.

9	a	No teachers that have punished the organizers of school pranks have ever expected support from the liberal principal.
9	b	The teachers that have punished no organizers of school pranks have ever expected support from the liberal principal.
9	c	The teachers that haven't punished the organizers of school pranks have ever expected support from the liberal principal.
9	d	The teachers that haven't punished any organizers of school pranks have ever expected support from the liberal principal.
9	e	The teachers that have punished the organizers of school pranks have ever expected support from the liberal principal.
10	a	No judges that have dismissed the jurors with clear biases have ever been challenged on their rulings.
10	b	The judges that have dismissed no jurors with clear biases have ever been challenged on their rulings.
10	c	The judges that haven't dismissed the jurors with clear biases have ever been challenged on their rulings.
10	d	The judges that haven't dismissed any jurors with clear biases have ever been challenged on their rulings.
10	e	The judges that have dismissed the jurors with clear biases have ever been challenged on their rulings.
11	a	No parents that have criticized the teachers of difficult classes have ever expressed concerns about students' work ethic.
11	b	The parents that have criticized no teachers of difficult classes have ever expressed concerns about students' work ethic.
11	c	The parents that haven't criticized the teachers of difficult classes have ever expressed concerns about students' work ethic.
11	d	The parents that haven't criticized any teachers of difficult classes have ever expressed concerns about students' work ethic.
11	e	The parents that have criticized the teachers of difficult classes have ever expressed concerns about students' work ethic.
12	a	No coaches that have commended the athletes with bad sportsmanship have ever valued building camaraderie among the teammates.
12	b	The coaches that have commended no athletes with bad sportsmanship have ever valued building camaraderie among the teammates.
12	c	The coaches that haven't commended the athletes with bad sportsmanship have ever valued building camaraderie among the teammates.

12	d	The coaches that haven't commended any athletes with bad sportsmanship have ever valued building camaraderie among the teammates.
12	e	The coaches that have commended the athletes with bad sportsmanship have ever valued building camaraderie among the teammates.
13	a	No moms that have consulted the doctors with great credentials have ever placed absolute faith in medical science.
13	b	The moms that have consulted no doctors with great credentials have ever placed absolute faith in medical science.
13	c	The moms that haven't consulted the doctors with great credentials have ever placed absolute faith in medical science.
13	d	The moms that haven't consulted any doctors with great credentials have ever placed absolute faith in medical science.
13	e	The moms that have consulted the doctors with great credentials have ever placed absolute faith in medical science.
14	a	No policemen that have injured the suspects of drug crimes have ever received criticism for implicit racial bias.
14	b	The policemen that have injured no suspects of drug crimes have ever received criticism for implicit racial bias.
14	c	The policemen that haven't injured the suspects of drug crimes have ever received criticism for implicit racial bias.
14	d	The policemen that haven't injured any suspects of drug crimes have ever received criticism for implicit racial bias.
14	e	The policemen that have injured the suspects of drug crimes have ever received criticism for implicit racial bias.
15	a	No supervisors that have praised the workers for strong performance have ever defended the employees in managerial meetings.
15	b	The supervisors that have praised no workers for strong performance have ever defended the employees in managerial meetings.
15	c	The supervisors that haven't praised the workers for strong performance have ever defended the employees in managerial meetings.
15	d	The supervisors that haven't praised any workers for strong performance have ever defended the employees in managerial meetings.

15	e	The supervisors that have praised the workers for strong performance have ever defended the employees in managerial meetings.
16	a	No producers that have signed the singers with sexist attitudes have ever been criticized by the fans.
16	b	The producers that have signed no singers with sexist attitudes have ever been criticized by the fans.
16	c	The producers that haven't signed the singers with sexist attitudes have ever been criticized by the fans.
16	d	The producers that haven't signed any singers with sexist attitudes have ever been criticized by the fans.
16	e	The producers that have signed the singers with sexist attitudes have ever been criticized by the fans.
17	a	No moviegoers that have favored the celebrities with Oscar nominations have ever shown interest in an independent film.
17	b	The moviegoers that have favored no celebrities with Oscar nominations have ever shown interest in an independent film.
17	c	The moviegoers that haven't favored the celebrities with Oscar nominations have ever shown interest in an independent film.
17	d	The moviegoers that haven't favored any celebrities with Oscar nominations have ever shown interest in an independent film.
17	e	The moviegoers that have favored the celebrities with Oscar nominations have ever shown interest in an independent film.
18	a	No nurses that have treated the patients with infectious diseases have ever expressed anxiety about the required vaccinations
18	b	The nurses that have treated no patients with infectious diseases have ever expressed anxiety about the required vaccinations
18	c	The nurses that haven't treated the patients with infectious diseases have ever expressed anxiety about the required vaccinations
18	d	The nurses that haven't treated any patients with infectious diseases have ever expressed anxiety about the required vaccinations
18	e	The nurses that have treated the patients with infectious diseases have ever expressed anxiety about the required vaccinations
19	a	No detectives that have intimidated the witnesses of horrific crimes have ever received training in conflict de-escalation.
19	b	The detectives that have intimidated no witnesses of horrific crimes have ever received training in conflict de-escalation.

19	c	The detectives that haven't intimidated the witnesses of horrific crimes have ever received training in conflict de-escalation.
19	d	The detectives that haven't intimidated any witnesses of horrific crimes have ever received training in conflict de-escalation.
19	e	The detectives that have intimidated the witnesses of horrific crimes have ever received training in conflict de-escalation.
20	a	No mechanics that have overcharged the customers with minimal expertise have ever received positive online reviews.
20	b	The mechanics that have overcharged no customers with minimal expertise have ever received positive online reviews.
20	c	The mechanics that haven't overcharged the customers with minimal expertise have ever received positive online reviews.
20	d	The mechanics that haven't overcharged any customers with minimal expertise have ever received positive online reviews.
20	e	The mechanics that have overcharged the customers with minimal expertise have ever received positive online reviews.
21	a	No billionaires that have bankrolled the candidates with Republican support have ever expressed dissatisfaction with conservative fiscal policies.
21	b	The billionaires that have bankrolled no candidates with Republican support have ever expressed dissatisfaction with conservative fiscal policies.
21	c	The billionaires that haven't bankrolled the candidates with Republican support have ever expressed dissatisfaction with conservative fiscal policies.
21	d	The billionaires that haven't bankrolled any candidates with Republican support have ever expressed dissatisfaction with conservative fiscal policies.
21	e	The billionaires that have bankrolled the candidates with Republican support have ever expressed dissatisfaction with conservative fiscal policies.
22	a	No lawyers that have assisted the clients with hopeless cases have ever developed impressive track records.
22	b	The lawyers that have assisted no clients with hopeless cases have ever developed impressive track records.
22	c	The lawyers that haven't assisted the clients with hopeless cases have ever developed impressive track records.
22	d	The lawyers that haven't assisted any clients with hopeless cases have ever developed impressive track records.
22	e	The lawyers that haven assisted the clients with hopeless cases have ever developed impressive track records.

23	a	No landlords that have approved the tenants with large dogs have ever worried about noise levels in the apartments.
23	b	The landlords that have approved no tenants with large dogs have ever worried about noise levels in the apartments.
23	c	The landlords that haven't approved the tenants with large dogs have ever worried about noise levels in the apartments.
23	d	The landlords that haven't approved any tenants with large dogs have ever worried about noise levels in the apartments.
23	e	The landlords that have approved the tenants with large dogs have ever worried about noise levels in the apartments.
24	a	No authors that have engaged the readers with refined taste have ever gained a broad audience.
24	b	The authors that have engaged no readers with refined taste have ever gained a broad audience.
24	c	The authors that haven't engaged the readers with refined taste have ever gained a broad audience.
24	d	The authors that haven't engaged any readers with refined taste have ever gained a broad audience.
24	e	The authors that have engaged the readers with refined taste have ever gained a broad audience.
25	a	No coaches that have drafted the players from top-tier schools have ever felt nervousness before a championship game.
25	b	The coaches that have drafted no players from top-tier schools have ever felt nervousness before a championship game.
25	c	The coaches that haven't drafted the players from top-tier schools have ever felt nervousness before a championship game.
25	d	The coaches that haven't drafted any players from top-tier schools have ever felt nervousness before a championship game.
25	e	The coaches that have drafted the players from top-tier schools have ever felt nervousness before a championship game.
26	a	No students that have visited the professors during office hours have ever been stressed about exams.
26	b	The students that have visited no professors during office hours have ever been stressed about exams.
26	c	The students that haven't visited the professors during office hours have ever been stressed about exams.
26	d	The students that haven't visited any professors during office hours have ever been stressed about exams.
26	e	The students that have visited the professors during office hours have ever been stressed about exams.
27	a	No researchers that have published the papers with stunning findings have ever been frustrated with the public's apathy.

27	b	The researchers that have published no papers with stunning findings have ever been frustrated with the public's apathy.
27	c	The researchers that haven't published the papers with stunning findings have ever been frustrated with the public's apathy.
27	d	The researchers that haven't published any papers with stunning findings have ever been frustrated with the public's apathy.
27	e	The researchers that have published the papers with stunning findings have ever been frustrated with the public's apathy.
28	a	No actors that have impressed the directors of blockbuster films have ever been nervous before an important audition.
28	b	The actors that have impressed no directors of blockbuster films have ever been nervous before an important audition.
28	c	The actors that haven't impressed the directors of blockbuster films have ever been nervous before an important audition.
28	d	The actors that haven't impressed any directors of blockbuster films have ever been nervous before an important audition.
28	e	The actors that have impressed the directors of blockbuster films have ever been nervous before an important audition.
29	a	No sprinters that have faced the competitors from top teams have ever been over-confident before an important race.
29	b	The sprinters that have faced no competitors from top teams have ever been over-confident before an important race.
29	c	The sprinters that haven't faced the competitors from top teams have ever been over-confident before an important race.
29	d	The sprinters that haven't faced any competitors from top teams have ever been over-confident before an important race.
29	e	The sprinters that have faced the competitors from top teams have ever been over-confident before an important race.
30	a	No collectors that have endorsed the artists with controversial themes have ever wanted the art world to be more political.
30	b	The collectors that have endorsed no artists with controversial themes have ever wanted the art world to be more political.

30	c	The collectors that haven't endorsed the artists with controversial themes have ever wanted the art world to be more political.
30	d	The collectors that haven't endorsed any artists with controversial themes have ever wanted the art world to be more political.
30	e	The collectors that have endorsed the artists with controversial themes have ever wanted the art world to be more political.
31	a	No journalists that have challenged the editors of prestigious newspapers have ever expected difficulty on the job market.
31	b	The journalists that have challenged no editors of prestigious newspapers have ever expected difficulty on the job market.
31	c	The journalists that haven't challenged the editors of prestigious newspapers have ever expected difficulty on the job market.
31	d	The journalists that haven't challenged any editors of prestigious newspapers have ever expected difficulty on the job market.
31	e	The journalists that have challenged the editors of prestigious newspapers have ever expected difficulty on the job market.
32	a	No teachers that have suspended the teenagers with behavioral problems have ever expected cooperation from the parents.
32	b	The teachers that have suspended no teenagers with behavioral problems have ever expected cooperation from the parents.
32	c	The teachers that haven't suspended the teenagers with behavioral problems have ever expected cooperation from the parents.
32	d	The teachers that haven't suspended any teenagers with behavioral problems have ever expected cooperation from the parents.
32	e	The teachers that have suspended the teenagers with behavioral problems have ever expected cooperation from the parents.
33	a	No tutors that have helped the students in challenging classes have ever suggested improvements in the student to teacher ratio.
33	b	The tutors that have helped no students in challenging classes have ever suggested improvements in the student to teacher ratio.
33	c	The tutors that haven't helped the students in challenging classes have ever suggested improvements in the student to teacher ratio.

33	d	The tutors that haven't helped any students in challenging classes have ever suggested improvements in the student to teacher ratio.
33	e	The tutors that have helped the students in challenging classes have ever suggested improvements in the student to teacher ratio.
34	a	No children that have respected the rules at community pools have ever caused trouble for the lifeguards.
34	b	The children that have respected no rules at community pools have ever caused trouble for the lifeguards.
34	c	The children that haven't respected the rules at community pools have ever caused trouble for the lifeguards.
34	d	The children that haven't respected any rules at community pools have ever caused trouble for the lifeguards.
34	e	The children that have respected the rules at community pools have ever caused trouble for the lifeguards.
35	a	No students that have completed the assignments during school hours have ever turned in their homework late.
35	b	The students that have completed no assignments during school hours have ever turned in their homework late.
35	c	The students that haven't completed the assignments during school hours have ever turned in their homework late.
35	d	The students that haven't completed any assignments during school hours have ever turned in their homework late.
35	e	The students that have completed the assignments during school hours have ever turned in their homework late.
36	a	No families that have visited the resorts during summer break have ever worried about the weather.
36	b	The families that have toured no resorts during summer break have ever worried about the weather.
36	c	The families that haven't toured the resorts during summer break have ever worried about the weather.
36	d	The families that haven't toured any resorts during summer break have ever worried about the weather.
36	e	The families that have toured the resorts during summer break have ever worried about the weather.
37	a	No swimmers that have done the warmups before morning practice have ever worried about strained muscles.
37	b	The swimmers that have done no warmups before morning practice have ever worried about strained muscles.
37	c	The swimmers that haven't done the warmups before morning practice have ever worried about strained muscles.

37	d	The swimmers that haven't done any warmups before morning practice have ever worried about strained muscles.
37	e	The swimmers that have done the warmups before morning practice have ever worried about strained muscles.
38	a	No actors that have memorized the lines before final auditions have ever given a compelling performance.
38	b	The actors that have memorized no lines before final auditions have ever given a compelling performance.
38	c	The actors that haven't memorized the lines before final auditions have ever given a compelling performance.
38	d	The actors that haven't memorized any lines before final auditions have ever given a compelling performance.
38	e	The actors that have memorized the lines before final auditions have ever given a compelling performance.
39	a	No graduates that have attended the rehearsals before the ceremony have ever been confused about the procession.
39	b	The graduates that have attended no rehearsals before the ceremony have ever been confused about the procession.
39	c	The graduates that haven't attended the rehearsals before the ceremony have ever been confused about the procession.
39	d	The graduates that haven't attended any rehearsals before the ceremony have ever been confused about the procession.
39	e	The graduates that have attended the rehearsals before the ceremony have ever been confused about the procession.
40	a	No teachers that have graded the exams before spring break have ever had students complain about an incomplete semester grade.
40	b	The teachers that have graded no exams before spring break have ever had students complain about an incomplete semester grade.
40	c	The teachers that haven't graded the exams before spring break have ever had students complain about an incomplete semester grade.
40	d	The teachers that haven't graded any exams before spring break have ever had students complain about an incomplete semester grade.
40	e	The teachers that have graded the exams before spring break have ever had students complain about an incomplete semester grade.

Table A.6: Full experimental stimuli for Experiment 6

A.7 Experiment 7

1	a	No authors that the critics have recommended in their reviews have ever received acknowledgment for a best-selling novel.
1	b	The authors that no critics have recommended in their reviews have ever received acknowledgment for a best-selling novel.
1	c	The authors that the critics have recommended in their reviews have ever received acknowledgment for a best-selling novel.
1	d	No authors that the critics have recommended have ever received acknowledgment for a best-selling novel.
1	e	The authors that no critics have recommended have ever received acknowledgment for a best-selling novel.
1	f	The authors that the critics have recommended have ever received acknowledgment for a best-selling novel.
2	a	No soldiers that the diplomats have supported in the trial have ever shown respect to the war victims.
2	b	The soldiers that no diplomats have supported in the trial have ever shown respect to the war victims.
2	c	The soldiers that the diplomats have supported in the trial have ever shown respect to the war victims.
2	d	No soldiers that the diplomats have supported have ever shown respect to the war victims.
2	e	The soldiers that no diplomats have supported have ever shown respect to the war victims.
2	f	The soldiers that the diplomats have supported have ever shown respect to the war victims.
3	a	No ambassadors that the diplomats have consulted about the treaty have ever seen brutality in the foreign war.
3	b	The ambassadors that no diplomats have consulted about the treaty have ever seen brutality in the foreign war.
3	c	The ambassadors that the diplomats have consulted about the treaty have ever seen brutality in the foreign war.
3	d	No ambassadors that the diplomats have consulted have ever seen brutality in the foreign war.
3	e	The ambassadors that no diplomats have consulted have ever seen brutality in the foreign war.
3	f	The ambassadors that the diplomats have consulted have ever seen brutality in the foreign war.
4	a	No professors that the students have challenged over low grades have ever wanted negativity in a class debate.
4	b	The professors that no students have challenged over low grades have ever wanted negativity in a class debate.
4	c	The professors that the students have challenged over low grades have ever wanted negativity in a class debate.

4	d	No professors that the students have challenged have ever wanted negativity in a class debate.
4	e	The professors that no students have challenged have ever wanted negativity in a class debate.
4	f	The professors that the students have challenged have ever wanted negativity in a class debate.
5	a	No customers that the salesmen have assisted in the outlet have ever expressed optimism for a full refund.
5	b	The customers that no salesmen have assisted in the outlet have ever expressed optimism for a full refund.
5	c	The customers that the salesmen have assisted in the outlet have ever expressed optimism for a full refund.
5	d	No customers that the salesmen have assisted have ever expressed optimism for a full refund.
5	e	The customers that no salesmen have assisted have ever expressed optimism for a full refund.
5	f	The customers that the salesmen have assisted have ever expressed optimism for a full refund.
6	a	No diplomats that the politicians have informed of recent policies have ever caused controversy in the liberal newspapers.
6	b	The diplomats that no politicians have informed of recent policies have ever caused controversy in the liberal newspapers.
6	c	The diplomats that the politicians have informed of recent policies have ever caused controversy in the liberal newspapers.
6	d	No diplomats that the politicians have informed have ever caused controversy in the liberal newspapers.
6	e	The diplomats that no politicians have informed have ever caused controversy in the liberal newspapers.
6	f	The diplomats that the politicians have informed have ever caused controversy in the liberal newspapers.
7	a	No maids that the housewives have thanked for their work have ever caused damage to the delicate clothing.
7	b	The maids that no housewives have thanked for their work have ever caused damage to the delicate clothing.
7	c	The maids that the housewives have thanked for their work have ever caused damage to the delicate clothing.
7	d	No maids that the housewives have thanked have ever caused damage to the delicate clothing.
7	e	The maids that no housewives have thanked have ever caused damage to the delicate clothing.
7	f	The maids that the housewives have thanked have ever caused damage to the delicate clothing.
8	a	No lawyers that the businessmen have hired for legal advice have ever received criticism for lost fraud trials.
8	b	The lawyers that no businessmen have hired for legal advice have ever received criticism for lost fraud trials.
8	c	The lawyers that the businessmen have hired for legal advice have ever received criticism for lost fraud trials.
8	d	No lawyers that the businessmen have hired have ever received criticism for lost fraud trials.
8	e	The lawyers that no businessmen have hired have ever received criticism for lost fraud trials.
8	f	The lawyers that the businessmen have hired have ever received criticism for lost fraud trials.
9	a	No students that the teachers have punished for bad behavior have ever expected friendliness from the strict principal.

9	b	The students that no teachers have punished for bad behavior have ever expected friendliness from the strict principal.
9	c	The students that the teachers have punished for bad behavior have ever expected friendliness from the strict principal.
9	d	No students that the teachers have punished have ever expected friendliness from the strict principal.
9	e	The students that no teachers have punished have ever expected friendliness from the strict principal.
9	f	The students that the teachers have punished have ever expected friendliness from the strict principal.
10	a	No babysitters that the children have disobeyed during an outing have ever expected gratitude from the disappointed parents.
10	b	The babysitters that no children have disobeyed during an outing have ever expected gratitude from the disappointed parents.
10	c	The babysitters that the children have disobeyed during an outing have ever expected gratitude from the disappointed parents.
10	d	No babysitters that the children have disobeyed have ever expected gratitude from the disappointed parents.
10	e	The babysitters that no children have disobeyed have ever expected gratitude from the disappointed parents.
10	f	The babysitters that the children have disobeyed have ever expected gratitude from the disappointed parents.
11	a	No actors that the producers have solicited for their films have ever experienced injuries from a dangerous stunt.
11	b	The actors that no producers have solicited for their films have ever experienced injuries from a dangerous stunt.
11	c	The actors that the producers have solicited for their films have ever experienced injuries from a dangerous stunt.
11	d	No actors that the producers have solicited have ever experienced injuries from a dangerous stunt.
11	e	The actors that no producers have solicited have ever experienced injuries from a dangerous stunt.
11	f	The actors that the producers have solicited have ever experienced injuries from a dangerous stunt.
12	a	No teachers that the parents have commended in PTA meetings have ever caused problems with the school administrators.
12	b	The teachers that no parents have commended in PTA meetings have ever caused problems with the school administrators.
12	c	The teachers that the parents have commended in PTA meetings have ever caused problems with the school administrators.

12	d	No teachers that the parents have commended have ever caused problems with the school administrators.
12	e	The teachers that no parents have commended have ever caused problems with the school administrators.
12	f	The teachers that the parents have commended have ever caused problems with the school administrators.
13	a	No students that the librarians have helped with book reports have ever made progress on the difficult assignment.
13	b	The students that no librarians have helped with book reports have ever made progress on the difficult assignment.
13	c	The students that the librarians have helped with book reports have ever made progress on the difficult assignment.
13	d	No students that the librarians have helped have ever made progress on the difficult assignment.
13	e	The students that no librarians have helped have ever made progress on the difficult assignment.
13	f	The students that the librarians have helped have ever made progress on the difficult assignment.
14	a	No nurses that the doctors have requested for the surgery have ever shown clumsiness in the operating room.
14	b	The nurses that no doctors have requested for the surgery have ever shown clumsiness in the operating room.
14	c	The nurses that the doctors have requested for the surgery have ever shown clumsiness in the operating room.
14	d	No nurses that the doctors have requested have ever shown clumsiness in the operating room.
14	e	The nurses that no doctors have requested have ever shown clumsiness in the operating room.
14	f	The nurses that the doctors have requested have ever shown clumsiness in the operating room.
15	a	No criminals that the policemen have caught in drug raids have ever felt satisfaction from a petty crime.
15	b	The criminals that no policemen have caught in drug raids have ever felt satisfaction from a petty crime.
15	c	The criminals that the policemen have caught in drug raids have ever felt satisfaction from a petty crime.
15	d	No criminals that the policemen have caught have ever felt satisfaction from a petty crime.
15	e	The criminals that no policemen have caught have ever felt satisfaction from a petty crime.
15	f	The criminals that the policemen have caught have ever felt satisfaction from a petty crime.
16	a	No employees that the managers have recommended for a raise have ever expressed frustration with the rude customers.
16	b	The employees that no managers have recommended for a raise have ever expressed frustration with the rude customers.

16	c	The employees that the managers have recommended for a raise have ever expressed frustration with the rude customers.
16	d	No employees that the managers have recommended have ever expressed frustration with the rude customers.
16	e	The employees that no managers have recommended have ever expressed frustration with the rude customers.
16	f	The employees that the managers have recommended have ever expressed frustration with the rude customers.
17	a	No accountants that the managers have blamed for company losses have ever seen rises in the quarterly profits.
17	b	The accountants that no managers have blamed for company losses have ever seen rises in the quarterly profits.
17	c	The accountants that the managers have blamed for company losses have ever seen rises in the quarterly profits.
17	d	No accountants that the managers have blamed have ever seen rises in the quarterly profits.
17	e	The accountants that no managers have blamed have ever seen rises in the quarterly profits.
17	f	The accountants that the managers have blamed have ever seen rises in the quarterly profits.
18	a	No candidates that the voters have supported during the election have ever shown friendliness to the rude journalists.
18	b	The candidates that no voters have supported during the election have ever shown friendliness to the rude journalists.
18	c	The candidates that the voters have supported during the election have ever shown friendliness to the rude journalists.
18	d	No candidates that the voters have supported have ever shown friendliness to the rude journalists.
18	e	The candidates that no voters have supported have ever shown friendliness to the rude journalists.
18	f	The candidates that the voters have supported have ever shown friendliness to the rude journalists.
19	a	No surgeons that the patients have consulted about the operation have ever expressed dissatisfaction with the hospital staff.
19	b	The surgeons that no patients have consulted about the operation have ever expressed dissatisfaction with the hospital staff.
19	c	The surgeons that the patients have consulted about the operation have ever expressed dissatisfaction with the hospital staff.

19	d	No surgeons that the patients have consulted have ever expressed dissatisfaction with the hospital staff.
19	e	The surgeons that no patients have consulted have ever expressed dissatisfaction with the hospital staff.
19	f	The surgeons that the patients have consulted have ever expressed dissatisfaction with the hospital staff.
20	a	No suspects that the witnesses have identified in photo lineups have ever shown nervousness in the court room.
20	b	The suspects that no witnesses have identified in photo lineups have ever shown nervousness in the court room.
20	c	The suspects that the witnesses have identified in photo lineups have ever shown nervousness in the court room.
20	d	No suspects that the witnesses have identified have ever shown nervousness in the court room.
20	e	The suspects that no witnesses have identified have ever shown nervousness in the court room.
20	f	The suspects that the witnesses have identified have ever shown nervousness in the court room.
21	a	No actresses that the moviegoers have praised for their performance have ever caused excitement at a film festival.
21	b	The actresses that no moviegoers have praised for their performance have ever caused excitement at a film festival.
21	c	The actresses that the moviegoers have praised for their performance have ever caused excitement at a film festival.
21	d	No actresses that the moviegoers have praised have ever caused excitement at a film festival.
21	e	The actresses that no moviegoers have praised have ever caused excitement at a film festival.
21	f	The actresses that the moviegoers have praised have ever caused excitement at a film festival.
22	a	No senators that the billionaires have supported with campaign donations have ever received hostility from online news media.
22	b	The senators that no billionaires have supported with campaign donations have ever received hostility from online news media.
22	c	The senators that the billionaires have supported with campaign donations have ever received hostility from online news media.
22	d	No senators that the billionaires have supported have ever received hostility from online news media.
22	e	The senators that no billionaires have supported have ever received hostility from online news media.
22	f	The senators that the billionaires have supported have ever received hostility from online news media.
23	a	No politicians that the journalists have endorsed in the media have ever earned trust from the rural communities.

23	b	The politicians that no journalists have endorsed in the media have ever earned trust from the rural communities.
23	c	The politicians that the journalists have endorsed in the media have ever earned trust from the rural communities.
23	d	No politicians that the journalists have endorsed have ever earned trust from the rural communities.
23	e	The politicians that no journalists have endorsed have ever earned trust from the rural communities.
23	f	The politicians that the journalists have endorsed have ever earned trust from the rural communities.
24	a	No teenagers that the parents have trusted with a car have ever expressed impatience with their rambunctious siblings.
24	b	The teenagers that no parents have trusted with a car have ever expressed impatience with their rambunctious siblings.
24	c	The teenagers that the parents have trusted with a car have ever expressed impatience with their rambunctious siblings.
24	d	No teenagers that the parents have trusted have ever expressed impatience with their rambunctious siblings.
24	e	The teenagers that no parents have trusted have ever expressed impatience with their rambunctious siblings.
24	f	The teenagers that the parents have trusted have ever expressed impatience with their rambunctious siblings.
25	a	No survivors that the medics have cured of their injuries have ever felt regret for their military service.
25	b	The survivors that no medics have cured of their injuries have ever felt regret for their military service.
25	c	The survivors that the medics have cured of their injuries have ever felt regret for their military service.
25	d	No survivors that the medics have cured have ever felt regret for their military service.
25	e	The survivors that no medics have cured have ever felt regret for their military service.
25	f	The survivors that the medics have cured have ever felt regret for their military service.
26	a	No players that the coaches have drafted for the team have ever felt nervousness before a championship game.
26	b	The players that no coaches have drafted for the team have ever felt nervousness before a championship game.
26	c	The players that the coaches have drafted for the team have ever felt nervousness before a championship game.
26	d	No players that the coaches have drafted have ever felt nervousness before a championship game.
26	e	The players that no coaches have drafted have ever felt nervousness before a championship game.
26	f	The players that the coaches have drafted have ever felt nervousness before a championship game.

27	a	No voters that the senators have courted at campaign rallies have ever caused controversy in a major election.
27	b	The voters that no senators have courted at campaign rallies have ever caused controversy in a major election.
27	c	The voters that the senators have courted at campaign rallies have ever caused controversy in a major election.
27	d	No voters that the senators have courted have ever caused controversy in a major election.
27	e	The voters that no senators have courted have ever caused controversy in a major election.
27	f	The voters that the senators have courted have ever caused controversy in a major election.
28	a	No professors that the students have visited during office hours have ever experienced tiredness after a long lecture.
28	b	The professors that no students have visited during office hours have ever experienced tiredness after a long lecture.
28	c	The professors that the students have visited during office hours have ever experienced tiredness after a long lecture.
28	d	No professors that the students have visited have ever experienced tiredness after a long lecture.
28	e	The professors that no students have visited have ever experienced tiredness after a long lecture.
28	f	The professors that the students have visited have ever experienced tiredness after a long lecture.
29	a	No actors that the judges have nominated for an award have ever experienced derision from the tabloid gossip.
29	b	The actors that no judges have nominated for an award have ever experienced derision from the tabloid gossip.
29	c	The actors that the judges have nominated for an award have ever experienced derision from the tabloid gossip.
29	d	No actors that the judges have nominated have ever experienced derision from the tabloid gossip.
29	e	The actors that no judges have nominated have ever experienced derision from the tabloid gossip.
29	f	The actors that the judges have nominated have ever experienced derision from the tabloid gossip.
30	a	No actresses that the directors have auditioned for the role have ever shown nervousness on a large stage.
30	b	The actresses that no directors have auditioned for the role have ever shown nervousness on a large stage.
30	c	The actresses that the directors have auditioned for the role have ever shown nervousness on a large stage.
30	d	No actresses that the directors have auditioned have ever shown nervousness on a large stage.
30	e	The actresses that no directors have auditioned have ever shown nervousness on a large stage.
30	f	The actresses that the directors have auditioned have ever shown nervousness on a large stage.

31	a	No champions that the competitors have defeated in important races have ever shown humility after a big win.
31	b	The champions that no competitors have defeated in important races have ever shown humility after a big win.
31	c	The champions that the competitors have defeated in important races have ever shown humility after a big win.
31	d	No champions that the competitors have defeated have ever shown humility after a big win.
31	e	The champions that no competitors have defeated have ever shown humility after a big win.
31	f	The champions that the competitors have defeated have ever shown humility after a big win.
32	a	No painters that the collectors have favored at prestigious exhibitions have ever depicted tranquility with very bright colors.
32	b	The painters that no collectors have favored at prestigious exhibitions have ever depicted tranquility with very bright colors.
32	c	The painters that the collectors have favored at prestigious exhibitions have ever depicted tranquility with very bright colors.
32	d	No painters that the collectors have favored have ever depicted tranquility with very bright colors.
32	e	The painters that no collectors have favored have ever depicted tranquility with very bright colors.
32	f	The painters that the collectors have favored have ever depicted tranquility with very bright colors.
33	a	No editors that the journalists have insulted in stressful meetings have ever had patience for a missed deadline.
33	b	The editors that no journalists have insulted in stressful meetings have ever had patience for a missed deadline.
33	c	The editors that the journalists have insulted in stressful meetings have ever had patience for a missed deadline.
33	d	No editors that the journalists have insulted have ever had patience for a missed deadline.
33	e	The editors that no journalists have insulted have ever had patience for a missed deadline.
33	f	The editors that the journalists have insulted have ever had patience for a missed deadline.
34	a	No teenagers that the teachers have scolded for their chattiness have ever experienced loneliness in the large class.
34	b	The teenagers that no teachers have scolded for their chattiness have ever experienced loneliness in the large class.

34	c	The teenagers that the teachers have scolded for their chattiness have ever experienced loneliness in the large class.
34	d	No teenagers that the teachers have scolded have ever experienced loneliness in the large class.
34	e	The teenagers that no teachers have scolded have ever experienced loneliness in the large class.
34	f	The teenagers that the teachers have scolded have ever experienced loneliness in the large class.
35	a	No students that the professors have tutored on the weekends have ever had trouble in a math class.
35	b	The students that no professors have tutored on the weekends have ever had trouble in a math class.
35	c	The students that the professors have tutored on the weekends have ever had trouble in a math class.
35	d	No students that the professors have tutored have ever had trouble in a math class.
35	e	The students that no professors have tutored have ever had trouble in a math class.
35	f	The students that the professors have tutored have ever had trouble in a math class.
36	a	No actors that the children have seen in family films have ever shown gore during a fight scene.
36	b	The actors that no children have seen in family films have ever shown gore during a fight scene.
36	c	The actors that the children have seen in family films have ever shown gore during a fight scene.
36	d	No actors that the children have seen have ever shown gore during a fight scene.
36	e	The actors that no children have seen have ever shown gore during a fight scene.
36	f	The actors that the children have seen have ever shown gore during a fight scene.

Table A.7: Full experimental stimuli for Experiment 7

A.8 Experiment 8

1	a	No surgeons that the patients trusted have ever prescribed experimental treatments.
1	b	The surgeons that no patients trusted have ever prescribed experimental treatments.
1	c	The surgeons that the patients trusted have ever prescribed experimental treatments.
1	d	No surgeons that the patients trusted to heal injuries have ever prescribed experimental treatments.
1	e	The surgeons that no patients trusted to heal injuries have ever prescribed experimental treatments.
1	f	The surgeons that the patients trusted to heal injuries have ever prescribed experimental treatments.
1	g	No surgeons that the patients trusted have healed any injuries with experimental treatments.
1	h	The surgeons that no patients trusted have healed any injuries with experimental treatments.
1	i	The surgeons that the patients trusted have healed any injuries with experimental treatments.
2	a	No singers that the choirs hired have ever lost their voices before a show.
2	b	The singers that no choirs hired have ever lost their voices before a show.

2	c	The singers that the choirs hired have ever lost their voices before a show.
2	d	No singers that the choirs hired to create harmonies have ever lost their voices before a show.
2	e	The singers that no choirs hired to create harmonies have ever lost their voices before a show.
2	f	The singers that the choirs hired to create harmonies have ever lost their voices before a show.
2	g	No singers that the choirs hired have created any harmonies without help from the director.
2	h	The singers that no choirs hired have created any harmonies without help from the director.
2	i	The singers that the choirs hired have created any harmonies without help from the director.
3	a	No dogs that the families adopted have ever ran away from the house.
3	b	The dogs that no families adopted have ever ran away from the house.
3	c	The dogs that the families adopted have ever ran away from the house.
3	d	No dogs that the families adopted to bring themselves joy have ever ran away from the house.
3	e	The dogs that no families adopted to bring themselves joy have ever ran away from the house.
3	f	The dogs that the families adopted to bring themselves joy have ever ran away from the house.
3	g	No dogs that the families adopted have brought any joy to the neighborhood kids.
3	h	The dogs that no families adopted have brought any joy to the neighborhood kids.
3	i	The dogs that the families adopted have brought any joy to the neighborhood kids.
4	a	No professors that the students chose have ever met their expectations.
4	b	The professors that no students chose have ever met their expectations.
4	c	The professors that the students chose have ever met their expectations.
4	d	No professors that the students chose to inspire motivation have ever met their expectations.
4	e	The professors that no students chose to inspire motivation have ever met their expectations.
4	f	The professors that the students chose to inspire motivation have ever met their expectations.
4	g	No professors that the students chose have inspired any motivation in the students' studies.
4	h	The professors that no students chose have inspired any motivation in the students' studies.
4	i	The professors that the students chose have inspired any motivation in the students' studies.
5	a	No watches that the businessmen wore have ever been bought at a department store.
5	b	The watches that no businessmen wore have ever been bought at a department store.
5	c	The watches that the businessmen wore have ever been bought at a department store.
5	d	No watches that the businessmen wore to signify wealth have ever been bought at a department store.
5	e	The watches that no businessmen wore to signify wealth have ever been bought at a department store.
5	f	The watches that the businessmen wore to signify wealth have ever been bought at a department store.
5	g	No watches that the businessmen wore have signified any wealth that encouraged investors.

5	h	The watches that no businessmen wore have signified any wealth that encouraged investors.
5	i	The watches that the businessmen wore have signified any wealth that encouraged investors.
6	a	No comments that the politicians published have ever improved the politician's ratings.
6	b	The comments that no politicians published have ever improved the politician's ratings.
6	c	The comments that the politicians published have ever improved the politician's ratings.
6	d	No comments that the politicians published to incite protests have ever improved the politician's ratings.
6	e	The comments that no politicians published to incite protests have ever improved the politician's ratings.
6	f	The comments that the politicians published to incite protests have ever improved the politician's ratings.
6	g	No comments that the politicians published have incited any protests against the other party.
6	h	The comments that no politicians published have incited any protests against the other party.
6	i	The comments that the politicians published have incited any protests against the other party.
7	a	No self-help books that the women bought will ever be based in science.
7	b	The self-help books that no women bought will ever be based in science.
7	c	The self-help books that the women bought will ever be based in science.
7	d	No self-help books that the women bought to teach them confidence will ever be based in science.
7	e	The self-help books that no women bought to teach them confidence will ever be based in science.
7	f	The self-help books that the women bought to teach them confidence will ever be based in science.
7	g	No self-help books that the women bought will teach any confidence boosting techniques.
7	h	The self-help books that no women bought will teach any confidence boosting techniques.
7	i	The self-help books that the women bought will teach any confidence boosting techniques.
8	a	No lawyers that the newspapers contacted will ever reveal their clients.
8	b	The lawyers that no newspapers contacted will ever reveal their clients.
8	c	The lawyers that the newspapers contacted will ever reveal their clients.
8	d	No lawyers that the newspapers contacted to give moral guidance will ever reveal their clients.
8	e	The lawyers that no newspapers contacted to give moral guidance will ever reveal their clients.
8	f	The lawyers that the newspapers contacted to give moral guidance will ever reveal their clients.
8	g	No lawyers that the newspapers contacted will give any moral guidance that doesn't follow the law.
8	h	The lawyers that no newspapers contacted will give any moral guidance that doesn't follow the law.
8	i	The lawyers that the newspapers contacted will give any moral guidance that doesn't follow the law.
9	a	No students that the teachers guided have ever been sent to the principal.
9	b	The students that no teachers guided have ever been sent to the principal.
9	c	The students that the teachers guided have ever been sent to the principal.

9	d	No students that the teachers guided to expect friendliness have ever been sent to the principal.
9	e	The students that no teachers guided to expect friendliness have ever been sent to the principal.
9	f	The students that the teachers guided to expect friendliness have ever been sent to the principal.
9	g	No students that the teachers guided have expected any friendliness from the strict principal.
9	h	The students that no teachers guided have expected any friendliness from the strict principal.
9	i	The students that the teachers guided have expected any friendliness from the strict principal.
10	a	No babysitters that the parents wanted have ever told their friends about the family.
10	b	The babysitters that no parents wanted have ever told their friends about the family.
10	c	The babysitters that the parents wanted have ever told their friends about the family.
10	d	No babysitters that the parents wanted to show gratitude have ever told their friends about the family.
10	e	The babysitters that no parents wanted to show gratitude have ever told their friends about the family.
10	f	The babysitters that the parents wanted to show gratitude have ever told their friends about the family.
10	g	No babysitters that the parents wanted have shown any gratitude for good recommendations.
10	h	The babysitters that no parents wanted have shown any gratitude for good recommendations.
10	i	The babysitters that the parents wanted have shown any gratitude for good recommendations.
11	a	No exercises that the trainers taught have ever increased muscle mass.
11	b	The exercises that no trainers taught have ever increased muscle mass.
11	c	The exercises that the trainers taught have ever increased muscle mass.
11	d	No exercises that the trainers taught to localize soreness have ever increased muscle mass.
11	e	The exercises that no trainers taught to localize soreness have ever increased muscle mass.
11	f	The exercises that the trainers taught to localize soreness have ever increased muscle mass.
11	g	No exercises that the trainers taught have localized any soreness in the athletes.
11	h	The exercises that no trainers taught have localized any soreness in the athletes.
11	i	The exercises that the trainers taught have localized any soreness in the athletes.
12	a	No coffees that the baristas recommended have ever been ordered by the regular customers.
12	b	The coffees that no baristas recommended have ever been ordered by the regular customers.
12	c	The coffees that the baristas recommended have ever been ordered by the regular customers.
12	d	No coffees that the baristas recommended to provide energy have ever been ordered by the regular customers.
12	e	The coffees that no baristas recommended to provide energy have ever been ordered by the regular customers.

12	f	The coffees that the baristas recommended to provide energy have ever been ordered by the regular customers.
12	g	No coffees that the baristas recommended have provided any energy to the regular customers.
12	h	The coffees that no baristas recommended have provided any energy to the regular customers.
12	i	The coffees that the baristas recommended have provided any energy to the regular customers.
13	a	No posters that the librarians put up have ever stayed up longer than a week.
13	b	The posters that no librarians put up have ever stayed up longer than a week.
13	c	The posters that the librarians put up have ever stayed up longer than a week.
13	d	No posters that the librarians put up to convey progress have ever stayed up longer than a week.
13	e	The posters that no librarians put up to convey progress have ever stayed up longer than a week.
13	f	The posters that the librarians put up to convey progress have ever stayed up longer than a week.
13	g	No posters that the librarians put up have conveyed any progress to the patrons who visited.
13	h	The posters that no librarians put up have conveyed any progress to the patrons who visited.
13	i	The posters that the librarians put up have conveyed any progress to the patrons who visited.
14	a	No nurses that the doctors instructed have ever been assigned to the night shift.
14	b	The nurses that no doctors instructed have ever been assigned to the night shift.
14	c	The nurses that the doctors instructed have ever been assigned to the night shift.
14	d	No nurses that the doctors instructed to show patience have ever been assigned to the night shift.
14	e	The nurses that no doctors instructed to show patience have ever been assigned to the night shift.
14	f	The nurses that the doctors instructed to show patience have ever been assigned to the night shift.
14	g	No nurses that the doctors instructed have shown any patience in the operating room.
14	h	The nurses that no doctors instructed have shown any patience in the operating room.
14	i	The nurses that the doctors instructed have shown any patience in the operating room.
15	a	No criminals that the policemen pressured have ever felt remorse for their crimes.
15	b	The criminals that no policemen pressured have ever felt remorse for their crimes.
15	c	The criminals that the policemen pressured have ever felt remorse for their crimes.
15	d	No criminals that the policemen pressured to admit guilt have ever felt remorse for their crimes.
15	e	The criminals that no policemen pressured to admit guilt have ever felt remorse for their crimes.
15	f	The criminals that the policemen pressured to admit guilt have ever felt remorse for their crimes.
15	g	No criminals that the policemen pressured have admitted any guilt for the petty crime.
15	h	The criminals that no policemen pressured have admitted any guilt for the petty crime.
15	i	The criminals that the policemen pressured have admitted any guilt for the petty crime.

16	a	No birds that the zookeepers bought have ever interested the children.
16	b	The birds that no zookeepers bought have ever interested the children.
16	c	The birds that the zookeepers bought have ever interested the children.
16	d	No birds that the zookeepers bought to improve business have ever interested the children.
16	e	The birds that no zookeepers bought to improve business have ever interested the children.
16	f	The birds that the zookeepers bought to improve business have ever interested the children.
16	g	No birds that the zookeepers bought have improved any business for the zoo.
16	h	The birds that no zookeepers bought have improved any business for the zoo.
16	i	The birds that the zookeepers bought have improved any business for the zoo.
17	a	No cats that the women purchased have ever lived longer than ten years.
17	b	The cats that no women purchased have ever lived longer than ten years.
17	c	The cats that the women purchased have ever lived longer than ten years.
17	d	No cats that the women purchased to provide company have ever lived longer than ten years.
17	e	The cats that no women purchased to provide company have ever lived longer than ten years.
17	f	The cats that the women purchased to provide company have ever lived longer than ten years.
17	g	No cats that the women purchased have provided any company during the evening.
17	h	The cats that no women purchased have provided any company during the evening.
17	i	The cats that the women purchased have provided any company during the evening.
18	a	No candidates that the voters supported have ever written controversial laws.
18	b	The candidates that no voters supported have ever written controversial laws.
18	c	The candidates that the voters supported have ever written controversial laws.
18	d	No candidates that the voters supported to bring peace have ever written controversial laws.
18	e	The candidates that no voters supported to bring peace have ever written controversial laws.
18	f	The candidates that the voters supported to bring peace have ever written controversial laws.
18	g	No candidates that the voters supported have brought any peace to the troubled country.
18	h	The candidates that no voters supported have brought any peace to the troubled country.
18	i	The candidates that the voters supported have brought any peace to the troubled country.
19	a	No models that the artists used have ever liked the paintings of them.
19	b	The models that no artists used have ever liked the paintings of them.
19	c	The models that the artists used have ever liked the paintings of them.
19	d	No models that the artists used to provide inspiration have ever liked the paintings of them.
19	e	The models that no artists used to provide inspiration have ever liked the paintings of them.

19	f	The models that the artists used to provide inspiration have ever liked the paintings of them.
19	g	No models that the artists used have provided any inspiration for new paintings.
19	h	The models that no artists used have provided any inspiration for new paintings.
19	i	The models that the artists used have provided any inspiration for new paintings.
20	a	No trees that the gardeners selected have ever grown taller than expected.
20	b	The trees that no gardeners selected have ever grown taller than expected.
20	c	The trees that the gardeners selected have ever grown taller than expected.
20	d	No trees that the gardeners selected to add eco friendliness have ever grown taller than expected.
20	e	The trees that no gardeners selected to add eco friendliness have ever grown taller than expected.
20	f	The trees that the gardeners selected to add eco friendliness have ever grown taller than expected.
20	g	No trees that the gardeners selected have added any eco friendliness to the gardens.
20	h	The trees that no gardeners selected have added any eco friendliness to the gardens.
20	i	The trees that the gardeners selected have added any eco friendliness to the gardens.
21	a	No cats that the landlords bought have ever used the litter boxes inside.
21	b	The cats that no landlords bought have ever used the litter boxes inside.
21	c	The cats that the landlords bought have ever used the litter boxes inside.
21	d	No cats that the landlords bought to protect against pests have ever used the litter boxes inside.
21	e	The cats that no landlords bought to protect against pests have ever used the litter boxes inside.
21	f	The cats that the landlords bought to protect against pests have ever used the litter boxes inside.
21	g	No cats that the landlords bought have protected against any pests outside the house.
21	h	The cats that no landlords bought have protected against any pests outside the house.
21	i	The cats that the landlords bought have protected against any pests outside the house.
22	a	No ambassadors that the government officials approached have ever returned to the country.
22	b	The ambassadors that no government officials approached have ever returned to the country.
22	c	The ambassadors that the government officials approached have ever returned to the country.
22	d	No ambassadors that the government officials approached to provoke hostility have ever returned to the country.
22	e	The ambassadors that no government officials approached to provoke hostility have ever returned to the country.
22	f	The ambassadors that the government officials approached to provoke hostility have ever returned to the country.

22	g	No ambassadors that the government officials approached have provoked any hostility from their governments abroad.
22	h	The ambassadors that no government officials approached have provoked any hostility from their governments abroad.
22	i	The ambassadors that the government officials approached have provoked any hostility from their governments abroad.
23	a	No politicians that the journalists endorsed have ever passed laws benefiting farmers.
23	b	The politicians that no journalists endorsed have ever passed laws benefiting farmers.
23	c	The politicians that the journalists endorsed have ever passed laws benefiting farmers.
23	d	No politicians that the journalists endorsed to earn trust from rural communities have ever passed laws benefiting farmers.
23	e	The politicians that no journalists endorsed to earn trust from rural communities have ever passed laws benefiting farmers.
23	f	The politicians that the journalists endorsed to earn trust from rural communities have ever passed laws benefiting farmers.
23	g	No politicians that the journalists endorsed have earned any trust from the rural communities.
23	h	The politicians that no journalists endorsed have earned any trust from the rural communities.
23	i	The politicians that the journalists endorsed have earned any trust from the rural communities.
24	a	No teenagers that the parents trusted have ever gone out of their way to help with chores.
24	b	The teenagers that no parents trusted have ever gone out of their way to help with chores.
24	c	The teenagers that the parents trusted have ever gone out of their way to help with chores.
24	d	No teenagers that the parents trusted to express kindness have ever gone out of their way to help with chores.
24	e	The teenagers that no parents trusted to express kindness have ever gone out of their way to help with chores.
24	f	The teenagers that the parents trusted to express kindness have ever gone out of their way to help with chores.
24	g	No teenagers that the parents trusted have expressed any kindness to a younger sibling.
24	h	The teenagers that no parents trusted have expressed any kindness to a younger sibling.
24	i	The teenagers that the parents trusted have expressed any kindness to a younger sibling.
25	a	No survivors that the medics encouraged have ever returned to thank their doctors.
25	b	The survivors that no medics encouraged have ever returned to thank their doctors.

25	c	The survivors that the medics encouraged have ever returned to thank their doctors.
25	d	No survivors that the medics encouraged to have bravery have ever returned to thank their doctors.
25	e	The survivors that no medics encouraged to have bravery have ever returned to thank their doctors.
25	f	The survivors that the medics encouraged to have bravery have ever returned to thank their doctors.
25	g	No survivors that the medics encouraged have had any bravery in the face of their long recoveries.
25	h	The survivors that no medics encouraged have had any bravery in the face of their long recoveries.
25	i	The survivors that the medics encouraged have had any bravery in the face of their long recoveries.
26	a	No football players that the coaches drafted have ever been injured in a big game.
26	b	The football players that no coaches drafted have ever been injured in a big game.
26	c	The football players that the coaches drafted have ever been injured in a big game.
26	d	No football players that the coaches drafted to add aggression to the team have ever been injured in a big game.
26	e	The football players that no coaches drafted to add aggression to the team have ever been injured in a big game.
26	f	The football players that the coaches drafted to add aggression to the team have ever been injured in a big game.
26	g	No football players that the coaches drafted have added any aggression to the team dynamics.
26	h	The football players that no coaches drafted have added any aggression to the team dynamics.
26	i	The football players that the coaches drafted have added any aggression to the team dynamics.
27	a	No rebels that the citizens trusted have ever been arrested for violence.
27	b	The rebels that no citizens trusted have ever been arrested for violence.
27	c	The rebels that the citizens trusted have ever been arrested for violence.
27	d	No rebels that the citizens trusted to cause chaos have ever been arrested for violence.
27	e	The rebels that no citizens trusted to cause chaos have ever been arrested for violence.
27	f	The rebels that the citizens trusted to cause chaos have ever been arrested for violence.
27	g	No rebels that the citizens trusted have caused any chaos at a public event.
27	h	The rebels that no citizens trusted have caused any chaos at a public event.
27	i	The rebels that the citizens trusted have caused any chaos at a public event.
28	a	No pencils that the teachers distributed have ever improved test scores.
28	b	The pencils that no teachers distributed have ever improved test scores.
28	c	The pencils that the teachers distributed have ever improved test scores.
28	d	No pencils that the teachers distributed to bring luck on the exam have ever improved test scores.

28	e	The pencils that no teachers distributed to bring luck on the exam have ever improved test scores.
28	f	The pencils that the teachers distributed to bring luck on the exam have ever improved test scores.
28	g	No pencils that the teachers distributed have brought any luck on the final exam.
28	h	The pencils that no teachers distributed have brought any luck on the final exam.
28	i	The pencils that the teachers distributed have brought any luck on the final exam.
29	a	No plants that the botanists studied have ever been native to North America.
29	b	The plants that no botanists studied have ever been native to North America.
29	c	The plants that the botanists studied have ever been native to North America.
29	d	No plants that the botanists studied to increase knowledge have ever been native to North America.
29	e	The plants that no botanists studied to increase knowledge have ever been native to North America.
29	f	The plants that the botanists studied to increase knowledge have ever been native to North America.
29	g	No plants that the botanists studied have increased any knowledge about the soil in North America.
29	h	The plants that no botanists studied have increased any knowledge about the soil in North America.
29	i	The plants that the botanists studied have increased any knowledge about the soil in North America.
30	a	No actresses that the directors auditioned have ever been cast in the play.
30	b	The actresses that no directors auditioned have ever been cast in the play.
30	c	The actresses that the directors auditioned have ever been cast in the play.
30	d	No actresses that the directors auditioned to depict elegance have ever been cast in the play.
30	e	The actresses that no directors auditioned to depict elegance have ever been cast in the play.
30	f	The actresses that the directors auditioned to depict elegance have ever been cast in the play.
30	g	No actresses that the directors auditioned have depicted any elegance in their auditions.
30	h	The actresses that no directors auditioned have depicted any elegance in their auditions.
30	i	The actresses that the directors auditioned have depicted any elegance in their auditions.
31	a	No fighters that the competitors defeated have ever talked to the press after the fight.
31	b	The fighters that no competitors defeated have ever talked to the press after the fight.
31	c	The fighters that the competitors defeated have ever talked to the press after the fight.
31	d	No fighters that the competitors defeated to assert dominance have ever talked to the press after the fight.
31	e	The fighters that no competitors defeated to assert dominance have ever talked to the press after the fight.
31	f	The fighters that the competitors defeated to assert dominance have ever talked to the press after the fight.
31	g	No fighters that the competitors defeated have asserted any dominance during a big game.
31	h	The fighters that no competitors defeated have asserted any dominance during a big game.
31	i	The fighters that the competitors defeated have asserted any dominance during a big game.

32	a	No paintings that the museum curators preferred have ever been voted best in show by the public.
32	b	The paintings that no museum curators preferred have ever been voted best in show by the public.
32	c	The paintings that the museum curators preferred have ever been voted best in show by the public.
32	d	No paintings that the museum curators preferred to depict tranquility have ever been voted best in show by the public.
32	e	The paintings that no museum curators preferred to depict tranquility have ever been voted best in show by the public.
32	f	The paintings that the museum curators preferred to depict tranquility have ever been voted best in show by the public.
32	g	No paintings that the museum curators preferred have depicted any tranquility using very bright colors.
32	h	The paintings that no museum curators preferred have depicted any tranquility using very bright colors.
32	i	The paintings that the museum curators preferred have depicted any tranquility using very bright colors.
33	a	No couches that the couples bought have ever cost less than \$100.
33	b	The couches that no couples bought have ever cost less than \$100.
33	c	The couches that the couples bought have ever cost less than \$100.
33	d	No couches that the couples bought to bring themselves comfort have ever cost less than \$100.
33	e	The couches that no couples bought to bring themselves comfort have ever cost less than \$100.
33	f	The couches that the couples bought to bring themselves comfort have ever cost less than \$100.
33	g	No couches that the couples bought have brought any comfort after a long day at work.
33	h	The couches that no couples bought have brought any comfort after a long day at work.
33	i	The couches that the couples bought have brought any comfort after a long day at work.
34	a	No service dogs that the breeders trained have ever been given to veterans.
34	b	The service dogs that no breeders trained have ever been given to veterans.
34	c	The service dogs that the breeders trained have ever been given to veterans.
34	d	No service dogs that the breeders trained to prevent loneliness have ever been given to veterans.
34	e	The service dogs that no breeders trained to prevent loneliness have ever been given to veterans.
34	f	The service dogs that the breeders trained to prevent loneliness have ever been given to veterans.
34	g	No service dogs that the breeders trained have prevented any loneliness for veterans.
34	h	The service dogs that no breeders trained have prevented any loneliness for veterans.
34	i	The service dogs that the breeders trained have prevented any loneliness for veterans.
35	a	No students that the professors would tutor will ever have trouble in a math class.
35	b	The students that no professors would tutor will ever have trouble in a math class.

35	c	The students that the professors would tutor will ever have trouble in a math class.
35	d	No students that the professors would tutor to increase understanding will ever have trouble in a math class.
35	e	The students that no professors would tutor to increase understanding will ever have trouble in a math class.
35	f	The students that the professors would tutor to increase understanding will ever have trouble in a math class.
35	g	No students that the professors would tutor will increase any understanding of the subject matter.
35	h	The students that no professors would tutor will increase any understanding of the subject matter.
35	i	The students that the professors would tutor will increase any understanding of the subject matter.
36	a	No movies that the children could watch have ever shown gore during a fight scene.
36	b	The movies that no children could watch have ever shown gore during a fight scene.
36	c	The movies that the children could watch have ever shown gore during a fight scene.
36	d	No movies that the children could watch to provide entertainment have ever shown gore during a fight scene.
36	e	The movies that no children could watch to provide entertainment have ever shown gore during a fight scene.
36	f	The movies that the children could watch to provide entertainment have ever shown gore during a fight scene.
36	g	No movies that the children could watch have provided any entertainment on the rainy day.
36	h	The movies that no children could watch have provided any entertainment on the rainy day.
36	i	The movies that the children could watch have provided any entertainment on the rainy day.

Table A.8: Full experimental stimuli for Experiment 8

A.9 Experiment 9

1	a	No surgeons that the patients trusted have ever prescribed experimental treatments.
1	b	The surgeons that no patients trusted have ever prescribed experimental treatments.
1	c	The surgeons that the patients trusted have ever prescribed experimental treatments
1	d	No surgeons that the patients trusted to heal injuries have ever prescribed experimental treatments.
1	e	The surgeons that no patients trusted to heal injuries have ever prescribed experimental treatments.
1	f	The surgeons that the patients trusted to heal injuries have ever prescribed experimental treatments.

1	g	No surgeons that the patients trusted have healed any injuries with experimental treatments.
1	h	The surgeons that no patients trusted have healed any injuries with experimental treatments.
1	i	The surgeons that the patients trused have healed any injuries with experimental treatments.
2	a	No singers that the choirs hired have ever lost their voices before a show.
2	b	The singers that no choirs hired have ever lost their voices before a show.
2	c	The singers that the choirs hired have ever lost their voices before a show.
2	d	No singers that the choirs hired to create harmonies have ever lost their voices before a show.
2	e	The singers that no choirs hired to create harmonies have ever lost their voices before a show.
2	f	The singers that the choirs hired to create harmonies have ever lost their voices before a show.
2	g	No singers that the choirs hired have created any harmonies without help from the director.
2	h	The singers that no choirs hired have created any harmonies without help from the director.
2	i	The singers that the choirs hired have created any harmonies without help from the director.
3	a	No dogs that the families adopted have ever ran away from the house.
3	b	The dogs that no families adopted have ever ran away from the house.
3	c	The dogs that the families adopted have ever ran away from the house.
3	d	No dogs that the families adopted to bring themselves joy have ever ran away from the house.
3	e	The dogs that no families adopted to bring themselves joy have ever ran away from the house.
3	f	The dogs that the families adopted to bring themselves joy have ever ran away from the house.
3	g	No dogs that the families adopted have brought any joy to the neighborhood kids.
3	h	The dogs that no families adopted have brought any joy to the neighborhood kids.
3	i	The dogs that the families adopted have brought any joy to the neighborhood kids.
4	a	No professors that the students chose have ever met their expectations.
4	b	The professors that no students chose have ever met their expectations.
4	c	The professors that the students chose have ever met their expectations.
4	d	No professors that the students chose to inspire motivation have ever met their expectations.
4	e	The professors that no students chose to inspire motivation have ever met their expectations.
4	f	The professors that the students chose to inspire motivation have ever met their expectations.
4	g	No professors that the students chose have inspired any motivation in the students' studies.
4	h	The professors that no students chose have inspired any motivation in the students' studies.
4	i	The professors that the students chose have inspired any motivation in the students' studies.
5	a	No watches that the businessmen wore have ever been bought at a department store.
5	b	The watches that no businessmen wore have ever been bought at a department store.

5	c	The watches that the businessmen wore have ever been bought at a department store.
5	d	No watches that the businessmen wore to signify wealth have ever been bought at a department store.
5	e	The watches that no businessmen wore to signify wealth have ever been bought at a department store.
5	f	The watches that the businessmen wore to signify wealth have ever been bought at a department store.
5	g	No watches that the businessmen wore have signified any wealth that encouraged investors.
5	h	The watches that no businessmen wore have signified any wealth that encouraged investors.
5	i	The watches that the businessmen wore have signified any wealth that encouraged investors.
6	a	No comments that the politicians published have ever improved the politician's ratings.
6	b	The comments that no politicians published have ever improved the politician's ratings.
6	c	The comments that the politicians published have ever improved the politician's ratings.
6	d	No comments that the politicians published to incite protests have ever improved the politician's ratings.
6	e	The comments that no politicians published to incite protests have ever improved the politician's ratings.
6	f	The comments that the politicians published to incite protests have ever improved the politician's ratings.
6	g	No comments that the politicians published have incited any protests against the other party.
6	h	The comments that no politicians published have incited any protests against the other party.
6	i	The comments that the politicians published have incited any protests against the other party.
7	a	No self-help books that the women bought will ever be based in science.
7	b	The self-help books that no women bought will ever be based in science.
7	c	The self-help books that the women bought will ever be based in science.
7	d	No self-help books that the women bought to teach them confidence will ever be based in science.
7	e	The self-help books that no women bought to teach them confidence will ever be based in science.
7	f	The self-help books that the women bought to teach them confidence will ever be based in science.
7	g	No self-help books that the women bought will teach any confidence boosting techniques.
7	h	The self-help books that no women bought will teach any confidence boosting techniques.
7	i	The self-help books that the women bought will teach any confidence boosting techniques.
8	a	No lawyers that the newspapers contacted will ever reveal their clients.
8	b	The lawyers that no newspapers contacted will ever reveal their clients.
8	c	The lawyers that the newspapers contacted will ever reveal their clients.
8	d	No lawyers that the newspapers contacted to give moral guidance will ever reveal their clients.
8	e	The lawyers that no newspapers contacted to give moral guidance will ever reveal their clients.
8	f	The lawyers that the newspapers contacted to give moral guidance will ever reveal their clients.
8	g	No lawyers that the newspapers contacted will give any moral guidance that doesn't follow the law.

8	h	The lawyers that no newspapers contacted will give any moral guidance that doesn't follow the law.
8	i	The lawyers that the newspapers contacted will give any moral guidance that doesn't follow the law.
9	a	No students that the teachers guided have ever been sent to the principal.
9	b	The students that no teachers guided have ever been sent to the principal.
9	c	The students that the teachers guided have ever been sent to the principal.
9	d	No students that the teachers guided to expect friendliness have ever been sent to the principal.
9	e	The students that no teachers guided to expect friendliness have ever been sent to the principal.
9	f	The students that the teachers guided to expect friendliness have ever been sent to the principal.
9	g	No students that the teachers guided have expected any friendliness from the strict principal.
9	h	The students that no teachers guided have expected any friendliness from the strict principal.
9	i	The students that the teachers guided have expected any friendliness from the strict principal.
10	a	No babysitters that the parents wanted have ever told their friends about the family.
10	b	The babysitters that no parents wanted have ever told their friends about the family.
10	c	The babysitters that the parents wanted have ever told their friends about the family.
10	d	No babysitters that the parents wanted to show gratitude have ever told their friends about the family.
10	e	The babysitters that no parents wanted to show gratitude have ever told their friends about the family.
10	f	The babysitters that the parents wanted to show gratitude have ever told their friends about the family.
10	g	No babysitters that the parents wanted have shown any gratitude for good recommendations.
10	h	The babysitters that no parents wanted have shown any gratitude for good recommendations.
10	i	The babysitters that the parents wanted have shown any gratitude for good recommendations.
11	a	No exercises that the trainers taught have ever increased muscle mass.
11	b	The exercises that no trainers taught have ever increased muscle mass.
11	c	The exercises that the trainers taught have ever increased muscle mass.
11	d	No exercises that the trainers taught to localize soreness have ever increased muscle mass.
11	e	The exercises that no trainers taught to localize soreness have ever increased muscle mass.
11	f	The exercises that the trainers taught to localize soreness have ever increased muscle mass.
11	g	No exercises that the trainers taught have localized any soreness in the athletes.
11	h	The exercises that no trainers taught have localized any soreness in the athletes.
11	i	The exercises that the trainers taught have localized any soreness in the athletes.
12	a	No coffees that the baristas recommended have ever been ordered by the regular customers.
12	b	The coffees that no baristas recommended have ever been ordered by the regular customers.
12	c	The coffees that the baristas recommended have ever been ordered by the regular customers.

12	d	No coffees that the baristas recommended to provide energy have ever been ordered by the regular customers.
12	e	The coffees that no baristas recommended to provide energy have ever been ordered by the regular customers.
12	f	The coffees that the baristas recommended to provide energy have ever been ordered by the regular customers.
12	g	No coffees that the baristas recommended have provided any energy to the regular customers.
12	h	The coffees that no baristas recommended have provided any energy to the regular customers.
12	i	The coffees that the baristas recommended have provided any energy to the regular customers.
13	a	No posters that the librarians put up have ever stayed up longer than a week.
13	b	The posters that no librarians put up have ever stayed up longer than a week.
13	c	The posters that the librarians put up have ever stayed up longer than a week.
13	d	No posters that the librarians put up to convey progress have ever stayed up longer than a week.
13	e	The posters that no librarians put up to convey progress have ever stayed up longer than a week.
13	f	The posters that the librarians put up to convey progress have ever stayed up longer than a week.
13	g	No posters that the librarians put up have conveyed any progress to the patrons who visited.
13	h	The posters that no librarians put up have conveyed any progress to the patrons who visited.
13	i	The posters that the librarians put up have conveyed any progress to the patrons who visited.
14	a	No nurses that the doctors instructed have ever been assigned to the night shift.
14	b	The nurses that no doctors instructed have ever been assigned to the night shift.
14	c	The nurses that the doctors instructed have ever been assigned to the night shift.
14	d	No nurses that the doctors instructed to show patience have ever been assigned to the night shift.
14	e	The nurses that no doctors instructed to show patience have ever been assigned to the night shift.
14	f	The nurses that the doctors instructed to show patience have ever been assigned to the night shift.
14	g	No nurses that the doctors instructed have shown any patience in the operating room.
14	h	The nurses that no doctors instructed have shown any patience in the operating room.
14	i	The nurses that the doctors instructed have shown any patience in the operating room.
15	a	No criminals that the policemen pressured have ever felt remorse for their crimes.
15	b	The criminals that no policemen pressured have ever felt remorse for their crimes.
15	c	The criminals that the policemen pressured have ever felt remorse for their crimes.
15	d	No criminals that the policemen pressured to admit guilt have ever felt remorse for their crimes.
15	e	The criminals that no policemen pressured to admit guilt have ever felt remorse for their crimes.

15	f	The criminals that the policemen pressured to admit guilt have ever felt remorse for their crimes.
15	g	No criminals that the policemen pressured have admitted any guilt for the petty crime.
15	h	The criminals that no policemen pressured have admitted any guilt for the petty crime.
15	i	The criminals that the policemen pressured have admitted any guilt for the petty crime.
16	a	No birds that the zookeepers bought have ever interested the children.
16	b	The birds that no zookeepers bought have ever interested the children.
16	c	The birds that the zookeepers bought have ever interested the children.
16	d	No birds that the zookeepers bought to improve business have ever interested the children.
16	e	The birds that no zookeepers bought to improve business have ever interested the children.
16	f	The birds that the zookeepers bought to improve business have ever interested the children.
16	g	No birds that the zookeepers bought have improved any business for the zoo.
16	h	The birds that no zookeepers bought have improved any business for the zoo.
16	i	The birds that the zookeepers bought have improved any business for the zoo.
17	a	No cats that the women purchased have ever lived longer than ten years.
17	b	The cats that no women purchased have ever lived longer than ten years.
17	c	The cats that the women purchased have ever lived longer than ten years.
17	d	No cats that the women purchased to provide company have ever lived longer than ten years.
17	e	The cats that no women purchased to provide company have ever lived longer than ten years.
17	f	The cats that the women purchased to provide company have ever lived longer than ten years.
17	g	No cats that the women purchased have provided any company during the evening.
17	h	The cats that no women purchased have provided any company during the evening.
17	i	The cats that the women purchased have provided any company during the evening.
18	a	No candidates that the voters supported have ever written controversial laws.
18	b	The candidates that no voters supported have ever written controversial laws.
18	c	The candidates that the voters supported have ever written controversial laws.
18	d	No candidates that the voters supported to bring peace have ever written controversial laws.
18	e	The candidates that no voters supported to bring peace have ever written controversial laws.
18	f	The candidates that the voters supported to bring peace have ever written controversial laws.
18	g	No candidates that the voters supported have brought any peace to the troubled country.
18	h	The candidates that no voters supported have brought any peace to the troubled country.
18	i	The candidates that the voters supported have brought any peace to the troubled country.
19	a	No models that the artists used have ever liked the paintings of them.

19	b	The models that no artists used have ever liked the paintings of them.
19	c	The models that the artists used have ever liked the paintings of them.
19	d	No models that the artists used to provide inspiration have ever liked the paintings of them.
19	e	The models that no artists used to provide inspiration have ever liked the paintings of them.
19	f	The models that the artists used to provide inspiration have ever liked the paintings of them.
19	g	No models that the artists used have provided any inspiration for new paintings.
19	h	The models that no artists used have provided any inspiration for new paintings.
19	i	The models that the artists used have provided any inspiration for new paintings.
20	a	No trees that the gardeners selected have ever grown taller than expected.
20	b	The trees that no gardeners selected have ever grown taller than expected.
20	c	The trees that the gardeners selected have ever grown taller than expected.
20	d	No trees that the gardeners selected to add eco friendliness have ever grown taller than expected.
20	e	The trees that no gardeners selected to add eco friendliness have ever grown taller than expected.
20	f	The trees that the gardeners selected to add eco friendliness have ever grown taller than expected.
20	g	No trees that the gardeners selected have added any eco friendliness to the gardens.
20	h	The trees that no gardeners selected have added any eco friendliness to the gardens.
20	i	The trees that the gardeners selected have added any eco friendliness to the gardens.
21	a	No cats that the landlords bought have ever used the litter boxes inside.
21	b	The cats that no landlords bought have ever used the litter boxes inside.
21	c	The cats that the landlords bought have ever used the litter boxes inside.
21	d	No cats that the landlords bought to protect against pests have ever used the litter boxes inside.
21	e	The cats that no landlords bought to protect against pests have ever used the litter boxes inside.
21	f	The cats that the landlords bought to protect against pests have ever used the litter boxes inside.
21	g	No cats that the landlords bought have protected against any pests outside the house.
21	h	The cats that no landlords bought have protected against any pests outside the house.
21	i	The cats that the landlords bought have protected against any pests outside the house.
22	a	No ambassadors that the government officials approached have ever returned to the country.
22	b	The ambassadors that no government officials approached have ever returned to the country.
22	c	The ambassadors that the government officials approached have ever returned to the country.
22	d	No ambassadors that the government officials approached to provoke hostility have ever returned to the country.

22	e	The ambassadors that no government officials approached to provoke hostility have ever returned to the country.
22	f	The ambassadors that the government officials approached to provoke hostility have ever returned to the country.
22	g	No ambassadors that the government officials approached have provoked any hostility from their governments abroad.
22	h	The ambassadors that no government officials approached have provoked any hostility from their governments abroad.
22	i	The ambassadors that the government officials approached have provoked any hostility from their governments abroad.
23	a	No politicians that the journalists endorsed have ever passed laws benefiting farmers.
23	b	The politicians that no journalists endorsed have ever passed laws benefiting farmers.
23	c	The politicians that the journalists endorsed have ever passed laws benefiting farmers.
23	d	No politicians that the journalists endorsed to earn trust from rural communities have ever passed laws benefiting farmers.
23	e	The politicians that no journalists endorsed to earn trust from rural communities have ever passed laws benefiting farmers.
23	f	The politicians that the journalists endorsed to earn trust from rural communities have ever passed laws benefiting farmers.
23	g	No politicians that the journalists endorsed have earned any trust from the rural communities.
23	h	The politicians that no journalists endorsed have earned any trust from the rural communities.
23	i	The politicians that the journalists endorsed have earned any trust from the rural communities.
24	a	No teenagers that the parents trusted have ever gone out of their way to help with chores.
24	b	The teenagers that no parents trusted have ever gone out of their way to help with chores.
24	c	The teenagers that the parents trusted have ever gone out of their way to help with chores.
24	d	No teenagers that the parents trusted to express kindness have ever gone out of their way to help with chores.
24	e	The teenagers that no parents trusted to express kindness have ever gone out of their way to help with chores.
24	f	The teenagers that the parents trusted to express kindness have ever gone out of their way to help with chores.
24	g	No teenagers that the parents trusted have expressed any kindness to a younger sibling.

24	h	The teenagers that no parents trusted have expressed any kindness to a younger sibling.
24	i	The teenagers that the parents trusted have expressed any kindness to a younger sibling.
25	a	No survivors that the medics encouraged have ever returned to thank their doctors.
25	b	The survivors that no medics encouraged have ever returned to thank their doctors.
25	c	The survivors that the medics encouraged have ever returned to thank their doctors.
25	d	No survivors that the medics encouraged to have bravery have ever returned to thank their doctors.
25	e	The survivors that no medics encouraged to have bravery have ever returned to thank their doctors.
25	f	The survivors that the medics encouraged to have bravery have ever returned to thank their doctors.
25	g	No survivors that the medics encouraged have had any bravery in the face of their long recoveries.
25	h	The survivors that no medics encouraged have had any bravery in the face of their long recoveries.
25	i	The survivors that the medics encouraged have had any bravery in the face of their long recoveries.
26	a	No football players that the coaches drafted have ever been injured in a big game.
26	b	The football players that no coaches drafted have ever been injured in a big game.
26	c	The football players that the coaches drafted have ever been injured in a big game.
26	d	No football players that the coaches drafted to add aggression to the team have ever been injured in a big game.
26	e	The football players that no coaches drafted to add aggression to the team have ever been injured in a big game.
26	f	The football players that the coaches drafted to add aggression to the team have ever been injured in a big game.
26	g	No football players that the coaches drafted have added any aggression to the team dynamics.
26	h	The football players that no coaches drafted have added any aggression to the team dynamics.
26	i	The football players that the coaches drafted have added any aggression to the team dynamics.
27	a	No rebels that the citizens trusted have ever been arrested for violence.
27	b	The rebels that no citizens trusted have ever been arrested for violence.
27	c	The rebels that the citizens trusted have ever been arrested for violence.
27	d	No rebels that the citizens trusted to cause chaos have ever been arrested for violence.
27	e	The rebels that no citizens trusted to cause chaos have ever been arrested for violence.
27	f	The rebels that the citizens trusted to cause chaos have ever been arrested for violence.
27	g	No rebels that the citizens trusted have caused any chaos at a public event.
27	h	The rebels that no citizens trusted have caused any chaos at a public event.
27	i	The rebels that the citizens trusted have caused any chaos at a public event.

28	a	No pencils that the teachers distributed have ever improved test scores.
28	b	The pencils that no teachers distributed have ever improved test scores.
28	c	The pencils that the teachers distributed have ever improved test scores.
28	d	No pencils that the teachers distributed to bring luck on the exam have ever improved test scores.
28	e	The pencils that no teachers distributed to bring luck on the exam have ever improved test scores.
28	f	The pencils that the teachers distributed to bring luck on the exam have ever improved test scores.
28	g	No pencils that the teachers distributed have brought any luck on the final exam.
28	h	The pencils that no teachers distributed have brought any luck on the final exam.
28	i	The pencils that the teachers distributed have brought any luck on the final exam.
29	a	No plants that the botanists studied have ever been native to North America.
29	b	The plants that no botanists studied have ever been native to North America.
29	c	The plants that the botanists studied have ever been native to North America.
29	d	No plants that the botanists studied to increase knowledge have ever been native to North America.
29	e	The plants that no botanists studied to increase knowledge have ever been native to North America.
29	f	The plants that the botanists studied to increase knowledge have ever been native to North America.
29	g	No plants that the botanists studied have increased any knowledge about the soil in North America.
29	h	The plants that no botanists studied have increased any knowledge about the soil in North America.
29	i	The plants that the botanists studied have increased any knowledge about the soil in North America.
30	a	No actresses that the directors auditioned have ever been cast in the play.
30	b	The actresses that no directors auditioned have ever been cast in the play.
30	c	The actresses that the directors auditioned have ever been cast in the play.
30	d	No actresses that the directors auditioned to depict elegance have ever been cast in the play.
30	e	The actresses that no directors auditioned to depict elegance have ever been cast in the play.
30	f	The actresses that the directors auditioned to depict elegance have ever been cast in the play.
30	g	No actresses that the directors auditioned have depicted any elegance in their auditions.
30	h	The actresses that no directors auditioned have depicted any elegance in their auditions.
30	i	The actresses that the directors auditioned have depicted any elegance in their auditions.
31	a	No fighters that the competitors defeated have ever talked to the press after the fight.
31	b	The fighters that no competitors defeated have ever talked to the press after the fight.
31	c	The fighters that the competitors defeated have ever talked to the press after the fight.
31	d	No fighters that the competitors defeated to assert dominance have ever talked to the press after the fight.
31	e	The fighters that no competitors defeated to assert dominance have ever talked to the press after the fight.

31	f	The fighters that the competitors defeated to assert dominance have ever talked to the press after the fight.
31	g	No fighters that the competitors defeated have asserted any dominance during a big game.
31	h	The fighters that no competitors defeated have asserted any dominance during a big game.
31	i	The fighters that the competitors defeated have asserted any dominance during a big game.
32	a	No paintings that the museum curators preferred have ever been voted best in show by the public.
32	b	The paintings that no museum curators preferred have ever been voted best in show by the public.
32	c	The paintings that the museum curators preferred have ever been voted best in show by the public.
32	d	No paintings that the museum curators preferred to depict tranquility have ever been voted best in show by the public.
32	e	The paintings that no museum curators preferred to depict tranquility have ever been voted best in show by the public.
32	f	The paintings that the museum curators preferred to depict tranquility have ever been voted best in show by the public.
32	g	No paintings that the museum curators preferred have depicted any tranquility using very bright colors.
32	h	The paintings that no museum curators preferred have depicted any tranquility using very bright colors.
32	i	The paintings that the museum curators preferred have depicted any tranquility using very bright colors.
33	a	No couches that the couples bought have ever cost less than \$100.
33	b	The couches that no couples bought have ever cost less than \$100.
33	c	The couches that the couples bought have ever cost less than \$100.
33	d	No couches that the couples bought to bring themselves comfort have ever cost less than \$100.
33	e	The couches that no couples bought to bring themselves comfort have ever cost less than \$100.
33	f	The couches that the couples bought to bring themselves comfort have ever cost less than \$100.
33	g	No couches that the couples bought have brought any comfort after a long day at work.
33	h	The couches that no couples bought have brought any comfort after a long day at work.
33	i	The couches that the couples bought have brought any comfort after a long day at work.
34	a	No service dogs that the breeders trained have ever been given to veterans.
34	b	The service dogs that no breeders trained have ever been given to veterans.
34	c	The service dogs that the breeders trained have ever been given to veterans.
34	d	No service dogs that the breeders trained to prevent loneliness have ever been given to veterans.
34	e	The service dogs that no breeders trained to prevent loneliness have ever been given to veterans.
34	f	The service dogs that the breeders trained to prevent loneliness have ever been given to veterans.
34	g	No service dogs that the breeders trained have prevented any loneliness for veterans.

34	h	The service dogs that no breeders trained have prevented any loneliness for veterans.
34	i	The service dogs that the breeders trained have prevented any loneliness for veterans.
35	a	No students that the professors would tutor will ever have trouble in a math class.
35	b	The students that no professors would tutor will ever have trouble in a math class.
35	c	The students that the professors would tutor will ever have trouble in a math class.
35	d	No students that the professors would tutor to increase understanding will ever have trouble in a math class.
35	e	The students that no professors would tutor to increase understanding will ever have trouble in a math class.
35	f	The students that the professors would tutor to increase understanding will ever have trouble in a math class.
35	g	No students that the professors would tutor will increase any understanding of the subject matter.
35	h	The students that no professors would tutor will increase any understanding of the subject matter.
35	i	The students that the professors would tutor will increase any understanding of the subject matter.
36	a	No movies that the children could watch have ever shown gore during a fight scene.
36	b	The movies that no children could watch have ever shown gore during a fight scene.
36	c	The movies that the children could watch have ever shown gore during a fight scene.
36	d	No movies that the children could watch to provide entertainment have ever shown gore during a fight scene.
36	e	The movies that no children could watch to provide entertainment have ever shown gore during a fight scene.
36	f	The movies that the children could watch to provide entertainment have ever shown gore during a fight scene.
36	g	No movies that the children could watch have provided any entertainment on the rainy day.
36	h	The movies that no children could watch have provided any entertainment on the rainy day.
36	i	The movies that the children could watch have provided any entertainment on the rainy day.

Table A.9: Full experimental stimuli for Experiment 9

A.10 Experiment 10

1	a	Very few critics that have recommended the authors of alternative genres have ever objected to mainstream literary trends.
---	---	--

1	b	The critics that have recommended very few authors of alternative genres have ever objected to mainstream literary trends.
1	c	The critics that haven't recommended the authors of alternative genres have ever objected to mainstream literary trends.
1	d	The critics that have recommended the authors of alternative genres have ever objected to mainstream literary trends.
2	a	Very few diplomats that have supported the refugees from war-torn countries have ever gained the respect of far-right journalists.
2	b	The diplomats that have supported very few refugees from war-torn countries have ever gained the respect of far-right journalists.
2	c	The diplomats that haven't supported the refugees from war-torn countries have ever gained the respect of far-right journalists.
2	d	The diplomats that have supported the refugees from war-torn countries have ever gained the respect of far-right journalists.
3	a	Very few paramedics that have revived the victims of heart attacks have ever underestimated the importance of a healthy lifestyle.
3	b	The paramedics that have revived very few victims of heart attacks have ever underestimated the importance of a healthy lifestyle.
3	c	The paramedics that haven't revived the victims of heart attacks have ever underestimated the importance of a healthy lifestyle.
3	d	The paramedics that have revived the victims of heart attacks have ever underestimated the importance of a healthy lifestyle.
4	a	Very few professors that have taught the children of needy families have ever supported the plans for privatizing the school system.
4	b	The professors that have taught very few children of needy families have ever supported the plans for privatizing the school system.
4	c	The professors that haven't taught the children of needy families have ever supported the plans for privatizing the school system.
4	d	The professors that have taught the children of needy families have ever supported the plans for privatizing the school system.
5	a	Very few salesmen that have assisted the customers with bad attitudes have ever requested additional shifts during the holiday season.

5	b	The salesmen that have assisted very few customers with bad attitudes have ever requested additional shifts during the holiday season.
5	c	The salesmen that haven't assisted the customers with bad attitudes have ever requested additional shifts during the holiday season.
5	d	The salesmen that have assisted the customers with bad attitudes have ever requested additional shifts during the holiday season.
6	a	Very few politicians that have addressed the reporters from socialist websites have ever caused controversy over a democratic bill.
6	b	The politicians that have addressed very few reporters from socialist websites have ever caused controversy over a democratic bill.
6	c	The politicians that haven't addressed the reporters from socialist websites have ever caused controversy over a democratic bill.
6	d	The politicians that have addressed the reporters from socialist websites have ever caused controversy over a democratic bill.
7	a	Very few designers that have hired the models from top-tier agencies have ever expressed concern about eating disorders in the fashion industry.
7	b	The designers that have hired very few models from top-tier agencies have ever expressed concern about eating disorders in the fashion industry.
7	c	The designers that haven't hired the models from top-tier agencies have ever expressed concern about eating disorders in the fashion industry.
7	d	The designers that have hired the models from top-tier agencies have ever expressed concern about eating disorders in the fashion industry.
8	a	Very few businessmen that have fired the salesmen with poor performance have ever received criticism from other powerful CEOs.
8	b	The businessmen that have fired very few salesmen with poor performance have ever received criticism from other powerful CEOs.
8	c	The businessmen that haven't fired the salesmen with poor performance have ever received criticism from other powerful CEOs.
8	d	The businessmen that have fired the salesmen with poor performance have ever received criticism from other powerful CEOs.
9	a	Very few teachers that have punished the organizers of school pranks have ever expected support from the liberal principal.

9	b	The teachers that have punished very few organizers of school pranks have ever expected support from the liberal principal.
9	c	The teachers that haven't punished the organizers of school pranks have ever expected support from the liberal principal.
9	d	The teachers that have punished the organizers of school pranks have ever expected support from the liberal principal.
10	a	Very few administrators that have congratulated the parents of graduating seniors have ever demonstrated interest in the students' success.
10	b	The administrators that have congratulated very few parents of graduating seniors have ever demonstrated interest in the students' success.
10	c	The administrators that haven't congratulated the parents of graduating seniors have ever demonstrated interest in the students' success.
10	d	The administrators that have congratulated the parents of graduating seniors have ever demonstrated interest in the students' success.
11	a	Very few judges that have dismissed the jurors with clear biases have ever been challenged on their rulings.
11	b	The judges that have dismissed very few jurors with clear biases have ever been challenged on their rulings.
11	c	The judges that haven't dismissed the jurors with clear biases have ever been challenged on their rulings.
11	d	The judges that have dismissed the jurors with clear biases have ever been challenged on their rulings.
12	a	Very few parents that have criticized the teachers of difficult classes have ever expressed concerns about students' work ethic.
12	b	The parents that have criticized very few teachers of difficult classes have ever expressed concerns about students' work ethic.
12	c	The parents that haven't criticized the teachers of difficult classes have ever expressed concerns about students' work ethic.
12	d	The parents that have criticized the teachers of difficult classes have ever expressed concerns about students' work ethic.
13	a	Very few coaches that have commended the athletes with bad sportsmanship have ever valued building camaraderie among the teammates.
13	b	The coaches that have commended very few athletes with bad sportsmanship have ever valued building camaraderie among the teammates.
13	c	The coaches that haven't commended the athletes with bad sportsmanship have ever valued building camaraderie among the teammates.

13	d	The coaches that have commended the athletes with bad sportsmanship have ever valued building camaraderie among the teammates.
14	a	Very few moms that have consulted the doctors with impeccable credentials have ever placed absolute faith in medical science.
14	b	The moms that have consulted very few doctors with impeccable credentials have ever placed absolute faith in medical science.
14	c	The moms that haven't consulted the doctors with impeccable credentials have ever placed absolute faith in medical science.
14	d	The moms that have consulted the doctors with impeccable credentials have ever placed absolute faith in medical science.
15	a	Very few policemen that have injured the suspects of drug crimes have ever received criticism for implicit racial bias.
15	b	The policemen that have injured very few suspects of drug crimes have ever received criticism for implicit racial bias.
15	c	The policemen that haven't injured the suspects of drug crimes have ever received criticism for implicit racial bias.
15	d	The policemen that have injured the suspects of drug crimes have ever received criticism for implicit racial bias.
16	a	Very few supervisors that have praised the workers for strong performance have ever defended the employees in managerial meetings.
16	b	The supervisors that have praised very few workers for strong performance have ever defended the employees in managerial meetings.
16	c	The supervisors that haven't praised the workers for strong performance have ever defended the employees in managerial meetings.
16	d	The supervisors that have praised the workers for strong performance have ever defended the employees in managerial meetings.
17	a	Very few producers that have signed the singers with sexist attitudes have ever been criticized by the fans.
17	b	The producers that have signed very few singers with sexist attitudes have ever been criticized by the fans.
17	c	The producers that haven't signed the singers with sexist attitudes have ever been criticized by the fans.
17	d	The producers that have signed the singers with sexist attitudes have ever been criticized by the fans.
18	a	Very few moviegoers that have favored the celebrities with Oscar nominations have ever shown interest in an independent film.

18	b	The moviegoers that have favored very few celebrities with Oscar nominations have ever shown interest in an independent film.
18	c	The moviegoers that haven't favored the celebrities with Oscar nominations have ever shown interest in an independent film.
18	d	The moviegoers that have favored the celebrities with Oscar nominations have ever shown interest in an independent film.
19	a	Very few nurses that have treated the patients with infectious diseases have ever expressed anxiety about the required vaccinations.
19	b	The nurses that have treated very few patients with infectious diseases have ever expressed anxiety about the required vaccinations.
19	c	The nurses that haven't treated the patients with infectious diseases have ever expressed anxiety about the required vaccinations.
19	d	The nurses that have treated the patients with infectious diseases have ever expressed anxiety about the required vaccinations.
20	a	Very few detectives that have intimidated the witnesses of horrific crimes have ever received training in conflict de-escalation.
20	b	The detectives that have intimidated very few witnesses of horrific crimes have ever received training in conflict de-escalation.
20	c	The detectives that haven't intimidated the witnesses of horrific crimes have ever received training in conflict de-escalation.
20	d	The detectives that have intimidated the witnesses of horrific crimes have ever received training in conflict de-escalation.
21	a	Very few mechanics that have overcharged the customers with minimal expertise have ever received positive online reviews.
21	b	The mechanics that have overcharged very few customers with minimal expertise have ever received positive online reviews.
21	c	The mechanics that haven't overcharged the customers with minimal expertise have ever received positive online reviews.
21	d	The mechanics that have overcharged the customers with minimal expertise have ever received positive online reviews.
22	a	Very few billionaires that have bankrolled the candidates with Republican support have ever expressed dissatisfaction with conservative fiscal policies.

22	b	The billionaires that have bankrolled very few candidates with Republican support have ever expressed dissatisfaction with conservative fiscal policies.
22	c	The billionaires that haven't bankrolled the candidates with Republican support have ever expressed dissatisfaction with conservative fiscal policies.
22	d	The billionaires that have bankrolled the candidates with Republican support have ever expressed dissatisfaction with conservative fiscal policies.
23	a	Very few lawyers that have assisted the clients with hopeless cases have ever developed impressive track records.
23	b	The lawyers that have assisted very few clients with hopeless cases have ever developed impressive track records.
23	c	The lawyers that haven't assisted the clients with hopeless cases have ever developed impressive track records.
23	d	The lawyers that haven assisted the clients with hopeless cases have ever developed impressive track records.
24	a	Very few landlords that have approved the tenants with large dogs have ever insisted on a quiet atmosphere in the apartments.
24	b	The landlords that have approved very few tenants with large dogs have ever insisted on a quiet atmosphere in the apartments.
24	c	The landlords that haven't approved the tenants with large dogs have ever insisted on a quiet atmosphere in the apartments.
24	d	The landlords that have approved the tenants with large dogs have ever insisted on a quiet atmosphere in the apartments.
25	a	Very few authors that have engaged the readers with refined taste have ever gained a broad audience.
25	b	The authors that have engaged very few readers with refined taste have ever gained a broad audience.
25	c	The authors that haven't engaged the readers with refined taste have ever gained a broad audience.
25	d	The authors that have engaged the readers with refined taste have ever gained a broad audience.
26	a	Very few coaches that have drafted the players from top-tier schools have ever felt nervousness before a championship game.
26	b	The coaches that have drafted very few players from top-tier schools have ever felt nervousness before a championship game.
26	c	The coaches that haven't drafted the players from top-tier schools have ever felt nervousness before a championship game.
26	d	The coaches that have drafted the players from top-tier schools have ever felt nervousness before a championship game.

27	a	Very few voters that have supported the candidates with unrealistic proposals have ever paid close attention to policy debates.
27	b	The voters that have supported very few candidates with unrealistic proposals have ever paid close attention to policy debates.
27	c	The voters that haven't supported the candidates with unrealistic proposals have ever paid close attention to policy debates.
27	d	The voters that have supported the candidates with unrealistic proposals have ever paid close attention to policy debates.
28	a	Very few students that have visited the professors during office hours have ever been stressed about exams.
28	b	The students that have visited very few professors during office hours have ever been stressed about exams.
28	c	The students that haven't visited the professors during office hours have ever been stressed about exams.
28	d	The students that have visited the professors during office hours have ever been stressed about exams.
29	a	Very few researchers that have dismissed the apprentices with lazy attitudes have ever experienced derision for their work ethics.
29	b	The researchers that have dismissed very few apprentices with lazy attitudes have ever experienced derision for their work ethics.
29	c	The researchers that haven't dismissed the apprentices with lazy attitudes have ever experienced derision for their work ethics.
29	d	The researchers that have dismissed the apprentices with lazy attitudes have ever experienced derision for their work ethics.
30	a	Very few actors that have impressed the directors of blockbuster films have ever been nervous before an important audition.
30	b	The actors that have impressed very few directors of blockbuster films have ever been nervous before an important audition.
30	c	The actors that haven't impressed the directors of blockbuster films have ever been nervous before an important audition.
30	d	The actors that have impressed the directors of blockbuster films have ever been nervous before an important audition.
31	a	Very few sprinters that have faced the competitors from Caribbean nations have ever expressed confidence before an important race.
31	b	The sprinters that have faced very few competitors from Caribbean nations have ever expressed confidence before an important race.

31	c	The sprinters that haven't faced the competitors from Caribbean nations have ever expressed confidence before an important race.
31	d	The sprinters that have faced the competitors from Caribbean nations have ever expressed confidence before an important race.
32	a	Very few collectors that have endorsed the artists with controversial themes have ever wanted the art world to be more political.
32	b	The collectors that have endorsed very few artists with controversial themes have ever wanted the art world to be more political.
32	c	The collectors that haven't endorsed the artists with controversial themes have ever wanted the art world to be more political.
32	d	The collectors that have endorsed the artists with controversial themes have ever wanted the art world to be more political.
33	a	Very few journalists that have challenged the editors of prestigious newspapers have ever expected difficulty on the job market.
33	b	The journalists that have challenged very few editors of prestigious newspapers have ever expected difficulty on the job market.
33	c	The journalists that haven't challenged the editors of prestigious newspapers have ever expected difficulty on the job market.
33	d	The journalists that have challenged the editors of prestigious newspapers have ever expected difficulty on the job market.
34	a	Very few teachers that have suspended the teenagers with behavioral problems have ever expected cooperation from the parents.
34	b	The teachers that have suspended very few teenagers with behavioral problems have ever expected cooperation from the parents.
34	c	The teachers that haven't suspended the teenagers with behavioral problems have ever expected cooperation from the parents.
34	d	The teachers that have suspended the teenagers with behavioral problems have ever expected cooperation from the parents.
35	a	Very few tutors that have helped the students in challenging classes have ever suggested improvements in the student to teacher ratio.
35	b	The tutors that have helped very few students in challenging classes have ever suggested improvements in the student to teacher ratio.

35	c	The tutors that haven't helped the students in challenging classes have ever suggested improvements in the student to teacher ratio.
35	d	The tutors that have helped the students in challenging classes have ever suggested improvements in the student to teacher ratio.
36	a	Very few children that have respected the lifeguards at community pools have ever broken the rules about running on the pool deck.
36	b	The children that have respected very few lifeguards at community pools have ever broken the rules about running on the pool deck.
36	c	The children that haven't respected the lifeguards at community pools have ever broken the rules about running on the pool deck.
36	d	The children that have respected the lifeguards at community pools have ever broken the rules about running on the pool deck.

Table A.10: Full experimental stimuli for Experiment 10

A.11 Experiment 11

1	a	No authors that the critics recommended have received any acknowledgment for a best-selling novel.
1	b	The authors that no critics recommended have received any acknowledgment for a best-selling novel.
1	c	The authors that the critics recommended have received any acknowledgment for a best-selling novel.
1	d	No authors that the critics recommended have ever received acknowledgment for a best-selling novel.
1	e	The authors that no critics recommended have ever received acknowledgment for a best-selling novel.
1	f	The authors that the critics recommended have ever received acknowledgment for a best-selling novel.
1	g	No authors that the critics recommended received any acknowledgement for a best-selling novel
1	h	The authors that no critics recommended received any acknowledgement for a best-selling novel
1	i	The authors that the critics recommended received any acknowledgment for a best-selling novel.
1	j	No authors that the critics recommended would have ever received acknowledgment for a best-selling novel.
1	k	The authors that no critics recommended would have ever received acknowledgment for a best-selling novel.
1	l	The authors that the critics recommended would have ever received acknowledgment for a best-selling novel.
2	a	No soldiers that the diplomats supported have shown any bravery in the controversial war.
2	b	The soldiers that no diplomats supported have shown any bravery in the controversial war.

2	c	The soldiers that the diplomats supported have shown any bravery in the controversial war.
2	d	No soldiers that the diplomats supported have ever shown bravery in the controversial war.
2	e	The soldiers that no diplomats supported have ever shown bravery in the controversial war.
2	f	The soldiers that the diplomats supported have ever shown bravery in the controversial war.
2	g	No soldiers that the diplomats supported showed any bravery in the controversial war.
2	h	The soldiers that no diplomats supported showed any bravery in the controversial war.
2	i	The solders that the diplomats supported showed any bravery in the controversial war.
2	j	No soldiers that the diplomats supported would have shown bravery in the controversial war.
2	k	The soldiers that no diplomats supported would have ever shown bravery in the controversial war.
2	l	The soldiers that the diplomats supported would have ever shown bravery in the controversial war.
3	a	No ambassadors that the diplomats consulted have seen any brutality in the foreign war.
3	b	The ambassadors that no diplomats consulted have seen any brutality in the foreign war.
3	c	The ambassadors that the diplomats consulted have seen any brutality in the foreign war.
3	d	No ambassadors that the diplomats consulted have ever seen brutality in the foreign war.
3	e	The ambassadors that no diplomats consulted have ever seen brutality in the foreign war.
3	f	The ambassadors that the diplomats consulted have ever seen brutality in the foreign war.
3	g	No ambassadors that the diplomats consulted saw any brutality in the foreign war.
3	h	The ambassadors that no diplomats consulted saw any brutality in the foreign war.
3	i	The ambassadors that the diplomats consulted saw any brutality in the foreign war.
3	j	No ambassadors that the diplomats consulted would have ever seen brutality in the foreign war.
3	k	The ambassadors that no diplomats consulted would have ever seen brutality in the foreign war.
3	l	The ambassadors that the diplomats consulted would have ever seen brutality in the foreign war.
4	a	No professors that the students respected have wanted any negativity in a class debate.
4	b	The professors that no students respected have wanted any negativity in a class debate.
4	c	The professors that the students respected have wanted any negativity in a class debate.
4	d	No professors that the students respected have ever wanted negativity in a class debate.
4	e	The professors that no students respected have ever wanted negativity in a class debate.
4	f	The professors that the students respected have ever wanted negativity in a class debate.
4	g	No professors that the students respected wanted any negativity in a class debate.
4	h	The professors that no students respected wanted any negativity in a class debate.
4	i	The professors that the students respected wanted any negativity in a class debate.
4	j	No professors that the students respected would have ever wanted negativity in a class debate.

4	k	The professors that no students respected would have ever wanted negativity in a class debate.
4	l	The professors that the students respected would have ever wanted negativity in a class debate.
5	a	No customers that the salesmen assisted have expressed any optimism for a full refund.
5	b	The customers that no salesmen assisted have expressed any optimism for a full refund.
5	c	The customers that the salesmen assisted have expressed any optimism for a full refund.
5	d	No customers that the salesmen assisted have ever expressed optimism for a full refund.
5	e	The customers that no salesmen assisted have ever expressed optimism for a full refund.
5	f	The customers that the salesmen assisted have ever expressed optimism for a full refund.
5	g	No customers that the salesmen assisted expressed any optimism for a full refund.
5	h	The customers that no salesmen assisted expressed any optimism for a full refund.
5	i	The customers that the salesmen assisted expressed any optimism for a full refund.
5	j	No customers that the salesmen assisted would have ever expressed optimism for a full refund.
5	k	The customers that no salesmen assisted would have ever expressed optimism for a full refund.
5	l	The customers that the salesmen assisted would have ever expressed optimism for a full refund.
6	a	No comments that the politicians ignored have caused any bitterness toward the liberal newspapers.
6	b	The comments that no politicians ignored have caused any bitterness toward the liberal newspapers.
6	c	The comments that the politicians ignored have caused any bitterness toward the liberal newspapers.
6	d	No comments that the politicians ignored have ever caused bitterness toward the liberal newspapers.
6	e	The comments that no politicians ignored have ever caused bitterness toward the liberal newspapers.
6	f	The comments that the politicians ignored have ever caused bitterness toward the liberal newspapers.
6	g	No comments that the politicians ignored caused any bitterness toward the liberal newspapers.
6	h	The comments that no politicians ignored caused any bitterness toward the liberal newspapers.
6	i	The comments that the politicians ignored caused any bitterness toward the liberal newspapers.
6	j	No comments that the politicians ignored would have ever caused bitterness toward the liberal newspapers.
6	k	The comments that no politicians ignored would have ever caused bitterness toward the liberal newspapers.
6	l	The comments that the politicians ignored would have ever caused bitterness toward the liberal newspapers.
7	a	No detergents that the housewives used have caused any damage to the delicate clothing.
7	b	The detergents that no housewives used have caused any damage to the delicate clothing.
7	c	The detergents that the housewives used have caused any damage to the delicate clothing.
7	d	No detergents that the housewives used have ever caused damage to the delicate clothing.
7	e	The detergents that no housewives used have ever caused damage to the delicate clothing.
7	f	The detergents that the housewives used have ever caused damage to the delicate clothing.

7	g	No detergents that the housewives used caused any damage to the delicate clothing.
7	h	The detergents that no housewives used caused any damage to the delicate clothing.
7	i	The detergents that the housewives used caused any damage to the delicate clothing.
7	j	No detergents that the housewives used would have ever caused damage to the delicate clothing.
7	k	The detergents that no housewives used would have ever caused damage to the delicate clothing.
7	l	The detergents that the housewives used would have ever caused damage to the delicate clothing.
8	a	No lawyers that the businessmen respected have received any criticism for a bad trial.
8	b	The lawyers that no businessmen respected have received any criticism for a bad trial.
8	c	The lawyers that the businessmen respected have received any criticism for a bad trial.
8	d	No lawyers that the businessmen respected have ever received criticism for a bad trial.
8	e	The lawyers that no businessmen respected have ever received criticism for a bad trial.
8	f	The lawyers that the businessmen respected have ever received criticism for a bad trial.
8	g	No lawyers that the businessmen respected received any criticism for a bad trial.
8	h	The lawyers that no businessmen respected received any criticism for a bad trial.
8	i	The lawyers that the businessmen respected received any criticism for a bad trial.
8	j	No lawyers that the businessmen respected would have ever received criticism for a bad trial.
8	k	The lawyers that no businessmen respected would have ever received criticism for a bad trial.
8	l	The lawyers that the businessmen respected would have ever received criticism for a bad trial.
9	a	No students that the teachers punished could expect any friendliness from the strict principal.
9	b	The students that no teachers punished could expect any friendliness from the strict principal.
9	c	The students that the teachers punished could expect any friendliness from the strict principal.
9	d	No students that the teachers punished could ever expect friendliness from the strict principal.
9	e	The students that no teachers punished could ever expect friendliness from the strict principal.
9	f	The students that the teachers punished could ever expect friendliness from the strict principal.
9	g	No students that the teachers punished expect any friendliness from the strict principal.
9	h	The students that no teachers punished expect any friendliness from the strict principal.
9	i	The students that the teachers punished expect any friendliness from the strict principal.
9	j	No students that the teachers punished would ever expect friendliness from the strict principal.
9	k	The students that no teachers punished would ever expect friendliness from the strict principal.
9	l	The students that the teachers punished would ever expect friendliness from the strict principal.
10	a	No babysitters that the children obeyed have shown any gratitude to the disappointed parents.
10	b	The babysitters that no children obeyed have shown any gratitude to the disappointed parents.

10	c	The babysitters that the children obeyed have shown any gratitude to the disappointed parents.
10	d	No babysitters that the children obeyed have ever shown gratitude to the disappointed parents.
10	e	The babysitters that no children obeyed have ever shown gratitude to the disappointed parents.
10	f	The babysitters that the children obeyed have ever shown gratitude to the disappointed parents.
10	g	No babysitters that the children obeyed showed any gratitude to the disappointed parents.
10	h	The babysitters that no children obeyed showed any gratitude to the disappointed parents.
10	i	The babysitters that the children obeyed showed any gratitude to the disappointed parents.
10	j	No babysitters that the children obeyed would have ever shown gratitude to the disappointed parents.
10	k	The babysitters that no children obeyed would have ever shown gratitude to the disappointed parents.
10	l	The babysitters that the children obeyed would have ever shown gratitude to the disappointed parents.
11	a	No actors that the fans recognized have experienced any soreness after a dangerous stunt.
11	b	The actors that no fans recognized have experienced any soreness after a dangerous stunt.
11	c	The actors that the fans recognized have experienced any soreness after a dangerous stunt.
11	d	No actors that the fans recognized have ever experienced soreness after a dangerous stunt.
11	e	The actors that no fans recognized have ever experienced soreness after a dangerous stunt.
11	f	The actors that the fans recognized have ever experienced soreness after a dangerous stunt.
11	g	No actors that the fans recognized experienced any soreness after a dangerous stunt.
11	h	The actors that no fans recognized experienced any soreness after a dangerous stunt.
11	i	The actors that the fans recognized experienced any soreness after a dangerous stunt.
11	j	No actors that the fans recognized would have ever experienced soreness after a dangerous stunt.
11	k	The actors that no fans recognized would have ever experienced soreness after a dangerous stunt.
11	l	The actors that the fans recognized would have ever experienced soreness after a dangerous stunt.
12	a	No teachers that the parents recommended have sparked any inspiration in the new students.
12	b	The teachers that no parents recommended have sparked any inspiration in the new students.
12	c	The teachers that the parents recommended have sparked any inspiration in the new students.
12	d	No teachers that the parents recommended have ever sparked inspiration in the new students.
12	e	The teachers that no parents recommended have ever sparked inspiration in the new students.
12	f	The teachers that the parents recommended have ever sparked inspiration in the new students.
12	g	No teachers that the parents recommended sparked any inspiration in the new students.
12	h	The teachers that no parents recommended sparked any inspiration in the new students.
12	i	The teachers that the parents recommended sparked any inspiration in the new students.
12	j	No teachers that the parents recommended would have ever sparked inspiration in the new students.

12	k	The teachers that no parents recommended would have ever sparked inspiration in the new students.
12	l	The teachers that the parents recommended would have ever sparked any inspiration in the new students.
13	a	No students that the librarians could help have made any progress on the difficult homework assignment.
13	b	The students that no librarians could help have made any progress on the difficult homework assignment.
13	c	The students that the librarians could help have made any progress on the difficult homework assignment.
13	d	No students that the librarians could help have ever made progress on the difficult homework assignment.
13	e	The students that no librarians could help have ever made progress on the difficult homework assignment.
13	f	The students that the librarians could help have ever made progress on the difficult homework assignment.
13	g	No students that the librarians could help made any progress on the difficult homework assignment.
13	h	The students that no librarians could help made any progress on the difficult homework assignment.
13	i	The students that the librarians could help made any progress on the difficult homework assignment.
13	j	No students that the librarians could help would have ever made any progress on the difficult homework assignment.
13	k	The students that no librarians could help would have ever made any progress on the difficult homework assignment.
13	l	The students that the librarians could help would have ever made any progress on the difficult homework assignment.
14	a	No nurses that the doctors appreciated have shown any patience in the operating room.
14	b	The nurses that no doctors appreciated have shown any patience in the operating room.
14	c	The nurses that the doctors appreciated have shown any patience in the operating room.
14	d	No nurses that the doctors appreciated have ever shown patience in the operating room.
14	e	The nurses that no doctors appreciated have ever shown patience in the operating room.
14	f	The nurses that the doctors appreciated have ever shown patience in the operating room.
14	g	No nurses that the doctors appreciated showed any patience in the operating room.
14	h	The nurses that no doctors appreciated showed any patience in the operating room.
14	i	The nurses that the doctors appreciated showed any patience in the operating room.
14	j	No nurses that the doctors appreciated would have ever shown patience in the operating room.
14	k	The nurses that no doctors appreciated would have ever shown any patience in the operating room.
14	l	The nurses that the doctors appreciated would have ever shown any patience in the operating room.
15	a	No criminals that the policemen could catch have felt any satisfaction from a petty crime.
15	b	The criminals that no policemen could catch have felt any satisfaction from a petty crime.
15	c	The criminals that the policemen could catch have felt any satisfaction from a petty crime.

15	d	No criminals that the policemen could catch have ever felt satisfaction from a petty crime.
15	e	The criminals that no policemen could catch have ever felt satisfaction from a petty crime.
15	f	The criminals that the policemen could catch have ever felt satisfaction from a petty crime.
15	g	No criminals that the policemen could catch felt any satisfaction from a petty crime.
15	h	The criminals that no policemen could catch felt any satisfaction from a petty crime.
15	i	The criminals that the policemen could catch felt any satisfaction from a petty crime.
15	j	No criminals that the policemen could catch would have ever felt satisfaction from a petty crime.
15	k	The criminals that no policemen could catch would have ever felt satisfaction from a petty crime.
15	l	The criminals that the policemen could catch would have ever felt satisfaction from a petty crime.
16	a	No employees that the managers recommended have shown any kindness to the rude customers.
16	b	The employees that no managers recommended have shown any kindness to the rude customers.
16	c	The employees that the managers recommended have shown any kindness to the rude customers.
16	d	No employees that the managers recommended have ever shown kindness to the rude customers.
16	e	The employees that no managers recommended have ever shown kindness to the rude customers.
16	f	The employees that the managers recommended have ever shown kindness to the rude customers.
16	g	No employees that the managers recommended showed any kindness to the rude customers.
16	h	The employees that no managers recommended showed any kindness to the rude customers.
16	i	The employees that the managers recommended showed any kindness to the rude customers.
16	j	No employees that the managers recommended would have ever showed kindness to the rude customers.
16	k	The employees that no managers recommended would have ever showed kindness to the rude customers.
16	l	The employees that the managers recommended would have ever showed kindness to the rude customers.
17	a	No accountants that the managers trusted have seen any rise in the quarterly profits.
17	b	The accountants that no managers trusted have seen any rise in the quarterly profits.
17	c	The accountants that the managers trusted have seen any rise in the quarterly profits.
17	d	No accountants that the managers trusted have ever seen a rise in the quarterly profits.
17	e	The accountants that no managers trusted have ever seen a rise in the quarterly profits.
17	f	The accountants that the managers trusted have ever seen a rise in the quarterly profits.
17	g	No accountants that the managers trusted saw any rise in the quarterly profits.
17	h	The accountants that no managers trusted saw any rise in the quarterly profits.
17	i	The accountants that the managers trusted saw any rise in the quarterly profits.
17	j	No accountants that the managers trusted would have ever seen any rise in the quarterly profits.
17	k	The accountants that no managers trusted would have ever seen any rise in the quarterly profits.

17	l	The accountants that the managers trusted would have ever seen any rise in the quarterly profits.
18	a	No candidates that the voters supported have shown any friendliness to the rude journalists.
18	b	The candidates that no voters supported have shown any friendliness to the rude journalists.
18	c	The candidates that the voters supported have shown any friendliness to the rude journalists.
18	d	No candidates that the voters supported have ever shown friendliness to the rude journalists.
18	e	The candidates that no voters supported have ever shown friendliness to the rude journalists.
18	f	The candidates that the voters supported have ever shown friendliness to the rude journalists.
18	g	No candidates that the voters supported showed any friendliness to the rude journalists.
18	h	The candidates that no voters supported showed any friendliness to the rude journalists.
18	i	The candidates that the voters supported showed any friendliness to the rude journalists.
18	j	No candidates that the voters supported would have ever shown friendliness to the rude journalists.
18	k	The candidates that no voters supported would have ever shown friendliness to the rude journalists.
18	l	The candidates that the voters supported would have ever shown friendliness to the rude journalists.
19	a	No surgeons that the patients trusted have shown any appreciation for the hospital staff.
19	b	The surgeons that no patients trusted have shown any appreciation for the hospital staff.
19	c	The surgeons that the patients trusted have shown any appreciation for the hospital staff.
19	d	No surgeons that the patients trusted have ever shown appreciation for the hospital staff.
19	e	The surgeons that no patients trusted have ever shown appreciation for the hospital staff.
19	f	The surgeons that the patients trusted have ever shown appreciation for the hospital staff.
19	g	No surgeons that the patients trusted showed any appreciation for the hospital staff.
19	h	The surgeons that no patients trusted showed any appreciation for the hospital staff.
19	i	The surgeons that the patients trusted showed any appreciation for the hospital staff.
19	j	No surgeons that the patients trusted would have ever shown appreciation for the hospital staff.
19	k	The surgeons that no patients trusted would have ever shown appreciation for the hospital staff.
19	l	The surgeons that the patients trusted would have ever shown appreciation for the hospital staff.
20	a	No suspects that the witnesses identified have shown any nervousness in the court room.
20	b	The suspects that no witnesses identified have shown any nervousness in the court room.
20	c	The suspects that the witnesses identified have shown any nervousness in the court room.
20	d	No suspects that the witnesses identified have ever shown nervousness in the court room.
20	e	The suspects that no witnesses identified have ever shown nervousness in the court room.
20	f	The suspects that the witnesses identified have ever shown nervousness in the court room.
20	g	No suspects that the witnesses identified showed any nervousness in the court room.

20	h	The suspects that no witnesses identified showed any nervousness in the court room.
20	i	The suspects that the witnesses identified showed any nervousness in the court room.
20	j	No suspects that the witnesses identified would have ever shown nervousness in the court room.
20	k	The suspects that no witnesses identified would have ever nervousness in the court room.
20	l	The suspects that the witnesses identified would have ever nervousness in the court room.
21	a	No actresses that the housewives admired have caused any commotion at a film festival.
21	b	The actresses that no housewives admired have caused any commotion at a film festival.
21	c	The actresses that the housewives admired have caused any commotion at a film festival.
21	d	No actresses that the housewives admired have ever caused a commotion at a film festival.
21	e	The actresses that no housewives admired have ever caused a commotion at a film festival.
21	f	The actresses that the housewives admired have ever caused a commotion at a film festival.
21	g	No actresses that the housewives admired caused any commotion at a film festival.
21	h	The actresses that no housewives admired caused any commotion at a film festival.
21	i	The actresses that the housewives admired caused any commotion at a film festival.
21	j	No actresses that the housewives admired would have ever caused commotion at a film festival.
21	k	The actresses that no housewives admired would have ever caused commotion at a film festival.
21	l	The actresses that the housewives admired would have ever caused commotion at a film festival.
22	a	No ambassadors that the government officials consulted have provoked any hostility from the liberal media.
22	b	The ambassadors that no government officials consulted have provoked any hostility from the liberal media.
22	c	The ambassadors that the government officials consulted have provoked any hostility from the liberal media.
22	d	No ambassadors that the government officials consulted have ever provoked hostility from the liberal media.
22	e	The ambassadors that no government officials consulted have ever provoked hostility from the liberal media.
22	f	The ambassadors that the government officials consulted have ever provoked hostility from the liberal media.
22	g	No ambassadors that the government officials consulted provoked any hostility from the liberal media.
22	h	The ambassadors that no government officials consulted provoked any hostility from the liberal media.
22	i	The ambassadors that the government officials consulted provoked any hostility from the liberal media.
22	j	No ambassadors that the government officials consulted would have ever provoked hostility from the liberal media.

22	k	The ambassadors that no government officials consulted would have ever provoked hostility from the liberal media.
22	l	The ambassadors that the government officials consulted would have ever provoked hostility from the liberal media.
23	a	No politicians that the journalists endorsed have earned any trust from the rural communities.
23	b	The politicians that no journalists endorsed have earned any trust from the rural communities.
23	c	The politicians that the journalists endorsed have earned any trust from the rural communities.
23	d	No politicians that the journalists endorsed have ever earned trust from the rural communities.
23	e	The politicians that no journalists endorsed have ever earned trust from the rural communities.
23	f	The politicians that the journalists endorsed have ever earned trust from the rural communities.
23	g	No politicians that the journalists endorsed earned any trust from the rural communities.
23	h	The politicians that no journalists endorsed earned any trust from the rural communities.
23	i	The politicians that the journalists endorsed earned any trust from the rural communities.
23	j	No politicians that the journalists endorsed would have ever earned trust from the rural communities.
23	k	The politicians that no journalists endorsed would have ever earned trust from the rural communities.
23	l	The politicians that the journalists endorsed would have ever earned trust from the rural communities.
24	a	No teenagers that the parents trusted have shown any kindness to a younger sibling.
24	b	The teenagers that no parents trusted have shown any kindness to a younger sibling.
24	c	The teenagers that the parents trusted have shown any kindness to a younger sibling.
24	d	No teenagers that the parents trusted have ever shown kindness to a younger sibling.
24	e	The teenagers that no parents trusted have ever shown kindness to a younger sibling.
24	f	The teenagers that the parents trusted have ever shown kindness to a younger sibling.
24	g	No teenagers that the parents trusted showed any kindness to a younger sibling.
24	h	The teenagers that no parents trusted showed any kindness to a younger sibling.
24	i	The teenagers that the parents trusted showed any kindness to a younger sibling.
24	j	No teenagers that the parents trusted would have ever shown kindness to a younger sibling.
24	k	The teenagers that no parents trusted would have ever shown any kindness to a younger sibling.
24	l	The teenagers that the parents trusted would have ever shown any kindness to a younger sibling.
25	a	No survivors that the medics could treat have shown any courage in an extreme emergency.
25	b	The survivors that no medics could treat have shown any courage in an extreme emergency.
25	c	The survivors that the medics could treat have shown any courage in an extreme emergency.
25	d	No survivors that the medics could treat have ever shown courage in an extreme emergency.

25	e	The survivors that no medics could treat have ever shown courage in an extreme emergency.
25	f	The survivors that the medics could treat have ever shown courage in an extreme emergency.
25	g	No survivors that the medics could treat showed any courage in an extreme emergency.
25	h	The survivors that no medics could treat showed any courage in an extreme emergency.
25	i	The survivors that the medics could treat showed any courage in an extreme emergency.
25	j	No survivors that the medics could treat would have ever shown courage in an extreme emergency.
25	k	The survivors that no medics could treat would have ever shown courage in an extreme emergency.
25	l	The survivors that the medics could treat would have ever shown courage in an extreme emergency.
26	a	No football players that the coaches drafted have felt any nervousness before a championship game.
26	b	The football players that no coaches drafted have felt any nervousness before a championship game.
26	c	The football players that the coaches drafted have felt any nervousness before a championship game.
26	d	No football players that the coaches drafted have ever felt nervousness before a championship game.
26	e	The football players that no coaches drafted have ever felt nervousness before a championship game.
26	f	The football players that the coaches drafted have ever felt nervousness before a championship game.
26	g	No football players that the coaches drafted felt any nervousness before a championship.
26	h	The football players that no coaches drafted felt any nervousness before a championship.
26	i	The football players that the coaches drafted felt any nervousness before a championship.
26	j	No football players that the coaches drafted would have ever felt nervousness before a championship.
26	k	The football players that no coaches drafted would have ever felt nervousness before a championship.
26	l	The football players that the coaches drafted would have ever felt nervousness before a championship.
27	a	No dictators that the citizens trusted have caused any chaos at a public event.
27	b	The dictators that no citizens trusted have caused any chaos at a public event.
27	c	The dictators that the citizens trusted have caused any chaos at a public event.
27	d	No dictators that the citizens trusted have ever caused chaos at a public event.
27	e	The dictators that no citizens trusted have ever caused chaos at a public event.
27	f	The dictators that the citizens trusted have ever caused chaos at a public event.
27	g	No dictators that the citizens trusted caused any chaos at a public event.
27	h	The dictators that no citizens trusted caused any chaos at a public event.
27	i	The dictators that the citizens trusted caused any chaos at a public event.
27	j	No dictators that the citizens trusted would have ever caused chaos at a public event.
27	k	The dictators that no citizens trusted would have ever caused chaos at a public event.
27	l	The dictators that the citizens trusted would have ever caused chaos at a public event.

28	a	No professors that the students could understand have experienced any tiredness after a long lecture.
28	b	The professors that no students could understand have experienced any tiredness after a long lecture.
28	c	The professors that the students could understand have experienced any tiredness after a long lecture.
28	d	No professors that the students could understand have ever experienced tiredness after a long lecture.
28	e	The professors that no students could understand have ever experienced tiredness after a long lecture.
28	f	The professors that the students could understand have ever experienced tiredness after a long lecture.
28	g	No professors that the students could understand experienced tiredness after a long lecture.
28	h	The professors that no students could understand experienced tiredness after a long lecture.
28	i	The professors that the students could understand experienced tiredness after a long lecture.
28	j	No professors that the students could understand would have ever experienced tiredness after a long lecture.
28	k	The professors that no students could understand would have ever experienced tiredness after a long lecture.
28	l	The professors that the students could understand would have ever experienced tiredness after a long lecture.
29	a	No actors that the judges nominated have had any luck at the award ceremonies.
29	b	The actors that no judges nominated have had any luck at the award ceremonies.
29	c	The actors that the judges nominated have had any luck at the award ceremonies.
29	d	No actors that the judges nominated have ever had much luck at the award ceremonies.
29	e	The actors that no judges nominated have ever had much luck at the award ceremonies.
29	f	The actors that the judges nominated have ever had much luck at the award ceremonies.
29	g	No actors that the judges nominated had any luck at the awards ceremonies.
29	h	The actors that no judges nominated had any luck at the awards ceremonies.
29	i	The actors that the judges nominated had any luck at the awards ceremonies.
29	j	No actors that the judges nominated would have ever had luck at the awards ceremonies.
29	k	The actors that no judges nominated would have ever had luck at the awards ceremonies.
29	l	The actors that the judges nominated would have ever had luck at the awards ceremonies.
30	a	No actresses that the directors auditioned have shown any elegance on a large stage.
30	b	The actresses that no directors auditioned have shown any elegance on a large stage.
30	c	The actresses that the directors auditioned have shown any elegance on a large stage.
30	d	No actresses that the directors auditioned have ever shown elegance on a large stage.
30	e	The actresses that no directors auditioned have ever shown elegance on a large stage.
30	f	The actresses that the directors auditioned have ever shown elegance on a large stage.
30	g	No actresses that the directors auditioned showed any elegance on a large stage.

30	h	The actresses that no directors auditioned showed any elegance on a large stage.
30	i	The actresses that the directors auditioned showed any elegance on a large stage.
30	j	No actresses that the directors auditioned would have ever showed elegance on a large stage.
30	k	The actresses that no directors auditioned would have ever shown elegance on a large stage.
30	l	The actresses that the directors auditioned would have ever shown elegance on a large stage.
31	a	No champions that the competitors defeated have shown any humility after a big game.
31	b	The champions that no competitors defeated have shown any humility after a big game.
31	c	The champions that the competitors defeated have shown any humility after a big game.
31	d	No champions that the competitors defeated have ever shown humility after a big game.
31	e	The champions that no competitors defeated have ever shown humility after a big game.
31	f	The champions that the competitors defeated have ever shown humility after a big game.
31	g	No champions that the competitors defeated showed any humility after a big game.
31	h	The champions that no champions defeated showed any humility after a big game.
31	i	The champions that the champions defeated showed any humility after a big game.
31	j	No champions that the competitors defeated would have ever shown humility after a big game.
31	k	The champions that no champions defeated would have ever shown humility after a big game.
31	l	The champions that the champions defeated would have ever shown humility after a big game.
32	a	No paintings that the collectors liked have depicted any tranquility with very bright colors.
32	b	The paintings that no collectors liked have depicted any tranquility with very bright colors.
32	c	The paintings that the collectors liked have depicted any tranquility with very bright colors.
32	d	No paintings that the collectors liked have ever depicted tranquility with very bright colors.
32	e	The paintings that no collectors liked have ever depicted tranquility with very bright colors.
32	f	The paintings that the collectors liked have ever depicted tranquility with very bright colors.
32	g	No paintings that the collectors liked depicted tranquility with very bright colors.
32	h	The paintings that no collectors liked depicted tranquility with very bright colors.
32	i	The paintings that the collectors liked depicted tranquility with very bright colors.
32	j	No paintings that the collectors liked would have ever depicted tranquility with very bright colors.
32	k	The paintings that no collectors liked would have ever depicted tranquility with very bright colors.
32	l	The paintings that the collectors liked would have ever depicted tranquility with very bright colors.
33	a	No editors that the journalists respected have had any patience for a missed deadline.
33	b	The editors that no journalists respected have had any patience for a missed deadline.
33	c	The editors that the journalists respected have had any patience for a missed deadline.

- 33 d No editors that the journalists respected have ever had much patience for a missed deadline.
- 33 e The editors that no journalists respected have ever had much patience for a missed deadline.
- 33 f The editors that the journalists respected have ever had much patience for a missed deadline.
- 33 g No editors that the journalists respected had much patience for a missed deadline.
- 33 h The editors that no journalists respected had much patience for a missed deadline.
- 33 i The editors that the journalists respected had much patience for a missed deadline.
- 33 j No editors that the journalists respected would have ever had much patience for a missed deadline.
- 33 k The editors that no journalists respected would have ever had much patience for a missed deadline.
- 33 l The editors that the journalists respected would have ever had much patience for a missed deadline.

- 34 a No teenagers that the teachers motivated have experienced any loneliness in the large class.
- 34 b The teenagers that no teachers motivated have experienced any loneliness in the large class.
- 34 c The teenagers that the teachers motivated have experienced any loneliness in the large class.
- 34 d No teenagers that the teachers motivated have ever experienced loneliness in the large class.
- 34 e The teenagers that no teachers motivated have ever experienced loneliness in the large class.
- 34 f The teenagers that the teachers motivated have ever experienced loneliness in the large class.
- 34 g No teenagers that the teachers motivated experienced any loneliness in the large class.
- 34 h The teenagers that no teachers motivated experienced any loneliness in the large class.
- 34 i The teenagers that the teachers motivated experienced any loneliness in the large class.
- 34 j No teenagers that the teachers motivated would have ever experienced loneliness in the large class.
- 34 k The teenagers that no teachers motivated would have ever experienced loneliness in the large class.
- 34 l The teenagers that the teachers motivated would have ever experienced loneliness in the large class.

- 35 a No students that the professors would tutor will have any trouble in a math class.
- 35 b The students that no professors would tutor will have any trouble in a math class.
- 35 c The students that the professors would tutor will have any trouble in a math class.
- 35 d No students that the professors would tutor will ever have trouble in a math class.
- 35 e The students that no professors would tutor will ever have trouble in a math class.
- 35 f The students that the professors would tutor will ever have trouble in a math class.
- 35 g No runners that the coaches trained drank any water during a competition.
- 35 h The runners that no coaches trained drank any water during a competition.
- 35 i The runners that the coaches trained drank any water during a competition.
- 35 j No runners that the coaches trained would have ever drunk water during a competition
- 35 k The runners that no coaches trained would have ever drunk water during a competition.

35	l	The runners that the coaches trained would have ever drunk water during a competition.
36	a	No movies that the children could watch have shown any gore during a fight scene.
36	b	The movies that no children could watch have shown any gore during a fight scene.
36	c	The movies that the children could watch have shown any gore during a fight scene.
36	d	No movies that the children could watch have ever shown gore during a fight scene.
36	e	The movies that no children could watch have ever shown gore during a fight scene.
36	f	The movies that the children could watch have ever shown gore during a fight scene
36	g	No movies that the children could watch showed any gore during a fight scene.
36	h	The movies that no children could watch showed any gore during a fight scene.
36	i	The movies that the children could watch showed any gore during a fight scene.
36	j	No movies that the children could watch would have ever shown gore during a fight scene.
36	k	The movies that no children could watch would have ever shown gore during a fight scene.
36	l	The movies that the children could watch would have ever shown gore during a fight scene.

Table A.11: Full experimental stimuli for Experiment 11

A.12 Experiment 12

1	a	No authors that the critics have recommended in their reviews have ever received acknowledgment for a best-selling novel.
1	b	The authors that no critics have recommended in their reviews have ever received acknowledgment for a best-selling novel.
1	c	The authors that the critics haven't recommended in their reviews have ever received acknowledgment for a best-selling novel.
1	d	The authors that the critics have recommended in their reviews have ever received acknowledgment for a best-selling novel.
1	e	No authors that the critics recommended in any of their reviews have ever received acknowledgment for a best-selling novel.
1	f	The authors that no critics recommended in any of their reviews have ever received acknowledgment for a best-selling novel.
1	g	The authors that the critics haven't recommended in any of their reviews have ever received acknowledgment for a best-selling novel.

1	h	The authors that the critics recommended in any of their reviews have ever received acknowledgment for a best-selling novel.
2	a	No soldiers that the diplomats have supported in the trial have ever shown respect to the war victims.
2	b	The soldiers that no diplomats have supported in the trial have ever shown respect to the war victims.
2	c	The soldiers that the diplomats haven't supported in the trial have ever shown respect to the war victims.
2	d	The soldiers that the diplomats have supported in the trial have ever shown respect to the war victims.
2	e	No soldiers that the diplomats have supported at all in the trial have ever shown respect to the war victims.
2	f	The soldiers that no diplomats have supported at all in the trial have ever shown respect to the war victims.
2	g	The soldiers that the diplomats haven't supported at all in the trial have ever shown respect to the war victims.
2	h	The soldiers that the diplomats have supported at all in the trial have ever shown respect to the war victims.
3	a	No ambassadors that the diplomats have consulted about the treaty have ever seen brutality in the foreign war.
3	b	The ambassadors that no diplomats have consulted about the treaty have ever seen brutality in the foreign war.
3	c	The ambassadors that the diplomats haven't consulted about the treaty have ever seen brutality in the foreign war.
3	d	The ambassadors that the diplomats have consulted about the treaty have ever seen brutality in the foreign war.
3	e	No ambassadors that the diplomats consulted about any aspects of the treaty at all have ever seen brutality in the foreign war.
3	f	The ambassadors that no diplomats consulted about any aspects of the treaty at all have ever seen brutality in the foreign war.
3	g	The ambassadors that the diplomats haven't consulted about any aspects of the treaty at all have ever seen brutality in the foreign war.
3	h	The ambassadors that the diplomats consulted about any aspects of the treaty at all have ever seen brutality in the foreign war.
4	a	No professors that the students have challenged over low grades have ever wanted negativity in a class debate.
4	b	The professors that no students have challenged over low grades have ever wanted negativity in a class debate.

4	c	The professors that the students haven't challenged over low grades have ever wanted negativity in a class debate.
4	d	The professors that the students have challenged over low grades have ever wanted negativity in a class debate.
4	e	No professors that the students have at any point challenged over low grades have ever wanted negativity in a class debate.
4	f	The professors that no students have at any point challenged over low grades have ever wanted negativity in a class debate.
4	g	The professors that the students haven't at any point challenged over low grades have ever wanted negativity in a class debate.
4	h	The professors that the students have at any point challenged over low grades have ever wanted negativity in a class debate.
5	a	No customers that the salesmen have assisted in the outlet have ever expressed optimism for a full refund.
5	b	The customers that no salesmen have assisted in the outlet have ever expressed optimism for a full refund.
5	c	The customers that the salesmen haven't assisted in the outlet have ever expressed optimism for a full refund.
5	d	The customers that the salesmen have assisted in the outlet have ever expressed optimism for a full refund.
5	e	No customers that the salesmen have assisted even a little bit have ever expressed optimism for a full refund.
5	f	The customers that no salesmen have assisted even a little bit have ever expressed optimism for a full refund.
5	g	The customers that the salesmen haven't assisted even a little bit have ever expressed optimism for a full refund.
5	h	The customers that the salesmen have assisted even a little bit have ever expressed optimism for a full refund.
6	a	No diplomats that the politicians have informed of recent policies have ever caused controversy in the liberal newspapers.
6	b	The diplomats that no politicians have informed of recent policies have ever caused controversy in the liberal newspapers.
6	c	The diplomats that the politicians haven't informed of recent policies have ever caused controversy in the liberal newspapers.
6	d	The diplomats that the politicians have informed of recent policies have ever caused controversy in the liberal newspapers.
6	e	No diplomats that the politicians have informed of any of the recent policies have ever caused controversy in the liberal newspapers.

6	f	The diplomats that no politicians have informed of any of the recent policies have ever caused controversy in the liberal newspapers.
6	g	The diplomats that the politicians haven't informed of any of the recent policies have ever caused controversy in the liberal newspapers.
6	h	The diplomats that the politicians have informed of any of the recent policies have ever caused controversy in the liberal newspapers.
7	a	No maids that the housewives have thanked for their work have ever caused damage to the delicate clothing.
7	b	The maids that no housewives have thanked for their work have ever caused damage to the delicate clothing.
7	c	The maids that the housewives haven't thanked for their work have ever caused damage to the delicate clothing.
7	d	The maids that the housewives have thanked for their work have ever caused damage to the delicate clothing.
7	e	No maids that the housewives thanked at all for their work have ever caused damage to the delicate clothing.
7	f	The maids that no housewives thanked at all for their work have ever caused damage to the delicate clothing.
7	g	The maids that the housewives haven't thanked at all for their work have ever caused damage to the delicate clothing.
7	h	The maids that the housewives thanked at all for their work have ever caused damage to the delicate clothing.
8	a	No lawyers that the businessmen have hired for legal advice have ever received criticism for lost fraud trials.
8	b	The lawyers that no businessmen have hired for legal advice have ever received criticism for lost fraud trials.
8	c	The lawyers that the businessmen haven't hired for legal advice have ever received criticism for lost fraud trials.
8	d	The lawyers that the businessmen have hired for legal advice have ever received criticism for lost fraud trials.
8	e	No lawyers that the businessmen would hire for any legal advice have ever received criticism for lost fraud trials.
8	f	The lawyers that no businessmen would hire for any legal advice have ever received criticism for lost fraud trials.
8	g	The lawyers that the businessmen wouldn't hire for any legal advice have ever received criticism for lost fraud trials.
8	h	The lawyers that the businessmen would hire for any legal advice have ever received criticism for lost fraud trials.
9	a	No students that the teachers have punished for bad behavior have ever expected friendliness from the strict principal.

9	b	The students that no teachers have punished for bad behavior have ever expected friendliness from the strict principal.
9	c	The students that the teachers haven't punished for bad behavior have ever expected friendliness from the strict principal.
9	d	The students that the teachers have punished for bad behavior have ever expected friendliness from the strict principal.
9	e	No students that the teachers have punished at all for their bad behavior have ever expected friendliness from the strict principal.
9	f	The students that no teachers have punished at all for their bad behavior have ever expected friendliness from the strict principal.
9	g	The students that the teachers haven't punished at all for their bad behavior have ever expected friendliness from the strict principal.
9	h	The students that the teachers have punished at all for their bad behavior have ever expected friendliness from the strict principal.
10	a	No babysitters that the children have disobeyed during an outing have ever expected gratitude from the disappointed parents.
10	b	The babysitters that no children have disobeyed during an outing have ever expected gratitude from the disappointed parents.
10	c	The babysitters that the children haven't disobeyed during an outing have ever expected gratitude from the disappointed parents.
10	d	The babysitters that the children have disobeyed during an outing have ever expected gratitude from the disappointed parents.
10	e	No babysitters that the children have disobeyed even a little bit have ever expected gratitude from the disappointed parents.
10	f	The babysitters that no children have disobeyed even a little bit have ever expected gratitude from the disappointed parents.
10	g	The babysitters that the children haven't disobeyed even a little bit have ever expected gratitude from the disappointed parents.
10	h	The babysitters that the children have disobeyed even a little bit have ever expected gratitude from the disappointed parents.
11	a	No actors that the producers have solicited for their films have ever experienced injuries from a dangerous stunt.

11	b	The actors that no producers have solicited for their films have ever experienced injuries from a dangerous stunt.
11	c	The actors that the producers haven't solicited for their films have ever experienced injury from a dangerous stunt.
11	d	The actors that the producers have solicited for their films have ever experienced injuries from a dangerous stunt.
11	e	No actors that the producers have solicited for any of their films have ever experienced injuries from a dangerous stunt.
11	f	The actors that no producers have solicited for any of their films have ever experienced injuries from a dangerous stunt.
11	g	The actors that the producers haven't solicited for any of their films have ever experienced injury from a dangerous stunt.
11	h	The actors that the producers have solicited for any of their films have ever experienced injuries from a dangerous stunt.
12	a	No students that the librarians have helped with book reports have ever made progress on the difficult assignment.
12	b	The students that no librarians have helped with book reports have ever made progress on the difficult assignment.
12	c	The students that the librarians haven't helped with book reports have ever made progress on the difficult assignment.
12	d	The students that the librarians have helped with book reports have ever made progress on the difficult assignment.
12	e	No students that the librarians have helped at all with their book reports have ever made progress on the difficult assignment.
12	f	The students that no librarians have helped at all with their book reports have ever made progress on the difficult assignment.
12	g	The students that the librarians haven't helped at all with their book reports have ever made progress on the difficult assignment.
12	h	The students that the librarians have helped at all with their book reports have ever made progress on the difficult assignment.
13	a	No nurses that the doctors have requested for the surgery have ever shown clumsiness in the operating room.

13	b	The nurses that no doctors have requested for the surgery have ever shown clumsiness in the operating room.
13	c	The nurses that the doctors haven't requested for the surgery have ever shown clumsiness in the operating room.
13	d	The nurses that the doctors have requested for the surgery have ever shown clumsiness in the operating room.
13	e	No nurses that the doctors have requested for any part of a surgery have ever shown clumsiness in the operating room.
13	f	The nurses that no doctors have requested for any part of a surgery have ever shown clumsiness in the operating room.
13	g	The nurses that the doctors haven't requested for any part of a surgery have ever shown clumsiness in the operating room.
13	h	The nurses that the doctors have requested for any part of a surgery have ever shown clumsiness in the operating room.
14	a	No criminals that the policemen have caught in drug raids have ever felt satisfaction from a petty crime.
14	b	The criminals that no policemen have caught in drug raids have ever felt satisfaction from a petty crime.
14	c	The criminals that the policemen haven't caught in drug raids have ever felt satisfaction from a petty crime.
14	d	The criminals that the policemen have caught in drug raids have ever felt satisfaction from a petty crime.
14	e	No criminals that the policemen could catch in any of the drug raids have ever felt satisfaction from a petty crime.
14	f	The criminals that no policemen could catch in any of the drug raids have ever felt satisfaction from a petty crime.
14	g	The criminals that the policemen couldn't catch in any of the drug raids have ever felt satisfaction from a petty crime.
14	h	The criminals that the policemen could catch in any of the drug raids have ever felt satisfaction from a petty crime.
15	a	No employees that the managers have recommended for a raise have ever expressed frustration with the rude customers.
15	b	The employees that no managers have recommended for a raise have ever expressed frustration with the rude customers.
15	c	The employees that the managers haven't recommended for a raise have ever expressed frustration with the rude customers.

15	d	The employees that the managers have recommended for a raise have ever expressed frustration with the rude customers.
15	e	No employees that the managers would even consider recommending for a raise have ever expressed frustration with the rude customers.
15	f	The employees that no managers would even consider recommending for a raise have ever expressed frustration with the rude customers.
15	g	The employees that the managers wouldn't even consider recommending for a raise have ever expressed frustration with the rude customers.
15	h	The employees that the managers would even consider recommending for a raise have ever expressed frustration with the rude customers.
16	a	No accountants that the managers have blamed for company losses have ever seen rises in the quarterly profits.
16	b	The accountants that no managers have blamed for company losses have ever seen rises in the quarterly profits.
16	c	The accountants that the managers haven't blamed for company losses have ever seen rises in the quarterly profits.
16	d	The accountants that the managers have blamed for company losses have ever seen rises in the quarterly profits.
16	e	No accountants that the managers have in any way blamed for company losses have ever seen rises in the quarterly profits.
16	f	The accountants that no managers have in any way blamed for company losses have ever seen rises in the quarterly profits.
16	g	The accountants that the managers haven't in any way blamed for company losses have ever seen rises in the quarterly profits.
16	h	The accountants that the managers have in any way blamed for company losses have ever seen rises in the quarterly profits.
17	a	No candidates that the voters have supported during the election have ever shown friendliness to the rude journalists.
17	b	The candidates that no voters have supported during the election have ever shown friendliness to the rude journalists.
17	c	The candidates that the voters haven't supported during the election have ever shown friendliness to the rude journalists.

17	d	The candidates that the voters have supported during the election have ever shown friendliness to the rude journalists.
17	e	No candidates that the voters have supported at all during the election have ever shown friendliness to the rude journalists.
17	f	The candidates that no voters have supported at all during the election have ever shown friendliness to the rude journalists.
17	g	The candidates that the voters haven't supported at all during the election have ever shown friendliness to the rude journalists.
17	h	The candidates that the voters have supported at all during the election have ever shown friendliness to the rude journalists.
18	a	No surgeons that the patients have consulted about the operation have ever expressed dissatisfaction with the hospital staff.
18	b	The surgeons that no patients have consulted about the operation have ever expressed dissatisfaction with the hospital staff.
18	c	The surgeons that the patients haven't consulted about the operation have ever expressed dissatisfaction with the hospital staff.
18	d	The surgeons that the patients have consulted about the operation have ever expressed dissatisfaction with the hospital staff.
18	e	No surgeons that the patients have even consulted about any of their operations have ever expressed dissatisfaction with the hospital staff.
18	f	The surgeons that no patients have even consulted about any of their operations have ever expressed dissatisfaction with the hospital staff.
18	g	The surgeons that the patients haven't even consulted about any of their operations have ever expressed dissatisfaction with the hospital staff.
18	h	The surgeons that the patients have even consulted about any of their operations have ever expressed dissatisfaction with the hospital staff.
19	a	No suspects that the witnesses have identified in photo line-ups have ever shown nervousness in the court room.
19	b	The suspects that no witnesses have identified in photo line-ups have ever shown nervousness in the court room.
19	c	The suspects that the witnesses haven't identified in photo line-ups have ever shown nervousness in the court room.

19	d	The suspects that the witnesses have identified in photo line-ups have ever shown nervousness in the court room.
19	e	No suspects that the witnesses could identify at all in any photo line-ups have ever shown nervousness in the court room.
19	f	The suspects that no witnesses could identify at all in any photo line-ups have ever shown nervousness in the court room.
19	g	The suspects that the witnesses couldn't identify at all in any photo line-ups have ever shown nervousness in the court room.
19	h	The suspects that the witnesses could identify at all in any photo line-ups have ever shown nervousness in the court room.
20	a	No actresses that the moviegoers have praised for their performance have ever caused excitement at a film festival.
20	b	The actresses that no moviegoers have praised for their performance have ever caused excitement at a film festival.
20	c	The actresses that the moviegoers haven't praised for their performance have ever caused excitement at a film festival.
20	d	The actresses that the moviegoers have praised for their performance have ever caused excitement at a film festival.
20	e	No actresses that the moviegoers have praised for any aspect of their performance have ever caused excitement at a film festival.
20	f	The actresses that no moviegoers have praised for any aspect of their performance have ever caused excitement at a film festival.
20	g	The actresses that the moviegoers haven't praised for any aspect of their performance have ever caused excitement at a film festival.
20	h	The actresses that the moviegoers have praised for any aspect of their performance have ever caused excitement at a film festival.
21	a	No senators that the billionaires have supported with campaign donations have ever received hostility from online news media.
21	b	The senators that no billionaires have supported with campaign donations have ever received hostility from online news media.
21	c	The senators that the billionaires haven't supported with campaign donations have ever received hostility from online news media.

21	d	The senators that the billionaires have supported with campaign donations have ever received hostility from online news media.
21	e	No senators that the billionaires have supported in any way have ever received hostility from online news media.
21	f	The senators that no billionaires have supported in any way have ever received hostility from online news media.
21	g	The senators that the billionaires haven't supported in any way have ever received hostility from online news media.
21	h	The senators that the billionaires have supported in any way have ever received hostility from online news media.
22	a	No politicians that the journalists have endorsed in the media have ever earned trust from the rural communities.
22	b	The politicians that no journalists have endorsed in the media have ever earned trust from the rural communities.
22	c	The politicians that the journalists haven't endorsed in the media have ever earned trust from the rural communities.
22	d	The politicians that the journalists have endorsed in the media have ever earned trust from the rural communities.
22	e	No politicians that the journalists would even consider endorsing in the media have ever earned trust from the rural communities.
22	f	The politicians that no journalists would even consider endorsing in the media have ever earned trust from the rural communities.
22	g	The politicians that the journalists wouldn't even consider endorsing in the media have ever earned trust from the rural communities.
22	h	The politicians that the journalists would even consider endorsing in the media have ever earned trust from the rural communities.
23	a	No survivors that the medics have cured of their injuries have ever felt regret for their military service
23	b	The survivors that no medics have cured of their injuries have ever felt regret for their military service
23	c	The survivors that the medics haven't cured of their injuries have ever felt regret for their military service
23	d	The survivors that the medics have cured of their injuries have ever felt regret for their military service
23	e	No survivors that the medics have cured of any injuries have ever felt regret for their military service
23	f	The survivors that no medics have cured of any injuries have ever felt regret for their military service

23	g	The survivors that the medics haven't cured of any injuries have ever felt regret for their military service
23	h	The survivors that the medics have cured of any injuries have ever felt regret for their military service
24	a	No voters that the senators have courted at campaign rallies have ever caused controversy in a major election
24	b	The voters that no senators have courted at campaign rallies have ever caused controversy in a major election
24	c	The voters that the senators haven't courted at campaign rallies have ever caused controversy in a major election.
24	d	The voters that the senators have courted at campaign rallies have ever caused controversy in a major election
24	e	No voters that the senators have courted at any of their campaign rallies have ever caused controversy in a major election
24	f	The voters that no senators have courted at any of their campaign rallies have ever caused controversy in a major election
24	g	The voters that the senators haven't courted at any of their campaign rallies have ever caused controversy in a major election.
24	h	The voters that the senators have courted at any of their campaign rallies have ever caused controversy in a major election
25	a	No professors that the students have visited during office hours have ever experienced tiredness after a long lecture.
25	b	The professors that no students have visited during office hours have ever experienced tiredness after a long lecture.
25	c	The professors that the students haven't visited during office hours have ever experienced tiredness after a long lecture.
25	d	The professors that the students have visited during office hours have ever experienced tiredness after a long lecture.
25	e	No professors that the students have visited at all during office hours have ever experienced tiredness after a long lecture.
25	f	The professors that no students have visited at all during office hours have ever experienced tiredness after a long lecture.
25	g	The professors that the students haven't visited at all during office hours have ever experienced tiredness after a long lecture.
25	h	The professors that the students have visited at all during office hours have ever experienced tiredness after a long lecture.

26	a	No actors that the judges have nominated for an award have ever experienced derision from the tabloid gossip.
26	b	The actors that no judges have nominated for an award have ever experienced derision from the tabloid gossip.
26	c	The actors that the judges haven't nominated for an award have ever experienced derision from the tabloid gossip.
26	d	The actors that the judges have nominated for an award have ever experienced derision from the tabloid gossip.
26	e	No actors that the judges would even consider nominating for any of the awards have ever experienced derision from the tabloid gossip.
26	f	The actors that no judges would even consider nominating for any of the awards have ever experienced derision from the tabloid gossip.
26	g	The actors that the judges wouldn't even consider nominating for any of the awards have ever experienced derision from the tabloid gossip.
26	h	The actors that the judges would even consider nominating for any of the awards have ever experienced derision from the tabloid gossip.
27	a	No champions that the competitors have defeated in important races have ever shown humility after a big win.
27	b	The champions that no competitors have defeated in important races have ever shown humility after a big win.
27	c	The champions that the competitors haven't defeated in important races have ever shown humility after a big win.
27	d	The champions that the competitors have defeated in important races have ever shown humility after a big win.
27	e	No champions that the competitors have at any point defeated in important an race have ever shown humility after a big win.
27	f	The champions that no competitors have at any point defeated in important an race have ever shown humility after a big win.
27	g	The champions that the competitors haven't at any point defeated in important an race have ever shown humility after a big win.
27	h	The champions that the competitors have at any point defeated in important an race have ever shown humility after a big win.

28	a	No painters that the collectors have favored at prestigious exhibitions have ever depicted tranquility with very bright colors.
28	b	The painters that no collectors have favored at prestigious exhibitions have ever depicted tranquility with very bright colors.
28	c	The painters that the collectors haven't favored at prestigious exhibitions have ever depicted tranquility with very bright colors.
28	d	The painters that the collectors have favored at prestigious exhibitions have ever depicted tranquility with very bright colors.
28	e	No painters that the collectors have in any way favored at prestigious exhibitions have ever depicted tranquility with very bright colors.
28	f	The painters that no collectors have in any way favored at prestigious exhibitions have ever depicted tranquility with very bright colors.
28	g	The painters that the collectors haven't in any way favored at prestigious exhibitions have ever depicted tranquility with very bright colors.
28	h	The painters that the collectors have in any way favored at prestigious exhibitions have ever depicted tranquility with very bright colors.
29	a	No editors that the journalists have disrespected in stressful meetings have ever had patience for a missed deadline.
29	b	The editors that no journalists have disrespected in stressful meetings have ever had patience for a missed deadline.
29	c	The editors that the journalists haven't disrespected in stressful meetings have ever had patience for a missed deadline.
29	d	The editors that the journalists have disrespected in stressful meetings have ever had patience for a missed deadline.
29	e	No editors that the journalists would even consider disrespecting in any way have ever had patience for a missed deadline.
29	f	The editors that no journalists would even consider disrespecting in any way have ever had patience for a missed deadline.
29	g	The editors that the journalists wouldn't even consider disrespecting in any way have ever had patience for a missed deadline.
29	h	The editors that the journalists would even consider disrespecting in any way have ever had patience for a missed deadline.

30	a	No teenagers that the teachers have scolded for their chattiness have ever experienced loneliness in the large class.
30	b	The teenagers that no teachers have scolded for their chattiness have ever experienced loneliness in the large class.
30	c	The teenagers that the teachers haven't scolded for their chattiness have ever experienced loneliness in the large class.
30	d	The teenagers that the teachers have scolded for their chattiness have ever experienced loneliness in the large class.
30	e	No teenagers that the teachers have scolded at any point for their chattiness have ever experienced loneliness in the large class.
30	f	The teenagers that no teachers have scolded at any point for their chattiness have ever experienced loneliness in the large class.
30	g	The teenagers that the teachers haven't scolded at any point for their chattiness have ever experienced loneliness in the large class.
30	h	The teenagers that the teachers have scolded at any point for their chattiness have ever experienced loneliness in the large class.
31	a	No students that the professors have tutored on the weekends have ever had trouble in a math class.
31	b	The students that no professors have tutored on the weekends have ever had trouble in a math class.
31	c	The students that the professors haven't tutored on the weekends have ever had trouble in a math class.
31	d	The students that the professors have tutored on the weekends have ever had trouble in a math class.
31	e	No students that the professors have at any point tutored on the weekends have ever had trouble in a math class.
31	f	The students that no professors have at any point tutored on the weekends have ever had trouble in a math class.
31	g	The students that the professors haven't at any point tutored on the weekends have ever had trouble in a math class.
31	h	The students that the professors have at any point tutored on the weekends have ever had trouble in a math class.
32	a	No actors that the children have seen in family films have ever shown gore during a fight scene.
32	b	The actors that no children have seen in family films have ever shown gore during a fight scene.
32	c	The actors that the children haven't seen in family films have ever shown gore during a fight scene.
32	d	The actors that the children have seen in family films have ever shown gore during a fight scene.

32	e	No actors that the children have seen in any family films at all have ever shown gore during a fight scene.
32	f	The actors that no children have seen in any family films at all have ever shown gore during a fight scene.
32	g	The actors that the children haven't seen in any family films at all have ever shown gore during a fight scene.
32	h	The actors that the children have seen in any family films at all have ever shown gore during a fight scene.

Table A.12: Full experimental stimuli for Experiment 12

A.13 Experiment 13

1	a	No legislators that the lobbyists could persuade have, at any time, received illegal funds.
1	b	The legislators that no lobbyists could persuade have, at any time, received illegal funds.
1	c	The legislators that the lobbyists could persuade have, at any time, received illegal funds.
1	d	No legislators that the lobbyists could persuade have received illegal funds at any time.
1	e	The legislators that no lobbyists could persuade have received illegal funds at any time.
1	f	The legislators that the lobbyists could persuade have received illegal funds at any time.
1	g	No legislators that the lobbyists could persuade have ever received illegal funds.
1	h	The legislators that no lobbyists could persuade have ever received illegal funds.
1	i	The legislators that the lobbyists could persuade have ever received illegal funds.
2	a	No ambassadors that the congressmen consulted have, at any time, been to the Middle East.
2	b	The ambassadors that no congressmen consulted have, at any time, been to the Middle East.
2	c	The ambassadors that the congressmen consulted have, at any time, been to the Middle East.
2	d	No ambassadors that the congressmen consulted have been to the Middle East at any time.
2	e	The ambassadors that no congressmen consulted have been to the Middle East at any time.
2	f	The ambassadors that the congressmen consulted have been to the Middle East at any time.
2	g	No ambassadors that the congressmen consulted have ever been to the Middle East.
2	h	The ambassadors that no congressmen consulted have ever been to the Middle East.
2	i	The ambassadors that the congressmen consulted have ever been to the Middle East.
3	a	No judges that the lawyers regarded very highly have, at any time, given the death penalty.
3	b	The judges that no lawyers regarded very highly have, at any time, given the death penalty.
3	c	The judges that the lawyers regarded very highly have, at any time, given the death penalty.
3	d	No judges that the lawyers regarded very highly have given the death penalty at any time.
3	e	The judges that no lawyers regarded very highly have given the death penalty at any time.
3	f	The judges that the lawyers regarded very highly have given the death penalty at any time.

3	g	No judges that the lawyers regarded very highly have ever given the death penalty.
3	h	The judges that no lawyers regarded very highly have ever given the death penalty.
3	i	The judges that the lawyers regarded very highly have ever given the death penalty.
4	a	No students that the teachers punished have, at any time, been under consideration for expulsion.
4	b	The students that no teachers punished have, at any time, been under consideration for expulsion.
4	c	The students that the teachers punished have, at any time, been under consideration for expulsion.
4	d	No students that the teachers punished have been under consideration for expulsion at any time.
4	e	The students that no teachers punished have been under consideration for expulsion at any time.
4	f	The students that the teachers punished have been under consideration for expulsion at any time.
4	g	No students that the teachers punished have ever been under consideration for expulsion.
4	h	The students that no teachers punished have ever been under consideration for expulsion.
4	i	The students that the teachers punished have ever been under consideration for expulsion.
5	a	No classes that the senior professors would teach have, at any time, been offered online.
5	b	The classes that no senior professors would teach have, at any time, been offered online.
5	c	The classes that the senior professors would teach have, at any time, been offered online
5	d	No classes that the senior professors would teach have been offered online at any time.
5	e	The classes that no senior professors would teach have been offered online at any time.
5	f	The classes that the senior professors would teach have been offered online at any time.
5	g	No classes that the senior professors would teach have ever been offered online.
5	h	The classes that no senior professors would teach have ever been offered online.
5	i	The classes that the senior professors would teach have ever been offered online
6	a	No stock brokers that the auditors investigated have, at any time, lost money on the stock market.
6	b	The stock brokers that no auditors investigated have, at any time, lost money on the stock market.
6	c	The stock brokers that the auditors investigated have, at any time, lost money on the stock market.
6	d	No stock brokers that the auditors investigated have lost money on the stock market at any time.
6	e	The stock brokers that no auditors investigated have lost money on the stock market at any time.
6	f	The stock brokers that the auditors investigated have lost money on the stock market at any time.
6	g	No stock brokers that the auditors investigated have ever lost money on the stock market.
6	h	The stock brokers that no auditors investigated have ever lost money on the stock market.
6	i	The stock brokers that the auditors investigated have ever lost money on the stock market.
7	a	No smugglers that the federal agents could find have, at any time, used a fake alias.
7	b	The smugglers that no federal agents could find have, at any time, used a fake alias.

7	c	The smugglers that the federal agents could find have, at any time, used a fake alias.
7	d	No smugglers that the federal agents could find have used a fake alias at any time.
7	e	The smugglers that no federal agents could find have used a fake alias at any time.
7	f	The smugglers that the federal agents could find have used a fake alias at any time.
7	g	No smugglers that the federal agents could find have ever used a fake alias.
7	h	The smugglers that no federal agents could find have ever used a fake alias.
7	i	The smugglers that the federal agents could find have ever used a fake alias.
8	a	No politicians that the journalists endorsed have, at any time, served in the senate before.
8	b	The politicians that no journalists endorsed have, at any time, served in the senate before.
8	c	The politicians that the journalists endorsed have, at any time, served in the senate before.
8	d	No politicians that the journalists endorsed have served in the senate at any time.
8	e	The politicians that no journalists endorsed have served in the senate at any time.
8	f	The politicians that the journalists endorsed have served in the senate at any time.
8	g	No politicians that the journalists endorsed have ever served in the senate before.
8	h	The politicians that no journalists endorsed have ever served in the senate before.
8	i	The politicians that the journalists endorsed have ever served in the senate before.
9	a	No professors that the students liked have, at any time, wanted to teach an introductory course.
9	b	The professors that no students liked have, at any time, wanted to teach an introductory course.
9	c	The professors that the students liked have, at any time, wanted to teach an introductory course.
9	d	No professors that the students liked have wanted to teach an introductory course at any time.
9	e	The professors that no students liked have wanted to teach an introductory course at any time.
9	f	The professors that the students liked have wanted to teach an introductory course at any time.
9	g	No professors that the students liked have ever wanted to teach an introductory course.
9	h	The professors that no students liked have ever wanted to teach an introductory course.
9	i	The professors that the students liked have ever wanted to teach an introductory course.
10	a	No neighborhoods that the police would patrol have, at any time, had a serious drug problem.
10	b	The neighborhoods that no police would patrol have, at any time, had a serious drug problem
10	c	The neighborhoods that the police would patrol have, at any time, had a serious drug problem.
10	d	No neighborhoods that the police would patrol have had a serious drug problem at any time.
10	e	The neighborhoods that no police would patrol have had a serious drug problem at any time.
10	f	The neighborhoods that the police would patrol have had a serious drug problem at any time.
10	g	No neighborhoods that the police would patrol have ever had a serious drug problem.

10	h	The neighborhoods that no police would patrol have ever had a serious drug problem
10	i	The neighborhoods that the police would patrol have ever had a serious drug problem.
11	a	No actors that the fans recognized have, at any time, needed a body-guard during an awards show.
11	b	The actors that no fans recognized have, at any time, needed a body-guard during an awards show.
11	c	The actors that the fans recognized have, at any time, needed a body-guard during an awards show.
11	d	No actors that the fans recognized have brought a body-guard to an awards show at any time.
11	e	The actors that no fans recognized have brought a body-guard to an awards show at any time.
11	f	The actors that the fans recognized have brought a body-guard to an awards show at any time.
11	g	No actors that the fans recognized have ever needed a body-guard during an awards show.
11	h	The actors that no fans recognized have ever needed a body-guard during an awards show.
11	i	The actors that the fans recognized have ever needed a body-guard during an awards show.
12	a	No interns that the managers recommended have, at any time, received equal benefits during their employment.
12	b	The interns that no managers recommended have, at any time, received equal benefits during their employment.
12	c	The interns that the managers recommended have, at any time, received equal benefits during their employment.
12	d	No interns that the managers recommended have received equal employment benefits at any time.
12	e	The interns that no managers recommended have received equal employment benefits at any time.
12	f	The interns that the managers recommended have received equal employment benefits at any time.
12	g	No interns that the managers recommended have ever received equal benefits during their employment.
12	h	The interns that no managers recommended have ever received equal benefits during their employment.
12	i	The interns that the managers recommended have ever received equal benefits during their employment.
13	a	No customers that the salesmen assisted have, in any way, been satisfied with their experience.
13	b	The customers that no salesmen assisted have, in any way, been satisfied with their experience.
13	c	The customers that the salesmen assisted have, in any way, been satisfied with their experience.
13	d	No customers that the salesmen assisted have been dissatisfied with their experience in any way.
13	e	The customers that no salesmen assisted have been dissatisfied with their experience in any way.
13	f	The customers that the salesmen assisted have been dissatisfied with their experience in any way.
13	g	No customers that the salesmen assisted have ever been satisfied with their experience.
13	h	The customers that no salesmen assisted have ever been satisfied with their experience.
13	i	The customers that the salesmen assisted have ever been satisfied with their experience.

14	a	No authors that the critics recommended have, in any way, been considered for an award.
14	b	The authors that no critics recommended have, in any way, been considered for an award.
14	c	The authors that the critics recommended have, in any way, been considered for an award.
14	d	No authors that the critics recommended have been tempted by plagiarism in any way.
14	e	The authors that no critics recommended have been tempted by plagiarism in any way.
14	f	The authors that the critics recommended have been tempted by plagiarism in any way.
14	g	No authors that the critics recommended have ever been considered for an award.
14	h	The authors that no critics recommended have ever been considered for an award.
14	i	The authors that the critics recommended have ever been considered for an award.
15	a	No secret codes that the amateur hacker could crack have, in any way, used computer encryption.
15	b	The secret codes that no amateur hacker could crack have, in any way, used computer encryption.
15	c	The secret codes that the amateur hacker could crack have, in any way, used computer encryption.
15	d	No secret codes that the amateur hacker could crack have been protected by computer encryption in any way.
15	e	The secret codes that no amateur hacker could crack have been protected by computer encryption in any way.
15	f	The secret codes that the amateur hacker could crack have been protected by computer encryption in any way.
15	g	No secret codes that the amateur hacker could crack have ever used computer encryption.
15	h	The secret codes that no amateur hacker could crack have ever used computer encryption.
15	i	The secret codes that the amateur hacker could crack have ever used computer encryption.
16	a	No controversial topics that the politicians would discuss have, in any way, been addressed in congress.
16	b	The controversial topics that no politicians would discuss have, in any way, been addressed in congress.
16	c	The controversial topics that the politicians would discuss have, in any way, been addressed in congress.
16	d	No controversial topics that the politicians discussed have been addressed by congress in any way.
16	e	The controversial topics that no politicians discussed have been addressed by congress in any way.
16	f	The controversial topics that the politicians discussed have been addressed by congress in any way.
16	g	No controversial topics that the politicians would discuss have ever been addressed in congress.
16	h	The controversial topics that no politicians would discuss have ever been addressed in congress.
16	i	The controversial topics that the politicians would discuss have ever been addressed in congress.
17	a	No books that the critics enjoyed have, in any way, appealed to female readers.
17	b	The books that no critics enjoyed have, in any way, appealed to female readers.

17	c	The books that the critics enjoyed have, in any way, appealed to female readers.
17	d	No books that the critics enjoyed have appealed to female readers in any way.
17	e	The books that no critics enjoyed have appealed to female readers in any way.
17	f	The books that the critics enjoyed have appealed to female readers in any way.
17	g	No books that the critics enjoyed have ever appealed to female readers.
17	h	The books that no critics enjoyed have ever appealed to female readers.
17	i	The books that the critics enjoyed have ever appealed to female readers.
18	a	No doctors that the nurses liked have, in any way, made an incorrect diagnosis.
18	b	The doctors that no nurses liked have, in any way, made an incorrect diagnosis.
18	c	The doctors that the nurses liked have, in any way, made an incorrect diagnosis.
18	d	No doctors that the nurses liked have enabled problematic family members in any way.
18	e	The doctors that no nurses liked have enabled problematic family members in any way.
18	f	The doctors that the nurses liked have enabled problematic family members in any way.
18	g	No doctors that the nurses liked have ever made an incorrect diagnosis.
18	h	The doctors that no nurses liked have ever made an incorrect diagnosis.
18	i	The doctors that the nurses liked have ever made an incorrect diagnosis.
19	a	No actresses that the housewives admired have, in any way, had cosmetic surgery.
19	b	The actresses that no housewives admired have, in any way, had cosmetic surgery.
19	c	The actresses that the housewives admired have, in any way, had cosmetic surgery.
19	d	No actresses that the housewives admired have entertained the idea of cosmetic surgery in any way.
19	e	The actresses that no housewives admired have entertained the idea of cosmetic surgery in any way.
19	f	The actresses that the housewives admired have entertained the idea of cosmetic surgery in any way.
19	g	No actresses that the housewives admired have ever had cosmetic surgery.
19	h	The actresses that no housewives admired have ever had cosmetic surgery.
19	i	The actresses that the housewives admired have ever had cosmetic surgery.
20	a	No criminals that the policemen could catch have, in any way, assumed a false identity.
20	b	The criminals that no policemen could catch have, in any way, assumed a false identity.
20	c	The criminals that the policemen could catch have, in any way, assumed a false identity.
20	d	No criminals that the policemen could catch have obscured their identities in any way.
20	e	The criminals that no policemen could catch have obscured their identities in any way.
20	f	The criminals that the policemen could catch have obscured their identities in any way.
20	g	No criminals that the policemen could catch have ever assumed a false identity.

20	h	The criminals that no policemen could catch have ever assumed a false identity.
20	i	The criminals that the policemen could catch have ever assumed a false identity.
21	a	No movies that the bored housewives liked have, in any way, exhibited excessive gore.
21	b	The movies that no bored housewives liked have, in any way, exhibited excessive gore.
21	c	The movies that the bored housewives liked have, in any way, exhibited excessive gore.
21	d	No movies that the bored housewives liked have exhibited violence in any way.
21	e	The movies that no bored housewives liked have exhibited violence in any way.
21	f	The movies that the bored housewives liked have exhibited violence in any way.
21	g	No movies that the bored housewives liked have ever exhibited excessive gore.
21	h	The movies that no bored housewives liked have ever exhibited excessive gore.
21	i	The movies that the bored housewives liked have ever exhibited excessive gore.
22	a	No suspects that the witnesses could identify have, in any way, been in trouble with the law.
22	b	The suspects that no witnesses could identify have, in any way, been in trouble with the law.
22	c	The suspects that the witnesses could identify have, in any way, been in trouble with the law.
22	d	No suspects that the witnesses could identify have resorted to criminal activity in any way.
22	e	The suspects that no witnesses could identify have resorted to criminal activity in any way.
22	f	The suspects that the witnesses could identify have resorted to criminal activity in any way.
22	g	No suspects that the witnesses could identify have ever been in trouble with the law.
22	h	The suspects that no witnesses could identify have ever been in trouble with the law.
22	i	The suspects that the witnesses could identify have ever been in trouble with the law.
23	a	No teachers that the parents recommended have, in any way, tolerated disruptions in the classroom.
23	b	The teachers that no parents recommended have, in any way, tolerated disruptions in the classroom.
23	c	The teachers that the parents recommended have, in any way, tolerated disruptions in the classroom.
23	d	No teachers that the parents recommended have encouraged disruptive behavior in any way.
23	e	The teachers that no parents recommended have encouraged disruptive behavior in any way.
23	f	The teachers that the parents recommended have encouraged disruptive behavior in any way.
23	g	No teachers that the parents recommended have ever tolerated disruptions in the classroom.
23	h	The teachers that no parents recommended have ever tolerated disruptions in the classroom.
23	i	The teachers that the parents recommended have ever tolerated disruptions in the classroom.
24	a	No candidates that the democratic voters supported have, in any way, accepted campaign donations.
24	b	The candidates that no democratic voters supported have, in any way, accepted campaign donations.
24	c	The candidates that the democratic voters supported have, in any way, accepted campaign donations.

24	d	No candidates that the republican voters supported have volunteered their time to the community in any way.
24	e	The candidates that no democratic voters supported have volunteered their time to the community in any way.
24	f	The candidates that the democratic voters supported have volunteered their time to the community in any way.
24	g	No candidates that the democratic voters supported have ever accepted campaign donations.
24	h	The candidates that no democratic voters supported have ever accepted campaign donations.
24	i	The candidates that the democratic voters supported have ever accepted campaign donations.
25	a	No babysitters that the children disobeyed have, at any time, requested a tip.
25	b	The babysitters that no children disobeyed have, at any time, requested a tip.
25	c	The babysitters that the children disobeyed have, at any time, requested a tip.
25	d	No babysitters that the children disobeyed have requested a tip at any time.
25	e	The babysitters that no children disobeyed have requested a tip at any time.
25	f	The babysitters that the children disobeyed have requested a tip at any time.
25	g	No babysitters that the children disobeyed have ever requested a tip.
25	h	The babysitters that no children disobeyed have ever requested a tip.
25	i	The babysitters that the children disobeyed have ever requested a tip.
26	a	No athletes that the competitors defeated have, in any way, shown good sportsmanship.
26	b	The athletes that no competitors defeated have, in any way, shown good sportsmanship.
26	c	The athletes that the competitors defeated have, in any way, shown good sportsmanship.
26	d	No athletes that the competitors defeated have shown good sportsmanship in any way.
26	e	The athletes that no competitors defeated have shown good sportsmanship in any way.
26	f	The athletes that the competitors defeated have shown good sportsmanship in any way.
26	g	No athletes that the competitors defeated have ever shown good sportsmanship.
26	h	The athletes that no competitors defeated have ever shown good sportsmanship.
26	i	The athletes that the competitors defeated have ever shown good sportsmanship.
27	a	No surgeons that the patients consulted have, at any time, suggested unnecessary operations.
27	b	The surgeons that no patients consulted have, at any time, suggested unnecessary operations.
27	c	The surgeons that the patients consulted have, at any time, suggested unnecessary operations.
27	d	No surgeons that the patients consulted have suggested unnecessary operations at any time.
27	e	The surgeons that no patients consulted have suggested unnecessary operations at any time.

- | | | |
|----|---|---|
| 27 | f | The surgeons that the patients consulted have suggested unnecessary operations at any time. |
| 27 | g | No surgeons that the patients consulted have ever suggested unnecessary operations. |
| 27 | h | The surgeons that no patients consulted have ever suggested unnecessary operations. |
| 27 | i | The surgeons that the patients consulted have ever suggested unnecessary operations. |

Table A.13: Full experimental stimuli for Experiment 13

A.14 Experiment 14

- | | | |
|---|---|---|
| 1 | a | No journalists that the editors recommended for the assignment argued that the readers would ever understand the complicated situation. |
| 1 | b | The journalists that no editors recommended for the assignment argued that the readers would ever understand the complicated situation. |
| 1 | c | The journalists that the editors recommended for the assignment argued that the readers would ever understand the complicated situation. |
| 1 | d | No journalists that the editors recommended for the assignment have ever argued that the readers would understand the complicated situation. |
| 1 | e | The journalists that no editors recommended for the assignment have ever argued that the readers would understand the complicated situation. |
| 1 | f | The journalists that the editors recommended for the assignment have ever argued that the readers would understand the complicated situation. |
| 1 | g | No journalists that the editors recommended for the assignment argued that the readers would in any way understand the complicated situation. |
| 1 | h | The journalists that no editors recommended for the assignment argued that the readers would in any way understand the complicated situation. |
| 1 | i | The journalists that the editors recommended for the assignment argued that the readers would in any way understand the complicated situation. |
| 1 | j | No journalists that the editors recommended for the assignment have in any way argued that the readers would understand the complicated situation. |
| 1 | k | The journalists that no editors recommended for the assignment have in any way argued that the readers would understand the complicated situation. |
| 1 | l | The journalists that the editors recommended for the assignment have in any way argued that the readers would understand the complicated situation. |

1	m	No journalists that the editors recommended for the assignment argued that the readers would understand the complicated situation in any way.
1	n	The journalists that no editors recommended for the assignment argued that the readers would understand the complicated situation in any way.
1	o	The journalists that the editors recommended for the assignment argued that the readers would understand the complicated situation in any way.
2	a	No investors that the businessmen informed about the recession predicted that the stock would ever drop below the initial offering price.
2	b	The investors that no businessmen informed about the recession predicted that the stock would ever drop below the initial offering price.
2	c	The investors that the businessmen informed about the recession predicted that the stock would ever drop below the initial offering price.
2	d	No investors that the businessmen informed about the recession have ever predicted that the stock would drop below the initial offering price.
2	e	The investors that no businessmen informed about the recession have ever predicted that the stock would drop below the initial offering price.
2	f	The investors that the businessmen informed about the recession have ever predicted that the stock would drop below the initial offering price.
2	g	No investors that the businessmen informed about the recession predicted that the stock would at any time drop below the initial offering price.
2	h	The investors that no businessmen informed about the recession predicted that the stock would at any time drop below the initial offering price.
2	i	The investors that the businessmen informed about the recession predicted that the stock would at any time drop below the initial offering price.
2	j	No investors that the businessmen informed about the recession have at any time predicted that the stock would drop below the initial offering price.
2	k	The investors that no businessmen informed about the recession have at any time predicted that the stock would drop below the initial offering price.
2	l	The investors that the businessmen informed about the recession have at any time predicted that the stock would drop below the initial offering price.
2	m	No investors that the businessmen informed about the recession predicted that the stock would drop below the initial offering price at any time.

2	n	The investors that no businessmen informed about the recession predicted that the stock would drop below the initial offering price at any time.
2	o	The investors that the businessmen informed about the recession predicted that the stock would drop below the initial offering price at any time.
3	a	No ambassadors that the diplomats consulted about the treaty suspected that the journalists would ever reveal the truth about election.
3	b	The ambassadors that no diplomats consulted about the treaty suspected that the journalists would ever reveal the truth about election.
3	c	The ambassadors that the diplomats consulted about the treaty suspected that the journalists would ever reveal the truth about election.
3	d	No ambassadors that the diplomats consulted about the treaty have ever suspected that the journalists would reveal the truth about election.
3	e	The ambassadors that no diplomats consulted about the treaty have ever suspected that the journalists would reveal the truth about election.
3	f	The ambassadors that the diplomats consulted about the treaty have ever suspected that the journalists would reveal the truth about election.
3	g	No ambassadors that the diplomats consulted about the treaty suspected that the journalists would at any time reveal the truth about election.
3	h	The ambassadors that no diplomats consulted about the treaty suspected that the journalists would at any time reveal the truth about election.
3	i	The ambassadors that the diplomats consulted about the treaty suspected that the journalists would at any time reveal the truth about election.
3	j	No ambassadors that the diplomats consulted about the treaty have at any time suspected that the journalists would reveal the truth about election.
3	k	The ambassadors that no diplomats consulted about the treaty have at any time suspected that the journalists would reveal the truth about election.
3	l	The ambassadors that the diplomats consulted about the treaty have at any time suspected that the journalists would reveal the truth about election.
3	m	No ambassadors that the diplomats consulted about the treaty suspected that the journalists would reveal the truth about election at any time.
3	n	The ambassadors that no diplomats consulted about the treaty suspected that the journalists would reveal the truth about election at any time.

3	o	The ambassadors that the diplomats consulted about the treaty suspected that the journalists would reveal the truth about election at any time.
4	a	No professors that the students trusted at the college thought that the administrators would ever increase the yearly tuition.
4	b	The professors that no students trusted at the college thought that the administrators would ever increase the yearly tuition.
4	c	The professors that the students trusted at the college thought that the administrators would ever increase the yearly tuition.
4	d	No professors that the students trusted at the college have ever thought that the administrators would increase the yearly tuition.
4	e	The professors that no students trusted at the college have ever thought that the administrators would increase the yearly tuition.
4	f	The professors that the students trusted at the college have ever thought that the administrators would increase the yearly tuition.
4	g	No professors that the students trusted at the college thought that the administrators would at any time increase the yearly tuition.
4	h	The professors that no students trusted at the college thought that the administrators would at any time increase the yearly tuition.
4	i	The professors that the students trusted at the college thought that the administrators would at any time increase the yearly tuition.
4	j	No professors that the students trusted at the college have at any time thought that the administrators would increase the yearly tuition.
4	k	The professors that no students trusted at the college have at any time thought that the administrators would increase the yearly tuition.
4	l	The professors that the students trusted at the college have at any time thought that the administrators would increase the yearly tuition.
4	m	No professors that the students trusted at the college thought that the administrators would increase the yearly tuition at any time.
4	n	The professors that no students trusted at the college thought that the administrators would increase the yearly tuition at any time.
4	o	The professors that the students trusted at the college thought that the administrators would increase the yearly tuition at any time.

5	a	No customers that the salesmen assisted in the showroom expected that the manager would ever consider their lowest offer.
5	b	The customers that no salesmen assisted in the showroom expected that the manager would ever consider their lowest offer.
5	c	The customers that the salesmen assisted in the showroom expected that the manager would ever consider their lowest offer.
5	d	No customers that the salesmen assisted in the showroom have ever expected that the manager would consider their lowest offer.
5	e	The customers that no salesmen assisted in the showroom have ever expected that the manager would consider their lowest offer.
5	f	The customers that the salesmen assisted in the showroom have ever expected that the manager would consider their lowest offer.
5	g	No customers that the salesmen assisted in the showroom expected that the manager would in any way consider their lowest offer.
5	h	The customers that no salesmen assisted in the showroom expected that the manager would in any way consider their lowest offer.
5	i	The customers that the salesmen assisted in the showroom expected that the manager would in any way consider their lowest offer.
5	j	No customers that the salesmen assisted in the showroom have in any way expected that the manager would consider their lowest offer.
5	k	The customers that no salesmen assisted in the showroom have in any way expected that the manager would consider their lowest offer.
5	l	The customers that the salesmen assisted in the showroom have in any way expected that the manager would consider their lowest offer.
5	m	No customers that the salesmen assisted in the showroom expected that the manager would consider their lowest offer in any way.
5	n	The customers that no salesmen assisted in the showroom expected that the manager would consider their lowest offer in any way.
5	o	The customers that the salesmen assisted in the showroom expected that the manager would consider their lowest offer in any way.
6	a	No protestors that the journalists interviewed at the rally implied that the legislators could ever pass the necessary laws.

6	b	The protestors that no journalists interviewed at the rally implied that the legislators could ever pass the necessary laws.
6	c	The protestors that the journalists interviewed at the rally implied that the legislators could ever pass the necessary laws.
6	d	No protestors that the journalists interviewed at the rally have ever implied that the legislators could pass the necessary laws.
6	e	The protestors that no journalists interviewed at the rally have ever implied that the legislators could pass the necessary laws.
6	f	The protestors that the journalists interviewed at the rally have ever implied that the legislators could pass the necessary laws.
6	g	No protestors that the journalists interviewed at the rally implied that the legislators could at any time pass the necessary laws.
6	h	The protestors that no journalists interviewed at the rally implied that the legislators could at any time pass the necessary laws.
6	i	The protestors that the journalists interviewed at the rally implied that the legislators could at any time pass the necessary laws.
6	j	No protestors that the journalists interviewed at the rally have at any time implied that the legislators could pass the necessary laws.
6	k	The protestors that no journalists interviewed at the rally have at any time implied that the legislators could pass the necessary laws.
6	l	The protestors that the journalists interviewed at the rally have at any time implied that the legislators could pass the necessary laws.
6	m	No protestors that the journalists interviewed at the rally implied that the legislators could pass the necessary laws at any time.
6	n	The protestors that no journalists interviewed at the rally implied that the legislators could pass the necessary laws at any time.
6	o	The protestors that the journalists interviewed at the rally implied that the legislators could pass the necessary laws at any time.
7	a	No senators that the corporations supported with campaign donations imagined that the lobbyists would ever accept the sly bribe.
7	b	The senators that no corporations supported with campaign donations imagined that the lobbyists would ever accept the sly bribe.

7	c	The senators that the corporations supported with campaign donations imagined that the lobbyists would ever accept the sly bribe.
7	d	No senators that the corporations supported with campaign donations have ever imagined that the lobbyists would accept the sly bribe.
7	e	The senators that no corporations supported with campaign donations have ever imagined that the lobbyists would accept the sly bribe.
7	f	The senators that the corporations supported with campaign donations have ever imagined that the lobbyists would accept the sly bribe.
7	g	No senators that the corporations supported with campaign donations imagined that the lobbyists would in any way accept the sly bribe.
7	h	The senators that no corporations supported with campaign donations imagined that the lobbyists would in any way accept the sly bribe.
7	i	The senators that the corporations supported with campaign donations imagined that the lobbyists would in any way accept the sly bribe.
7	j	No senators that the corporations supported with campaign donations have in any way imagined that the lobbyists would accept the sly bribe.
7	k	The senators that no corporations supported with campaign donations have in any way imagined that the lobbyists would accept the sly bribe.
7	l	The senators that the corporations supported with campaign donations have in any way imagined that the lobbyists would accept the sly bribe.
7	m	No senators that the corporations supported with campaign donations imagined that the lobbyists would accept the sly bribe in any way.
7	n	The senators that no corporations supported with campaign donations imagined that the lobbyists would accept the sly bribe in any way.
7	o	The senators that the corporations supported with campaign donations imagined that the lobbyists would accept the sly bribe in any way.
8	a	No lawyers that the policemen respected after the trial anticipated that the judge would ever deliver such a harsh sentence.
8	b	The lawyers that no policemen respected after the trial anticipated that the judge would ever deliver such a harsh sentence.
8	c	The lawyers that the policemen respected after the trial anticipated that the judge would ever deliver such a harsh sentence.

8	d	No lawyers that the policemen respected after the trial have ever anticipated that the judge would deliver such a harsh sentence.
8	e	The lawyers that no policemen respected after the trial have ever anticipated that the judge would deliver such a harsh sentence.
8	f	The lawyers that the policemen respected after the trial have ever anticipated that the judge would deliver such a harsh sentence.
8	g	No lawyers that the policemen respected after the trial anticipated that the judge would at any time deliver such a harsh sentence.
8	h	The lawyers that no policemen respected after the trial anticipated that the judge would at any time deliver such a harsh sentence.
8	i	The lawyers that the policemen respected after the trial anticipated that the judge would at any time deliver such a harsh sentence.
8	j	No lawyers that the policemen respected after the trial have at any time anticipated that the judge would deliver such a harsh sentence.
8	k	The lawyers that no policemen respected after the trial have at any time anticipated that the judge would deliver such a harsh sentence.
8	l	The lawyers that the policemen respected after the trial have at any time anticipated that the judge would deliver such a harsh sentence.
8	m	No lawyers that the policemen respected after the trial anticipated that the judge would deliver such a harsh sentence at any time.
8	n	The lawyers that no policemen respected after the trial anticipated that the judge would deliver such a harsh sentence at any time.
8	o	The lawyers that the policemen respected after the trial anticipated that the judge would deliver such a harsh sentence at any time.
9	a	No students that the teachers punished for bad behavior expected that the principal would ever hear about the incident.
9	b	The students that no teachers punished for bad behavior expected that the principal would ever hear about the incident.
9	c	The students that the teachers punished for bad behavior expected that the principal would ever hear about the incident.
9	d	No students that the teachers punished for bad behavior have ever expected that the principal would hear about the incident.

9	e	The students that no teachers punished for bad behavior have ever expected that the principal would hear about the incident.
9	f	The students that the teachers punished for bad behavior have ever expected that the principal would hear about the incident.
9	g	No students that the teachers punished for bad behavior expected that the principal would at any time hear about the incident.
9	h	The students that no teachers punished for bad behavior expected that the principal would at any time hear about the incident.
9	i	The students that the teachers punished for bad behavior expected that the principal would at any time hear about the incident.
9	j	No students that the teachers punished for bad behavior have at any time expected that the principal would hear about the incident.
9	k	The students that no teachers punished for bad behavior have at any time expected that the principal would hear about the incident.
9	l	The students that the teachers punished for bad behavior have at any time expected that the principal would hear about the incident.
9	m	No students that the teachers punished for bad behavior expected that the principal would hear about the incident at any time.
9	n	The students that no teachers punished for bad behavior expected that the principal would hear about the incident at any time.
9	o	The students that the teachers punished for bad behavior expected that the principal would hear about the incident at any time.
10	a	No accountants that the inspectors audited in the past year thought that the IRS would ever follow up on the scandal.
10	b	The accountants that no inspectors audited in the past year thought that the IRS would ever follow up on the scandal.
10	c	The accountants that the inspectors audited in the past year thought that the IRS would ever follow up on the scandal.
10	d	No accountants that the inspectors audited in the past year have ever thought that the IRS would follow up on the scandal.
10	e	The accountants that no inspectors audited in the past year have ever thought that the IRS would follow up on the scandal.

- 10 f The accountants that the inspectors audited in the past year have ever thought that the IRS would follow up on the scandal.
- 10 g No accountants that the inspectors audited in the past year thought that the IRS would in any way follow up on the scandal.
- 10 h The accountants that no inspectors audited in the past year thought that the IRS would in any way follow up on the scandal.
- 10 i The accountants that the inspectors audited in the past year thought that the IRS would in any way follow up on the scandal.
- 10 j No accountants that the inspectors audited in the past year have in any way thought that the IRS would follow up on the scandal.
- 10 k The accountants that no inspectors audited in the past year have in any way thought that the IRS would follow up on the scandal.
- 10 l The accountants that the inspectors audited in the past year have in any way thought that the IRS would follow up on the scandal.
- 10 m No accountants that the inspectors audited in the past year thought that the IRS would follow up on the scandal in any way.
- 10 n The accountants that no inspectors audited in the past year thought that the IRS would follow up on the scandal in any way.
- 10 o The accountants that the inspectors audited in the past year thought that the IRS would follow up on the scandal in any way.
-
- 11 a No actors that the fans recognized at the after-party believed that the paparazzi would ever find out about the affair.
- 11 b The actors that no fans recognized at the after-party believed that the paparazzi would ever find out about the affair.
- 11 c The actors that the fans recognized at the after-party believed that the paparazzi would ever find out about the affair.
- 11 d No actors that the fans recognized at the after-party have ever believed that the paparazzi would find out about the affair.
- 11 e The actors that no fans recognized at the after-party have ever believed that the paparazzi would find out about the affair.
- 11 f The actors that the fans recognized at the after-party have ever believed that the paparazzi would find out about the affair.

11	g	No actors that the fans recognized at the after-party believed that the paparazzi would at any time find out about the affair.
11	h	The actors that no fans recognized at the after-party believed that the paparazzi would at any time find out about the affair.
11	i	The actors that the fans recognized at the after-party believed that the paparazzi would at any time find out about the affair.
11	j	No actors that the fans recognized at the after-party have at any time believed that the paparazzi would find out about the affair.
11	k	The actors that no fans recognized at the after-party have at any time believed that the paparazzi would find out about the affair.
11	l	The actors that the fans recognized at the after-party have at any time believed that the paparazzi would find out about the affair.
11	m	No actors that the fans recognized at the after-party believed that the paparazzi would find out about the affair at any time.
11	n	The actors that no fans recognized at the after-party believed that the paparazzi would find out about the affair at any time.
11	o	The actors that the fans recognized at the after-party believed that the paparazzi would find out about the affair at any time.
12	a	No teachers that the parents recommended for the award expected that the faculty would ever be honored for their work.
12	b	The teachers that no parents recommended for the award expected that the faculty would ever be honored for their work.
12	c	The teachers that the parents recommended for the award expected that the faculty would ever be honored for their work.
12	d	No teachers that the parents recommended for the award have ever expected that the faculty would be honored for their work.
12	e	The teachers that no parents recommended for the award have ever expected that the faculty would be honored for their work.
12	f	The teachers that the parents recommended for the award have ever expected that the faculty would be honored for their work.
12	g	No teachers that the parents recommended for the award expected that the faculty would in any way be honored for their work.

12	h	The teachers that no parents recommended for the award expected that the faculty would in any way be honored for their work.
12	i	The teachers that the parents recommended for the award expected that the faculty would in any way be honored for their work.
12	j	No teachers that the parents recommended for the award have in any way expected that the faculty would be honored for their work.
12	k	The teachers that no parents recommended for the award have in any way expected that the faculty would be honored for their work.
12	l	The teachers that the parents recommended for the award have in any way expected that the faculty would be honored for their work.
12	m	No teachers that the parents recommended for the award expected that the faculty would be honored for their work in any way.
12	n	The teachers that no parents recommended for the award expected that the faculty would be honored for their work in any way.
12	o	The teachers that the parents recommended for the award expected that the faculty would be honored for their work in any way.
13	a	No students that the librarians could help in the afternoon expected that the teacher would ever adjust expectations.
13	b	The students that no librarians could help in the afternoon expected that the teacher would ever adjust expectations.
13	c	The students that the librarians could help in the afternoon expected that the teacher would ever adjust expectations.
13	d	No students that the librarians could help in the afternoon have ever expected that the teacher would adjust expectations.
13	e	The students that no librarians could help in the afternoon have ever expected that the teacher would adjust expectations.
13	f	The students that the librarians could help in the afternoon have ever expected that the teacher would adjust expectations.
13	g	No students that the librarians could help in the afternoon expected that the teacher would in any way adjust expectations.
13	h	The students that no librarians could help in the afternoon expected that the teacher would in any way adjust expectations.

13	i	The students that the librarians could help in the afternoon expected that the teacher would in any way adjust expectations.
13	j	No students that the librarians could help in the afternoon have in any way expected that the teacher would adjust expectations.
13	k	The students that no librarians could help in the afternoon have in any way expected that the teacher would adjust expectations.
13	l	The students that the librarians could help in the afternoon have in any way expected that the teacher would adjust expectations.
13	m	No students that the librarians could help in the afternoon expected that the teacher would adjust expectations in any way.
13	n	The students that no librarians could help in the afternoon expected that the teacher would adjust expectations in any way.
13	o	The students that the librarians could help in the afternoon expected that the teacher would adjust expectations in any way.
14	a	No children that the bullies picked on at recess thought that the teacher would ever condone the behavior.
14	b	The children that no bullies picked on at recess thought that the teacher would ever condone the behavior.
14	c	The children that the bullies picked on at recess thought that the teacher would ever condone the behavior.
14	d	No children that the bullies picked on at recess have ever thought that the teacher would condone the behavior.
14	e	The children that no bullies picked on at recess have ever thought that the teacher would condone the behavior.
14	f	The children that the bullies picked on at recess have ever thought that the teacher would condone the behavior.
14	g	No children that the bullies picked on at recess thought that the teacher would in any way condone the behavior.
14	h	The children that no bullies picked on at recess thought that the teacher would in any way condone the behavior.
14	i	The children that the bullies picked on at recess thought that the teacher would in any way condone the behavior.
14	j	No children that the bullies picked on at recess have in any way thought that the teacher would condone the behavior.
14	k	The children that no bullies picked on at recess have in any way thought that the teacher would condone the behavior.

14	l	The children that the bullies picked on at recess have in any way thought that the teacher would condone the behavior.
14	m	No children that the bullies picked on at recess thought that the teacher would condone the behavior in any way.
14	n	The children that no bullies picked on at recess thought that the teacher would condone the behavior in any way.
14	o	The children that the bullies picked on at recess thought that the teacher would condone the behavior in any way.
15	a	No criminals that the policemen could catch in the raid expected that the judge would ever accept a plea bargain.
15	b	The criminals that no policemen could catch in the raid expected that the judge would ever accept a plea bargain.
15	c	The criminals that the policemen could catch in the raid expected that the judge would ever accept a plea bargain.
15	d	No criminals that the policemen could catch in the raid have ever expected that the judge would accept a plea bargain.
15	e	The criminals that no policemen could catch in the raid have ever expected that the judge would accept a plea bargain.
15	f	The criminals that the policemen could catch in the raid have ever expected that the judge would accept a plea bargain.
15	g	No criminals that the policemen could catch in the raid expected that the judge would at any time accept a plea bargain.
15	h	The criminals that no policemen could catch in the raid expected that the judge would at any time accept a plea bargain.
15	i	The criminals that the policemen could catch in the raid expected that the judge would at any time accept a plea bargain.
15	j	No criminals that the policemen could catch in the raid have at any time expected that the judge would accept a plea bargain.
15	k	The criminals that no policemen could catch in the raid have at any time expected that the judge would accept a plea bargain.
15	l	The criminals that the policemen could catch in the raid have at any time expected that the judge would accept a plea bargain.

- | | | |
|----|---|---|
| 15 | m | No criminals that the policemen could catch in the raid expected that the judge would accept a plea bargain at any time. |
| 15 | n | The criminals that no policemen could catch in the raid expected that the judge would accept a plea bargain at any time. |
| 15 | o | The criminals that the policemen could catch in the raid expected that the judge would accept a plea bargain at any time. |
| 16 | a | No employees that the managers recommended for the promotion anticipated that the boss would ever disregard their qualifications. |
| 16 | b | The employees that no managers recommended for the promotion anticipated that the boss would ever disregard their qualifications. |
| 16 | c | The employees that the managers recommended for the promotion anticipated that the boss would ever disregard their qualifications. |
| 16 | d | No employees that the managers recommended for the promotion have ever anticipated that the boss would disregard their qualifications. |
| 16 | e | The employees that no managers recommended for the promotion have ever anticipated that the boss would disregard their qualifications. |
| 16 | f | The employees that the managers recommended for the promotion have ever anticipated that the boss would disregard their qualifications. |
| 16 | g | No employees that the managers recommended for the promotion anticipated that the boss would in any way disregard their qualifications. |
| 16 | h | The employees that no managers recommended for the promotion anticipated that the boss would in any way disregard their qualifications. |
| 16 | i | The employees that the managers recommended for the promotion anticipated that the boss would in any way disregard their qualifications. |
| 16 | j | No employees that the managers recommended for the promotion have in any way anticipated that the boss would disregard their qualifications. |
| 16 | k | The employees that no managers recommended for the promotion have in any way anticipated that the boss would disregard their qualifications. |
| 16 | l | The employees that the managers recommended for the promotion have in any way anticipated that the boss would disregard their qualifications. |
| 16 | m | No employees that the managers recommended for the promotion anticipated that the boss would disregard their qualifications in any way. |

16	n	The employees that no managers recommended for the promotion anticipated that the boss would disregard their qualifications in any way.
16	o	The employees that the managers recommended for the promotion anticipated that the boss would disregard their qualifications in any way.
17	a	No investors that the managers trusted with the money thought that the stock prices would ever increase drastically.
17	b	The investors that no managers trusted with the money thought that the stock prices would ever increase drastically.
17	c	The investors that the managers trusted with the money thought that the stock prices would ever increase drastically.
17	d	No investors that the managers trusted with the money have ever thought that the stock prices would increase drastically.
17	e	The investors that no managers trusted with the money have ever thought that the stock prices would increase drastically.
17	f	The investors that the managers trusted with the money have ever thought that the stock prices would increase drastically.
17	g	No investors that the managers trusted with the money thought that the stock prices would at any time increase drastically.
17	h	The investors that no managers trusted with the money thought that the stock prices would at any time increase drastically.
17	i	The investors that the managers trusted with the money thought that the stock prices would at any time increase drastically.
17	j	No investors that the managers trusted with the money have at any time thought that the stock prices would increase drastically.
17	k	The investors that no managers trusted with the money have at any time thought that the stock prices would increase drastically.
17	l	The investors that the managers trusted with the money have at any time thought that the stock prices would increase drastically.
17	m	No investors that the managers trusted with the money thought that the stock prices would increase drastically at any time.
17	n	The investors that no managers trusted with the money thought that the stock prices would increase drastically at any time.

17	o	The investors that the managers trusted with the money thought that the stock prices would increase drastically at any time.
18	a	No candidates that the voters supported during the election believed that the mayor would ever fulfill his campaign promises.
18	b	The candidates that no voters supported during the election believed that the mayor would ever fulfill his campaign promises.
18	c	The candidates that the voters supported during the election believed that the mayor would ever fulfill his campaign promises.
18	d	No candidates that the voters supported during the election have ever believed that the mayor would fulfill his campaign promises.
18	e	The candidates that no voters supported during the election have ever believed that the mayor would fulfill his campaign promises.
18	f	The candidates that the voters supported during the election have ever believed that the mayor would fulfill his campaign promises.
18	g	No candidates that the voters supported during the election believed that the mayor would at any time fulfill his campaign promises.
18	h	The candidates that no voters supported during the election believed that the mayor would at any time fulfill his campaign promises.
18	i	The candidates that the voters supported during the election believed that the mayor would at any time fulfill his campaign promises.
18	j	No candidates that the voters supported during the election have at any time believed that the mayor would fulfill his campaign promises.
18	k	The candidates that no voters supported during the election have at any time believed that the mayor would fulfill his campaign promises.
18	l	The candidates that the voters supported during the election have at any time believed that the mayor would fulfill his campaign promises.
18	m	No candidates that the voters supported during the election believed that the mayor would fulfill his campaign promises at any time.
18	n	The candidates that no voters supported during the election believed that the mayor would fulfill his campaign promises at any time.
18	o	The candidates that the voters supported during the election believed that the mayor would fulfill his campaign promises at any time.

19	a	No doctors that the nurses assisted during the operation assumed that the insurance company would ever deny the claim.
19	b	The doctors that no nurses assisted during the operation assumed that the insurance company would ever deny the claim.
19	c	The doctors that the nurses assisted during the operation assumed that the insurance company would ever deny the claim.
19	d	No doctors that the nurses assisted during the operation have ever assumed that the insurance company would deny the claim.
19	e	The doctors that no nurses assisted during the operation have ever assumed that the insurance company would deny the claim.
19	f	The doctors that the nurses assisted during the operation have ever assumed that the insurance company would deny the claim.
19	g	No doctors that the nurses assisted during the operation assumed that the insurance company would at any time deny the claim.
19	h	The doctors that no nurses assisted during the operation assumed that the insurance company would at any time deny the claim.
19	i	The doctors that the nurses assisted during the operation assumed that the insurance company would at any time deny the claim.
19	j	No doctors that the nurses assisted during the operation have at any time assumed that the insurance company would deny the claim.
19	k	The doctors that no nurses assisted during the operation have at any time assumed that the insurance company would deny the claim.
19	l	The doctors that the nurses assisted during the operation have at any time assumed that the insurance company would deny the claim.
19	m	No doctors that the nurses assisted during the operation assumed that the insurance company would deny the claim at any time.
19	n	The doctors that no nurses assisted during the operation assumed that the insurance company would deny the claim at any time.
19	o	The doctors that the nurses assisted during the operation assumed that the insurance company would deny the claim at any time.
20	a	No criminals that the witnesses could identify in the courtroom suspected that the jury would ever find out about the evidence.

20	b	The criminals that no witnesses could identify in the courtroom suspected that the jury would ever find out about the evidence.
20	c	The criminals that the witnesses could identify in the courtroom suspected that the jury would ever find out about the evidence.
20	d	No criminals that the witnesses could identify in the courtroom have ever suspected that the jury would find out about the evidence.
20	e	The criminals that no witnesses could identify in the courtroom have ever suspected that the jury would find out about the evidence.
20	f	The criminals that the witnesses could identify in the courtroom have ever suspected that the jury would find out about the evidence.
20	g	No criminals that the witnesses could identify in the courtroom suspected that the jury would at any time find out about the evidence.
20	h	The criminals that no witnesses could identify in the courtroom suspected that the jury would at any time find out about the evidence.
20	i	The criminals that the witnesses could identify in the courtroom suspected that the jury would at any time find out about the evidence.
20	j	No criminals that the witnesses could identify in the courtroom have at any time suspected that the jury would find out about the evidence.
20	k	The criminals that no witnesses could identify in the courtroom have at any time suspected that the jury would find out about the evidence.
20	l	The criminals that the witnesses could identify in the courtroom have at any time suspected that the jury would find out about the evidence.
20	m	No criminals that the witnesses could identify in the courtroom suspected that the jury would find out about the evidence at any time.
20	n	The criminals that no witnesses could identify in the courtroom suspected that the jury would find out about the evidence at any time.
20	o	The criminals that the witnesses could identify in the courtroom suspected that the jury would find out about the evidence at any time.
21	a	No actresses that the critics liked in the movie expected that the director would ever dismiss their contributions.
21	b	The actresses that no critics liked in the movie expected that the director would ever dismiss their contributions.

21	c	The actresses that the critics liked in the movie expected that the director would ever dismiss their contributions.
21	d	No actresses that the critics liked in the movie have ever expected that the director would dismiss their contributions.
21	e	The actresses that no critics liked in the movie have ever expected that the director would dismiss their contributions.
21	f	The actresses that the critics liked in the movie have ever expected that the director would dismiss their contributions.
21	g	No actresses that the critics liked in the movie expected that the director would in any way dismiss their contributions.
21	h	The actresses that no critics liked in the movie expected that the director would in any way dismiss their contributions.
21	i	The actresses that the critics liked in the movie expected that the director would in any way dismiss their contributions.
21	j	No actresses that the critics liked in the movie have in any way expected that the director would dismiss their contributions.
21	k	The actresses that no critics liked in the movie have in any way expected that the director would dismiss their contributions.
21	l	The actresses that the critics liked in the movie have in any way expected that the director would dismiss their contributions.
21	m	No actresses that the critics liked in the movie expected that the director would dismiss their contributions in any way.
21	n	The actresses that no critics liked in the movie expected that the director would dismiss their contributions in any way.
21	o	The actresses that the critics liked in the movie expected that the director would dismiss their contributions in any way.
22	a	No legislators that the congressmen consulted about the proposal suggested that the government should ever increase military spending for the war.
22	b	The legislators that no congressmen consulted about the proposal suggested that the government should ever increase military spending for the war.
22	c	The legislators that the congressmen consulted about the proposal suggested that the government should ever increase military spending for the war.

22	d	No legislators that the congressmen consulted about the proposal have ever suggested that the government should increase military spending for the war.
22	e	The legislators that no congressmen consulted about the proposal have ever suggested that the government should increase military spending for the war.
22	f	The legislators that the congressmen consulted about the proposal have ever suggested that the government should increase military spending for the war.
22	g	No legislators that the congressmen consulted about the proposal suggested that the government should at any time increase military spending for the war.
22	h	The legislators that no congressmen consulted about the proposal suggested that the government should at any time increase military spending for the war.
22	i	The legislators that the congressmen consulted about the proposal suggested that the government should at any time increase military spending for the war.
22	j	No legislators that the congressmen consulted about the proposal have at any time suggested that the government should increase military spending for the war.
22	k	The legislators that no congressmen consulted about the proposal have at any time suggested that the government should increase military spending for the war.
22	l	The legislators that the congressmen consulted about the proposal have at any time suggested that the government should increase military spending for the war.
22	m	No legislators that the congressmen consulted about the proposal suggested that the government should increase military spending for the war at any time.
22	n	The legislators that no congressmen consulted about the proposal suggested that the government should increase military spending for the war at any time.
22	o	The legislators that the congressmen consulted about the proposal suggested that the government should increase military spending for the war at any time.
23	a	No politicians that the journalists endorsed in the newspaper thought that the election would ever cause such a huge scandal.
23	b	The politicians that no journalists endorsed in the newspaper thought that the election would ever cause such a huge scandal.
23	c	The politicians that the journalists endorsed in the newspaper thought that the election would ever cause such a huge scandal.
23	d	No politicians that the journalists endorsed in the newspaper have ever thought that the election would cause such a huge scandal.

23	e	The politicians that no journalists endorsed in the newspaper have ever thought that the election would cause such a huge scandal.
23	f	The politicians that the journalists endorsed in the newspaper have ever thought that the election would cause such a huge scandal.
23	g	No politicians that the journalists endorsed in the newspaper thought that the election would in any way cause such a huge scandal.
23	h	The politicians that no journalists endorsed in the newspaper thought that the election would in any way cause such a huge scandal.
23	i	The politicians that the journalists endorsed in the newspaper thought that the election would in any way cause such a huge scandal.
23	j	No politicians that the journalists endorsed in the newspaper have in any way thought that the election would cause such a huge scandal.
23	k	The politicians that no journalists endorsed in the newspaper have in any way thought that the election would cause such a huge scandal.
23	l	The politicians that the journalists endorsed in the newspaper have in any way thought that the election would cause such a huge scandal.
23	m	No politicians that the journalists endorsed in the newspaper thought that the election would cause such a huge scandal in any way.
23	n	The politicians that no journalists endorsed in the newspaper thought that the election would cause such a huge scandal in any way.
23	o	The politicians that the journalists endorsed in the newspaper thought that the election would cause such a huge scandal in any way.
24	a	No teenagers that the parents trusted with a car expected that driving would ever be stressful.
24	b	The teenagers that no parents trusted with a car expected that driving would ever be stressful.
24	c	The teenagers that the parents trusted with a car expected that driving would ever be stressful.
24	d	No teenagers that the parents trusted with a car have ever expected that driving would be stressful.
24	e	The teenagers that no parents trusted with a car have ever expected that driving would be stressful.
24	f	The teenagers that the parents trusted with a car have ever expected that driving would be stressful.
24	g	No teenagers that the parents trusted with a car expected that driving would in any way be stressful.
24	h	The teenagers that no parents trusted with a car expected that driving would in any way be stressful.
24	i	The teenagers that the parents trusted with a car expected that driving would in any way be stressful.
24	j	No teenagers that the parents trusted with a car have in any way expected that driving would be stressful.
24	k	The teenagers that no parents trusted with a car have in any way expected that driving would be stressful.

24	l	The teenagers that the parents trusted with a car have in any way expected that driving would be stressful.
24	m	No teenagers that the parents trusted with a car expected that driving would be stressful in any way.
24	n	The teenagers that no parents trusted with a car expected that driving would be stressful in any way.
24	o	The teenagers that the parents trusted with a car expected that driving would be stressful in any way.
25	a	No survivors that the medics could treat with a first-aid kit expected that a full recovery would ever become possible.
25	b	The survivors that no medics could treat with a first-aid kit expected that a full recovery would ever become possible.
25	c	The survivors that the medics could treat with a first-aid kit expected that a full recovery would ever become possible.
25	d	No survivors that the medics could treat with a first-aid kit have ever expected that a full recovery would become possible.
25	e	The survivors that no medics could treat with a first-aid kit have ever expected that a full recovery would become possible.
25	f	The survivors that the medics could treat with a first-aid kit have ever expected that a full recovery would become possible.
25	g	No survivors that the medics could treat with a first-aid kit expected that a full recovery would at any time become possible.
25	h	The survivors that no medics could treat with a first-aid kit expected that a full recovery would at any time become possible.
25	i	The survivors that the medics could treat with a first-aid kit expected that a full recovery would at any time become possible.
25	j	No survivors that the medics could treat with a first-aid kit have at any time expected that a full recovery would become possible.
25	k	The survivors that no medics could treat with a first-aid kit have at any time expected that a full recovery would become possible.
25	l	The survivors that the medics could treat with a first-aid kit have at any time expected that a full recovery would become possible.
25	m	No survivors that the medics could treat with a first-aid kit expected that a full recovery would become possible at any time.
25	n	The survivors that no medics could treat with a first-aid kit expected that a full recovery would become possible at any time.

25	o	The survivors that the medics could treat with a first-aid kit expected that a full recovery would become possible at any time.
26	a	No athletes that the coaches recruited for the team anticipated that the scandal would ever receive so much media coverage.
26	b	The athletes that no coaches recruited for the team anticipated that the scandal would ever receive so much media coverage.
26	c	The athletes that the coaches recruited for the team anticipated that the scandal would ever receive so much media coverage.
26	d	No athletes that the coaches recruited for the team have ever anticipated that the scandal would receive so much media coverage.
26	e	The athletes that no coaches recruited for the team have ever anticipated that the scandal would receive so much media coverage.
26	f	The athletes that the coaches recruited for the team have ever anticipated that the scandal would receive so much media coverage.
26	g	No athletes that the coaches recruited for the team anticipated that the scandal would at any time receive so much media coverage.
26	h	The athletes that no coaches recruited for the team anticipated that the scandal would at any time receive so much media coverage.
26	i	The athletes that the coaches recruited for the team anticipated that the scandal would at any time receive so much media coverage.
26	j	No athletes that the coaches recruited for the team have at any time anticipated that the scandal would receive so much media coverage.
26	k	The athletes that no coaches recruited for the team have at any time anticipated that the scandal would receive so much media coverage.
26	l	The athletes that the coaches recruited for the team have at any time anticipated that the scandal would receive so much media coverage.
26	m	No athletes that the coaches recruited for the team anticipated that the scandal would receive so much media coverage at any time.
26	n	The athletes that no coaches recruited for the team anticipated that the scandal would receive so much media coverage at any time.
26	o	The athletes that the coaches recruited for the team anticipated that the scandal would receive so much media coverage at any time.

27	a	No congressmen that the citizens supported during the crisis assumed that the treasury would ever address the national debt.
27	b	The congressmen that no citizens supported during the crisis assumed that the treasury would ever address the national debt.
27	c	The congressmen that the citizens supported during the crisis assumed that the treasury would ever address the national debt.
27	d	No congressmen that the citizens supported during the crisis have ever assumed that the treasury would address the national debt.
27	e	The congressmen that no citizens supported during the crisis have ever assumed that the treasury would address the national debt.
27	f	The congressmen that the citizens supported during the crisis have ever assumed that the treasury would address the national debt.
27	g	No congressmen that the citizens supported during the crisis assumed that the treasury would in any way address the national debt.
27	h	The congressmen that no citizens supported during the crisis assumed that the treasury would in any way address the national debt.
27	i	The congressmen that the citizens supported during the crisis assumed that the treasury would in any way address the national debt.
27	j	No congressmen that the citizens supported during the crisis have in any way assumed that the treasury would address the national debt.
27	k	The congressmen that no citizens supported during the crisis have in any way assumed that the treasury would address the national debt.
27	l	The congressmen that the citizens supported during the crisis have in any way assumed that the treasury would address the national debt.
27	m	No congressmen that the citizens supported during the crisis assumed that the treasury would address the national debt in any way.
27	n	The congressmen that no citizens supported during the crisis assumed that the treasury would address the national debt in any way.
27	o	The congressmen that the citizens supported during the crisis assumed that the treasury would address the national debt in any way.
28	a	No professors that the students visited during office hours anticipated that the workload would ever be so overwhelming.

28	b	The professors that no students visited during office hours anticipated that the workload would ever be so overwhelming.
28	c	The professors that the students visited during office hours anticipated that the workload would ever be so overwhelming.
28	d	No professors that the students visited during office hours have ever anticipated that the workload would be so overwhelming.
28	e	The professors that no students visited during office hours have ever anticipated that the workload would be so overwhelming.
28	f	The professors that the students visited during office hours have ever anticipated that the workload would be so overwhelming.
28	g	No professors that the students visited during office hours anticipated that the workload would in any way be so overwhelming.
28	h	The professors that no students visited during office hours anticipated that the workload would in any way be so overwhelming.
28	i	The professors that the students visited during office hours anticipated that the workload would in any way be so overwhelming.
28	j	No professors that the students visited during office hours have in any way anticipated that the workload would be so overwhelming.
28	k	The professors that no students visited during office hours have in any way anticipated that the workload would be so overwhelming.
28	l	The professors that the students visited during office hours have in any way anticipated that the workload would be so overwhelming.
28	m	No professors that the students visited during office hours anticipated that the workload would be so overwhelming in any way.
28	n	The professors that no students visited during office hours anticipated that the workload would be so overwhelming in any way.
28	o	The professors that the students visited during office hours anticipated that the workload would be so overwhelming in any way.
29	a	No actors that the judges nominated for an award expected that the movie would ever be a blockbuster hit.
29	b	The actors that no judges nominated for an award expected that the movie would ever be a blockbuster hit.
29	c	The actors that the judges nominated for an award expected that the movie would ever be a blockbuster hit.

29	d	No actors that the judges nominated for an award have ever expected that the movie would be a blockbuster hit.
29	e	The actors that no judges nominated for an award have ever expected that the movie would be a blockbuster hit.
29	f	The actors that the judges nominated for an award have ever expected that the movie would be a blockbuster hit.
29	g	No actors that the judges nominated for an award expected that the movie would in any way be a blockbuster hit.
29	h	The actors that no judges nominated for an award expected that the movie would in any way be a blockbuster hit.
29	i	The actors that the judges nominated for an award expected that the movie would in any way be a blockbuster hit.
29	j	No actors that the judges nominated for an award have in any way expected that the movie would be a blockbuster hit.
29	k	The actors that no judges nominated for an award have in any way expected that the movie would be a blockbuster hit.
29	l	The actors that the judges nominated for an award have in any way expected that the movie would be a blockbuster hit.
29	m	No actors that the judges nominated for an award expected that the movie would be a blockbuster hit in any way.
29	n	The actors that no judges nominated for an award expected that the movie would be a blockbuster hit in any way.
29	o	The actors that the judges nominated for an award expected that the movie would be a blockbuster hit in any way.
30	a	No actresses that the directors auditioned for the role thought that the movie would ever cause controversy.
30	b	The actresses that no directors auditioned for the role thought that the movie would ever cause controversy.
30	c	The actresses that the directors auditioned for the role thought that the movie would ever cause controversy.
30	d	No actresses that the directors auditioned for the role have ever thought that the movie would cause controversy.
30	e	The actresses that no directors auditioned for the role have ever thought that the movie would cause controversy.
30	f	The actresses that the directors auditioned for the role have ever thought that the movie would cause controversy.

30	g	No actresses that the directors auditioned for the role thought that the movie would in any way cause controversy.
30	h	The actresses that no directors auditioned for the role thought that the movie would in any way cause controversy.
30	i	The actresses that the directors auditioned for the role thought that the movie would in any way cause controversy.
30	j	No actresses that the directors auditioned for the role have in any way thought that the movie would cause controversy.
30	k	The actresses that no directors auditioned for the role have in any way thought that the movie would cause controversy.
30	l	The actresses that the directors auditioned for the role have in any way thought that the movie would cause controversy.
30	m	No actresses that the directors auditioned for the role thought that the movie would cause controversy in any way.
30	n	The actresses that no directors auditioned for the role thought that the movie would cause controversy in any way.
30	o	The actresses that the directors auditioned for the role thought that the movie would cause controversy in any way.
31	a	No champions that the competitors defeated in the race expected that the coach would ever receive a life-time achievement award.
31	b	The champions that no competitors defeated in the race expected that the coach would ever receive a life-time achievement award.
31	c	The champions that the competitors defeated in the race expected that the coach would ever receive a life-time achievement award.
31	d	No champions that the competitors defeated in the race have ever expected that the coach would receive a life-time achievement award.
31	e	The champions that no competitors defeated in the race have ever expected that the coach would receive a life-time achievement award.
31	f	The champions that the competitors defeated in the race have ever expected that the coach would receive a life-time achievement award.
31	g	No champions that the competitors defeated in the race expected that the coach would at any time receive a life-time achievement award.

31	h	The champions that no competitors defeated in the race expected that the coach would at any time receive a life-time achievement award.
31	i	The champions that the competitors defeated in the race expected that the coach would at any time receive a life-time achievement award.
31	j	No champions that the competitors defeated in the race have at any time expected that the coach would receive a life-time achievement award.
31	k	The champions that no competitors defeated in the race have at any time expected that the coach would receive a life-time achievement award.
31	l	The champions that the competitors defeated in the race have at any time expected that the coach would receive a life-time achievement award.
31	m	No champions that the competitors defeated in the race expected that the coach would receive a life-time achievement award at any time.
31	n	The champions that no competitors defeated in the race expected that the coach would receive a life-time achievement award at any time.
31	o	The champions that the competitors defeated in the race expected that the coach would receive a life-time achievement award at any time.
32	a	No artists that the collectors regarded very highly suggested that the gallery should ever use cheap frames for the expensive paintings.
32	b	The artists that no collectors regarded very highly suggested that the gallery should ever use cheap frames for the expensive paintings.
32	c	The artists that the collectors regarded very highly suggested that the gallery should ever use cheap frames for the expensive paintings.
32	d	No artists that the collectors regarded very highly have ever suggested that the gallery should use cheap frames for the expensive paintings.
32	e	The artists that no collectors regarded very highly have ever suggested that the gallery should use cheap frames for the expensive paintings.
32	f	The artists that the collectors regarded very highly have ever suggested that the gallery should use cheap frames for the expensive paintings.
32	g	No artists that the collectors regarded very highly suggested that the gallery should at any time use cheap frames for the expensive paintings.
32	h	The artists that no collectors regarded very highly suggested that the gallery should at any time use cheap frames for the expensive paintings.

32	i	The artists that the collectors regarded very highly suggested that the gallery should at any time use cheap frames for the expensive paintings.
32	j	No artists that the collectors regarded very highly have at any time suggested that the gallery should use cheap frames for the expensive paintings.
32	k	The artists that no collectors regarded very highly have at any time suggested that the gallery should use cheap frames for the expensive paintings.
32	l	The artists that the collectors regarded very highly have at any time suggested that the gallery should use cheap frames for the expensive paintings.
32	m	No artists that the collectors regarded very highly suggested that the gallery should use cheap frames for the expensive paintings at any time.
32	n	The artists that no collectors regarded very highly suggested that the gallery should use cheap frames for the expensive paintings at any time.
32	o	The artists that the collectors regarded very highly suggested that the gallery should use cheap frames for the expensive paintings at any time.
33	a	No scientists that the reporters cited in the story believed that the public would ever care about the new discovery.
33	b	The scientists that no reporters cited in the story believed that the public would ever care about the new discovery.
33	c	The scientists that the reporters cited in the story believed that the public would ever care about the new discovery.
33	d	No scientists that the reporters cited in the story have ever believed that the public would care about the new discovery.
33	e	The scientists that no reporters cited in the story have ever believed that the public would care about the new discovery.
33	f	The scientists that the reporters cited in the story have ever believed that the public would care about the new discovery.
33	g	No scientists that the reporters cited in the story believed that the public would in any way care about the new discovery.
33	h	The scientists that no reporters cited in the story believed that the public would in any way care about the new discovery.
33	i	The scientists that the reporters cited in the story believed that the public would in any way care about the new discovery.

33	j	No scientists that the reporters cited in the story have in any way believed that the public would care about the new discovery.
33	k	The scientists that no reporters cited in the story have in any way believed that the public would care about the new discovery.
33	l	The scientists that the reporters cited in the story have in any way believed that the public would care about the new discovery.
33	m	No scientists that the reporters cited in the story believed that the public would care about the new discovery in any way.
33	n	The scientists that no reporters cited in the story believed that the public would care about the new discovery in any way.
33	o	The scientists that the reporters cited in the story believed that the public would care about the new discovery in any way.
34	a	No teenagers that the teachers motivated before the test claimed that the parents should ever help with assignments.
34	b	The teenagers that no teachers motivated before the test claimed that the parents should ever help with assignments.
34	c	The teenagers that the teachers motivated before the test claimed that the parents should ever help with assignments.
34	d	No teenagers that the teachers motivated before the test have ever claimed that the parents should help with assignments.
34	e	The teenagers that no teachers motivated before the test have ever claimed that the parents should help with assignments.
34	f	The teenagers that the teachers motivated before the test have ever claimed that the parents should help with assignments.
34	g	No teenagers that the teachers motivated before the test claimed that the parents should in any way help with assignments.
34	h	The teenagers that no teachers motivated before the test claimed that the parents should in any way help with assignments.
34	i	The teenagers that the teachers motivated before the test claimed that the parents should in any way help with assignments.
34	j	No teenagers that the teachers motivated before the test have in any way claimed that the parents should help with assignments.

34	k	The teenagers that no teachers motivated before the test have in any way claimed that the parents should help with assignments.
34	l	The teenagers that the teachers motivated before the test have in any way claimed that the parents should help with assignments.
34	m	No teenagers that the teachers motivated before the test claimed that the parents should help with assignments in any way.
34	n	The teenagers that no teachers motivated before the test claimed that the parents should help with assignments in any way.
34	o	The teenagers that the teachers motivated before the test claimed that the parents should help with assignments in any way.
35	a	No students that the professors could tutor on the weekend thought that the assignments would ever be useful outside of school.
35	b	The students that no professors could tutor on the weekend thought that the assignments would ever be useful outside of school.
35	c	The students that the professors could tutor on the weekend thought that the assignments would ever be useful outside of school.
35	d	No students that the professors could tutor on the weekend have ever thought that the assignments would be useful outside of school.
35	e	The students that no professors could tutor on the weekend have ever thought that the assignments would be useful outside of school.
35	f	The students that the professors could tutor on the weekend have ever thought that the assignments would be useful outside of school.
35	g	No students that the professors could tutor on the weekend thought that the assignments would in any way be useful outside of school.
35	h	The students that no professors could tutor on the weekend thought that the assignments would in any way be useful outside of school.
35	i	The students that the professors could tutor on the weekend thought that the assignments would in any way be useful outside of school.
35	j	No students that the professors could tutor on the weekend have in any way thought that the assignments would be useful outside of school.
35	k	The students that no professors could tutor on the weekend have in any way thought that the assignments would be useful outside of school.

35	l	The students that the professors could tutor on the weekend have in any way thought that the assignments would be useful outside of school.
35	m	No students that the professors could tutor on the weekend thought that the assignments would be useful outside of school in any way.
35	n	The students that no professors could tutor on the weekend thought that the assignments would be useful outside of school in any way.
35	o	The students that the professors could tutor on the weekend thought that the assignments would be useful outside of school in any way.
36	a	No protestors that the reporters interviewed on live television expected that the mayor would ever give in to the numerous demands.
36	b	The protestors that no reporters interviewed on live television expected that the mayor would ever give in to the numerous demands.
36	c	The protestors that the reporters interviewed on live television expected that the mayor would ever give in to the numerous demands.
36	d	No protestors that the reporters interviewed on live television have ever expected that the mayor would give in to the numerous demands.
36	e	The protestors that no reporters interviewed on live television have ever expected that the mayor would give in to the numerous demands.
36	f	The protestors that the reporters interviewed on live television have ever expected that the mayor would give in to the numerous demands.
36	g	No protestors that the reporters interviewed on live television expected that the mayor would at any time give in to the numerous demands.
36	h	The protestors that no reporters interviewed on live television expected that the mayor would at any time give in to the numerous demands.
36	i	The protestors that the reporters interviewed on live television expected that the mayor would at any time give in to the numerous demands.
36	j	No protestors that the reporters interviewed on live television have at any time expected that the mayor would give in to the numerous demands.
36	k	The protestors that no reporters interviewed on live television have at any time expected that the mayor would give in to the numerous demands.
36	l	The protestors that the reporters interviewed on live television have at any time expected that the mayor would give in to the numerous demands.

36	m	No protestors that the reporters interviewed on live television expected that the mayor would give in to the numerous demands at any time.
36	n	The protestors that no reporters interviewed on live television expected that the mayor would give in to the numerous demands at any time.
36	o	The protestors that the reporters interviewed on live television expected that the mayor would give in to the numerous demands at any time.
37	a	No waitresses that the cooks berated during the dinner rush expected that the stress of the job would ever lessen.
37	b	The waitresses that no cooks berated during the dinner rush expected that the stress of the job would ever lessen.
37	c	The waitresses that the cooks berated during the dinner rush expected that the stress of the job would ever lessen.
37	d	No waitresses that the cooks berated during the dinner rush have ever expected that the stress of the job would lessen.
37	e	The waitresses that no cooks berated during the dinner rush have ever expected that the stress of the job would lessen.
37	f	The waitresses that the cooks berated during the dinner rush have ever expected that the stress of the job would lessen.
37	g	No waitresses that the cooks berated during the dinner rush expected that the stress of the job would in any way lessen.
37	h	The waitresses that no cooks berated during the dinner rush expected that the stress of the job would in any way lessen.
37	i	The waitresses that the cooks berated during the dinner rush expected that the stress of the job would in any way lessen.
37	j	No waitresses that the cooks berated during the dinner rush have in any way expected that the stress of the job would lessen.
37	k	The waitresses that no cooks berated during the dinner rush have in any way expected that the stress of the job would lessen.
37	l	The waitresses that the cooks berated during the dinner rush have in any way expected that the stress of the job would lessen.
37	m	No waitresses that the cooks berated during the dinner rush expected that the stress of the job would lessen in any way.

37	n	The waitresses that no cooks berated during the dinner rush expected that the stress of the job would lessen in any way.
37	o	The waitresses that the cooks berated during the dinner rush expected that the stress of the job would lessen in any way.
38	a	No workers that the researchers relied on to provide data anticipated that the website would ever become unavailable.
38	b	The workers that no researchers relied on to provide data anticipated that the website would ever become unavailable.
38	c	The workers that the researchers relied on to provide data anticipated that the website would ever become unavailable.
38	d	No workers that the researchers relied on to provide data have ever anticipated that the website would become unavailable.
38	e	The workers that no researchers relied on to provide data have ever anticipated that the website would become unavailable.
38	f	The workers that the researchers relied on to provide data have ever anticipated that the website would become unavailable.
38	g	No workers that the researchers relied on to provide data anticipated that the website would at any time become unavailable.
38	h	The workers that no researchers relied on to provide data anticipated that the website would at any time become unavailable.
38	i	The workers that the researchers relied on to provide data anticipated that the website would at any time become unavailable.
38	j	No workers that the researchers relied on to provide data have at any time anticipated that the website would become unavailable.
38	k	The workers that no researchers relied on to provide data have at any time anticipated that the website would become unavailable.
38	l	The workers that the researchers relied on to provide data have at any time anticipated that the website would become unavailable.
38	m	No workers that the researchers relied on to provide data anticipated that the website would become unavailable at any time.
38	n	The workers that no researchers relied on to provide data anticipated that the website would become unavailable at any time.

38	o	The workers that the researchers relied on to provide data anticipated that the website would become unavailable at any time.
39	a	No athletes that the opponents defeated in the big game imagined that the fans would ever turn against the team.
39	b	The athletes that no opponents defeated in the big game imagined that the fans would ever turn against the team.
39	c	The athletes that the opponents defeated in the big game imagined that the fans would ever turn against the team.
39	d	No athletes that the opponents defeated in the big game have ever imagined that the fans would turn against the team.
39	e	The athletes that no opponents defeated in the big game have ever imagined that the fans would turn against the team.
39	f	The athletes that the opponents defeated in the big game have ever imagined that the fans would turn against the team.
39	g	No athletes that the opponents defeated in the big game imagined that the fans would in any way turn against the team.
39	h	The athletes that no opponents defeated in the big game imagined that the fans would in any way turn against the team.
39	i	The athletes that the opponents defeated in the big game imagined that the fans would in any way turn against the team.
39	j	No athletes that the opponents defeated in the big game have in any way imagined that the fans would turn against the team.
39	k	The athletes that no opponents defeated in the big game have in any way imagined that the fans would turn against the team.
39	l	The athletes that the opponents defeated in the big game have in any way imagined that the fans would turn against the team.
39	m	No athletes that the opponents defeated in the big game imagined that the fans would turn against the team in any way.
39	n	The athletes that no opponents defeated in the big game imagined that the fans would turn against the team in any way.
39	o	The athletes that the opponents defeated in the big game imagined that the fans would turn against the team in any way.

40	a	No engineers that the clients trusted to evaluate the construction predicted that the building would ever become structurally unsound.
40	b	The engineers that no clients trusted to evaluate the construction predicted that the building would ever become structurally unsound.
40	c	The engineers that the clients trusted to evaluate the construction predicted that the building would ever become structurally unsound.
40	d	No engineers that the clients trusted to evaluate the construction have ever predicted that the building would become structurally unsound.
40	e	The engineers that no clients trusted to evaluate the construction have ever predicted that the building would become structurally unsound.
40	f	The engineers that the clients trusted to evaluate the construction have ever predicted that the building would become structurally unsound.
40	g	No engineers that the clients trusted to evaluate the construction predicted that the building would at any time become structurally unsound.
40	h	The engineers that no clients trusted to evaluate the construction predicted that the building would at any time become structurally unsound.
40	i	The engineers that the clients trusted to evaluate the construction predicted that the building would at any time become structurally unsound.
40	j	No engineers that the clients trusted to evaluate the construction have at any time predicted that the building would become structurally unsound.
40	k	The engineers that no clients trusted to evaluate the construction have at any time predicted that the building would become structurally unsound.
40	l	The engineers that the clients trusted to evaluate the construction have at any time predicted that the building would become structurally unsound.
40	m	No engineers that the clients trusted to evaluate the construction predicted that the building would become structurally unsound at any time.
40	n	The engineers that no clients trusted to evaluate the construction predicted that the building would become structurally unsound at any time.
40	o	The engineers that the clients trusted to evaluate the construction predicted that the building would become structurally unsound at any time.
41	a	No publishers that the novelists criticized for their narrow focus thought that the readers would ever embrace an unconventional storyline.

41	b	The publishers that no novelists criticized for their narrow focus thought that the readers would ever embrace an unconventional storyline.
41	c	The publishers that the novelists criticized for their narrow focus thought that the readers would ever embrace an unconventional storyline.
41	d	No publishers that the novelists criticized for their narrow focus have ever thought that the readers would embrace an unconventional storyline.
41	e	The publishers that no novelists criticized for their narrow focus have ever thought that the readers would embrace an unconventional storyline.
41	f	The publishers that the novelists criticized for their narrow focus have ever thought that the readers would embrace an unconventional storyline.
41	g	No publishers that the novelists criticized for their narrow focus thought that the readers would in any way embrace an unconventional storyline.
41	h	The publishers that no novelists criticized for their narrow focus thought that the readers would in any way embrace an unconventional storyline.
41	i	The publishers that the novelists criticized for their narrow focus thought that the readers would in any way embrace an unconventional storyline.
41	j	No publishers that the novelists criticized for their narrow focus have in any way thought that the readers would embrace an unconventional storyline.
41	k	The publishers that no novelists criticized for their narrow focus have in any way thought that the readers would embrace an unconventional storyline.
41	l	The publishers that the novelists criticized for their narrow focus have in any way thought that the readers would embrace an unconventional storyline.
41	m	No publishers that the novelists criticized for their narrow focus thought that the readers would embrace an unconventional storyline in any way.
41	n	The publishers that no novelists criticized for their narrow focus thought that the readers would embrace an unconventional storyline in any way.
41	o	The publishers that the novelists criticized for their narrow focus thought that the readers would embrace an unconventional storyline in any way.
42	a	No baristas that the customers chatted with every day expected that the coffee shop would ever be forced to close.
42	b	The baristas that no customers chatted with every day expected that the coffee shop would ever be forced to close.

42	c	The baristas that the customers chatted with every day expected that the coffee shop would ever be forced to close.
42	d	No baristas that the customers chatted with every day have ever expected that the coffee shop would be forced to close.
42	e	The baristas that no customers chatted with every day have ever expected that the coffee shop would be forced to close.
42	f	The baristas that the customers chatted with every day have ever expected that the coffee shop would be forced to close.
42	g	No baristas that the customers chatted with every day expected that the coffee shop would at any time be forced to close.
42	h	The baristas that no customers chatted with every day expected that the coffee shop would at any time be forced to close.
42	i	The baristas that the customers chatted with every day expected that the coffee shop would at any time be forced to close.
42	j	No baristas that the customers chatted with every day have at any time expected that the coffee shop would be forced to close.
42	k	The baristas that no customers chatted with every day have at any time expected that the coffee shop would be forced to close.
42	l	The baristas that the customers chatted with every day have at any time expected that the coffee shop would be forced to close.
42	m	No baristas that the customers chatted with every day expected that the coffee shop would be forced to close at any time.
42	n	The baristas that no customers chatted with every day expected that the coffee shop would be forced to close at any time.
42	o	The baristas that the customers chatted with every day expected that the coffee shop would be forced to close at any time.
43	a	No CEOs that the environmentalists denounced in their protests suggested that the public outcry would ever influence business practices.
43	b	The CEOs that no environmentalists denounced in their protests suggested that the public outcry would ever influence business practices.
43	c	The CEOs that the environmentalists denounced in their protests suggested that the public outcry would ever influence business practices.

43	d	No CEOs that the environmentalists denounced in their protests have ever suggested that the public outcry would influence business practices.
43	e	The CEOs that no environmentalists denounced in their protests have ever suggested that the public outcry would influence business practices.
43	f	The CEOs that the environmentalists denounced in their protests have ever suggested that the public outcry would influence business practices.
43	g	No CEOs that the environmentalists denounced in their protests suggested that the public outcry would in any way influence business practices.
43	h	The CEOs that no environmentalists denounced in their protests suggested that the public outcry would in any way influence business practices.
43	i	The CEOs that the environmentalists denounced in their protests suggested that the public outcry would in any way influence business practices.
43	j	No CEOs that the environmentalists denounced in their protests have in any way suggested that the public outcry would influence business practices.
43	k	The CEOs that no environmentalists denounced in their protests have in any way suggested that the public outcry would influence business practices.
43	l	The CEOs that the environmentalists denounced in their protests have in any way suggested that the public outcry would influence business practices.
43	m	No CEOs that the environmentalists denounced in their protests suggested that the public outcry would influence business practices in any way.
43	n	The CEOs that no environmentalists denounced in their protests suggested that the public outcry would influence business practices in any way.
43	o	The CEOs that the environmentalists denounced in their protests suggested that the public outcry would influence business practices in any way.
44	a	No caterers that the guests approached after the wedding anticipated that the business would ever fail to find new customers.
44	b	The caterers that no guests approached after the wedding anticipated that the business would ever fail to find new customers.
44	c	The caterers that the guests approached after the wedding anticipated that the business would ever fail to find new customers.
44	d	No caterers that the guests approached after the wedding have ever anticipated that the business would fail to find new customers.

44	e	The caterers that no guests approached after the wedding have ever anticipated that the business would fail to find new customers.
44	f	The caterers that the guests approached after the wedding have ever anticipated that the business would fail to find new customers.
44	g	No caterers that the guests approached after the wedding anticipated that the business would at any time fail to find new customers.
44	h	The caterers that no guests approached after the wedding anticipated that the business would at any time fail to find new customers.
44	i	The caterers that the guests approached after the wedding anticipated that the business would at any time fail to find new customers.
44	j	No caterers that the guests approached after the wedding have at any time anticipated that the business would fail to find new customers.
44	k	The caterers that no guests approached after the wedding have at any time anticipated that the business would fail to find new customers.
44	l	The caterers that the guests approached after the wedding have at any time anticipated that the business would fail to find new customers.
44	m	No caterers that the guests approached after the wedding anticipated that the business would fail to find new customers at any time.
44	n	The caterers that no guests approached after the wedding anticipated that the business would fail to find new customers at any time.
44	o	The caterers that the guests approached after the wedding anticipated that the business would fail to find new customers at any time.
45	a	No bartenders that the customers flirted with to get free drinks suspected that the drunken crowd would ever become unruly.
45	b	The bartenders that no customers flirted with to get free drinks suspected that the drunken crowd would ever become unruly.
45	c	The bartenders that the customers flirted with to get free drinks suspected that the drunken crowd would ever become unruly.
45	d	No bartenders that the customers flirted with to get free drinks have ever suspected that the drunken crowd would become unruly.
45	e	The bartenders that no customers flirted with to get free drinks have ever suspected that the drunken crowd would become unruly.

45	f	The bartenders that the customers flirted with to get free drinks have ever suspected that the drunken crowd would become unruly.
45	g	No bartenders that the customers flirted with to get free drinks suspected that the drunken crowd would in any way become unruly.
45	h	The bartenders that no customers flirted with to get free drinks suspected that the drunken crowd would in any way become unruly.
45	i	The bartenders that the customers flirted with to get free drinks suspected that the drunken crowd would in any way become unruly.
45	j	No bartenders that the customers flirted with to get free drinks have in any way suspected that the drunken crowd would become unruly.
45	k	The bartenders that no customers flirted with to get free drinks have in any way suspected that the drunken crowd would become unruly.
45	l	The bartenders that the customers flirted with to get free drinks have in any way suspected that the drunken crowd would become unruly.
45	m	No bartenders that the customers flirted with to get free drinks suspected that the drunken crowd would become unruly in any way.
45	n	The bartenders that no customers flirted with to get free drinks suspected that the drunken crowd would become unruly in any way.
45	o	The bartenders that the customers flirted with to get free drinks suspected that the drunken crowd would become unruly in any way.

Table A.14: Full experimental stimuli for Experiment 14

A.15 Experiment 15

1	a	Mary thought that no authors that the critics recommended had ever written a best-selling novel.
1	b	Mary thought that the authors that no critics recommended had ever written a best-selling novel.
1	c	Mary didn't think that the authors that the critics recommended had ever written a best-selling novel.
1	d	Mary thought that the authors that the critics didn't recommend had ever written a best-selling novel.
1	e	Mary thought that the authors that the critics recommended had ever written a best-selling novel.
1	f	Mary thought that no authors that admired the critics had ever written a best-selling novel.
1	g	Mary thought that the authors that admired no critics had ever written a best-selling novel.
1	h	Mary didn't think that the authors that admired the critics had ever written a best-selling novel.

1	i	Mary thought that the authors that didn't admire the critics had ever written a best-selling novel.
1	j	Mary thought that the authors that admired the critics had ever written a best-selling novel.
2	a	John believed that no painters that the critics liked had ever created beautiful art.
2	b	John believed that the painters that no critics liked had ever created beautiful art.
2	c	John didn't believe that the painters that the critics liked had ever created beautiful art.
2	d	John believed that the painters that the critics didn't like had ever created beautiful art.
2	e	John believed that the painters that the critics liked had ever created beautiful art.
2	f	John believed that no painters that ignored the critics had ever created beautiful art.
2	g	John believed that the painters that ignored no critics had ever created beautiful art.
2	h	John didn't believe that the painters that ignored the critics had ever created beautiful art.
2	i	John believed that the painters that didn't ignore the critics had ever created beautiful art.
2	j	John believed that the painters that ignored the critics had ever created beautiful art.
3	a	Jennifer knew that no ambassadors that the diplomats consulted had ever seen the brutality of war.
3	b	Jennifer knew that the ambassadors that no diplomats consulted had ever seen the brutality of war.
3	c	Jennifer didn't know that the ambassadors that the diplomats consulted had ever seen the brutality of war.
3	d	Jennifer knew that the ambassadors that the diplomats didn't consult had ever seen the brutality of war.
3	e	Jennifer knew that the ambassadors that the diplomats consulted had ever seen the brutality of war.
3	f	Jennifer knew that no ambassadors that consulted the diplomats had ever seen the brutality of war.
3	g	Jennifer knew that the ambassadors that consulted no diplomats had ever seen the brutality of war.
3	h	Jennifer didn't know that the ambassadors that consulted the diplomats had ever seen the brutality of war.
3	i	Jennifer knew that the ambassadors that didn't consult the diplomats had ever seen the brutality of war.
3	j	Jennifer knew that the ambassadors that consulted the diplomats had ever seen the brutality of war.
4	a	Mark expected that no babysitters that the children disobeyed had ever requested a tip.
4	b	Mark expected that the babysitters that no children disobeyed had ever requested a tip.
4	c	Mark didn't expect that the babysitters that the children disobeyed had ever requested a tip.
4	d	Mark expected that the babysitters that the children didn't disobey had ever requested a tip.
4	e	Mark expected that the babysitters that the children disobeyed had ever requested a tip.
4	f	Mark expected that no babysitters that entertained the children had ever requested a tip.
4	g	Mark expected that the babysitters that entertained no children had ever requested a tip.
4	h	Mark didn't expect that the babysitters that entertained the children had ever requested a tip.
4	i	Mark expected that the babysitters that didn't entertain the children had ever requested a tip.
4	j	Mark expected that the babysitters that entertained the children had ever requested a tip.

5	a	Elizabeth reckoned that no customers that the salesmen assisted had ever complained to the manager.
5	b	Elizabeth reckoned that the customers that no salesmen assisted had ever complained to the manager.
5	c	Elizabeth didn't reckon that the customers that the salesmen assisted had ever complained to the manager.
5	d	Elizabeth reckoned that the customers that the salesmen didn't assist had ever complained to the manager.
5	e	Elizabeth reckoned that the customers that the salesmen assisted had ever complained to the manager.
5	f	Elizabeth reckoned that no customers that appreciated the salesmen had ever complained to the manager.
5	g	Elizabeth reckoned that the customers that appreciated no salesmen had ever complained to the manager.
5	h	Elizabeth didn't reckon that the customers that appreciated the salesmen had ever complained to the manager.
5	i	Elizabeth reckoned that the customers that didn't appreciate the salesmen had ever complained to the manager.
5	j	Elizabeth reckoned that the customers that appreciated the salesmen had ever complained to the manager.
6	a	Kimberly felt that no actresses that the directors favored had ever demonstrated true talent.
6	b	Kimberly felt that the actresses that no directors favored had ever demonstrated true talent.
6	c	Kimberly didn't feel that the actresses that the directors favored had ever demonstrated true talent.
6	d	Kimberly felt that the actresses that the directors didn't favor had ever demonstrated true talent.
6	e	Kimberly felt that the actresses that the directors favored had ever demonstrated true talent.
6	f	Kimberly felt that no actresses that awed the directors had ever demonstrated true talent.
6	g	Kimberly felt that the actresses that awed no directors had ever demonstrated true talent.
6	h	Kimberly didn't feel that the actresses that awed the directors had ever demonstrated true talent.
6	i	Kimberly felt that the actresses that didn't awe the directors had ever demonstrated true talent.
6	j	Kimberly felt that the actresses that awed the directors had ever demonstrated true talent.
7	a	James thought that no athletes that the competitors defeated had ever shown good sportsmanship.
7	b	James thought that the athletes that no competitors defeated had ever shown good sportsmanship.
7	c	James didn't think that the athletes that the competitors defeated had ever shown good sportsmanship.
7	d	James thought that the athletes that the competitors didn't defeat had ever shown good sportsmanship.
7	e	James thought that the athletes that the competitors defeated had ever shown good sportsmanship.
7	f	James thought that no athletes that defeated the competitors had ever shown good sportsmanship.
7	g	James thought that the athletes that defeated no competitors had ever shown good sportsmanship.
7	h	James didn't think that the athletes that defeated the competitors had ever shown good sportsmanship.
7	i	James thought that the athletes that didn't defeat the competitors had ever shown good sportsmanship.

7	j	James thought that the athletes that defeated the competitors had ever shown good sportsmanship.
8	a	Anthony believed that no lawyers that the businessmen hired had ever fabricated evidence.
8	b	Anthony believed that the lawyers that no businessmen hired had ever fabricated evidence.
8	c	Anthony didn't believe that the lawyers that the businessmen hired had ever fabricated evidence.
8	d	Anthony believed that the lawyers that the businessmen didn't hire had ever fabricated evidence.
8	e	Anthony believed that the lawyers that the businessmen hired had ever fabricated evidence.
8	f	Anthony believed that no lawyers that represented the businessmen had ever fabricated evidence.
8	g	Anthony believed that the lawyers that represented no businessmen had ever fabricated evidence.
8	h	Anthony didn't believe that the lawyers that represented the businessmen had ever fabricated evidence.
8	i	Anthony believed that the lawyers that didn't represent the businessmen had ever fabricated evidence.
8	j	Anthony believed that the lawyers that represented the businessmen had ever fabricated evidence.
9	a	Brian knew that no students that the teachers punished had ever caused a disruption.
9	b	Brian knew that the students that no teachers punished had ever caused a disruption.
9	c	Brian didn't know that the students that the teachers punished had ever caused a disruption.
9	d	Brian knew that the students that the teachers didn't punish had ever caused a disruption.
9	e	Brian knew that the students that the teachers punished had ever caused a disruption.
9	f	Brian knew that no students that valued the teachers had ever caused a disruption.
9	g	Brian knew that the students that valued no teachers had ever caused a disruption.
9	h	Brian didn't know that the students that valued the teachers had ever caused a disruption.
9	i	Brian knew that the students that didn't value the teachers had ever caused a disruption.
9	j	Brian knew that the students that valued the teachers had ever caused a disruption.
10	a	Karen expected that no employees that the managers promoted had ever publicly ridiculed the company.
10	b	Karen expected that the employees that no managers promoted had ever publicly ridiculed the company.
10	c	Karen didn't expect that the employees that the managers promoted had ever publicly ridiculed the company.
10	d	Karen expected that the employees that the managers didn't promote had ever publicly ridiculed the company.
10	e	Karen expected that the employees that the managers promoted had ever publicly ridiculed the company.
10	f	Karen expected that no employees that satisfied the managers had ever publicly ridiculed the company.
10	g	Karen expected that the employees that satisfied no managers had ever publicly ridiculed the company.
10	h	Karen didn't expect that the employees that satisfied the managers had ever publicly ridiculed the company.
10	i	Karen expected that the employees that didn't satisfy the managers had ever publicly ridiculed the company.

10	j	Karen expected that the employees that satisfied the managers had ever publicly ridiculed the company.
11	a	Nancy reckoned that no accountants that the clients criticized had ever mismanaged funds.
11	b	Nancy reckoned that the accountants that no clients criticized had ever mismanaged funds.
11	c	Nancy didn't reckon that the accountants that the clients criticized had ever mismanaged funds.
11	d	Nancy reckoned that the accountants that the clients didn't criticize had ever mismanaged funds.
11	e	Nancy reckoned that the accountants that the clients criticized had ever mismanaged funds.
11	f	Nancy reckoned that no accountants that appeased the clients had ever mismanaged funds.
11	g	Nancy reckoned that the accountants that appeased no clients had ever mismanaged funds.
11	h	Nancy didn't reckon that the accountants that appeased the clients had ever mismanaged funds.
11	i	Nancy reckoned that the accountants that didn't appease the clients had ever mismanaged funds.
11	j	Nancy reckoned that the accountants that appeased the clients had ever mismanaged funds.
12	a	Steven felt that no teachers that the parents appreciated had ever caused problems.
12	b	Steven felt that the teachers that no parents appreciated had ever caused problems.
12	c	Steven didn't feel that the teachers that the parents appreciated had ever caused problems.
12	d	Steven felt that the teachers that the parents didn't appreciate had ever caused problems.
12	e	Steven felt that the teachers that the parents appreciated had ever caused problems.
12	f	Steven felt that no teachers that consulted the parents had ever caused problems.
12	g	Steven felt that the teachers that consulted no parents had ever caused problems.
12	h	Steven didn't feel that the teachers that consulted the parents had ever caused problems.
12	i	Steven felt that the teachers that didn't consult the parents had ever caused problems.
12	j	Steven felt that the teachers that consulted the parents had ever caused problems.
13	a	Susan thought that no students that the librarians helped had ever completed the difficult assignment.
13	b	Susan thought that the students that no librarians helped had ever completed the difficult assignment.
13	c	Susan didn't think that the students that the librarians helped had ever completed the difficult assignment.
13	d	Susan thought that the students that the librarians didn't help had ever completed the difficult assignment.
13	e	Susan thought that the students that the librarians helped had ever completed the difficult assignment.
13	f	Susan thought that no students that pestered the librarians had ever completed the difficult assignment.
13	g	Susan thought that the students that pestered no librarians had ever completed the difficult assignment.
13	h	Susan didn't think that the students that pestered the librarians had ever completed the difficult assignment.
13	i	Susan thought that the students that didn't pester the librarians had ever completed the difficult assignment.
13	j	Susan thought that the students that pestered the librarians had ever completed the difficult assignment.

14	a	Jessica believed that no nurses that the doctors recommended had ever neglected the patients.
14	b	Jessica believed that the nurses that no doctors recommended had ever neglected the patients.
14	c	Jessica didn't believe that the nurses that the doctors recommended had ever neglected the patients.
14	d	Jessica believed that the nurses that the doctors didn't recommend had ever neglected the patients.
14	e	Jessica believed that the nurses that the doctors recommended had ever neglected the patients.
14	f	Jessica believed that no nurses that assisted the doctors had ever neglected the patients.
14	g	Jessica believed that the nurses that assisted no doctors had ever neglected the patients.
14	h	Jessica didn't believe that the nurses that assisted the doctors had ever neglected the patients.
14	i	Jessica believed that the nurses that didn't assist the doctors had ever neglected the patients.
14	j	Jessica believed that the nurses that assisted the doctors had ever neglected the patients.
15	a	Margaret knew that no candidates that the voters supported had ever proposed radical policies.
15	b	Margaret knew that the candidates that no voters supported had ever proposed radical policies.
15	c	Margaret didn't know that the candidates that the voters supported had ever proposed radical policies.
15	d	Margaret knew that the candidates that the voters didn't support had ever proposed radical policies.
15	e	Margaret knew that the candidates that the voters supported had ever proposed radical policies.
15	f	Margaret knew that no candidates that disappointed the voters had ever proposed radical policies.
15	g	Margaret knew that the candidates that disappointed no voters had ever proposed radical policies.
15	h	Margaret didn't know that the candidates that disappointed the voters had ever proposed radical policies.
15	i	Margaret knew that the candidates that didn't disappoint the voters had ever proposed radical policies.
15	j	Margaret knew that the candidates that disappointed the voters had ever proposed radical policies.
16	a	Michael expected that no teachers that the teenagers admired had ever imposed strict classroom rules.
16	b	Michael expected that the teachers that no teenagers admired had ever imposed strict classroom rules.
16	c	Michael didn't expect that the teachers that the teenagers admired had ever imposed strict classroom rules.
16	d	Michael expected that the teachers that the teenagers didn't admire had ever imposed strict classroom rules.
16	e	Michael expected that the teachers that the teenagers admired had ever imposed strict classroom rules.
16	f	Michael expected that no teachers that disciplined the teenagers had ever imposed strict classroom rules.
16	g	Michael expected that the teachers that disciplined no teenagers had ever imposed strict classroom rules.
16	h	Michael didn't expect that the teachers that disciplined the teenagers had ever imposed strict classroom rules.
16	i	Michael expected that the teachers that didn't discipline the teenagers had ever imposed strict classroom rules.
16	j	Michael expected that the teachers that disciplined the teenagers had ever imposed strict classroom rules.

17	a	Christopher reckoned that no politicians that the journalists endorsed had ever earned rural voters' trust.
17	b	Christopher reckoned that the politicians that no journalists endorsed had ever earned rural voters' trust.
17	c	Christopher didn't reckon that the politicians that the journalists endorsed had ever earned rural voters' trust.
17	d	Christopher reckoned that the politicians that the journalists didn't endorse had ever earned rural voters' trust.
17	e	Christopher reckoned that the politicians that the journalists endorsed had ever earned rural voters' trust.
17	f	Christopher reckoned that no politicians that undermined the journalists had ever earned rural voters' trust.
17	g	Christopher reckoned that the politicians that undermined no journalists had ever earned rural voters' trust.
17	h	Christopher didn't reckon that the politicians that undermined the journalists had ever earned rural voters' trust.
17	i	Christopher reckoned that the politicians that didn't undermine the journalists had ever earned rural voters' trust.
17	j	Christopher reckoned that the politicians that undermined the journalists had ever earned rural voters' trust.
18	a	Sarah felt that no criminals that the policemen caught had ever deserved prison time.
18	b	Sarah felt that the criminals that no policemen caught had ever deserved prison time.
18	c	Sarah didn't feel that the criminals that the policemen caught had ever deserved prison time.
18	d	Sarah felt that the criminals that the policemen didn't catch had ever deserved prison time.
18	e	Sarah felt that the criminals that the policemen caught had ever deserved prison time.
18	f	Sarah felt that no criminals that resisted the policemen had ever deserved prison time.
18	g	Sarah felt that the criminals that resisted no policemen had ever deserved prison time.
18	h	Sarah didn't feel that the criminals that resisted the policemen had ever deserved prison time.
18	i	Sarah felt that the criminals that didn't resist the policemen had ever deserved prison time.
18	j	Sarah felt that the criminals that resisted the policemen had ever deserved prison time.
19	a	Richard thought that no surgeons that the patients consulted had ever suggested unnecessary operations.
19	b	Richard thought that the surgeons that no patients consulted had ever suggested unnecessary operations.
19	c	Richard didn't think that the surgeons that the patients consulted had ever suggested unnecessary operations.
19	d	Richard thought that the surgeons that the patients didn't consult had ever suggested unnecessary operations.
19	e	Richard thought that the surgeons that the patients consulted had ever suggested unnecessary operations.

19	f	Richard thought that no surgeons that reassured the patients had ever suggested unnecessary operations.
19	g	Richard thought that the surgeons that reassured no patients had ever suggested unnecessary operations.
19	h	Richard didn't think that the surgeons that reassured the patients had ever suggested unnecessary operations.
19	i	Richard thought that the surgeons that didn't reassure the patients had ever suggested unnecessary operations.
19	j	Richard thought that the surgeons that reassured the patients had ever suggested unnecessary operations.
20	a	Joseph believed that no suspects that the witnesses identified had ever evaded questioning.
20	b	Joseph believed that the suspects that no witnesses identified had ever evaded questioning.
20	c	Joseph didn't believe that the suspects that the witnesses identified had ever evaded questioning.
20	d	Joseph believed that the suspects that the witnesses didn't identify had ever evaded questioning.
20	e	Joseph believed that the suspects that the witnesses identified had ever evaded questioning.
20	f	Joseph believed that no suspects that threatened the witnesses had ever evaded questioning.
20	g	Joseph believed that the suspects that threatened no witnesses had ever evaded questioning.
20	h	Joseph didn't believe that the suspects that threatened the witnesses had ever evaded questioning.
20	i	Joseph believed that the suspects that didn't threaten the witnesses had ever evaded questioning.
20	j	Joseph believed that the suspects that threatened the witnesses had ever evaded questioning.

Table A.15: Full experimental stimuli for Experiment 15

A.16 Experiment 16

1	a	No authors that the critics recommended had ever written a best-selling novel.
1	b	The authors that no critics recommended had ever written a best-selling novel.
1	d	The authors that the critics didn't recommend had ever written a best-selling novel.
1	e	The authors that the critics recommended had ever written a best-selling novel.
1	f	No authors that the critics recommended have ever written a best-selling novel.
1	g	The authors that no critics recommended have ever written a best-selling novel.
1	h	The authors that the critics didn't recommend have ever written a best-selling novel.
1	i	The authors that the critics recommended have ever written a best-selling novel.
2	a	No painters that the critics liked had ever created beautiful art.
2	b	The painters that no critics liked had ever created beautiful art.
2	d	The painters that the critics didn't like had ever created beautiful art.

2	e	The painters that the critics liked had ever created beautiful art.
2	f	No painters that the critics liked have ever created beautiful art.
2	g	The painters that no critics liked have ever created beautiful art.
2	h	The painters that the critics didn't like have ever created beautiful art.
2	i	The painters that the critics liked have ever created beautiful art.
3	a	No ambassadors that the diplomats consulted had ever seen the brutality of war.
3	b	The ambassadors that no diplomats consulted had ever seen the brutality of war.
3	d	The ambassadors that the diplomats didn't consult had ever seen the brutality of war.
3	e	The ambassadors that the diplomats consulted had ever seen the brutality of war.
3	f	No ambassadors that the diplomats consulted have ever seen the brutality of war.
3	g	The ambassadors that no diplomats consulted have ever seen the brutality of war.
3	h	The ambassadors that the diplomats didn't consult have ever seen the brutality of war.
3	i	The ambassadors that the diplomats consulted have ever seen the brutality of war.
4	a	No babysitters that the children disobeyed had ever requested a tip.
4	b	The babysitters that no children disobeyed had ever requested a tip.
4	d	The babysitters that the children didn't disobey had ever requested a tip.
4	e	The babysitters that the children disobeyed had ever requested a tip.
4	f	No babysitters that the children disobeyed have ever requested a tip.
4	g	The babysitters that no children disobeyed have ever requested a tip.
4	h	The babysitters that the children didn't disobey have ever requested a tip.
4	i	The babysitters that the children disobeyed have ever requested a tip.
5	a	No customers that the salesmen assisted had ever complained to the manager.
5	b	The customers that no salesmen assisted had ever complained to the manager.
5	d	The customers that the salesmen didn't assist had ever complained to the manager.
5	e	The customers that the salesmen assisted had ever complained to the manager.
5	f	No customers that the salesmen assisted have ever complained to the manager.
5	g	The customers that no salesmen assisted have ever complained to the manager.
5	h	The customers that the salesmen didn't assist have ever complained to the manager.
5	i	The customers that the salesmen assisted have ever complained to the manager.
6	a	No actresses that the directors favored had ever demonstrated true talent.
6	b	The actresses that no directors favored had ever demonstrated true talent.
6	d	The actresses that the directors didn't favor had ever demonstrated true talent.

6	e	The actresses that the directors favored had ever demonstrated true talent.
6	f	No actresses that the directors favored have ever demonstrated true talent.
6	g	The actresses that no directors favored have ever demonstrated true talent.
6	h	The actresses that the directors didn't favor have ever demonstrated true talent.
6	i	The actresses that the directors favored have ever demonstrated true talent.
7	a	No athletes that the competitors defeated had ever shown good sportsmanship.
7	b	The athletes that no competitors defeated had ever shown good sportsmanship.
7	d	The athletes that the competitors didn't defeat had ever shown good sportsmanship.
7	e	The athletes that the competitors defeated had ever shown good sportsmanship.
7	f	No athletes that the competitors defeated have ever shown good sportsmanship.
7	g	The athletes that no competitors defeated have ever shown good sportsmanship.
7	h	The athletes that the competitors didn't defeat have ever shown good sportsmanship.
7	i	The athletes that the competitors defeated have ever shown good sportsmanship.
8	a	No lawyers that the businessmen hired had ever fabricated evidence.
8	b	The lawyers that no businessmen hired had ever fabricated evidence.
8	d	The lawyers that the businessmen didn't hire had ever fabricated evidence.
8	e	The lawyers that the businessmen hired had ever fabricated evidence.
8	f	No lawyers that the businessmen hired have ever fabricated evidence.
8	g	The lawyers that no businessmen hired have ever fabricated evidence.
8	h	The lawyers that the businessmen didn't hire have ever fabricated evidence.
8	i	The lawyers that the businessmen hired have ever fabricated evidence.
9	a	No students that the teachers punished had ever caused a disruption.
9	b	The students that no teachers punished had ever caused a disruption.
9	d	The students that the teachers didn't punish had ever caused a disruption.
9	e	The students that the teachers punished had ever caused a disruption.
9	f	No students that the teachers punished have ever caused a disruption.
9	g	The students that no teachers punished have ever caused a disruption.
9	h	The students that the teachers didn't punish have ever caused a disruption.
9	i	The students that the teachers punished have ever caused a disruption.
10	a	No employees that the managers promoted had ever publicly ridiculed the company.
10	b	The employees that no managers promoted had ever publicly ridiculed the company.
10	d	The employees that the managers didn't promote had ever publicly ridiculed the company.

10	e	The employees that the managers promoted had ever publicly ridiculed the company.
10	f	No employees that the managers promoted have ever publicly ridiculed the company.
10	g	The employees that no managers promoted have ever publicly ridiculed the company.
10	h	The employees that the managers didn't promote have ever publicly ridiculed the company.
10	i	The employees that the managers promoted have ever publicly ridiculed the company.
11	a	No accountants that the clients criticized had ever mismanaged funds.
11	b	The accountants that no clients criticized had ever mismanaged funds.
11	d	The accountants that the clients didn't criticize had ever mismanaged funds.
11	e	The accountants that the clients criticized had ever mismanaged funds.
11	f	No accountants that the clients criticized have ever mismanaged funds.
11	g	The accountants that no clients criticized have ever mismanaged funds.
11	h	The accountants that the clients didn't criticize have ever mismanaged funds.
11	i	The accountants that the clients criticized have ever mismanaged funds.
12	a	No teachers that the parents appreciated had ever caused problems.
12	b	The teachers that no parents appreciated had ever caused problems.
12	d	The teachers that the parents didn't appreciate had ever caused problems.
12	e	The teachers that the parents appreciated had ever caused problems.
12	f	No teachers that the parents appreciated have ever caused problems.
12	g	The teachers that no parents appreciated have ever caused problems.
12	h	The teachers that the parents didn't appreciate have ever caused problems.
12	i	The teachers that the parents appreciated have ever caused problems.
13	a	No students that the librarians helped had ever completed the difficult assignment.
13	b	The students that no librarians helped had ever completed the difficult assignment.
13	d	The students that the librarians didn't help had ever completed the difficult assignment.
13	e	The students that the librarians helped had ever completed the difficult assignment.
13	f	No students that the librarians helped have ever completed the difficult assignment.
13	g	The students that no librarians helped have ever completed the difficult assignment.
13	h	The students that the librarians didn't help have ever completed the difficult assignment.
13	i	The students that the librarians helped have ever completed the difficult assignment.
14	a	No nurses that the doctors recommended had ever neglected the patients.
14	b	The nurses that no doctors recommended had ever neglected the patients.
14	d	The nurses that the doctors didn't recommend had ever neglected the patients.

14	e	The nurses that the doctors recommended had ever neglected the patients.
14	f	No nurses that the doctors recommended have ever neglected the patients.
14	g	The nurses that no doctors recommended have ever neglected the patients.
14	h	The nurses that the doctors didn't recommend have ever neglected the patients.
14	i	The nurses that the doctors recommended have ever neglected the patients.
15	a	No candidates that the voters supported had ever proposed radical policies.
15	b	The candidates that no voters supported had ever proposed radical policies.
15	d	The candidates that the voters didn't support had ever proposed radical policies.
15	e	The candidates that the voters supported had ever proposed radical policies.
15	f	No candidates that the voters supported have ever proposed radical policies.
15	g	The candidates that no voters supported have ever proposed radical policies.
15	h	The candidates that the voters didn't support have ever proposed radical policies.
15	i	The candidates that the voters supported have ever proposed radical policies.
16	a	No teachers that the teenagers admired had ever imposed strict classroom rules.
16	b	The teachers that no teenagers admired had ever imposed strict classroom rules.
16	d	The teachers that the teenagers didn't admire had ever imposed strict classroom rules.
16	e	The teachers that the teenagers admired had ever imposed strict classroom rules.
16	f	No teachers that the teenagers admired have ever imposed strict classroom rules.
16	g	The teachers that no teenagers admired have ever imposed strict classroom rules.
16	h	The teachers that the teenagers didn't admire have ever imposed strict classroom rules.
16	i	The teachers that the teenagers admired have ever imposed strict classroom rules.
17	a	No politicians that the journalists endorsed had ever earned rural voters' trust.
17	b	The politicians that no journalists endorsed had ever earned rural voters' trust.
17	d	The politicians that the journalists didn't endorse had ever earned rural voters' trust.
17	e	The politicians that the journalists endorsed had ever earned rural voters' trust.
17	f	No politicians that the journalists endorsed have ever earned rural voters' trust.
17	g	The politicians that no journalists endorsed have ever earned rural voters' trust.
17	h	The politicians that the journalists didn't endorse have ever earned rural voters' trust.
17	i	The politicians that the journalists endorsed have ever earned rural voters' trust.
18	a	No criminals that the policemen caught had ever deserved prison time.
18	b	The criminals that no policemen caught had ever deserved prison time.
18	d	The criminals that the policemen didn't catch had ever deserved prison time.

18	e	The criminals that the policemen caught had ever deserved prison time.
18	f	No criminals that the policemen caught have ever deserved prison time.
18	g	The criminals that no policemen caught have ever deserved prison time.
18	h	The criminals that the policemen didn't catch have ever deserved prison time.
18	i	The criminals that the policemen caught have ever deserved prison time.
19	a	No surgeons that the patients consulted had ever suggested unnecessary operations.
19	b	The surgeons that no patients consulted had ever suggested unnecessary operations.
19	d	The surgeons that the patients didn't consult had ever suggested unnecessary operations.
19	e	The surgeons that the patients consulted had ever suggested unnecessary operations.
19	f	No surgeons that the patients consulted have ever suggested unnecessary operations.
19	g	The surgeons that no patients consulted have ever suggested unnecessary operations.
19	h	The surgeons that the patients didn't consult have ever suggested unnecessary operations.
19	i	The surgeons that the patients consulted have ever suggested unnecessary operations.
20	a	No suspects that the witnesses identified had ever evaded questioning.
20	b	The suspects that no witnesses identified had ever evaded questioning.
20	d	The suspects that the witnesses didn't identify had ever evaded questioning.
20	e	The suspects that the witnesses identified had ever evaded questioning.
20	f	No suspects that the witnesses identified have ever evaded questioning.
20	g	The suspects that no witnesses identified have ever evaded questioning.
20	h	The suspects that the witnesses didn't identify have ever evaded questioning.
20	i	The suspects that the witnesses identified have ever evaded questioning.

Table A.16: Full experimental stimuli for Experiment 16

A.17 Experiment 17

1	impostor	After the spinning wheel pricked her, for how many years did Cinderella sleep before she was awakened by the kiss of a prince?
1	normal	After the spinning wheel pricked her, for how many years did Sleeping Beauty sleep before she was awakened by the kiss of a prince?
2	impostor	An airplane crashes on the boundary of two countries. Where should the survivors be buried?
2	normal	An airplane crashes on the boundary of two countries. Where should the dead be buried?
3	impostor	In the fairy tale of Red Riding Hood, who did the fox dress up as with the intention to eat her?

3	normal	In the fairy tale of Red Riding Hood, who did the wolf dress up as with the intention to eat her?
4	impostor	By how many imps was Snow White helped before the prince came to her?
4	normal	By how many dwarves was Snow White helped before the prince came to her?
5	impostor	How many stars are in the upper left corner of the Canadian flag?
5	normal	How many stars are in the upper left corner of the American flag?
6	impostor	What aspect of the Mona Lisa by the Italian painter Michelangelo made the painting famous?
6	normal	What aspect of the Mona Lisa by the Italian painter Da Vinci made the painting famous?
7	impostor	How did painter Vincent van Gogh lose his eye during his life?
7	normal	How did painter Vincent van Gogh lose his ear during his life?
8	impostor	Does the Chinese yen cost more or less than the US Dollar?
8	normal	Does the Chinese yuan cost more or less than the US Dollar?
9	impostor	From which country does JK Rowling originate, the author of the well-known Lord of the Rings books?
9	normal	From which country does JK Rowling originate, the author of the well-known Harry Potter books?
10	impostor	How many animals of each kind did Moses take on the Ark?
10	normal	How many animals of each kind did Noah take on the Ark?
11	impostor	In which Spanish city can one visit the Colosseum?
11	normal	In which Italian city can one visit the Colosseum?
12	impostor	In which story was the sea being split into two by Jehova?
12	normal	In which story was the sea being split into two by Moses?
13	impostor	How many apples does one need to make a cup of orange juice?
13	normal	How many oranges does one need to make a cup of orange juice?
14	impostor	Is the capital city of Hungary, Bucharest, located in the middle or in the south of the country?
14	normal	Is the capital city of Hungary, Budapest, located in the middle or in the south of the country?
15	impostor	The famous formula $E=mc^2$ was formulated by which Austrian researcher?
15	normal	The famous formula $E=mc^2$ was formulated by which German researcher?
16	impostor	What is the second largest star in our solar system, after Jupiter?
16	normal	What is the second largest planet in our solar system, after Jupiter?
17	impostor	What happened to Jesus after he was betrayed by Jonas, whom he considered to be his friend?
17	normal	What happened to Jesus after he was betrayed by Judas, whom he considered to be his friend?
18	impostor	What has been the greatest hit by the American singer Justin Bieber who became famous through YouTube?

18	normal	What has been the greatest hit by the Canadian singer Justin Bieber who became famous through YouTube?
19	impostor	What is equal to a-squared plus b-squared, using Archimedes' theorem?
19	normal	What is equal to a-squared plus b-squared, using Pythagoras' theorem?
20	impostor	What is the exact year in which Columbus discovered America in the fourteenth century?
20	normal	What is the exact year in which Columbus discovered America in the fifteenth century?
21	impostor	What is the name of the country that is located between France and Spain in the Alps?
21	normal	What is the name of the country that is located between France and Spain in the Pyrenees?
22	impostor	What is the name of the English rock band of the sixties and seventies that had Ringo Starr as their guitarist?
22	normal	What is the name of the English rock band of the sixties and seventies that had Ringo Starr as their drummer?
23	impostor	What is the name of the holiday during which children dress up and walk from door to door to give candy?
23	normal	What is the name of the holiday during which children dress up and walk from door to door to collect candy?
24	impostor	What is the name of the Jewish girl who became well-known after the First World War because of her diary?
24	normal	What is the name of the Jewish girl who became well-known after the Second World War because of her diary?
25	impostor	What is the name of the main character in Twilight who will later on change into a werewolf? ¹¹⁶
25	normal	What is the name of the main character in Twilight who will later on change into a vampire?
26	impostor	What is the name of the late King of Rock who became famous because of the moonwalk and songs like Thriller?
26	normal	What is the name of the late King of Pop who became famous because of the moonwalk and songs like Thriller?
27	impostor	What is the name of the mean character from the famous Disney film called 100 Dalmatians?
27	normal	What is the name of the mean character from the famous Disney film called 101 Dalmatians?
28	impostor	What is the name of the Mexican dip made of mashed artichoke and often eaten with nachos?
28	normal	What is the name of the Mexican dip made of mashed avocado and often eaten with nachos?
29	impostor	What is the name of the raised bumps on paper which enables deaf people to read?

¹¹⁶This item was excluded from our analyses.

29	normal	What is the name of the raised bumps on paper which enables blind people to read?
30	impostor	What is the next line in the play Macbeth after the famous statement “To be or not to be?”
30	normal	What is the next line in the play Hamlet after the famous statement “To be or not to be?”
31	impostor	What kind of fruit fell out of the tree after which Einstein discovered the laws of gravity?
31	normal	What kind of fruit fell out of the tree after which Newton discovered the laws of gravity?
32	impostor	When was the Taiwanese city Hiroshima hit by an atomic bomb?
32	normal	When was the Japanese city Hiroshima hit by an atomic bomb?
33	impostor	Which American president was stabbed in 1963 by a citizen?
33	normal	Which American president was shot in 1963 by a citizen?
34	impostor	Which animal is the best friend of the green creature from Disney’s Shrek in the famous film?
34	normal	Which animal is the best friend of the green creature from Dreamworks’ Shrek in the famous film?
35	impostor	Which city near Naples was destroyed by the Etna volcano in ancient times?
35	normal	Which city near Naples was destroyed by the Vesuvius volcano in ancient times?
36	impostor	Which British monarch formally opened the Olympic winter games in London in 2012?
36	normal	Which British monarch formally opened the Olympic summer games in London in 2012?
37	impostor	Which colour was the birthmark that the Russian general Yeltsin had on his head?
37	normal	Which colour was the birthmark that the Russian general Gorbachev had on his head?
38	impostor	Which company sells many iPhone computers every year?
38	normal	Which company sells many iPhone cell phones every year?
39	impostor	Which country did Margaret Thatcher serve for as president?
39	normal	Which country did Margaret Thatcher serve for as prime minister?
40	impostor	Which date only occurs in a leap year, which occurs once every three years?
40	normal	Which date only occurs in a leap year, which occurs once every four years?
41	impostor	Which fast food restaurant has a huge white M as their logo?
41	normal	Which fast food restaurant has a huge yellow M as their logo?
42	impostor	Which is the nationality of Thomas Edison, the inventor of the telephone?
42	normal	Which is the nationality of Alexander Bell, the inventor of the telephone?
43	impostor	What is the animal that hides acorns that fall from the elm tree?
43	normal	What is the animal that hides acorns that fall from the oak tree?
44	impostor	Who found the glass slipper that was left by Snow White at the ball?
44	normal	Who found the glass slipper that was left by Cinderella at the ball?
45	impostor	In which famous cartoon is Charlie Brown’s pet cat named Snoopy?

45	normal	In which famous cartoon is Charlie Brown's pet dog named Snoopy?
46	impostor	During which world war did the Germans attack Pearl Harbor?
46	normal	During which world war did the Japanese attack Pearl Harbor?
47	impostor	What kind of tree was chopped down by Lincoln as a young boy?
47	normal	What kind of tree was chopped down by Washington as a young boy?
48	impostor	Who wears a red suit and gives out birthday presents from his sleigh?
48	normal	Who wears a red suit and gives out Christmas presents from his sleigh?
49	impostor	In what movie is Ariel is friends with a lobster named Sebastian?
49	normal	In what movie is Ariel is friends with a crab named Sebastian?
50	impostor	In baseball, where does a player run after hitting the ball with a racket?
50	normal	In baseball, where does a player run after hitting the ball with a bat?

Table A.17: Full experimental stimuli for Experiment 17

A.18 Experiment 18

1	After the spinning wheel pricked her, Sleeping Beauty slept for 100 years before she was awakened by the kiss of a prince.
2	An airplane crashes on the boundary of two countries. The dead should be buried in their native countries.
3	In the fairy tale of Red Riding Hood, the wolf dressed up as Red Riding Hood's grandmother with the intention to eat her.
4	Snow White was helped by seven dwarves before the prince came to her.
5	There are 50 stars are in the upper left corner of the American flag.
6	The aspect of the Mona Lisa by the Italian painter Da Vinci that made the painting famous is the smile.
7	Painter Vincent van Gogh lost his ear during his life by cutting it off himself.
8	The Chinese Yuan costs less than the US Dollar.
9	JK Rowling, the author of the well-known Harry Potter books, originated in England.
10	Noah took two animals of each kind on the ark.
11	One can visit the Colosseum in the Italian city of Rome.
12	The sea is split into two by Moses in the story of the Crossing of the Red Sea.
13	One needs four oranges to make a cup of orange juice.
14	The capital city of Hungary, Budapest, is located in the middle of the country.

15	The famous formula $E=mc^2$ was formulated by the German researcher Einstein.
16	The second largest planet in our solar system, after Jupiter, is Saturn.
17	After Jesus was betrayed by Judas, whom he considered to be his friend, he was crucified.
18	One of the greatest hits by the Canadian singer Justin Bieber, who became famous through YouTube, is 'What do you mean'.
19	A-squared plus b-squared is equal to c-squared using Pythagoras' theorem.
20	The exact year in which Columbus discovered America in the fifteenth century is 1492.
21	The name of the country that is located between France and Spain in the Pyrenees is Andorra.
22	The name of the English rock band of the sixties and seventies that had Ringo Starr as their drummer is The Beatles.
23	The name of the holiday during which children dress up and walk from door to door to collect candy is Halloween.
24	The name of the Jewish girl who became well-known after the Second World War because of her diary is Anne Frank.
25	The name of the main character in Twilight who will later on change into a vampire is Bella.
26	The name of the late King of Pop who became famous because of the moonwalk and songs like Thriller is Michael Jackson.
27	The name of the mean character from the famous Disney film called 101 Dalmatians is Cruella De Vil.
28	The name of the Mexican dip made of mashed avocado and often eaten with nachos is guacamole.
29	The name of the raised bumps on paper which enables blind people to read is Braille.
30	The next line in the play Hamlet after the famous statement 'To be or not to be' is 'that is the question'.
31	The fruit that fell out of the tree after which Newton discovered the laws of gravity was an apple.
32	The Japanese city Hiroshima was hit by an atomic bomb in 1945.
33	The American president that was shot in 1963 by a citizen was Kennedy.
34	The animal that is the best friend of the green creature from Dreamworks' Shrek in the famous film is a donkey.
35	The city near Naples that was destroyed by the Vesuvius volcano in ancient times was Pompeii.
36	The British monarch who formally opened the Olympic summer games in London in 2012 was Queen Elizabeth II.
37	The colour of the birthmark that the Russian general Gorbachev had on his head was wine-red.
38	The company that sells many iPhone cell phones every year is Apple.
39	The country that Margaret Thatcher served for as Prime Minister is the United Kingdom.

40	The date that only occurs in a leap year, which occurs once every four years, is February 29.
41	The fast food restaurant that has a huge yellow M as their logo is McDonalds.
42	The nationality of Alexander Bell, the inventor of the telephone, was American.
43	The animal that hides acorns that fall from the oak tree is a squirrel.
44	The prince found the glass slipper that was left by Cinderella at the ball.
45	In the famous Peanuts cartoon, Charlie Brown's pet dog is named Snoopy.
46	During the Second World War, the Japanese attacked Pearl Harbor.
47	The kind of tree that was chopped down by Washington as a young boy is a cherry tree.
48	Santa Claus wears a red suit and gives out Christmas presents from his sleigh.
49	In the movie The Little Mermaid, Ariel is friends with a crab named Sebastian.
50	In baseball, after hitting the ball with a bat the player has to run to first base.

Table A.18: Full experimental stimuli for Experiment 18

A.19 Experiment 21

1	After the spinning wheel pricked her, for how many years did ...
2	An airplane crashes on the boundary of two countries. Where should the ...
3	In the fairy tale of Red Riding Hood, who did the ...
4	By how many ...
5	How many stars are in the upper left corner of the ...
6	What aspect of the Mona Lisa by the Italian painter ...
7	How did painter Vincent van Gogh lose his ...
8	Does the Chinese ...
9	From which country does JK Rowling originate, the author of the well-known ...
10	How many animals of each kind did ...
11	In which ...
12	In which story was the sea being split into two by ...
13	How many ...
14	Is the capital city of Hungary, ...
15	The famous formula $E=mc^2$ was formulated by which ...

16	What is the second largest ...
17	What happened to Jesus after he was betrayed by ...
18	What has been the greatest hit by the ...
19	What is equal to a-squared plus b-squared, using ...
20	What is the exact year in which Columbus discovered America in the ...
21	What is the name of the country that is located between France and Spain in the ...
22	What is the name of the English rock band of the sixties and seventies that had Ringo Starr as their ...
23	What is the name of the holiday during which children dress up and walk from door to door to ...
24	What is the name of the Jewish girl who became well-known after the ...
25	What is the name of the main character in Twilight who will later on change into a ...
26	What is the name of the late King of ...
27	What is the name of the mean character from the famous Disney film called ...
28	What is the name of the Mexican dip made of mashed ...
29	What is the name of the raised bumps on paper which enables ...
30	What is the next line in the play ...
31	What kind of fruit fell out of the tree after which ...
32	When was the ...
33	Which American president was ...
34	Which animal is the best friend of the green creature from ...
35	Which city near Naples was destroyed by the ...
36	Which British monarch formally opened the Olympic ...
37	Which colour was the birthmark that the Russian general ...
38	Which company sells many iPhone ...
39	Which country did Margaret Thatcher serve for as ...
40	Which date only occurs in a leap year, which occurs once every ...
41	Which fast food restaurant has a huge ...
42	Which is the nationality of ...
43	What is the animal that hides acorns that fall from the ...
44	Who found the glass slipper that was left by ...
45	In which famous cartoon is Charlie Brown's pet ...
46	During which world war did the ...

47	What kind of tree was chopped down by ...
48	Who wears a red suit and gives out ...
49	In what movie is Ariel is friends with a ...
50	In baseball, where does a player run after hitting the ball with a ...

Table A.19: Full experimental stimuli for Experiment 21

A.20 Experiment 22

1	impostor	After the spinning wheel pricked her, Cinderella slept for 100 years before she was awakened by the kiss of a prince.
1	normal	After the spinning wheel pricked her, Sleeping Beauty slept for 100 years before she was awakened by the kiss of a prince.
2	impostor	An airplane crashes on the boundary of two countries. The survivors should be buried in their native countries.
2	normal	An airplane crashes on the boundary of two countries. The dead should be buried in their native countries.
3	impostor	In the fairy tale of Red Riding Hood, the fox dressed up as Red Riding Hood's grandmother with the intention to eat her.
3	normal	In the fairy tale of Red Riding Hood, the wolf dressed up as Red Riding Hood's grandmother with the intention to eat her.
4	impostor	Snow White was helped by seven imps before the prince came to her.
4	normal	Snow White was helped by seven dwarves before the prince came to her.
5	impostor	There are 50 stars are in the upper left corner of the Canadian flag.
5	normal	There are 50 stars are in the upper left corner of the American flag.
6	impostor	The aspect of the Mona Lisa by the Italian painter Michelangelo that made the painting famous is the smile.
6	normal	The aspect of the Mona Lisa by the Italian painter Da Vinci that made the painting famous is the smile.
7	impostor	Painter Vincent van Gogh lost his eye during his life by cutting it himself.
7	normal	Painter Vincent van Gogh lost his ear during his life by cutting it himself.
8	impostor	The Chinese yen is worth less than the US Dollar.

8	normal	The Chinese yuan is worth less than the US Dollar.
9	impostor	JK Rowling, the author of the well-known Lord of the Rings books, is from the United Kingdom.
9	normal	JK Rowling, the author of the well-known Harry Potter books, is from the United Kingdom.
10	impostor	Moses took two animals of each kind on the ark.
10	normal	Noah took two animals of each kind on the ark.
11	impostor	One can visit the Colosseum in the Spanish city of Rome.
11	normal	One can visit the Colosseum in the Italian city of Rome.
12	impostor	The sea is split into two by Jehovah in the story of the Crossing of the Red Sea.
12	normal	The sea is split into two by Moses in the story of the Crossing of the Red Sea.
13	impostor	One needs four apples to make a cup of orange juice.
13	normal	One needs four oranges to make a cup of orange juice.
14	impostor	The capital city of Hungary, Bucharest, is located in the middle of the country.
14	normal	The capital city of Hungary, Budapest, is located in the middle of the country.
15	impostor	The famous formula $E=mc^2$ was formulated by the Austrian researcher Einstein.
15	normal	The famous formula $E=mc^2$ was formulated by the German researcher Einstein.
16	impostor	The second largest star in our solar system, after Jupiter, is Saturn.
16	normal	The second largest planet in our solar system, after Jupiter, is Saturn.
17	impostor	After Jesus was betrayed by Jonas, whom he considered to be his friend, he was crucified.
17	normal	After Jesus was betrayed by Judas, whom he considered to be his friend, he was crucified.
18	impostor	One of the greatest hits by the American singer Justin Bieber, who became famous through YouTube, is 'What do you mean'.
18	normal	One of the greatest hits by the Canadian singer Justin Bieber, who became famous through YouTube, is 'What do you mean'.
19	impostor	$A^2 + B^2 = C^2$ is equal to C^2 using Archimedes' theorem.
19	normal	$A^2 + B^2 = C^2$ is equal to C^2 using Pythagoras' theorem.
20	impostor	The exact year in which Columbus discovered America in the fourteenth century is 1492.
20	normal	The exact year in which Columbus discovered America in the fifteenth century is 1492.
21	impostor	The name of the country that is located between France and Spain in the Alps is Andorra.
21	normal	The name of the country that is located between France and Spain in the Pyrenees is Andorra.
22	impostor	The name of the English rock band of the sixties and seventies that had Ringo Starr as their guitarist is The Beatles.

22	normal	The name of the English rock band of the sixties and seventies that had Ringo Starr as their drummer is The Beatles.
23	impostor	The name of the holiday during which children dress up and walk from door to door to give candy is Halloween.
23	normal	The name of the holiday during which children dress up and walk from door to door to collect candy is Halloween.
24	impostor	The name of the Jewish girl who became well-known after the First World War because of her diary is Anne Frank.
24	normal	The name of the Jewish girl who became well-known after the Second World War because of her diary is Anne Frank.
25	impostor	The name of the main character in Twilight who will later on change into a werewolf is Bella.
25	normal	The name of the main character in Twilight who will later on change into a vampire is Bella.
26	impostor	The name of the late King of Rock who became famous because of the moonwalk and songs like Thriller is Michael Jackson.
26	normal	The name of the late King of Pop who became famous because of the moonwalk and songs like Thriller is Michael Jackson.
27	impostor	The name of the mean character from the famous Disney film called 100 Dalmatians is Cruella De Vil.
27	normal	The name of the mean character from the famous Disney film called 101 Dalmatians is Cruella De Vil.
28	impostor	The name of the Mexican dip made of mashed artichoke and often eaten with nachos is guacamole.
28	normal	The name of the Mexican dip made of mashed avocado and often eaten with nachos is guacamole.
29	impostor	The name of the raised bumps on paper which enable deaf people to read is Braille.
29	normal	The name of the raised bumps on paper which enable blind people to read is Braille.
30	impostor	The next line in the play Macbeth after the famous statement 'To be or not to be' is 'that is the question'.
30	normal	The next line in the play Hamlet after the famous statement 'To be or not to be' is 'that is the question'.
31	impostor	The fruit that fell out of the tree after which Einstein discovered the laws of gravity was an apple.
31	normal	The fruit that fell out of the tree after which Newton discovered the laws of gravity was an apple.
32	impostor	The Taiwanese city Hiroshima was hit by an atomic bomb in 1945.
32	normal	The Japanese city Hiroshima was hit by an atomic bomb in 1945.

33	impostor	The American president that was stabbed in 1963 by a citizen was Kennedy.
33	normal	The American president that was shot in 1963 by a citizen was Kennedy.
34	impostor	The animal that is the best friend of the green creature from Disney's Shrek in the famous film is a donkey.
34	normal	The animal that is the best friend of the green creature from DreamWorks' Shrek in the famous film is a donkey.
35	impostor	The city near Naples that was destroyed by the Etna volcano in ancient times was Pompeii.
35	normal	The city near Naples that was destroyed by the Vesuvius volcano in ancient times was Pompeii.
36	impostor	The British monarch who formally opened the Olympic winter games in London in 2012 was Queen Elizabeth II.
36	normal	The British monarch who formally opened the Olympic summer games in London in 2012 was Queen Elizabeth II.
37	impostor	The color of the birthmark that the Russian general Yeltsin had on his head was wine-red.
37	normal	The color of the birthmark that the Russian general Gorbachev had on his head was wine-red.
38	impostor	The company that sells many iPhone computers every year is Apple.
38	normal	The company that sells many iPhone cell phones every year is Apple.
39	impostor	The country that Margaret Thatcher served for as president is the United Kingdom.
39	normal	The country that Margaret Thatcher served for as prime minister is the United Kingdom.
40	impostor	The date that only occurs in a leap year, which occurs once every three years, is February 29.
40	normal	The date that only occurs in a leap year, which occurs once every four years, is February 29.
41	impostor	The fast food restaurant that has a huge white M as their logo is McDonalds.
41	normal	The fast food restaurant that has a huge yellow M as their logo is McDonalds.
42	impostor	The nationality of Thomas Edison, the inventor of the telephone, was American.
42	normal	The nationality of Alexander Bell, the inventor of the telephone, was American.
43	impostor	The animal that hides acorns that fall from the elm is a squirrel.
43	normal	The animal that hides acorns that fall from the oak is a squirrel.
44	impostor	The prince found the glass slipper that was left by Snow White at the ball.
44	normal	The prince found the glass slipper that was left by Cinderella at the ball.
45	impostor	In the famous Peanuts cartoon, Charlie Brown's pet cat is named Snoopy.
45	normal	In the famous Peanuts cartoon, Charlie Brown's pet dog is named Snoopy.
46	impostor	During the Second World War, the Germans attacked Pearl Harbor.

46	normal	During the Second World War, the Japanese attacked Pearl Harbor.
47	impostor	The kind of tree that was chopped down by Lincoln as a young boy is a cherry tree.
47	normal	The kind of tree that was chopped down by Washington as a young boy is a cherry tree.
48	impostor	Santa Claus wears a red suit and gives out birthday presents from his sleigh.
48	normal	Santa Claus wears a red suit and gives out Christmas presents from his sleigh.
49	impostor	In the movie The Little Mermaid, Ariel is friends with a lobster named Sebastian.
49	normal	In the movie The Little Mermaid, Ariel is friends with a crab named Sebastian.
50	impostor	In baseball, after hitting the ball with racket the player has to run to first base.
50	normal	In baseball, after hitting the ball with bat the player has to run to first base.
51	impostor	Water contains two atoms of helium and one atom of oxygen.
51	normal	Water contains two atoms of hydrogen and one atom of oxygen.
52	impostor	The biological term for the process in which humans exchange carbon monoxide for oxygen is 'respiration'.
52	normal	The biological term for the process in which humans exchange carbon dioxide for oxygen is 'respiration'.
53	impostor	During the first trimester, a single embryo can split to form fraternal twins.
53	normal	During the first trimester, a single embryo can split to form identical twins.
54	impostor	Animals classified as carnivores include those that act as prey and hunt other animals for food.
54	normal	Animals classified as carnivores include those that act as predators and hunt other animals for food.
55	impostor	A half-life refers to the number of years it takes for fifty percent of a sample of an isotope to grow.
55	normal	A half-life refers to the number of years it takes for fifty percent of a sample of an isotope to decay.
56	impostor	The excess of vitamin C, commonly found in citrus fruits, causes a disease commonly known as scurvy, characterized by impaired immunity and weakness.
56	normal	The deficiency of vitamin C, commonly found in citrus fruits, causes a disease commonly known as scurvy, characterized by impaired immunity and weakness.
57	impostor	The defining feature of primates, from the class of invertebrate mammals, is the opposable thumb, which enables them to grasp and handle objects.
57	normal	The defining feature of primates, from the class of vertebrate mammals, is the opposable thumb, which enables them to grasp and handle objects.
58	impostor	Photosynthesis is the process by which plants use the green chemical chlorine and sunlight to make food.

58	normal	Photosynthesis is the process by which plants use the green chemical chlorophyll and sunlight to make food.
59	impostor	Electrons are the negatively charged subatomic particles that orbit the atomic nucleus, which contains a mix of neutrons and positrons.
59	normal	Electrons are the negatively charged subatomic particles that orbit the atomic nucleus, which contains a mix of neutrons and protons.
60	impostor	Renewable resources such as coal, petroleum, and natural gas are known as 'fossil fuels'.
60	normal	Non-renewable resources such as coal, petroleum, and natural gas are known as 'fossil fuels'.
61	impostor	Metamorphosis is a profound change in an organism's form that occurs naturally during its life cycle, such as the change from a caterpillar to a butterfly or a tadpole to a lizard.
61	normal	Metamorphosis is a profound change in an organism's form that occurs naturally during its life cycle, such as the change from a caterpillar to a butterfly or a tadpole to a frog.
62	impostor	The digestive organ found on many aquatic animals is the gill.
62	normal	The respiratory organ found on many aquatic animals is the gill.
63	impostor	The human species, known as Homo Erectus, is thought to have originated on the African continent.
63	normal	The human species, known as Homo Sapiens, is thought to have originated on the African continent.
64	impostor	The rigid outer boundary of an animal cell is called a cell wall.
64	normal	The rigid outer boundary of a plant cell is called a cell wall.
65	impostor	Paul Revere is famously credited with alerting Colonial militia of approaching British forces during the Civil War.
65	normal	Paul Revere is famously credited with alerting Colonial militia of approaching British forces during the Revolutionary War.
66	impostor	The famous Rosetta Stone has allowed scholars to interpret ancient Egyptian writing known as Cuneiform.
66	normal	The famous Rosetta Stone has allowed scholars to interpret ancient Egyptian writing known as Hieroglyphics.
67	impostor	The British Congress imposed fees on the colonies' sugar, tea, and stamps, leading to the eventual Revolutionary War.
67	normal	The British Parliament imposed fees on the colonies' sugar, tea, and stamps, leading to the eventual Revolutionary War.
68	impostor	The series of economic programs known as the Square Deal was implemented by President Franklin D. Roosevelt, who was paralyzed by polio.

68	normal	The series of economic programs known as the New Deal was implemented by President Franklin D. Roosevelt, who was paralyzed by polio.
69	impostor	Thomas Jefferson wrote the Constitution in the year 1776.
69	normal	Thomas Jefferson wrote the Declaration of Independence in the year 1776.
70	impostor	The epic poem by Virgil which details Odysseus' journey is the Odyssey.
70	normal	The epic poem by Homer which details Odysseus' journey is the Odyssey.
71	impostor	Earth's highest mountain above sea level, located on the border of Nepal and Bhutan and part of the Himalayas, is Mount Everest.
71	normal	Earth's highest mountain above sea level, located on the border of Nepal and China and part of the Himalayas, is Mount Everest.
72	impostor	On the popular NBC television show Friends, the characters' main hangout is a bar called Central Perk.
72	normal	On the popular NBC television show Friends, the characters' main hangout is a coffee shop called Central Perk.
73	impostor	On the sitcom Seinfeld, George Costanza worked for the New York Mets for three years.
73	normal	On the sitcom Seinfeld, George Costanza worked for the New York Yankees for three years.
74	impostor	Contestants on the popular singing competition show American Idol audition in front of celebrity judges and, if selected, receive a 'golden ticket' to travel to New York and continue in the competition.
74	normal	Contestants on the popular singing competition show American Idol audition in front of celebrity judges and, if selected, receive a 'golden ticket' to travel to Hollywood and continue in the competition.
75	impostor	The Game of Thrones television series takes place primarily in the fictional Seven Kingdoms of Middle Earth, where various noble families fight for the Iron Throne.
75	normal	The Game of Thrones television series takes place primarily in the fictional Seven Kingdoms of Westeros, where various noble families fight for the Iron Throne.
76	impostor	The original Fellowship of the Ring, from the Lord of the Rings series, includes five hobbits, most notably Bilbo Baggins, who is tasked with carrying the One Ring.
76	normal	The original Fellowship of the Ring, from the Lord of the Rings series, includes five hobbits, most notably Frodo Baggins, who is tasked with carrying the One Ring.
77	impostor	Holden Caulfield, the protagonist of J.D. Salinger's The Catcher in the Rye, wonders at various points in the novel about what happens to the geese in Central Park in winter.

77	normal	Holden Caulfield, the protagonist of J.D. Salinger's <i>The Catcher in the Rye</i> , wonders at various points in the novel about what happens to the ducks in Central Park in winter.
78	impostor	In Harper Lee's classic novel, <i>To Kill a Mockingbird</i> , the main character learns about race and racism through her father's work as a doctor
78	normal	In Harper Lee's classic novel, <i>To Kill a Mockingbird</i> , the main character learns about race and racism through her father's work as a lawyer
79	impostor	Harry Potter learns that he is a wizard on his eleventh birthday, when he is visited by Rubeus Hagrid, a half-human half-goblin employee of Hogwarts.
79	normal	Harry Potter learns that he is a wizard on his eleventh birthday, when he is visited by Rubeus Hagrid, a half-human half-giant employee of Hogwarts.
80	impostor	Luke Skywalker learns about the Force and the Jedi Knights from Han Solo after the death of his aunt and uncle, who raised him.
80	normal	Luke Skywalker learns about the Force and the Jedi Knights from Obi-Wan Kenobi after the death of his aunt and uncle, who raised him.
81	impostor	In the original <i>Toy Story</i> movie, all the toys fear the brother named Sid, who mistreats his toys, often taking them apart and creating 'mutant' toys from their pieces.
81	normal	In the original <i>Toy Story</i> movie, all the toys fear the neighbor named Sid, who mistreats his toys, often taking them apart and creating 'mutant' toys from their pieces.
82	impostor	The movie <i>Finding Nemo</i> features a goldfish named Marlin who must search for his son Nemo with the help of a fish who suffers from memory problems.
82	normal	The movie <i>Finding Nemo</i> features a clownfish named Marlin who must search for his son Nemo with the help of a fish who suffers from memory problems.
83	impostor	In the film <i>Home Alone</i> , a child is accidentally left behind when his family goes to London for Christmas, leaving him to ultimately fend off burglars with a series of elaborate booby traps.
83	normal	In the film <i>Home Alone</i> , a child is accidentally left behind when his family goes to Paris for Christmas, leaving him to ultimately fend off burglars with a series of elaborate booby traps.
84	impostor	Gregor Mendel, the 19th century friar and scientist, is best known for his research on the inheritance of traits in corn plants.
84	normal	Gregor Mendel, the 19th century friar and scientist, is best known for his research on the inheritance of traits in pea plants.
85	impostor	The concept for which B.F. Skinner is famous is 'classical conditioning' such as a dog's salivation response to a bell after the bell has been repeatedly paired with the presentation of food.

85	normal	The concept for which Ivan Pavlov is famous is 'classical conditioning' such as a dog's salivation response to a bell after the bell has been repeatedly paired with the presentation of food.
86	impostor	The three types of rock are igneous, sedimentary, and mineral.
86	normal	The three types of rock are igneous, sedimentary, and metamorphic.
87	impostor	The water cycle includes processes like evaporation, in which liquid water becomes water vapor, and sublimation, in which water vapor condenses and falls to earth in the form of rain or snow.
87	normal	The water cycle includes processes like evaporation, in which liquid water becomes water vapor, and precipitation, in which water vapor condenses and falls to earth in the form of rain or snow.
88	impostor	Living things are made up of cells which themselves contain substructures known as Golgi bodies, such as the nucleus or mitochondria.
88	normal	Living things are made up of cells which themselves contain substructures known as organelles, such as the nucleus or mitochondria.
89	impostor	Boundaries between tectonic plates in the earth's core are associated with the creation of phenomena like earthquakes and volcanic eruptions.
89	normal	Boundaries between tectonic plates in the earth's crust are associated with the creation of phenomena like earthquakes and volcanic eruptions.
90	impostor	The first law of thermodynamics, which states that a body in motion remains in motion, and a body at rest remains at rest, until acted upon by a force, was discovered by Isaac Newton.
90	normal	The first law of motion, which states that a body in motion remains in motion, and a body at rest remains at rest, until acted upon by a force, was discovered by Isaac Newton.
91	impostor	In chemistry, a catalyst is a substance that increases the yield of a chemical reaction.
91	normal	In chemistry, a catalyst is a substance that increases the rate of a chemical reaction.
92	impostor	A molecule contains two or more atoms held together by electrical bonds.
92	normal	A molecule contains two or more atoms held together by chemical bonds.
93	impostor	Diamonds, which are made of calcium, are typically considered the hardest naturally occurring substance.
93	normal	Diamonds, which are made of carbon, are typically considered the hardest naturally occurring substance.
94	impostor	The first government structure of the United States following the American Revolutionary War was established by the Magna Carta.
94	normal	The first government structure of the United States following the American Revolutionary War was established by the Articles of Confederation.

95	impostor	The Boston Tea Party, in which chests of tea were thrown into the Plymouth Harbor, was a protest against Parliament's Tea Act.
95	normal	The Boston Tea Party, in which chests of tea were thrown into the Boston Harbor, was a protest against Parliament's Tea Act.
96	impostor	Richard Nixon, who ultimately resigned from office due to the Watergate scandal, was elected President in 1968 alongside Vice President Gerald Ford.
96	normal	Richard Nixon, who ultimately resigned from office due to the Watergate scandal, was elected President in 1968 alongside Vice President Spiro Agnew.
97	impostor	Mesopotamian civilization is thought to have emerged near the Indus and Euphrates rivers, a region now sometimes referred to as the 'Fertile Crescent'.
97	normal	Mesopotamian civilization is thought to have emerged near the Tigris and Euphrates rivers, a region now sometimes referred to as the 'Fertile Crescent'.
98	impostor	The mythical son of Jupiter and Juno and Greek god of war was called Mars.
98	normal	The mythical son of Jupiter and Juno and Roman god of war was called Mars.
99	impostor	The Berlin Wall was constructed by the government of West Germany to separate East Berlin and West Berlin during the Cold War.
99	normal	The Berlin Wall was constructed by the government of East Germany to separate East Berlin and West Berlin during the Cold War.
100	impostor	The military nobility of medieval Korean culture were known as samurai.
100	normal	The military nobility of medieval Japanese culture were known as samurai.
101	impostor	American singer-songwriter Taylor Swift has won numerous Tony awards for her music.
101	normal	American singer-songwriter Taylor Swift has won numerous Grammy awards for her music.
102	impostor	Yosemite National Park, known for its geothermal features like Old Faithful geyser is among the most-visited parks in the US.
102	normal	Yellowstone National Park, known for its geothermal features like Old Faithful geyser is among the most-visited parks in the US.
103	impostor	Honolulu is the capital of the state of Hawaii, which consists of a chain of islands in the Atlantic Ocean.
103	normal	Honolulu is the capital of the state of Hawaii, which consists of a chain of islands in the Pacific Ocean.
104	impostor	Lines of constant longitude, sometimes called 'parallels', run east-west as circles parallel to the equator.

104	normal	Lines of constant latitude, sometimes called 'parallels', run east–west as circles parallel to the equator.
105	impostor	The FIFA World Cup is an international baseball tournament held every four years.
105	normal	The FIFA World Cup is an international soccer tournament held every four years.
106	impostor	The Venus flytrap is a herbivorous plant known for its ability to trap and digest insects.
106	normal	The Venus flytrap is a carnivorous plant known for its ability to trap and digest insects.
107	impostor	In order to win a game of checkers, players aim to checkmate the opponent's king.
107	normal	In order to win a game of chess, players aim to checkmate the opponent's king.
108	impostor	In the board game Monopoly, players are allowed to collect 200 dollars of the game's currency when their game piece passes the Jail square.
108	normal	In the board game Monopoly, players are allowed to collect 200 dollars of the game's currency when their game piece passes the Go square.
109	impostor	Orcas, sometimes called 'killer whales', are a type of carnivorous oceanic fish and have a characteristic black and white coloring.
109	normal	Orcas, sometimes called 'killer whales', are a type of carnivorous oceanic mammal and have a characteristic black and white coloring.
110	impostor	Electric eels are known for their ability to generate electricity which they use to stun their predators, or animals they hunt for food.
110	normal	Electric eels are known for their ability to generate electricity which they use to stun their prey, or animals they hunt for food.
111	impostor	Coniferous trees, like pines and maples, typically have long, thin, needle-like leaves which remain on the tree throughout winter.
111	normal	Coniferous trees, like pines and firs, typically have long, thin, needle-like leaves which remain on the tree throughout winter.
112	impostor	The primary food source of the koala bear, native to China, is bamboo.
112	normal	The primary food source of the panda bear, native to China, is bamboo.
113	impostor	The Statue of Liberty was a gift to the United States from England and is now located in New York City.
113	normal	The Statue of Liberty was a gift to the United States from France and is now located in New York City.
114	impostor	Pride and Prejudice, written by Elizabeth Bennet, is considered one of the most famous romantic novels in history.

114	normal	Pride and Prejudice, written by Jane Austen, is considered one of the most famous romantic novels in history.
115	impostor	The famous speech that begins 'Four score and twenty years ago' was given by Abraham Lincoln.
115	normal	The famous speech that begins 'Four score and seven years ago' was given by Abraham Lincoln.
116	impostor	In the US, phone numbers consist of seven digits, preceded by a three-digit zip code.
116	normal	In the US, phone numbers consist of seven digits, preceded by a three-digit area code.
117	impostor	Neil Armstrong was the first person to walk on the sun.
117	normal	Neil Armstrong was the first person to walk on the moon.
118	impostor	In the comic book series, Clark Kent changes into Superman in a toll booth.
118	normal	In the comic book series, Clark Kent changes into Superman in a phone booth.
119	impostor	Batman is a superhero who, along with his sidekick Robin, protects the city of Metropolis.
119	normal	Batman is a superhero who, along with his sidekick Robin, protects the city of Gotham.
120	impostor	The circumference of a circle can be computed as pi times the radius squared.
120	normal	The area of a circle can be computed as pi times the radius squared.

Table A.20: Full experimental stimuli for Experiment 22

A.21 Experiment 23

1	primes	Moses (impostor), Noah (intended), kind (related), snarl (unrelated)
1	other lex. dec.	puzzler, righthunder, unpromptef
1	impostor	Moses took two animals of each kind on the ark.
1	normal	Noah took two animals of each kind on the ark.
2	primes	guitarist (impostor), drummer (intended), band (related), crackpots (unrelated)
2	other lex. dec.	Pluto, rulebool, ilolating
2	impostor	The name of the English rock band of the sixties and seventies that had Ringo Starr as their guitarist is The Beatles.
2	normal	The name of the English rock band of the sixties and seventies that had Ringo Starr as their drummer is The Beatles.
3	primes	give (impostor), collect (intended), door (related), mutt (unrelated)
3	other lex. dec.	shipman, proadcaster, flyem

3	impostor	The name of the holiday during which children dress up and walk from door to door to give candy is Halloween.
3	normal	The name of the holiday during which children dress up and walk from door to door to collect candy is Halloween.
4	primes	First (impostor), Second (intended), because (related), necks (unrelated)
4	other lex. dec.	shorty, gilts, purrender
4	impostor	The name of the Jewish girl who became well-known after the First World War because of her diary is Anne Frank.
4	normal	The name of the Jewish girl who became well-known after the Second World War because of her diary is Anne Frank.
5	primes	deaf (impostor), blind (intended), raised (related), sops (unrelated)
5	other lex. dec.	nevermore, santon, fandescant
5	impostor	The name of the raised bumps on paper which enable deaf people to read is Braille.
5	normal	The name of the raised bumps on paper which enable blind people to read is Braille.
6	primes	Macbeth (impostor), Hamlet (intended), statement (related), triceps (unrelated)
6	other lex. dec.	flustered, glandard, Pahrain
6	impostor	The next line in the play Macbeth after the famous statement 'To be or not to be' is 'that is the question'.
6	normal	The next line in the play Hamlet after the famous statement 'To be or not to be' is 'that is the question'.
7	primes	Einstein (impostor), Newton (intended), laws (related), cockpits (unrelated)
7	other lex. dec.	skippers, spoku, daldy
7	impostor	The fruit that fell out of the tree after which Einstein discovered the laws of gravity was an apple.
7	normal	The fruit that fell out of the tree after which Newton discovered the laws of gravity was an apple.
8	primes	Etna (impostor), Vesuvius (intended), ancient (related), agar (unrelated)
8	other lex. dec.	nibble, seguiled, bombints
8	impostor	The city near Naples that was destroyed by the Etna volcano in ancient times was Pompeii.
8	normal	The city near Naples that was destroyed by the Vesuvius volcano in ancient times was Pompeii.
9	primes	three (impostor), four (intended), occurs (related), wormy (unrelated)
9	other lex. dec.	Cuba, aerosop, sevenue
9	impostor	The date that only occurs in a leap year, which occurs once every three years, is February 29.
9	normal	The date that only occurs in a leap year, which occurs once every four years, is February 29.

10	primes	birthday (impostor), Christmas (intended), suit (related), swamping (unrelated)
10	other lex. dec.	cults, arboreel, leview
10	impostor	Santa Claus wears a red suit and gives out birthday presents from his sleigh.
10	normal	Santa Claus wears a red suit and gives out Christmas presents from his sleigh.
11	primes	lobster (impostor), crab (intended), friends (related), invests (unrelated)
11	other lex. dec.	grunting, agvancing, suarantee
11	impostor	In the movie The Little Mermaid, Ariel is friends with a lobster named Sebastian.
11	normal	In the movie The Little Mermaid, Ariel is friends with a crab named Sebastian.
12	primes	helium (impostor), hydrogen (intended), two (related), loafed (unrelated)
12	other lex. dec.	baldness, smire, dearl
12	impostor	Water contains two atoms of helium and one atom of oxygen.
12	normal	Water contains two atoms of hydrogen and one atom of oxygen.
13	primes	monoxide (impostor), dioxide (intended), term (related), typecast (unrelated)
13	other lex. dec.	nightclubs, elphasis, ratchup
13	impostor	The biological term for the process in which humans exchange carbon monoxide for oxygen is 'respiration'.
13	normal	The biological term for the process in which humans exchange carbon dioxide for oxygen is 'respiration'.
14	primes	fraternal (impostor), identical (intended), trimester (related), orangeade (unrelated)
14	other lex. dec.	stunk, gaking, merchunts
14	impostor	During the first trimester, a single embryo can split to form fraternal twins.
14	normal	During the first trimester, a single embryo can split to form identical twins.
15	primes	chlorine (impostor), chlorophyll (intended), food (related), minstrel (unrelated)
15	other lex. dec.	stroked, dariables, jorerunners
15	impostor	Photosynthesis is the process by which plants use the green chemical chlorine and sunlight to make food.
15	normal	Photosynthesis is the process by which plants use the green chemical chlorophyll and sunlight to make food.
16	primes	positrons (impostor), protons (intended), charged (related), apologist (unrelated)
16	other lex. dec.	champ, sequesper, sompute
16	impostor	Electrons are the negatively charged subatomic particles that orbit the atomic nucleus, which contains a mix of neutrons and positrons.

16	normal	Electrons are the negatively charged subatomic particles that orbit the atomic nucleus, which contains a mix of neutrons and protons.
17	primes	Civil (impostor), Revolutionary (intended), credited (related), spore (unrelated)
17	other lex. dec.	pouch, insanuty, darness
17	impostor	Paul Revere is famously credited with alerting Colonial militia of approaching British forces during the Civil War.
17	normal	Paul Revere is famously credited with alerting Colonial militia of approaching British forces during the Revolutionary War.
18	primes	Cuneiform (impostor), Hieroglyphics (intended), scholars (related), bowstring (unrelated)
18	other lex. dec.	swells, clith, risclosure
18	impostor	The famous Rosetta Stone has allowed scholars to interpret ancient Egyptian writing known as Cuneiform.
18	normal	The famous Rosetta Stone has allowed scholars to interpret ancient Egyptian writing known as Hieroglyphics.
19	primes	Square (impostor), New (intended), paralyzed (related), paunch (unrelated)
19	other lex. dec.	campfire, drittoon, lompression
19	impostor	The series of economic programs known as the Square Deal was implemented by President Franklin D. Roosevelt, who was paralyzed by polio.
19	normal	The series of economic programs known as the New Deal was implemented by President Franklin D. Roosevelt, who was paralyzed by polio.
20	primes	Virgil (impostor), Homer (intended), epic (related), blowup (unrelated)
20	other lex. dec.	closets, tredetermined, shromium
20	impostor	The epic poem by Virgil which details Odysseus' journey is the Odyssey.
20	normal	The epic poem by Homer which details Odysseus' journey is the Odyssey.
21	primes	yield (impostor), rate (intended), substance (related), roars (unrelated)
21	other lex. dec.	Geneva, eltered, wuffs
21	impostor	In chemistry, a catalyst is a substance that increases the yield of a chemical reaction.
21	normal	In chemistry, a catalyst is a substance that increases the rate of a chemical reaction.
22	primes	Yosemite (impostor), Yellowstone (intended), features (related), spinster (unrelated)
22	other lex. dec.	squeal, drinflint, horsage
22	impostor	Yosemite National Park, known for its geothermal features like Old Faithful geyser is among the most-visited parks in the US.

22	normal	Yellowstone National Park, known for its geothermal features like Old Faithful geyser is among the most-visited parks in the US.
23	primes	Atlantic (impostor), Pacific (intended), islands (related), misspell (unrelated)
23	other lex. dec.	smokes, refonances, firology
23	impostor	Honolulu is the capital of the state of Hawaii, which consists of a chain of islands in the Atlantic Ocean.
23	normal	Honolulu is the capital of the state of Hawaii, which consists of a chain of islands in the Pacific Ocean.
24	primes	longitude (impostor), latitude (intended), constant (related), hairbrush (unrelated)
24	other lex. dec.	mink, riographer, fecently
24	impostor	Lines of constant longitude, sometimes called 'parallels', run east,Äwest as circles parallel to the equator.
24	normal	Lines of constant latitude, sometimes called 'parallels', run east,Äwest as circles parallel to the equator.
25	primes	herbivorous (impostor), carnivorous (intended), ability (related), bricklaying (unrelated)
25	other lex. dec.	slickers, Marpist, ifpotence
25	impostor	The Venus flytrap is an herbivorous plant known for its ability to trap and digest insects.
25	normal	The Venus flytrap is a carnivorous plant known for its ability to trap and digest insects.
26	primes	fish (impostor), mammal (intended), coloring (related), axes (unrelated)
26	other lex. dec.	tailcoat, secoverable, begut
26	impostor	Orcas, sometimes called 'killer whales', are a type of carnivorous oceanic fish and have a characteristic black and white coloring.
26	normal	Orcas, sometimes called 'killer whales', are a type of carnivorous oceanic mammal and have a characteristic black and white coloring.
27	primes	maples (impostor), firs (intended), tree (related), butane (unrelated)
27	other lex. dec.	foxhole, chooping, veverely
27	impostor	Coniferous trees, like pines and maples, typically have long, thin, needle-like leaves which remain on the tree throughout winter.
27	normal	Coniferous trees, like pines and firs, typically have long, thin, needle-like leaves which remain on the tree throughout winter.
28	primes	koala (impostor), panda (intended), primary (related), deism (unrelated)
28	other lex. dec.	likable, thackle, otal
28	impostor	The primary food source of the koala bear, native to China, is bamboo.

28	normal	The primary food source of the panda bear, native to China, is bamboo.
29	primes	twenty (impostor), seven (intended), speech (related), croaks (unrelated)
29	other lex. dec.	waterproof, ircognito, upgreded
29	impostor	The famous speech that begins 'Four score and twenty years ago' was given by Abraham Lincoln.
29	normal	The famous speech that begins 'Four score and seven years ago' was given by Abraham Lincoln.
30	primes	zip (impostor), area (intended), digits (related), oaf (unrelated)
30	other lex. dec.	snuff, plexoglis, brisis
30	impostor	In the US, phone numbers consist of seven digits, preceded by a three-digit zip code.
30	normal	In the US, phone numbers consist of seven digits, preceded by a three-digit area code.
31	primes	toll (impostor), phone (intended), booth (related), bolo (unrelated)
31	other lex. dec.	obeying, heploysing, isduced
31	impostor	In the comic book series, Clark Kent changes into Superman in a toll booth.
31	normal	In the comic book series, Clark Kent changes into Superman in a phone booth.
32	primes	Metropolis (impostor), Gotham (intended), sidekick (related), overpowers (unrelated)
32	other lex. dec.	patties, narehouses, Abigoul
32	impostor	Batman is a superhero who, along with his sidekick Robin, protects the city of Metropolis.
32	normal	Batman is a superhero who, along with his sidekick Robin, protects the city of Gotham.

Table A.21: Full experimental stimuli for Experiment 23

Bibliography

- Acuña-Fariña, J Carlos, Enrique Meseguer, & Manuel Carreiras. 2014. Gender and number agreement in comprehension in spanish. *Lingua* 143. 108–128.
- Alcocer, Pedro & Colin Phillips. 2012. Using relational syntactic constraints in content-addressable memory architectures for sentence parsing. *Unpublished manuscript*.
- Allott, Nicholas & P Rubio Fernandez. 2002. This paper fills a much-needed gap. *Actes de l'Atelier des doctorants en linguistique*. 97–102.
- Almeida, Felipe & Geraldo Xexéo. 2019. Word embeddings: a survey. *arXiv preprint arXiv:1901.09069*.
- Alter, Adam L, Daniel M Oppenheimer, Nicholas Epley, & Rebecca N Eyre. 2007. Overcoming intuition: metacognitive difficulty activates analytic reasoning. *Journal of experimental psychology: General* 136(4). 569.
- Anderson, Catherine. 2004. *The structure and real-time comprehension of quantifier scope ambiguity*. Northwestern University dissertation.
- Anderson, John R, Daniel Bothell, Michael D Byrne, Scott Douglass, Christian Lebiere, & Yulin Qin. 2004. An integrated theory of the mind. *Psychological review* 111(4). 1036.
- Arts, Anja, Alfons Maes, Leo Noordman, & Carel Jansen. 2011. Overspecification facilitates object identification. *Journal of pragmatics* 43(1). 361–374.
- Auguste, Jeremy, Arnaud Rey, & Benoit Favre. 2017. Evaluation of word embeddings against cognitive processes: primed reaction times in lexical decision and naming tasks. In *Proceedings of the 2nd workshop on evaluating vector space representations for NLP*, 21–26. Copenhagen, Denmark: Association for Computational Linguistics. DOI: [10.18653/v1/W17-5304](https://doi.org/10.18653/v1/W17-5304). <https://aclanthology.org/W17-5304>.
- Baker, C Lee. 1970. Double negatives. *Linguistic inquiry* 1(2). 169–186.
- Baker, Linda. 1979. Comprehension monitoring: identifying and coping with text confusions. *Journal of reading behavior* 11(4). 365–374.

- Baker, Linda & Jody L Wagner. 1987. Evaluating information for truthfulness: the effects of logical subordination. *Memory & Cognition* 15(3). 247–255.
- Balota, David A, Melvin J Yap, Keith A Hutchison, Michael J Cortese, Brett Kessler, Bjorn Loftis, James H Neely, Douglas L Nelson, Greg B Simpson, & Rebecca Treiman. 2007. The english lexicon project. *Behavior research methods* 39(3). 445–459.
- Barker, Chris. 2018. Negative polarity as scope marking. *Linguistics and philosophy* 41(5). 483–510.
- Barr, Dale J, Roger Levy, Christoph Scheepers, & Harry J Tily. 2013. Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of memory and language* 68(3). 255–278.
- Barthel, Mathias. 2021. Speech planning interferes with language comprehension: evidence from semantic illusions in question-response sequences. In *Proceedings of the 25th workshop on the semantics and pragmatics of dialogue*.
- Bever, Thomas G. 1970. The cognitive basis for linguistic structures. *Cognition and the development of language*.
- Bock, Kathryn & Carol A Miller. 1991. Broken agreement. *Cognitive psychology* 23(1). 45–93.
- Bodner, Glen E & Michael EJ Masson. 2003. Beyond spreading activation: an influence of relatedness proportion on masked semantic priming. *Psychonomic Bulletin & Review* 10(3). 645–652.
- Bohan, Jason, Hartmut Leuthold, Yuko Hijikata, & Anthony J Sanford. 2012. The processing of good-fit semantic anomalies: an erp investigation. *Neuropsychologia* 50(14). 3174–3184.
- Bohan, Jason & Anthony Sanford. 2008. Semantic anomalies at the borderline of consciousness: an eye-tracking investigation. *Quarterly Journal of Experimental Psychology* 61(2). 232–239.
- Bohan, Jason Thomas. 2008. *Depth of processing and semantic anomalies*. University of Glasgow dissertation.
- Bott, Oliver & Fabian Schlotterbeck. 2015. The processing domain of scope interaction. *Journal of Semantics* 32(1). 39–92.
- Bottoms, Hayden C, Andrea N Eslick, & Elizabeth J Marsh. 2010. Memory and the mooses illusion: failures to detect contradictions with stored knowledge yield negative memorial consequences. *Memory* 18(6). 670–678.
- Bransford, John D & Jeffery J Franks. 1971. The abstraction of linguistic ideas. *Cognitive psychology* 2(4). 331–350.
- Brashier, Nadia M, Emmaline Drew Eliseev, & Elizabeth J Marsh. 2020. An initial accuracy focus prevents illusory truth. *Cognition* 194. 104054.

- Bredart, Serge & Karin Modolo. 1988. Moses strikes again: focalization effect on a semantic illusion. *Acta psychologica* 67(2). 135–144.
- Budiu, Raluca & John R Anderson. 2008. Negation in nonliteral sentences.
- Bürkner, Paul-Christian. 2017. Brms: an r package for bayesian multilevel models using stan. *Journal of statistical software* 80. 1–28.
- Büttner, Anke Caroline. 2007. Questions versus statements: challenging an assumption about semantic illusions. *Quarterly Journal of Experimental Psychology* 60(6). 779–789.
- Büttner, Anke Caroline. 2012. The effect of working memory load on semantic illusions: what the phonological loop and central executive have to contribute. *Memory* 20(8). 882–890.
- Cantor, Allison D & Elizabeth J Marsh. 2017. Expertise effects in the moses illusion: detecting contradictions with stored knowledge. *Memory* 25(2). 220–230.
- Carston, Robyn. 2012. Word meaning and concept expressed. *The Linguistic Review* 29(4). 607–623.
- Chemla, Emmanuel, Vincent Homer, & Daniel Rothschild. 2011. Modularity and intuitions in formal semantics: the case of polarity items. *Linguistics and Philosophy* 34(6). 537–570.
- Chierchia, Gennaro. 2006. Broaden your views: implicatures of domain widening and the “logicity” of language. *Linguistic inquiry* 37(4). 535–590.
- Chow, Wing-Yee, Shota Momma, Cybelle Smith, Ellen Lau, & Colin Phillips. 2016. Prediction as memory retrieval: timing and mechanisms. *Language, Cognition and Neuroscience* 31(5). 617–627.
- Christianson, Kiel. 2016. When language comprehension goes wrong for the right reasons: good-enough, underspecified, or shallow language processing. *Quarterly journal of experimental psychology* 69(5). 817–828.
- Clark, Herbert H & William G Chase. 1972. On the process of comparing sentences against pictures. *Cognitive psychology* 3(3). 472–517.
- Clifton, Charles & Lyn Frazier. 2010. When are downward entailing contexts identified? the case of the domain-widener ever. *Linguistic Inquiry (Online)* 41(4). 681.
- Clifton, Charles, Lyn Frazier, & Patricia Deevy. 1999. Feature manipulation in sentence comprehension: 2703. *Italian journal of linguistics* 11(1). 11–40.
- Coltheart, Max, Kathleen Rastle, Conrad Perry, Robyn Langdon, & Johannes Ziegler. 2001. Drc: a dual route cascaded model of visual word recognition and reading aloud. *Psychological review* 108(1). 204.
- Comrie, Bernard et al. 1985. *Tense*. Vol. 17. Cambridge university press.

- Cook, Anne E, Erinn K Walsh, Margaret AA Bills, John C Kircher, & Edward J O'Brien. 2018. Validation of semantic illusions independent of anomaly detection: evidence from eye movements. *Quarterly Journal of Experimental Psychology* 71(1). 113–121.
- Cummings, Ian & Claudia Felser. 2013. The role of working memory in the processing of reflexives. *Language and Cognitive Processes* 28(1-2). 188–219.
- Cummings, Ian & Patrick Sturt. 2014. Coargumenthood and the processing of reflexives. *Journal of Memory and Language* 75. 117–139.
- Davies, Mark. 2008. *The Corpus of Contemporary American English (COCA)*. Available online at <https://www.english-corpora.org/coca/>.
- Davis, Danielle K & Lise Abrams. 2016. Here's looking at you: visual similarity exacerbates the moose illusion for semantically similar celebrities. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 42(1). 75.
- Deckert, Mikolaj. 2015. Processing fluency and decision-making: the role of language structure. *Psychology of Language and Communication* 19(2). 149.
- Dhaene, Sara, Nicolas Dirix, Hélène Van Marcke, & Evy Woumans. 2021. Moses or noah? a case of 'potato-potahto' when using a foreign language. *Bilingualism: Language and Cognition*. 1–15.
- Dillon, Brian, Alan Mishler, Shayne Sloggett, & Colin Phillips. 2013. Contrasting intrusion profiles for agreement and anaphora: experimental and modeling evidence. *Journal of Memory and Language* 69(2). 85–103.
- de Dios Flores, Iria. 2019. Processing sentences with multiple negations: grammatical structures that are perceived as unacceptable. *Frontiers in Psychology*. 2346.
- de Dios Flores, Iria, Hanna Muller, & Colin Phillips. 2017. Negative polarity illusions: licensors that don't cause illusions, and blockers that do. Poster presented at the CUNY Conference on Human Sentence Processing, Boston.
- Drenhaus, Heiner, Douglas Saddy, & Stefan Frisch. 2005. Processing negative polarity items: when negation comes through the backdoor. *Linguistic evidence: Empirical, theoretical, and computational perspectives*. 145–165.
- Duffy, Susan A, John M Henderson, & Robin K Morris. 1989. Semantic facilitation of lexical access during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15(5). 791.
- Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier, & Jacques Mehler. 1999. Epenthetic vowels in Japanese: a perceptual illusion? *Journal of experimental psychology: human perception and performance* 25(6). 1568.

- Eager, Christopher & Joseph Roy. 2017. Mixed effects models are sometimes terrible. *arXiv preprint arXiv:1701.04858*.
- Erickson, Thomas D & Mark E Mattson. 1981. From words to meaning: a semantic illusion. *Journal of Verbal Learning and Verbal Behavior* 20(5). 540–551.
- Fauconnier, Gilles. 1975a. Polarity and the scale principle. In *Proceedings of the chicago linguistics society*, 11, 188–199.
- Fauconnier, Gilles. 1975b. Pragmatic scales and logical structure. *Linguistic inquiry* 6(3). 353–375.
- Ferreira, Fernanda. 2003. The misinterpretation of noncanonical sentences. *Cognitive psychology* 47(2). 164–203.
- Ferreira, Fernanda, Karl GD Bailey, & Vittoria Ferraro. 2002. Good-enough representations in language comprehension. *Current directions in psychological science* 11(1). 11–15.
- Ferreira, Fernanda & Nikole D Patson. 2007. The ‘good enough’ approach to language comprehension. *Language and linguistics compass* 1(1-2). 71–83.
- Fischler, Ira, Paul A Bloom, Donald G Childers, Salim E Roucos, & Nathan W Perry Jr. 1983. Brain potentials related to stages of sentence verification. *Psychophysiology* 20(4). 400–409.
- Foulsham, Tom, James Farley, & Alan Kingstone. 2013. Mind wandering in sentence reading: decoupling the link between mind and eye. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale* 67(1). 51.
- Franklin, Michael S, Jonathan Smallwood, & Jonathan W Schooler. 2011. Catching the mind in flight: using behavioral indices to detect mindless reading in real time. *Psychonomic bulletin & review* 18(5). 992–997.
- Frazier, Lyn & Keith Rayner. 1982. Making and correcting errors during sentence comprehension: eye movements in the analysis of structurally ambiguous sentences. *Cognitive psychology* 14(2). 178–210.
- Fromkin, Victoria A. 1973. The non-anomalous nature of anomalous utterances. *Speech errors as linguistic evidence*. 215–242.
- Gajewski, Jon Robert. 2007. Neg-raising and polarity. *Linguistics and Philosophy* 30(3). 289–328.
- Gajewski, Jon Robert. 2011. Licensing strong npis. *Natural Language Semantics* 19(2). 109–148.
- Gaston, Phoebe, Linnaea Stockall, Sarah VanWagenen, & Alec Marantz. 2021. Memory for affixes in a long-lag priming paradigm. *Glossa: a journal of general linguistics* 6(1).
- Gelman, Andrew & John Carlin. 2014. Beyond power calculations: assessing type s (sign) and type m

- (magnitude) errors. *Perspectives on Psychological Science* 9(6). 641–651.
- Getz, Laura M & Joseph C Toscano. 2021. Rethinking the mcgurk effect as a perceptual illusion. *Attention, Perception, & Psychophysics* 83(6). 2583–2598.
- Giannakidou, Anastasia. 1998. *Polarity sensitivity as (non) veridical dependency*. John Benjamins Publishing.
- Giannakidou, Anastasia. 2006. Only, emotive factive verbs, and the dual nature of polarity dependency. *Language*. 575–603.
- Gibson, Edward, Leon Bergen, & Steven T Piantadosi. 2013. Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences* 110(20). 8051–8056.
- Glanzberg, Michael. 2009. Descriptions, negation, and focus. In *Compositionality, context and semantic values*, 193–220. Springer.
- Goodman, Steven N & Jesse A Berlin. 1994. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Annals of internal medicine* 121(3). 200–206.
- Gordon, Peter C, Randall Hendrick, & Marcus Johnson. 2001. Memory interference during language processing. *Journal of experimental psychology: learning, memory, and cognition* 27(6). 1411.
- Gordon, Peter C, Randall Hendrick, & Marcus Johnson. 2004. Effects of noun phrase type on sentence complexity. *Journal of memory and Language* 51(1). 97–114.
- Green, Peter & Catriona J MacLeod. 2016. Simr: an r package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution* 7(4). 493–498.
- Grice, Herbert P. 1975. Logic and conversation. In *Speech acts*, 41–58. Brill.
- Gulordava, Kristina, Piotr Bojanowski, Edouard Grave, Tal Linzen, & Marco Baroni. 2018. Colorless green recurrent networks dream hierarchically. In Marilyn Walker, Heng Ji, & Amanda Stent (eds.), *Proceedings of the 2018 conference of the North American chapter of the association for computational linguistics: Human language technologies*, 1195–1205. New Orleans, Louisiana: Association for Computational Linguistics. DOI: [10.18653/v1/N18-1108](https://doi.org/10.18653/v1/N18-1108).
- Hagoort, Peter, Lea Hald, Marcel Bastiaansen, & Karl Magnus Petersson. 2004. Integration of word meaning and world knowledge in language comprehension. *science* 304(5669). 438–441.
- Hammerly, Christopher, Adrian Staub, & Brian Dillon. 2019. The grammaticality asymmetry in agreement attraction reflects response bias: experimental and modeling evidence. *Cognitive psychology* 110. 70–104.

- Hannon, Brenda & Meredyth Daneman. 2001. Susceptibility to semantic illusions: an individual-differences perspective. *Memory & cognition* 29(3). 449–461.
- Harley, Trevor A & Siobhan BG MacAndrew. 2001. Constraints upon word substitution speech errors. *Journal of Psycholinguistic Research* 30(4). 395–418.
- Harris, Zellig S. 1954. Distributional structure. *Word* 10(2-3). 146–162.
- Hasher, Lynn, David Goldstein, & Thomas Toppino. 1977. Frequency and the conference of referential validity. *Journal of verbal learning and verbal behavior* 16(1). 107–112.
- Hedge, Craig, Georgina Powell, & Petroc Sumner. 2018. The reliability paradox: why robust cognitive tasks do not produce reliable individual differences. *Behavior research methods* 50(3). 1166–1186.
- Herburger, Elena & Simon Mauck. 2013. The chance of being an npi. *Beyond ‘any’ and ‘ever’. New explorations in negative polarity sensitivity*. 213–240.
- Heyman, Tom, Anke Bruninx, Keith A Hutchison, & Gert Storms. 2018. The (un) reliability of item-level semantic priming effects. *Behavior Research Methods* 50(6). 2173–2183.
- Hildebrandt, Luis & Matthew Husband. 2017. Quantifiers, restrictors, and illusory npi licensing. Poster presented at the CUNY Conference on Human Sentence Processing, Boulder.
- Hoeksema, Jack. 2005. In days, weeks, months, years, ages: a class of temporal negative polarity items. *Rejected papers in linguistics etc* 1.
- Hoeksema, Jack. 2007. Parasitic licensing of negative polarity items. *The Journal of Comparative Germanic Linguistics* 10(3). 163–182.
- Homer, Vincent. 2008. Presuppositions can be disruptors too: a case against strawson-entailment. In *Proceedings of wccfl*, vol. 27, 220–228.
- Homer, Vincent. 2021. Domains of polarity items. *Journal of Semantics* 38(1). 1–48.
- Honnibal, Matthew & Ines Montani. 2017. spaCy 2: natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.
- Horn, Laurence Robert. 1972. *On the semantic properties of logical operators in english*. University of California, Los Angeles.
- Hornby, Peter A. 1974. Surface structure and presupposition. *Journal of verbal learning and verbal behavior* 13(5). 530–538.
- Huang, Kuan-Jung, Suhas Arehalli, Mari Kugemoto, Christian Muxica, Grusha Prasad, Brian Dillon, & Tal Linzen. 2022. Spr mega-benchmark shows surprisal tracks construction- but not item-level

- difficulty. Talk presented at the Human Sentence Processing Conference Satellite, Maryland.
- Hutchison, Keith A, James H Neely, & Jeffrey D Johnson. 2001. With great expectations, can two” wrongs” prime a” right”? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27(6). 1451.
- Israel, Michael. 1997. The scalar model of polarity sensitivity: the case of the aspectual operators. *Amsterdam studies in the theory and history of linguistic science series 4*. 209–230.
- Israel, Michael. 2011. *The grammar of polarity: pragmatics, sensitivity, and the logic of scales*. Vol. 127. Cambridge University Press.
- Izaute, Marie, Laurence Paire-Ficout, & Elisabeth Bacon. 2004. Benzodiazepines and semantic memory: effects of lorazepam on the moses illusion. *Psychopharmacology* 172(3). 309–315.
- van Jaarsveld, Henk J, Ton Dijkstra, & Daan Hermans. 1997. The detection of semantic illusions: task-specific effects for similarity and position of distorted terms. *Psychological research* 59(4). 219–230.
- Jackendoff, Ray S. 1969. An interpretive theory of negation. *Foundations of language*. 218–241.
- Jacobson, Pauline. 2018. What is — or, for that matter, isn’t — ‘experimental’ semantics? In *The science of meaning: essays on the metatheory of natural language semantics*, 46–72. Oxford University Press.
- Jäger, Lena A, Felix Engelmann, & Shravan Vasishth. 2017. Similarity-based interference in sentence comprehension: literature review and bayesian meta-analysis. *Journal of Memory and Language* 94. 316–339.
- Jäger, Lena A, Daniela Mertzen, Julie A Van Dyke, & Shravan Vasishth. 2020. Interference patterns in subject-verb agreement and reflexives revisited: a large-sample study. *Journal of Memory and Language* 111. 104063.
- Janouskova, Adela, Jakub Kocyan, Magdalena Simova, Kamila Uvirova, Kamila Zahradnickova, Martin Vaculik, & Jakub Prochazka. 2022. The effect of font readability on the moses illusion: a replication study. *Consciousness and Cognition* 99. 103284.
- Kadmon, Nirit & Fred Landman. 1993. Any. *Linguistics and philosophy*. 353–422.
- Kahneman, Daniel, Stewart Paul Slovic, Paul Slovic, & Amos Tversky. 1982. *Judgment under uncertainty: heuristics and biases*. Cambridge university press.
- Kamas, Eleen N, Lynne M Reder, & Michael S Ayers. 1996. Partial matching in the moses illusion: response bias not sensitivity. *Memory & Cognition* 24(6). 687–699.
- King, Jonathan & Marcel Adam Just. 1991. Individual differences in syntactic processing: the role of working memory. *Journal of memory and language* 30(5). 580–602.

- Klima, Edward. 1964. Negation in english. In Jerry Fodor & Jerrold Katz (eds.), *The structure of language*, 246–232. Prentice Hall.
- Krifka, Manfred. 1990. Four thousand ships passed through the lock: object-induced measure functions on events. *Linguistics and Philosophy*. 487–520.
- Krifka, Manfred. 1995. The semantics and pragmatics of polarity items. *Linguistic Analysis* 25(3–4). 209–257.
- Kush, Dave, Jeffrey Lidz, & Colin Phillips. 2015. Relation-sensitive retrieval: evidence from bound variable pronouns. *Journal of memory and language* 82. 18–40.
- Kush, Dave W. 2013. *Respecting relations: memory access and antecedent retrieval in incremental sentence processing*. University of Maryland, College Park dissertation.
- Kwon, Nayoung & Patrick Sturt. 2016. Attraction effects in honorific agreement in korean. *Frontiers in psychology* 7. 1302.
- Ladusaw, William. 1996. Negation and polarity items. In Shalom Lappin (ed.), *The handbook of contemporary semantic theory*, 321–341. Oxford: Blackwell Reference.
- Ladusaw, William. 1979. *Polarity sensitivity as inherent scope relations*. University of Texas at Austin dissertation.
- Lago, Sol, Carlos Acuña Fariña, & Enrique Meseguer. 2021. The reading signatures of agreement attraction. *Open Mind*. 1–22.
- Laka, Itziar. 1994. *On the syntax of negation*. Garland Publishing.
- Lakens, Daniël, Anne M Scheel, & Peder M Isager. 2018. Equivalence testing for psychological research: a tutorial. *Advances in Methods and Practices in Psychological Science* 1(2). 259–269.
- Lee, David S, Eunjung Kim, & Norbert Schwarz. 2015. Something smells fishy: olfactory suspicion cues improve performance on the moses illusion and wason rule discovery task. *Journal of Experimental Social Psychology* 59. 47–50.
- Lee, So Young, Aydogan Yanilmaz, Jiwon Yun, & John E Drury. 2018. The processing of turkish and korean npi licensing and intrusion: erp evidence. Poster presented at the CUNY Conference on Human Sentence Processing, Davis.
- Lenth, Russell, Henrik Singmann, Jonathon Love, Paul Buerkner, & Maxime Herve. 2018. Emmeans: estimated marginal means, aka least-squares means. *R package version* 1(1). 3.
- Levelt, Willem JM, Ardi Roelofs, & Antje S Meyer. 1999. A theory of lexical access in speech production. *Behavioral and brain sciences* 22(1). 1–38.

- Levine, Marc & Mary HH Ensom. 2001. Post hoc power analysis: an idea whose time has passed? *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy* 21(4). 405–409.
- Levy, Roger. 2008. A noisy-channel model of human sentence comprehension under uncertain input. In *Proceedings of the 2008 conference on empirical methods in natural language processing*, 234–243.
- Lewis, David. 1975. Adverbs of quantification. In Paul H. Portner & Barbara H. Partee (eds.), vol. 178, 188.
- Lewis, Richard L & Shravan Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive science* 29(3). 375–419.
- Linebarger, Marcia C. 1987. Negative polarity and grammatical representation. *Linguistics and philosophy*. 325–387.
- Liu, Feng-hsi. 1990. *Scope dependency in english and chinese*. University of California: Los Angeles dissertation.
- Lowder, Matthew W & Peter C Gordon. 2021. Relative clause effects at the matrix verb depend on type of intervening material. *Cognitive Science* 45(9). e13039.
- Lucas, Margery. 2000. Semantic priming without association: a meta-analytic review. *Psychonomic bulletin & review* 7(4). 618–630.
- Lund, Kevin & Curt Burgess. 1996. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior research methods, instruments, & computers* 28(2). 203–208.
- Maillat, Didier & Steve Oswald. 2009. Defining manipulative discourse: the pragmatics of cognitive illusions. *International Review of Pragmatics* 1(2). 348–370.
- Malko, Anton. 2018. *The role of structural information in the resolution of long-distance dependencies*. University of Maryland: College Park dissertation.
- Marr, David. 1982. *Vision: a computational investigation into the human representation and processing of visual information*. San Francisco: W. H. Freeman.
- McClelland, James L & Jeffrey L Elman. 1986. The trace model of speech perception. *Cognitive psychology* 18(1). 1–86.
- McCourt, Michael Sullivan. 2021. *Semantics and pragmatics in a modular mind*. University of Maryland, College Park dissertation.
- McElree, Brian. 2006. Accessing recent events. *Psychology of learning and motivation* 46. 155–200.
- McGurk, Harry & John MacDonald. 1976. Hearing lips and seeing voices. *Nature* 264(5588). 746–

- Mendia, Jon Ander, Ethan Poole, & Brian Dillon. 2018. Spurious np-i licensing and exhaustification. In *Semantics and linguistic theory*, vol. 28, 233–250.
- Meseguer, Enrique, Manuel Carreiras, & Charles Clifton. 2002. Overt reanalysis strategies and eye movements during the reading of mild garden path sentences. *Memory & cognition* 30(4). 551–561.
- Meyer, Andrew, Shane Frederick, Terence C Burnham, Juan D Guevara Pinto, Ty W Boyer, Linden J Ball, Gordon Pennycook, Rakefet Ackerman, Valerie A Thompson, & Jonathon P Schuldt. 2015. Disfluent fonts don't help people solve math problems. *Journal of Experimental Psychology: General* 144(2). e16.
- Meyer, David E & Roger W Schvaneveldt. 1971. Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of experimental psychology* 90(2). 227.
- Mikolov, Tomáš, Wen-tau Yih, & Geoffrey Zweig. 2013. Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: human language technologies*, 746–751.
- Miller, George A, Richard Beckwith, Christiane Fellbaum, Derek Gross, & Katherine J Miller. 1990. Introduction to wordnet: an on-line lexical database. *International journal of lexicography* 3(4). 235–244.
- Miller, George A & Stephen Isard. 1964. Free recall of self-embedded english sentences. *Information and control* 7(3). 292–303.
- Montalbetti, Mario M. 1984. *After binding: on the interpretation of pronouns*. Massachusetts Institute of Technology dissertation.
- Muller, Hanna, Iria de Dios Flores, & Colin Phillips. 2019. Not (just) any licensors cause negative polarity illusions. Poster presented at the CUNY Conference on Human Sentence Processing, Boulder.
- Neely, James H. 1977. Semantic priming and retrieval from lexical memory: roles of inhibitionless spreading activation and limited-capacity attention. *Journal of experimental psychology: general* 106(3). 226.
- Neely, James H, Dennis E Keefe, & Kent L Ross. 1989. Semantic priming in the lexical decision task: roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15(6). 1003.
- Ng, Anne & Matthew Husband. 2017. Interference effects across the at-issue/not-at-issue divide: agreement and np-i licensing. Poster presented at the CUNY Conference on Human Sentence Processing, Boston.

- Nicol, Janet & David Swinney. 1989. The role of structure in coreference assignment during sentence comprehension. *Journal of psycholinguistic research* 18(1). 5–19.
- Nicol, Janet L, Kenneth I Forster, & Csaba Veres. 1997. Subject–verb agreement processes in comprehension. *Journal of Memory and Language* 36(4). 569–587.
- Nieuwland, Mante S & Jos JA Van Berkum. 2006. When peanuts fall in love: n400 evidence for the power of discourse. *Journal of cognitive neuroscience* 18(7). 1098–1111.
- O’Brien, Edward J & Anne E Cook. 2016. Coherence threshold and the continuity of processing: the ri-val model of comprehension. *Discourse Processes* 53(5-6). 326–338.
- O’Connor, Ellen. 2015. *Comparative illusions at the syntax-semantics interface*. University of Southern California dissertation.
- van Oostendorp, Herre & Ineke Kok. 1990. Failing to notice errors in sentences. *Language and cognitive processes* 5(2). 105–113.
- van Oostendorp, Herre & Sjaak de Mul. 1990. Moses beats adam: a semantic relatedness effect on a semantic illusion. *Acta psychologica* 74(1). 35–46.
- Orth, Wesley, Masaya Yoshida, & Shayne Sloggett. 2020a. Illusions of ungrammaticality: evidence from ppi. Poster presented at the CUNY Conference on Human Sentence Processing, Amherst.
- Orth, Wesley, Masaya Yoshida, & Shayne Sloggett. 2020b. Tracking the time course of np_i illusion: why do only some appear online? Poster presented at the CUNY Conference on Human Sentence Processing, Amherst.
- Orth, Wesley, Masaya Yoshida, & Shayne Sloggett. 2021. Negative polarity item (np_i) illusion is a quantification phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 47(6). 906.
- Paape, Dario, Shravan Vasishth, & Titus von der Malsburg. 2020. Quadruplex negatio invertit? the on-line processing of depth charge sentences. *Journal of Semantics* 37(4). 509–555.
- Parker, Dan. 2019. Two minds are not always better than one: modeling evidence for a single sentence analyzer. *Glossa*.
- Parker, Dan. 2022. Ellipsis interference revisited: new evidence for feature markedness effects in retrieval. *Journal of Memory and Language* 124. 104314.
- Parker, Dan & Colin Phillips. 2011. Illusory negative polarity item licensing is selective. Poster presented at the CUNY Conference on Human Sentence Processing, Stanford.
- Parker, Dan & Colin Phillips. 2016. Negative polarity illusions and the format of hierarchical encodings

- in memory. *Cognition* 157. 321–339.
- Parker, Dan & Colin Phillips. 2017. Reflexive attraction in comprehension is selective. *Journal of Memory and Language* 94. 272–290.
- Patil, Umesh, Shravan Vasishth, & Richard L Lewis. 2016. Retrieval interference in syntactic processing: the case of reflexive binding in english. *Frontiers in Psychology* 7. 329.
- Patson, Nikole D & E Matthew Husband. 2016. Misinterpretations in agreement and agreement attraction. *Quarterly Journal of Experimental Psychology* 69(5). 950–971.
- Pearlmutter, Neal J, Susan M Garnsey, & Kathryn Bock. 1999. Agreement processes in sentence comprehension. *Journal of Memory and language* 41(3). 427–456.
- Perugini, Marco, Marcello Gallucci, & Giulio Costantini. 2014. Safeguard power as a protection against imprecise power estimates. *Perspectives on Psychological Science* 9(3). 319–332.
- Pietroski, Paul M. 2018. *Conjoining meanings: semantics without truth values*. Oxford University Press.
- Postal, Paul M. 2005. Suppose (if only for an hour) that negative polarity items are negation-containing phrases. *Ms., New York University*.
- Potter, Mary C & Linda Lombardi. 1990. Regeneration in the short-term recall of sentences. *Journal of Memory and Language* 29(6). 633–654.
- Radford, Alec, Jeff Wu, Rewon Child, David Luan, Dario Amodei, & Ilya Sutskever. 2019. Language models are unsupervised multitask learners.
- Rao, Rajesh PN & Dana H Ballard. 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience* 2(1). 79–87.
- Raposo, Ana & J Frederico Marques. 2013. The contribution of fronto-parietal regions to sentence comprehension: insights from the moose illusion. *NeuroImage* 83. 431–437.
- Reder, Lynne M & Axel Cleeremans. 1990. The role of partial matches in comprehension: the moose illusion revisited. In *Psychology of learning and motivation*, vol. 25, 233–258. Elsevier.
- Reder, Lynne M & Gail W Kusbit. 1991. Locus of the moose illusion: imperfect encoding, retrieval, or match? *Journal of Memory and Language* 30(4). 385–406.
- Reichenbach, Hans. 1947. *Elements of symbolic logic*.
- Rohde, Douglas LT, Laura M Gonnerman, & David C Plaut. 2006. An improved model of semantic similarity based on lexical co-occurrence. *Communications of the ACM* 49(627-633). 116.

- Rullmann, Hotze. 2003. Additive particles and polarity. *Journal of semantics* 20(4). 329–401.
- Russell, Bertrand. 1905. On denoting. *Mind* 14(56). 479–493.
- Salovich, Nikita A, Anya M Kirsch, & David N Rapp. 2022. Evaluative mindsets can protect against the influence of false information. *Cognition* 225. 105121.
- Salton, Gerard, Anita Wong, & Chung-Shu Yang. 1975. A vector space model for automatic indexing. *Communications of the ACM* 18(11). 613–620.
- Sanford, Anthony J, Hartmut Leuthold, Jason Bohan, & Alison JS Sanford. 2011. Anomalies at the borderline of awareness: an erp study. *Journal of Cognitive Neuroscience* 23(3). 514–523.
- Sanford, Anthony J, Linda M Moxey, & Kevin B Paterson. 1996. Attentional focusing with quantifiers in production and comprehension. *Memory & Cognition* 24(2). 144–155.
- Sanford, Anthony J & Patrick Sturt. 2002. Depth of processing in language comprehension: not noticing the evidence. *Trends in cognitive sciences* 6(9). 382–386.
- Schad, Daniel J, Antje Nuthmann, & Ralf Engbert. 2012. Your mind wanders weakly, your mind wanders deeply: objective measures reveal mindless reading at different levels. *Cognition* 125(2). 179–194.
- Schlueter, Zoe, Dan Parker, & Ellen Lau. 2019. Error-driven retrieval in agreement attraction rarely leads to misinterpretation. *Frontiers in psychology* 10. 1002.
- Sedivy, Julie C, Michael K Tanenhaus, Craig G Chambers, & Gregory N Carlson. 1999. Achieving incremental semantic interpretation through contextual representation. *Cognition* 71(2). 109–147.
- Sennrich, Rico, Barry Haddow, & Alexandra Birch. 2016. Neural machine translation of rare words with subword units. In *Proceedings of the 54th annual meeting of the association for computational linguistics*, 1715–1725. Berlin, Germany: Association for Computational Linguistics. DOI: [10.18653/v1/P16-1162](https://doi.org/10.18653/v1/P16-1162).
- Shafto, Meredith & Donald MacKay. 2000. The mooses, mega-mooses, and armstrong illusions: integrating language comprehension and semantic memory. *Psychological Science* 11(5). 372–378.
- Shafto, Meredith & Donald MacKay. 2010. Miscomprehension, meaning, and phonology: the unknown and phonological armstrong illusions. *European Journal of Cognitive Psychology* 22(4). 529–568.
- Shapiro, Kimron L, Jane E Raymond, & Karen M Arnell. 1997. The attentional blink. *Trends in cognitive sciences* 1(8). 291–296.
- Simons, Daniel J & Daniel T Levin. 1997. Change blindness. *Trends in cognitive sciences* 1(7). 261–267.

- Simonsohn, Uri. 2015. Small telescopes: detectability and the evaluation of replication results. *Psychological science* 26(5). 559–569.
- Slattery, Timothy J, Patrick Sturt, Kiel Christianson, Masaya Yoshida, & Fernanda Ferreira. 2013. Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language* 69(2). 104–120.
- Slevc, L Robert & Randi C Martin. 2016. Syntactic agreement attraction reflects working memory processes. *Journal of Cognitive Psychology* 28(7). 773–790.
- Slioussar, Natalia & Anton Malko. 2016. Gender agreement attraction in russian: production and comprehension evidence. *Frontiers in psychology* 7. 1651.
- Sloggett, Shayne. 2017. *When errors aren't: how comprehenders selectively violate binding theory*. University of Massachusetts: Amherst dissertation.
- Smallwood, Jonathan. 2011. Mind-wandering while reading: attentional decoupling, mindless reading and the cascade model of inattention. *Language and Linguistics Compass* 5(2). 63–77.
- Song, Hyunjin & Norbert Schwarz. 2008. Fluency and the detection of misleading questions: low processing fluency attenuates the moles illusion. *Social cognition* 26(6). 791–799.
- Sonnenschein, Susan & Grover J Whitehurst. 1982. The effects of redundant communications on the behavior of listeners: does a picture need a thousand words? *Journal of Psycholinguistic Research* 11(2). 115–125.
- Speckmann, Felix & Christian Unkelbach. 2021. Moles, money, and multiple-choice: the moles illusion in a multiple-choice format with high incentives. *Memory & Cognition* 49(4). 843–862.
- Staub, Adrian. 2010. Eye movements and processing difficulty in object relative clauses. *Cognition* 116(1). 71–86.
- Staub, Adrian, Brian Dillon, & Charles Clifton. 2017. The matrix verb as a source of comprehension difficulty in object relative sentences. *Cognitive Science* 41. 1353–1376.
- Sturt, Patrick. 2003. The time-course of the application of binding constraints in reference resolution. *Journal of memory and language* 48(3). 542–562.
- Sturt, Patrick, Anthony J Sanford, Andrew Stewart, & Eugene Dawydiak. 2004. Linguistic focus and good-enough representations: an application of the change-detection paradigm. *Psychonomic bulletin & review* 11(5). 882–888.
- Szabolcsi, Anna, Lewis Bott, & Brian McElree. 2008. The effect of negative polarity items on inference verification. *Journal of semantics* 25(4). 411–450.

- Tabor, Whitney, Bruno Galantucci, & Daniel Richardson. 2004. Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language* 50(4). 355–370.
- Tian, Ye & Richard Breheny. 2016. Dynamic pragmatic view of negation processing. In *Negation and polarity: experimental perspectives*, 21–43. Springer.
- Townsend, David J & Thomas G Bever. 2001. *Sentence comprehension: the integration of habits and rules*. MIT Press.
- Traxler, Matthew J. 2014. Trends in syntactic parsing: anticipation, bayesian estimation, and good-enough parsing. *Trends in cognitive sciences* 18(11). 605–611.
- Tune, Sarah, Matthias Schlesewsky, Steven L Small, Anthony J Sanford, Jason Bohan, Jona Sassenhagen, & Ina Bornkessel-Schlesewsky. 2014. Cross-linguistic variation in the neurophysiological response to semantic processing: evidence from anomalies at the borderline of awareness. *Neuropsychologia* 56. 147–166.
- Tunstall, Susanne Lynn. 1998. *The interpretation of quantifiers: semantics and processing*. University of Massachusetts: Amherst dissertation.
- Umanath, Sharda, Patrick O Dolan, & Elizabeth J Marsh. 2014. Ageing and the moses illusion: older adults fall for moses but if asked directly, stick with noah. *Memory* 22(5). 481–492.
- Uribe-Echevarria, Maria. 1994. *Interface licensing conditions on negative polarity items: a theory of polarity and tense interactions*. University of Connecticut dissertation.
- Vaessen, Rowie. 2017. *The effect of semantic illusions on l1 and l2 sentence processing: did moses speak two languages on the ark?* MA thesis.
- Vasishth, Shravan, Sven Brüssow, Richard L Lewis, & Heiner Drenhaus. 2008. Processing polarity: how the ungrammatical intrudes on the grammatical. *Cognitive Science* 32(4). 685–712.
- Vasishth, Shravan, Daniela Mertzen, Lena A Jäger, & Andrew Gelman. 2018. The statistical significance filter leads to overoptimistic expectations of replicability. *Journal of Memory and Language* 103. 151–175.
- Von der Malsburg, Titus & Shravan Vasishth. 2013. Scanpaths reveal syntactic underspecification and reanalysis strategies. *Language and Cognitive Processes* 28(10). 1545–1578.
- Von Fintel, Kai. 1999. Npi licensing, strawson entailment, and context dependency. *Journal of semantics* 16(2). 97–148.
- Wagers, Matthew W, Ellen F Lau, & Colin Phillips. 2009. Agreement attraction in comprehension: representations and processes. *Journal of memory and language* 61(2). 206–237.

- Wason, Peter C & Shuli S Reich. 1979. A verbal illusion. *Quarterly Journal of Experimental Psychology* 31(4). 591–597.
- Wellwood, Alexis, Roumyana Pancheva, Valentine Hacquard, & Colin Phillips. 2018. The anatomy of a comparative illusion. *Journal of Semantics* 35(3). 543–583.
- van der Wouden, Ton. 1997. *Negative contexts*. Routledge.
- Xiang, Ming, Brian Dillon, & Colin Phillips. 2006. Testing the strength of the spurious licensing effect for negative polarity items. Talk presented at the CUNY Conference on Human Sentence Processing, New York.
- Xiang, Ming, Brian Dillon, & Colin Phillips. 2009. Illusory licensing effects across dependency types: erp evidence. *Brain and Language* 108(1). 40–55.
- Xiang, Ming, Julian Grove, & Anastasia Giannakidou. 2013. Dependency-dependent interference: npI interference, agreement attraction, and global pragmatic inferences. *Frontiers in psychology* 4. 708.
- Xiang, Ming, Julian Grove, & Anastasia Giannakidou. 2016. Semantic and pragmatic processes in the comprehension of negation: an event related potential study of negative polarity sensitivity. *Journal of Neurolinguistics* 38. 71–88.
- Yanilmaz, Aydogan & John E Drury. 2018a. Intervening and non-intervening interference. Poster presented at the CUNY Conference on Human Sentence Processing, Davis.
- Yanilmaz, Aydogan & John E Drury. 2018b. Prospective npI licensing and intrusion in Turkish. *Language, Cognition and Neuroscience* 33(1). 111–138.
- Yun, Jiwon, So-Young Lee, & John E Drury. 2018. Negative polarity illusion in Korean. In Celeste Guillemot, Tomoyuki Yoshida, & Seunghun J Lee (eds.), *Proceedings of the 13th workshop on Altaic formal linguistics*. MIT Press.
- Zöllner, F. 1860. Über eine neue art von pseudoskopie und ihre beziehungen zu den von plateau und oppel beschriebenen bewegungsphanomenen. *Annalen der Physik und Chemie* 186. 500–523.
- Zwarts, Frans. 1998. Three types of polarity. In *Plurality and quantification*, 177–238. Springer.