

**SHIFTING THE PERSPECTIVAL LANDSCAPE:  
METHODS FOR ENCODING, IDENTIFYING, AND SELECTING  
PERSPECTIVES**

A Dissertation Presented

by

CAROLYN JANE ANDERSON

Submitted to the Graduate School of the  
University of Massachusetts Amherst in partial fulfillment  
of the requirements for the degree of

**DOCTOR OF PHILOSOPHY**

September 2020

Linguistics

© Copyright by Carolyn Jane Anderson 2020

All Rights Reserved

**SHIFTING THE PERSPECTIVAL LANDSCAPE:  
METHODS FOR ENCODING, IDENTIFYING, AND SELECTING  
PERSPECTIVES**

A Dissertation Presented

by

CAROLYN JANE ANDERSON

Approved as to style and content by:

---

Rajesh Bhatt, Co-chair

---

Brian Dillon, Co-chair

---

Daniel Altshuler, Member

---

Mohit Iyyer, External Member

---

Joe Pater, Department Chair  
Linguistics

## **DEDICATION**

*To Barbara and Charles Hill, who taught their daughters and grandchildren to value learning.*

## ACKNOWLEDGMENTS

The first fruits of my gratitude are owed to my committee members: Rajesh Bhatt, Brian Dillon, Daniel Altshuler, and Mohit Iyyer.

Brian took me in when I was a field semanticist and turned me into a computational psycholinguist. I am very grateful to him for agreeing to advise me despite my occasional claims that syntax is “just drawing pictures.” Despite his claim that he “isn’t a semanticist” he has been an excellent advisor to this semanticist.

Daniel’s advice has been both challenging and invigorating. I have always come out of our advising sessions with a narrower focus and renewed energy. I am also grateful to Daniel as a model of an active scholar, teacher, and advisor in an undergraduate-focused environment.

I started working with Rajesh more recently, but he played a pivotal role in guiding several seemingly unconnected projects into a (hopefully) cohesive whole. The range of his expertise has been invaluable in making connections between seemingly disparate areas of the project. I have also learned a lot about undergraduate advising from the resonant strains of Rajesh’s advising conversations that drift down the hallway and into my office.

I would also like to thank Mohit Iyyer for providing a perspective from the NLP community, which was particularly valuable in framing and developing the computational models presented in the second half of this dissertation.

Next, I owe gratitude towards my collaborators. I would like to thank my research assistant, Tessa Patapoutian, who coded the production data presented in Chapter 8 and spent many hours probing neural network representations of perspectival motion verbs for another project. Although it did not become a part of this dissertation, the insightful discussions that I had with Tessa while working on that project influenced my thinking about many aspects of the dissertation.

I also collaborated with Jack Duff on the issue of interspeaker variability in perspective identification. His own work on epithets and his insight into the similarities in our data were helpful in addressing one of the most puzzling findings of Chapter 4.

Throughout the development of this dissertation, I benefitted from discussions with a number of people at UMass and elsewhere. The analysis of *tomorrow* presented in Chapter 4 benefitted greatly from feedback from many people throughout its development, including John Kingston, Amy Rose Deal, Vincent Homer, Hazel Pearson, Kristen Syrett, Philippe Schlenker, Emar Maier, and Isabelle Charnavel. The models and experimental results discussed in Chapters 7 and 8 owe much to discussions with Isabelle Charnavel, Lyn Frazier, Judith Degen, Jesse Harris, and Richard Futrell.

In my time at UMass, I have learned so much from the members of the UMass Semantics workshop, particularly, Seth Cable, Angelika Kratzer, and Ana Arregui; and the members of the Psycholinguistics workshop, particularly Lyn Frazier, John Kingston, and Gaja Jarosz. I'd like to thank Gaja in particular for her advice on being a computational linguist outside of academia and on being a female computational linguist within academia.

I'd like to thank Kyle Johnson, for a most excellent cat,<sup>1</sup> Calvin Coolidge.<sup>2</sup>

I also owe many thanks to Tom Maxfield, whose hard work and friendliness make life in the UMass Linguistics department run much more smoothly.

I'd like to thank some of the UMass companions who have encouraged me during the last five years: Sakshi Bhatia, Rodica Ivan, Kimberly Johnson, Leah Chapman, Alex Göbel, Zahra Mirrazi, Erika Mayer, and Max Nelson. I would also particularly like to thank Andrew Lamont, Brandon Prickett, Maggie Baird, Chris Hammerly, Coral Hughto, Georgia Simon, Mike Clauss, Shay Hucklebridge, and the rest of the Sound Psychological Feelings Spiders for their support in both trivial and consequential pursuits.

---

<sup>1</sup>I use the indefinite article here to avoid slighting the other excellent cats that have kept me company throughout my PhD, Smurfie Libkind and Captain Haddock.

<sup>2</sup>Né Theodore.

I owe special thanks to Jyoti Iyer, who has been an excellent office mate for the last few years, at least until the onset of the pandemic made our office inaccessible. Thank you for puzzling over data with me, supporting me through the worst days of grad school, and putting up with the alarming expansion of my hazard zone of tea leaves and mugs.

My interest in formal semantics and computational approaches developed in tandem at Swarthmore under the respective guidance of Ted Fernald and Rich Wicentowski. I would like to thank Rich for his calm kindness and valuable advice at the most precarious moment of my PhD. I would also like to thank the whole Swarthmore CS department for creating such an unusually welcoming environment. I am a very stubborn person; without the intervention of such drastic friendliness, I would never have found the computational community in which I am becoming more and more at home.

I owe my first academic home, the Swarthmore Linguistics department, a large debt of gratitude. Ted Fernald was my very first academic advisor and I could not have asked for a better one. I would particularly like to thank the other two members of Ted's advanced semantics seminar for puzzling through Montague's PTQ with me and demonstrating the value of interdisciplinary collaboration. I would also like to thank Aaron Dinkin for his supervision of my undergraduate thesis. I have kept his balance of free exploration and structured supervision in mind as a reference model as I embark on my own advising relationships.

My academic path towards this dissertation and on to the future has been one of constant meandering back and forth between Computer Science and Linguistics. I owe Nate Foster thanks for many things, not the least of which is his confidence that the choice between disciplines is not as significant as it sometimes seems.<sup>3</sup> I also thank Leslie Saxon and Jessica Coon for welcoming me as I was learning about the many different ways of being a linguist.

I have had many advisors and mentors along my zigzagging path, but one constant has been the excellent mentorship that I have received from Brook Lillehaugen. Brook guided me through my first journal submission and through two years of fieldwork in Oaxaca. I

---

<sup>3</sup>The most of which is his sound selection of postdocs.

also owe warm thanks to Felipe Lopez, for his help during my time in Oaxaca, and to the community of San Lucas Quiaviní.

Next, I turn to the friends who have supported me. I would like to thank Chris Geissler, for his generosity in snack provision and other crucial aspects of friendship; Elizabeth Grumer, for her constant friendship over many years; and Kelly Larson, for her engaging correspondence and hospitality.

I also owe thanks to Jeremy Rapaport-Stein and Elena Kingston for many pleasant dinners in Boston, Northampton, and between, and for their gracious hospitality on a number of occasions, and to the rest of the friends of Sophie, for lending their sympathetic ears and their names. You are much better friends than your Lego alter egos suggest.

I would particularly like to thank Sophie Libkind, Athena Froehlich, and Kate Derosier for their friendship and support over the last decade. If I were to enumerate all of the debts of gratitude I owe them, the resulting list would be longer than this dissertation. I hope it is enough to say that I would not have become the person that I am now if we had never met.

I also owe many thanks to my family. In particular, I want to thank my parents, Pam and Alan Anderson, for teaching me to express myself coherently and, when necessary, firmly, and to my brother, Nicholas Anderson, for many years of shoe-rescuing and silly debates. As an ideal older brother, Nick has inspired much of my academic success as I chased after him, spurred on by his expert use of the elder sibling weapon of choice: relentless teasing.

I owe thanks to Barbara Hill, Alicia Anderson, Peggy Dean, Becky Dean, Jenny Clark, Harry Dean, and the rest of the growing family that I was born into, for book suggestions, board games, and merriment of various kinds, both in-person and, lately, virtual. I also owe many thanks to Indranie, Subhasish, and Jishnu Guha, for welcoming me so warmly into a new family.

Last of all, I would like to thank Arjun Guha for sharpening my sense of curiosity and wonder, for lending his Legos and engineering skills to help build the stimuli for Chapter 4, for explaining continuations to me about six different times, and for supporting me in innumerable other ways.

## **ABSTRACT**

### **SHIFTING THE PERSPECTIVAL LANDSCAPE: METHODS FOR ENCODING, IDENTIFYING, AND SELECTING PERSPECTIVES**

SEPTEMBER 2020

CAROLYN JANE ANDERSON

B.A., SWARTHMORE COLLEGE

M.A., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Rajesh Bhatt and Professor Brian Dillon

This dissertation explores the semantics and pragmatics of perspectival expressions. Perspective, or point-of-view, encompasses an individual's thoughts, perceptions, and location. Many expressions in natural language have components of their meanings that shift depending on whose perspective they are evaluated against. In this dissertation, I explore two sets of questions relating to perspective sensitivity. The first set of questions relate to how perspective is encoded in the semantics of perspectival expressions. The second set of questions relate to how conversation participants treat perspectival expressions: the speaker's selection of a perspective and the listener's identification of the speaker's perspective.

In Part I, I explore the landscape of perspectival expressions by exploring different semantic mechanisms for encoding the perspective holder. In Chapter 2, I introduce key properties

of perspectival expressions through a discussion of one canonical perspectival expression: the motion verb *come*. In Chapter 3, I discuss the various ways of encoding the perspective holder in the semantics of perspectival expressions. I contrast the predictions of these approaches and lay out a set of diagnostics to guide the analysis of perspectival expressions.

I present two case studies using this set of diagnostics. In Chapter 3, I probe the semantics of the well-studied perspectival expression *come* in American English, and argue in favor of a perspective-anaphoric analysis. In Chapter 4, I focus on an expression that has not previously been recognized as perspectival, the temporal adverbial *tomorrow*. Through a series of experimental studies, I make the case that *tomorrow* is perspective-sensitive for some American English speakers, and narrow the hypothesis space for a perspectival account of *tomorrow*. I sketch a perspective-anaphoric semantics for *tomorrow*, while leaving open the possibility of a logophoric analysis. I conclude Part I with a discussion of how perspectival expressions fit into the broader landscape of context sensitivity.

In Part II, I turn to a fresh set of questions about perspective: how do conversation participants select and identify perspectives? In Chapter 6, I discuss previous models of perspective production and comprehension, and factors that affect these processes, such as a bias towards the perspective of the speaker. I argue that although the selection and identification of perspective holders may be guided by simple heuristics some of the time, certain cases require a more involved reasoning system. In Chapters 7 and 8, I develop models of perspectival reasoning in comprehension and production rooted in a leading framework for pragmatic reasoning: the Rational Speech Acts framework.

In Chapter 7, I propose and implement a computational model of perspective identification. I posit that listeners reason jointly about the speaker’s intended message and their adopted perspective using a mental model of the speaker’s production process. I present two comprehension studies that support a key assumption of the proposed Perspectival Rational Speech Acts model: that listeners reason simultaneously over multiple perspectives to better understand the speaker’s intended meaning.

In Chapter 8, I propose a model of perspective selection that mirrors the Perspectival Rational Speech Acts comprehension model. I posit that speakers reason about the listener's comprehension process in order to pick a perspective and an utterance that will maximize their chance of being understood. However, the results of the production study do not match the model's predictions. I conclude with a discussion of the challenges that the attested asymmetry between speaker and listeners poses for the Rational Speech Acts framework.

The main contributions of this dissertation are as follows: (1) a comparison of four approaches to encoding the semantics of perspective, leading to a diagnostic toolkit for perspectival expressions; (2) an experimental case study that employs the diagnostics to identify a novel perspectival expression; (3) an implemented computational model of perspective identification, supported by experimental evidence; and (4) an implemented computational model of perspective selection, which reveals further challenges in perspective production.

## TABLE OF CONTENTS

	Page
<b>ACKNOWLEDGMENTS .....</b>	<b>v</b>
<b>ABSTRACT .....</b>	<b>ix</b>
<b>LIST OF TABLES .....</b>	<b>xiii</b>
<b>LIST OF FIGURES .....</b>	<b>xiv</b>
<b>LIST OF ABBREVIATIONS .....</b>	<b>xv</b>
 <b>CHAPTER</b>	
<b>1. INTRODUCTION .....</b>	<b>1</b>
 <b>PART I: ENCODING PERSPECTIVES</b>	
<b>2. PERSPECTIVAL MOTION VERBS .....</b>	<b>7</b>
2.1 The lexical semantics of perspectival motion verbs .....	8
2.2 Components of the licensing of <i>come</i> .....	12
2.3 Properties of perspectival motion verbs.....	20
2.4 Summary.....	28
<b>3. THE SEMANTICS OF PERSPECTIVE.....</b>	<b>29</b>
3.1 Perspectival phenomena .....	29

3.2	Possible semantics for perspectival expressions .....	41
3.3	A set of diagnostics for perspectival encoding .....	55
3.4	Applying perspectival diagnostics to <i>come</i> .....	66
3.5	Summary.....	78
<b>4.</b>	<b>EXPLORING TEMPORAL PERSPECTIVE .....</b>	<b>80</b>
4.1	Introduction .....	80
4.2	The semantics of <i>tomorrow</i> .....	82
4.3	Experiment 1: establishing a baseline .....	97
4.4	Experiment 2: testing the quotational account .....	104
4.5	Experiments 3a and 3b: testing the Free Indirect Discourse account.....	107
4.6	Experiment 4: testing the indexical shift account .....	115
4.7	Developing a non-indexical account of <i>tomorrow</i> .....	123
4.8	Towards a perspectival account of <i>tomorrow</i> .....	135
4.9	Summary.....	146
<b>5.</b>	<b>INTERIM CONCLUSION .....</b>	<b>150</b>
5.1	The theoretical landscape of perspectival expressions.....	150
5.2	The attested landscape of perspective sensitivity .....	155
5.3	Open questions .....	156
5.4	Taking stock .....	162
<b>PART II: TAKING PERSPECTIVES</b>		
<b>6.</b>	<b>PERSPECTIVE SELECTION AND IDENTIFICATION .....</b>	<b>166</b>
6.1	Approaches to perspective identification .....	168
6.2	Approaches to perspective selection .....	174
6.3	A bias towards the speaker .....	178
6.4	The cost of perspective shift .....	186

6.5	Discourse factors in perspective shift .....	190
6.6	Towards models of perspectival reasoning.....	198
<b>7.</b>	<b>A RATIONAL APPROACH TO PERSPECTIVE INFERENCE .....</b>	<b>200</b>
7.1	Rational approaches to pragmatics .....	202
7.2	Key properties of perspectival motion verb comprehension .....	211
7.3	An RSA model of perspectival expressions .....	215
7.4	Capturing key properties of perspective identification .....	218
7.5	How do listeners interpret perspectival expressions?.....	236
7.6	Summary.....	251
<b>8.</b>	<b>A RATIONAL APPROACH TO PERSPECTIVE SELECTION .....</b>	<b>253</b>
8.1	The Rational Speech Acts framework as a model of production .....	254
8.2	A RSA model of utterance and perspective selection .....	261
8.3	Capturing perspective shift with the PRSA production model .....	263
8.4	How do speakers select perspectival expressions? .....	273
8.5	Extensions to the PRSA model.....	285
8.6	Summary.....	292
<b>9.</b>	<b>CONCLUSION .....</b>	<b>295</b>
9.1	Main findings .....	296
9.2	Open questions .....	300
9.3	Last thoughts .....	305
<b>APPENDIX: DEMOGRAPHIC ANALYSIS OF NON-UTTERANCE TIME <i>TOMORROW</i> READINGS.....</b>	<b>307</b>	
<b>APPENDIX: LOGOPHORIC ANALYSES OF <i>TOMORROW</i> .....</b>	<b>313</b>	
<b>APPENDIX: <i>DE SE</i> JUDGMENTS AND NON-UTTERANCE TIME <i>TOMORROW</i> .....</b>	<b>331</b>	

APPENDIX: PRSA MODEL SPECIFICATION AND IMPLEMENTATION.....	333
APPENDIX: ADDITIONAL RSA FINDINGS .....	339
APPENDIX: EXPERIMENT 1 ITEMS .....	346
APPENDIX: EXPERIMENT 2 ITEMS .....	357
APPENDIX: EXPERIMENT 3A ITEMS.....	367
APPENDIX: EXPERIMENT 3B ITEMS.....	377
APPENDIX: EXPERIMENT 4 ITEMS .....	387
APPENDIX: QUANTIFICATIONAL BINDING TASK 1 ITEMS .....	398
APPENDIX: QUANTIFICATIONAL BINDING TASK 2 ITEMS .....	400
APPENDIX: EXPERIMENT 5A AND 5B ITEMS .....	402
APPENDIX: EXPERIMENT 6 ITEMS .....	414
BIBLIOGRAPHY .....	424

## LIST OF TABLES

<b>Table</b>	<b>Page</b>
2.1 Typology of perspective holders for <i>come</i> from Gathercole (1987), Nakazawa (2007), and Barlew (2017) (English and SLQ Zapotec data from my work) .....	13
3.1 Candidate perspectival expressions proposed by Bylinina et al. (2015), Fillmore (1966), Mitchell (1986), Speas and Tenny (2003), and Smith (2009), as summarized by Barlew (2017) .....	30
3.2 Perspective set cardinality predictions by account .....	56
3.3 <i>De se</i> predictions by account .....	58
3.4 Syntactic domain for perspective shift predictions by account .....	59
3.5 Shift Together predictions by account .....	61
3.6 Cross-utterance anchoring predictions by account .....	63
3.7 Quantificational binding predictions by account .....	64
3.8 Predictions by account .....	65
3.9 Predictions by account compared to <i>come</i> data .....	74
4.1 Experiment 1 predictions .....	97
4.2 Experiment 1 results .....	102
4.3 Experiment 1 mixed effects regression analysis, fixed effects (N=1440).....	102

4.4	Experiment 1 mixed effects regression analysis including medium fillers, fixed effects (N=1728) .....	103
4.5	Experiment 2 results .....	106
4.6	Experiment 1 and 2 mean scores .....	106
4.7	T-tests between main conditions in Experiments 1 and 2 using z-transformed means .....	106
4.8	Experiment 3 predictions .....	109
4.9	Experiment 3a mean ratings by condition .....	111
4.10	Experiment 3a mixed effects regression analysis, fixed effects (N=960).....	112
4.11	Experiment 1 and Experiment 3a comparison mixed-effects regression analysis, fixed effects and interactions (N=2400).....	112
4.12	Experiment 3b results.....	113
4.13	Experiment 3b mixed-effects analysis, fixed effects (N=960).....	113
4.14	Experiment 3b mixed-effects analysis with medium fillers, fixed effects (N=960) .....	113
4.15	Experiment 1 and Experiment 3b comparison mixed-effects regression analysis, fixed effects and interactions (N=2400).....	114
4.16	Experiment 4 predictions .....	117
4.17	Comparison of mean ratings across experiments .....	118
4.18	Experiment 4 mixed effects regression analysis, fixed effects (N=960).....	118
4.19	Experiment 4 mixed effects regression analysis with medium fillers, fixed effects (N=960) .....	119
4.20	Experiment 1 and Experiment 4 comparison mixed-effects regression analysis, fixed effects and interactions (N=2400).....	119

4.21 Experiment 4 results . . . . .	120
4.22 Quant. Binding Task 1 results . . . . .	129
4.23 Quantificational Binding Task 2 results . . . . .	132
4.24 Quant. Binding Task 2 mixed effects regression analysis, fixed effects (N=576) . . . . .	132
4.25 Predictions by account compared to <i>tomorrow</i> data . . . . .	135
4.26 Predictions by account compared to <i>tomorrow</i> data . . . . .	139
4.27 Semantic ontology . . . . .	140
5.1 Predictions by account . . . . .	151
5.2 Predictions by account . . . . .	152
6.1 Results from Harris and Potts (2009)'s Experiment 1 . . . . .	171
6.2 Cross-linguistic data on the orientation of evidentials from Korotkova (2016) . . . . .	184
7.1 Experiment 5a results . . . . .	243
7.2 Experiment 5a RT mixed effects regression analysis, fixed effects (N=3630) . . . . .	243
7.3 Experiment 5b results . . . . .	247
7.4 Experiment 5b RT mixed effects regression analysis, fixed effects (N=2581) . . . . .	248
7.5 Experiment 5b acceptance mixed effects regression analysis, fixed effects (N=3103) . . . . .	248
8.1 Experiment 6 inner-annotator agreement scores by category . . . . .	277

8.2	PRSA Pragmatic Speaker predicted proportion of utterances by scene.....	279
8.3	Experiment 6 proportion of all responses by condition and type .....	280
8.4	Experiment 6 proportion of motion responses by condition and type .....	280
8.5	Experiment 6 proportion of all deictic responses by condition and type .....	281
8.6	Experiment 6 <i>come</i> response mixed effects logistic regression analysis with Speaker treated as baseline .....	281
8.7	Experiment 6 <i>come</i> response mixed effects logistic regression analysis with Listener and Speaker as fixed-effects .....	281
9.1	Predictions by account .....	297

## LIST OF FIGURES

<b>Figure</b>	<b>Page</b>
3.1 Disagreements in spatial perspective (left) and mental perspective (right) .....	31
3.2 Quantifier approach to indexical shift (Schlenker, 2003, tree from Sundaresan (2020a)) .....	46
4.1 Predicted availability of non-UT readings of <i>tomorrow</i> by analysis .....	96
4.2 Experiment 1 example stimulus .....	99
4.3 Example medium filler .....	100
4.4 Experiment 1 participant means by condition .....	101
4.5 Experiment 1 means by participant .....	103
4.6 Attested versus predicted availability of non-UT readings of <i>tomorrow</i> by analysis .....	104
4.7 Experiment 2 example stimulus .....	105
4.8 Attested versus predicted availability of non-UT readings of <i>tomorrow</i> by analysis .....	107
4.9 Experiment 3a example stimulus .....	110
4.10 Experiment 3b example stimulus .....	111
4.11 Attested versus predicted availability of non-UT readings of <i>tomorrow</i> by analysis .....	114

4.12 Experiment 4 example stimulus .....	117
4.13 Experiment 4 main condition ratings.....	118
4.14 Experiment 4 <i>tomorrow</i> participant means in order of difference between highest and lowest rating (minimum and maximum ratings shown in red) .....	120
4.15 Experiment 4 <i>the next day</i> participant means in order of difference between highest and lowest rating (minimum and maximum ratings shown in red) .....	121
4.16 Attested versus predicted availability of non-UT readings of <i>tomorrow</i> by analysis .....	122
4.17 Predicted covariation of <i>tomorrow</i> in quant. binding environments .....	128
4.18 Quant. Binding Task 1 example stimuli .....	129
4.19 Predicted versus attested covariation of <i>tomorrow</i> in quant. binding environments .....	130
4.20 Quantificational Binding Task 2 naturally occurring <i>tomorrow</i> items .....	131
4.21 Quantificational Binding Task 2 perspectival <i>tomorrow</i> items .....	131
4.22 By-item ratings for items in Quant. Bind Task 2 .....	133
4.24 Predicted versus attested covariation of <i>tomorrow</i> in quant. binding environments .....	134
4.23 By-participant means for <i>tomorrow</i> ratings in Expt. 3b and in Quant. Bind Task 2 .....	134
4.25 No Shift Together effects for perspective-anaphoric <i>tomorrow</i> .....	145
4.26 No Shift Together effects for perspective-anaphoric <i>tomorrow</i> and <i>come</i> .....	146
5.1 Experimental item from Duff (2018a) .....	161

6.1	Example appositive stimuli from Harris and Potts (2009) . . . . .	171
6.2	Example texts from Child et al. (2020) . . . . .	187
6.3	Harris (2012) Experiment 7 stimulus . . . . .	188
6.4	Example stimulus from Kaiser and Lee (2017b) . . . . .	192
6.5	Results reported in Kaiser and Lee (2017b) . . . . .	193
6.6	Poll responses for (201) (left) and (202) (right) . . . . .	194
6.7	Candidate discourse factors in perspective selection . . . . .	197
7.1	Vanilla Rational Speech Acts model (Bergen et al., 2012) . . . . .	204
7.2	World set, utterance set, and interpretation function for scalar implicature example . . . . .	206
7.3	Vanilla Literal Listener calculation for scalar implicature example . . . . .	206
7.4	Pragmatic Speaker calculation for scalar implicature example . . . . .	207
7.5	Vanilla RSA calculation for <i>Some are red</i> . . . . .	208
7.6	World set, utterance set, and interpretation function for non-uniform scalar implicature example . . . . .	209
7.7	Vanilla Literal Listener calculation for non-uniform scalar implicature example . . . . .	210
7.8	Pragmatic Speaker calculation for non-uniform scalar implicature example . . . . .	210
7.9	Vanilla RSA calculation for <i>Some are red</i> in non-uniform model . . . . .	210
7.10	Perspectival Rational Speech Acts model (preliminary) . . . . .	217
7.11	Lexical semantics for the candidate utterances . . . . .	220

7.12 World set (destination indicated by yellow box).....	221
7.13 Speaker = mover dataset .....	222
7.14 Posterior probabilities of world-perspective pairs when mover = speaker and perspective cost = 0 .....	223
7.15 Dataset for Example (198).....	225
7.16 Pragmatic Listener posterior marginal probabilities over worlds for (198), plain semantics for <i>go</i> .....	226
7.17 World set for Example (199) .....	227
7.18 Utterance set for Example (199).....	227
7.19 Pragmatic Listener posterior marginal probabilities over perspectives for Figure 7.17 world set, plain semantics for <i>go</i> .....	228
7.20 Perspectival Rational Speech Acts model (revised) .....	229
7.21 Effect of perspective cost on posterior perspective probabilities, subject = Thelma (cost increasing from 0 (left) to 1 (right)) .....	230
7.22 Pragmatic Listener marginal posterior probabilities of utterances, mover = Thelma (cost parameters = 0, 0.5, 1 from left to right) .....	231
7.23 Poll responses for (201) (left) and (202) (right) .....	232
7.24 Dataset for marginal posterior probability prediction illustration .....	234
7.25 Posterior probabilities of world-perspective pairs, plain semantics, mover = Thelma and perspective cost = 0 .....	235
7.26 Pragmatic Listener marginal posterior probabilities of utterances, mover = Thelma (cost parameters = 0, 0.5, 1 from left to right) .....	235
7.27 PRSA experiment characters.....	237
7.28 Comprehension stimulus: Perspectival Both condition .....	238

7.29	Comprehension stimulus: Plain Speaker condition .....	238
7.30	Comprehension spatial stimulus .....	239
7.31	PRSA (top) and Speaker Default predictions (bottom) .....	241
7.32	PRSA and Speaker Default predictions for Experiment 5a .....	242
7.33	Experiment 5a participant mean differences between Perspectival and Plain RTs.....	244
7.34	Experiment 5a participant mean acceptance rates by condition .....	244
7.35	Pragmatically odd filler example: definiteness violation.....	246
7.36	Experiment 5b participant mean differences between Perspectival and Plain RTs.....	249
7.37	Experiment 5b participant mean acceptance rates by condition .....	249
8.1	Vanilla RSA model of production (initial) .....	255
8.2	Alternative models of production .....	256
8.3	RSA referring expression example .....	258
8.4	Vanilla RSA Literal Listener calculation for Figure 8.3 dataset .....	258
8.5	Vanilla RSA Pragmatic Speaker calculation for Figure 8.3 dataset .....	259
8.6	Vanilla RSA model of production (final) .....	260
8.7	Perspectival Rational Speech Acts model of production .....	262
8.8	Dataset for informativity-driven competition illustration .....	265
8.9	PRSA Pragmatic Speaker marginal posterior probabilities given Figure 8.8 dataset (plain semantics for <i>go</i> , cost parameters = 0, 0.5, 1 from left to right) .....	266

8.10 Dataset for marginal posterior probability prediction illustration .....	269
8.11 Pragmatic Speaker marginal posterior probabilities of utterances, mover = Thelma (cost parameters = 0, 0.5, 1 from left to right) .....	270
8.12 Pragmatic Listener marginal posterior probabilities of worlds, mover = Thelma (cost parameters = 0, 0.5, 1 from left to right) .....	271
8.13 Second-level Pragmatic Speaker marginal posterior probabilities of utterances, mover = Thelma (cost parameters = 0, 0.5, 1 from left to right) .....	272
8.14 Experiment 6 example stimulus .....	276
8.15 Experiment 6 data coding categories .....	277
8.16 PRSA predictions (top), Speaker Default predictions (middle), and experimental results (bottom) .....	282
8.17 Egocentric Perspectival Rational Speech Acts model .....	286
8.18 Egocentric Pragmatic Listener marginal posterior probabilities over worlds, subject = Thelma and plain <i>go</i> semantics (cost parameters = 0, 0.5, 1 from left to right) .....	287
8.19 Lexical Uncertainty Pragmatic Speaker marginal posterior probabilities for utterances, subject = Thelma (cost parameters = 0, 0.5, 1 from left to right) .....	289
8.20 Pragmatic Listener marginal posterior probabilities over worlds when Both and None scene priors are higher, subject = Thelma and plain <i>go</i> semantics (cost parameters = 0, 0.5, 1 from left to right) .....	290
8.21 Pragmatic Speaker marginal posterior probabilities for utterances when Both and None scene priors are higher, subject = Thelma and plain <i>go</i> semantics (cost parameters = 0, 0.5, 1 from left to right) .....	290
8.22 Pragmatic Listener marginal posterior probabilities over worlds when Speaker and Listener scene priors are higher, subject = Thelma and plain <i>go</i> semantics (cost parameters = 0, 0.5, 1 from left to right) .....	290

8.23 Pragmatic Listener marginal posterior probabilities over worlds with a Dirichlet prior over perspectives favoring the speaker, subject = Thelma and plain <i>go</i> semantics (cost parameters = 0, 0.5, 1 from left to right) .....	292
8.24 Pragmatic Speaker marginal posterior probabilities over utterances with a Dirichlet prior over perspectives favoring the speaker, subject = Thelma and plain <i>go</i> semantics (cost parameters = 0, 0.5, 1 from left to right) .....	292
A.1 Ratings for <i>tomorrow</i> items in all experiments by geographic region .....	308
A.2 Ratings for <i>tomorrow</i> items in Experiment 4 by geographic region .....	308
A.3 Ratings for <i>tomorrow</i> items in all experiments by gender .....	309
A.4 Ratings for <i>tomorrow</i> items in all experiments by age .....	309
A.5 Ratings for <i>tomorrow</i> items in all experiments by age group .....	309
A.6 Experiment 2 ratings for <i>tomorrow</i> items by Big Five Openness scores .....	310
A.7 Experiment 2 ratings for <i>tomorrow</i> items by Author Recognition scores .....	310
A.8 Experiment 2 ratings for <i>tomorrow</i> items by AQ Imagination scores .....	311
A.9 Experiment 2 ratings for <i>tomorrow</i> items by AQ Communication scores .....	311
B.1 Perspectival function applied to a non-perspectival argument by PFA .....	316
B.2 Non-perspectival function applied to a perspectival argument by PFA .....	316
B.3 Perspectival function applied to a perspectival argument by PFA, perspective identification .....	316
B.4 Perspectival function applied to a perspectival argument by PFA, no perspective identification .....	316
B.5 Derivation with two perspectival expressions combining via PPM .....	317

B.6	Derivation with two perspectival expressions combining via PFA .....	323
B.7	Failed logophoric projection over two non-unified perspectival variables .....	323
B.8	Derivation showing lack of Shift Together effects between logophoric and anaphoric perspectival variables .....	324
B.9	Non-utterance time derivation for Version 3 .....	325
B.10	No Shift Together effects under Perspective Pronominalization .....	327
B.11	Derivation showing lack of Shift Together effects in logophoric environments .....	329
E.1	Second-level Pragmatic Speaker marginal posterior probabilities of utterances, mover = Thelma and perspective cost at second level (cost parameters = 0, 0.5, 1 from left to right) .....	339
E.2	Second-level Pragmatic Speaker marginal posterior probabilities of utterances, mover = Thelma and perspective cost at first level (cost parameters = 0, 0.5, 1 from left to right) .....	340
E.3	Second-level Pragmatic Speaker marginal posterior probabilities of utterances, mover = Thelma and perspective cost at both levels (cost parameters = 0, 0.5, 1 from left to right) .....	340
E.4	Example non-uniform priors over worlds .....	341
E.5	Literal Listener calculation with uniform priors over worlds, plain semantics for <i>go</i> .....	342
E.6	Literal Listener calculation with non-uniform priors over worlds, plain semantics for <i>go</i> .....	343
E.7	Pragmatic Speaker calculation with uniform priors over worlds for sentences involving Thelma .....	344
E.8	Pragmatic Speaker calculation with non-uniform priors over worlds for sentences involving Thelma .....	345

## LIST OF ABBREVIATIONS

1	first person
3	third person
4	fourth person
ana	anaphoric
ANAPH	anaphor
COMP	complementizer
idx	indexical
log	logophoric
LOG	logophor
NOM	nominative
PERF	perfective
PRES	present
PST	past
Q	question
f	feminine
s	singular

## CHAPTER 1

### INTRODUCTION

Every person has their own perspective: on events, on themselves, on other people. This is a consequence of being located in time and space, as people are: from our particular standpoint, we are closer to some things than others, we have better access to some things than others. We may see some things more clearly than others.

This extends to the mental landscape as well. None of us are omniscient, which is another way of saying that we see some things more clearly than others. We have better access to some facts than others: for instance, we have privileged access to our own perceptions and beliefs, while we are dependent on other people to tell us about theirs. It is the limitations of our viewpoints that make them unique.

Our perspectives influence the way we communicate: both in what we talk about, and how we talk about it. When a tree is nearer to the side of me on which, as Merriam-Webster charmingly puts it, “the heart is mostly located,” I say that it is on the left. If you are standing face-to-face with me, you will likely describe it as on the right. When I eat a piece of key lime pie, I say that it is tasty. If your tastes differ from mine, you say that it is not. If a cat walks towards me, I say that it is coming; if we are standing across the yard from each other, you may say that it is going.

Or you may not. You might instead put yourself in my shoes, mentally orient yourself in my place. In that case, you can describe things as I might: you can say that the tree is on my left, the pie is tasty, the cat is coming. This is called **perspective shift**: the switch from one person’s perspective to that of someone else.

Perspective is encoded in natural language in many ways. Despite their various forms, perspectival phenomena share two traits: they are subjective, conveying information that

depends on an individual's limited point-of-view, and they are shiftable, allowing some choice in whose perspective they convey. In this dissertation, I explore the semantics and pragmatics of expressions with these characteristics.

This dissertation is divided into two parts, each of which focuses on a distinct set of questions about perspective. Part I explores the similarities and differences among different classes of perspectival expressions: expressions whose meaning depends in part on the unique experiences, location, sensations, or beliefs of an individual. In Part I, I explore different ways that the semantics of perspectival expressions can encode the perspective holder. I illustrate the theoretical landscape of perspectival expressions: the various mechanisms that natural language provides for encoding perspective, and their predicted consequences.

In Chapter 2, I lay out some of the general properties of perspectival expressions through the lens of a canonical perspectival expression: the perspectival motion verb *come*. Using *come* as an example expression, I provide an intuitive sense of how perspectival expressions behave in a variety of environments. Although the set of individuals allowed to serve as perspective holders varies across perspectival expressions, the dependence on a perspective holder is encoded in the semantics of all perspectival expressions.

In Chapter 3 I explore the different methods that natural language provides for encoding this dependency on a perspective holder in the semantics of perspectival expressions. I propose that there are four broad families of approaches: lexical stipulation, indexicality, logophoric binding, and perspective-anaphoricity. I illustrate the differing predictions made by these accounts in order to develop a set of diagnostics for perspectival expressions. I conclude with a demonstration of how the set of diagnostics can be applied, using American English *come* as a case study. I argue for a perspective-anaphoric analysis of American English *come*, but discuss how the attested cross-linguistic variation in the behavior of *come* may motivate different treatment of perspectival motion verbs in other languages.

In Chapter 4, I present a second, more complicated case study. I focus on an expression not previously considered perspectival: the temporal indexical *tomorrow* in American English. Through a series of experiments, I explore the acceptability of non-utterance time

interpretations of *tomorrow* in a variety of environments. I show that some American English speakers accept non-utterance time readings outside of the contexts where indexicals are expected to shift. I also show that *tomorrow* can covary with quantifiers in some, but not all, quantificational binding contexts. I argue that for the American English speakers who accept these readings, *tomorrow* is perspective-sensitive. Having narrowed the space of possible analyses for *tomorrow* to a logophoric or perspective-anaphoric account, I present some tentative data favoring the anaphoric approach, and develop a perspective-anaphoric semantics for *tomorrow*. I also uncover a surprising degree of interspeaker variability in judgments about *tomorrow* whose source remains to be identified.

In Chapter 5, I conclude Part I with a discussion of how perspectival expressions fit into the broader landscape of context sensitivity. The behavior of different classes of context-sensitive expressions overlaps in all but a few environments. I argue that care must be taken to test the particular environments that do discern between classes of context sensitivity, since, as highlighted by the case study on *tomorrow*, applying these diagnostics can reveal unexpected perspective sensitivity.

Part II addresses a separate set of questions about the semantics and pragmatics of perspective that focus on how listeners and speakers handle perspectival expressions. A central characteristic of perspectival expressions is their shiftability. This optionality about which perspective to use poses both an interpretative challenge for the listener and a selection problem for the speaker. In the second half of the dissertation, I explore how speakers select perspectives and how listeners identify the perspective that is being used.

In Chapter 6, I lay out the issues of perspective selection and perspective identification. I discuss previous approaches to perspective identification and selection: simple heuristic approaches, two-stage systems, and reasoning-based systems. I review some of the factors that may influence the prominence of perspectives, focusing in particular on two pressures identified by Harris (2012): the cognitive cost of perspective shift and bias towards the speaker's perspective. I argue that although conversation participants may rely on simple heuristics some of the time, there must also be a more sophisticated perspectival reasoning system available in certain contexts.

In Chapter 7, I propose a reasoning-based model of perspective identification. I develop and implement a model of the perspective comprehension process based in the Rational Speech Acts framework. I view this process as a Bayesian joint inference task: listeners reason simultaneously about the speaker’s intended meaning and their adopted perspective using a mental model of the speaker’s production process. I present a series of simulations showing that the model captures known perspectival interpretation behavior, as well as generating a novel prediction. I test this prediction in two comprehension studies. The results provide qualified support for a key assumption of the proposed model: that listeners consider multiple perspectives simultaneously when processing perspectival expressions.

In Chapter 8, I turn to the question of perspective selection. I argue that Rational Speech Acts models of production, while less common than comprehension models, capture several key principles of cooperative speaker behavior. I present a Rational Speech Acts model of perspective production in which speakers reason jointly over pairs of perspectives and utterances in order to maximize their chances of communicative success, using a mental model of how the listener will interpret their utterance. I present a production study that tests the predictions of the proposed model, but find that the results do not support the model. These findings suggests an asymmetry between speaker and listener behavior that is troubling for the Rational Speech Acts framework. I also explore a number of theoretically-motivated modifications to the Perspectival Rational Speech Acts system, but conclude that none of them resolve the observed asymmetry between speaker and listener behavior.

This dissertation builds on a large body of work on perspectival expressions in many languages. Despite the richness of the literature, however, there has been comparatively little work that probes across classes of perspectival expressions. My sense is that this is in part due to the incredible diversity of the phenomena that have been labeled perspectival. I hope that by laying out the various semantic mechanisms that natural language provides for encoding perspective, and the consequences of the different semantics, I have helped to illuminate the broader landscape of perspectival expressions. I see the set of perspectival diagnostics that I have proposed as a roadmap for navigating this landscape, and I hope

that Part I provides some useful tools for future work exploring the interactions among perspectival expressions.

Similarly, although there is a growing body of literature investigating the impact of various discourse factors on the perspective prominence (as discussed in Chapter 6), relatively little is known about whether the impact of these factors is consistent across classes of perspectival expressions. I hope that this dissertation can contribute to this area of inquiry in two ways: first, in Part I, by proposing a semantic taxonomy of perspectival expressions, and second, in Part II, by proposing computational models of perspective comprehension and production capable of generating gradient predictions. This gradient output is useful for exploring the discourse factors that affect the prominence of particular perspectives without ruling them out completely.

Perspective, or point-of-view, is an aspect of meaning that is pervasive across categories of linguistic expressions. Understanding how perspective is encoded semantically is critical to understanding context sensitivity more broadly. Understanding how speakers and listeners select and identify perspectives is central to understanding how shared and privileged information is tracked and used in discourse. This dissertation begins to answer two sets of questions about perspectival expressions, but many more remain.

## **PART I: ENCODING PERSPECTIVES**

## CHAPTER 2

### PERSPECTIVAL MOTION VERBS

Perspectival motion verbs describe motion relative to the location of a perspective holder. In American English, the set of perspectival motion verbs includes *come* and *go*.<sup>1</sup> In many varieties of English, it also includes the verbs *bring* and *take*; however, the perspectival component of these verbs seems to be fading for many American English speakers, so I will focus on *come* and *go* in this chapter.<sup>2</sup>

The verbs *come* and *go* and their cross-linguistic equivalents are sometimes called **basic motion verbs** because they are hypothesized to exist in all languages and to be some of the earliest acquired verbs of motion (Miller and Johnson-Laird, 1976), though these claims are controversial (Wilkins and Hill, 1995).

Another reason that these verbs are referred to as ‘basic’ is because they do not convey any manner-of-motion information. These verbs contribute two implications: (1) a motion event and (2) a perspectival component. The difference between *come* and *go* lies in the perspectival component: *come* describes motion towards the location of a perspective holder, and *go* describes motion that is not towards a perspective holder.

In (1a) and (1b), for instance, the motion is towards the location of the speaker, a common perspective holder, and *come* is preferred to *go*. In (1c), on the other hand, the motion is away from the speaker’s current location, and *go* is felicitous.

1. *Flavia de Luce is at her home in the village of Bishop’s Lacey. Winston Churchill lives at 10 Downing Street.*

---

<sup>1</sup>Whether the lexical semantics of *go* are actually perspectival is a point of contention that I will return to; however, because it is contentious, it should be included in the category for discussion purposes at least.

<sup>2</sup>See Clark and Garnica (1974) for documentation of this.

- (a) Flavia: Winston Churchill is coming to Bishop's Lacey.
- (b) Flavia: #Winston Churchill is going to Bishop's Lacey.
- (c) Flavia: I am going to Downing Street.

This gives an intuitive sense of the licensing conditions of *come* and *go*, but opens up several important questions. First, there is the question of how the two components of meaning (the motion implication and the perspectival implication) are encoded in the grammar. Up to this point, I have used the term *implication* to avoid committing to level of linguistic representation. Is the infelicity in (1b) because the sentence is false, due to presupposition failure, or merely dispreferred relative to (1a)?

Second, I have claimed that *come* is licensed in (1a) because the speaker is a perspective holder. What does this mean? What is a perspective holder, and how do we know that the speaker (but not, apparently, Winston Churchill) can be one?

Lastly, what is the spatial relationship between the perspective holder's location and the destination of motion? In (1a), Churchill is described as in motion to the village; is he necessarily in motion towards Flavia?

I address the first set of questions in Section 2.1 and the second two in Section 2.2, before turning to some general properties of perspectival motion verbs in Section 2.3.

## 2.1 The lexical semantics of perspectival motion verbs

My account of the lexical semantics of perspectival motion verbs follows mostly from the comprehensive analysis of *come* proposed by Barlew (2017). [In versus by. Check throughout. –Daniel] However, I translate his dynamic semantics into a simple event semantics here for illustrative purposes. Like Barlew (2017), I posit that the meaning of perspectival motion verbs consists of two parts: a motion implication and a perspectival implication.

### 2.1.1 Motion implication

I will assume the simple event semantics representation in (2) for the motion implication of *go* and *come*.<sup>3</sup>

2. (a) **Lexical semantics for *go* (preliminary):**

$$[[\text{go}]]^{C,g} = \lambda x. \lambda e. \text{MOVE}(e) \wedge \text{DEST}(e, x)$$

- (b) **Lexical semantics for *come* (preliminary):**

$$[[\text{come}]]^{C,g} = \lambda x. \lambda e. \text{MOVE}(e) \wedge \text{DEST}(e, x)$$

The destination of motion may be implicit, as in (3).

3. (a) I am going now.

- (b) I am coming now.

In this case, I assume that the destination argument ends up existentially bound, as in (4).

$$4. [[\text{I am going}]]^{C,g} = \exists x. \exists e. \text{MOVE}(e) \wedge \text{DEST}(e, x) \wedge \text{AGENT}(e, C_{\text{speaker}})$$

### 2.1.2 Perspectival implication

Both *come* and *go* share the same lexical semantics for their motion implication. However, their perspectival content varies. What is this perspectival content?

In his foundational work on perspectival motion verbs, Fillmore (1966) distinguished the meanings of *come* and *go* by exploring the licensing conditions for each verb. He proposed a set of rules that determine which is used in a given situation.

Winston (1988) proposed a reformulation of Fillmore (1966)'s licensing conditions around the centrality of the perspective holder, writing:

Fillmore identified certain space-time points [...] as particularly relevant to the use of *come* and *go*. I agree that these space-time points are the crucial ones, but I believe that they are crucial because they represent plausible perspectives for the speaker. They are the space-time locations of people or objects with which a speaker or narrator is either likely, or at least able, to associate himself

---

<sup>3</sup>In the denotations given throughout this dissertation, I assume that the thematic role of the subject is introduced in an intermediate projection as in Kratzer (1996), but nothing in my analysis rests on this assumption.

in thought or imagination; as such they constitute possible viewpoints for the speaker or narrator. (Winston, 1988, p. 27)

Her chief insight was that Fillmore's conditions of use can be distilled to the principle that *come* describes motion towards a perspective holder, while *go* does not.

How is this requirement encoded? Since Fillmore (1966), the perspectival implication has been treated as a presupposition. In addition to Fillmore (1966)'s early data, Oshima (2006b) presents the example in (5) as evidence that when the speaker or listener<sup>4</sup> is the perspective holder, the implication that they are located at the destination of motion projects out of attitude contexts, as expected for presuppositional content.

5. *John is speaking to Linda. Bob has been in New York for a month, and John believes that (Bob believes that) Bob has been in New York for a month.*

John: Bob believes that Chris came to San Jose two weeks ago.

*Presupposes:* John or Linda is in San Jose at utterance time or event time.

*Does not presuppose:* Bob believes that John or Linda is in San Jose at utterance time or event time. (Oshima, 2006b)

I will discuss various strategies for encoding the semantics of perspective more precisely in Chapter 3, where I discuss the issue in the broader context of perspectival phenomena. For this chapter, I will adopt Barlew (2017)'s approach and treat the perspectival implication as a presupposition arising from a free perspective variable. As shown in (6), I propose that there is a perspective variable that is free in the context of the sentence, which is resolved anaphorically.

6. **Lexical semantics for *come*:**

$[[\text{come}]]^{C,g} = \lambda x. \lambda e. \text{MOVE}(e) \wedge \text{DEST}(e, x) \wedge x = \text{LOC}(p)$ , where  $p$  represents the perspective holder.

---

<sup>4</sup>Throughout this dissertation, I will use the term *listener* rather than *addressee*. This is for consistency with the literature in the Rational Speech Acts framework which is important in Part II. However, I do not mean to assume anything about the modality of the speech act.

If there is no discourse-given individual in the Common Ground to whom the variable can be resolved, then a presupposition violation results, as for a pronoun whose referent cannot be determined in the discourse context. However, as I will show in Chapter 3, there are alternative ways of encoding the perspectival implication; I adopt this analysis as a convenience for the remainder of this chapter to illustrate the licensing conditions of *come*.

One consequence of adopting this strategy is that I will assume that there is a unique perspective holder for each instance of a perspectival motion verb, even when the perspective holder cannot be identified from the context. For instance, **come** in (7) below describes motion towards 221B Baker Street, where both the speaker and listener are located. I will assume that there is a unique but underspecified perspective holder: either the speaker or the listener.

7. *Holmes and Watson are both at 221B Baker Street.*

Holmes: The King of Bohemia is coming to call upon us.

By contrast, (8) is unambiguous. Since there is only one possible perspective holder at the destination of motion, the unique perspective holder can be identified.

8. *Miss Marple, alone in her cottage, is talking on the phone with her nephew.*

Marple: I'm so pleased that you are coming here for a visit, dear.

I will use the term **anchoring** to describe the relationship between the perspectival motion verb and the perspective holder: the verb *come* is **anchored to the speaker** in (9) because the presence of the speaker at the destination of motion is what licenses the use of *come*.

9. Please come meet me here.

A perspectival motion verb is also sometimes referred to as **oriented** to a particular perspective; (9) might also be described a **speaker-oriented** use of *come*.

The representation of the perspectival component of *go* has proved contentious. Some previous proposals represent it as part of the lexical semantics of *go*, as the inverse to *come* (Fillmore, 1966; Oshima, 2006a).

## 10. Perspectival lexical semantics for *go*:

$[[\text{go}]]^{C,g} = \lambda x. \lambda e. \text{MOVE}(e) \wedge \text{DEST}(e, x) \wedge x \neq \text{LOC}(p)$ , where  $p$  represents the perspective holder.

Other authors propose that the anti-perspectival implication arises pragmatically (Wilkins and Hill, 1995; Sudo, 2018). Because *go* is in competition with *come*, the listener who hears *go* draws an inference that *come* was not licensed. Since the motion implication of the two verbs is identical, the only reason that *come* could fail to be licensed in a situation where *go* is licensed is when the perspectival implication of *come* is not satisfied: the perspective holder is not at the destination of motion. This gives rise to the inference that when *go* is used, the perspective holder is not at the destination of motion. Under this analysis, the lexical semantics for *go* only contain the motion event implication.

## 11. Plain lexical semantics for *go*:

$[[\text{go}]]^{C,g} = \lambda x. \lambda e. \text{MOVE}(e) \wedge \text{DEST}(e, x)$

For illustrative purposes in this chapter, I will adopt a perspectival semantics for *go*. This allows me to speak of a unique perspective holder for *go* as well as *come*. However, I am not committed to this, and I show in Chapter 7 that the desired properties of *go* follow from either analysis.

## 2.2 Components of the licensing of *come*

The perspectival licensing conditions for perspectival motion verbs varies across languages. There are three components to consider: the set of possible perspective holders, the time at which the anchoring relation holds, and the allowed anchoring relations.

### 2.2.1 Perspective holders

Cross-linguistically, the speaker's perspective seems to be preferred, both because of the number of languages in which only speaker-anchoring is allowed, and because it is the speaker is the most common perspective holder in languages that allow other perspective holders as well. Table 2.1 shows the cross-linguistic anchoring facts, compiled from Gath-

ercole (1987), Nakazawa (2007), and Barlew (2017). As can be seen, all attested languages allow the speaker to be the perspective holder of *come*.

**Table 2.1.** Typology of perspective holders for *come* from Gathercole (1987), Nakazawa (2007), and Barlew (2017) (English and SLQ Zapotec data from my work)

Language	UT speaker	ET speaker	Listener	Home	Accompaniment	Attitude holder
Shibe	✓	X	X	?	?	?
Spanish	✓	X	X	✓	✓	?
Abaza	✓	X	%	?	?	?
Botin	✓	?	X	?	?	?
Jacaltec	✓	?	X	?	?	?
Thai	✓	?	X	?	✓	?
Mandarin	✓	%	X	X	%	?
Chindali	✓	✓	X	?	?	X
Catalan	✓	✓	?	?	?	?
Chagga	✓	✓	?	?	?	?
Italian	✓	✓	?	?	?	?
Palauan	✓	✓	?	?	?	?
Tangut	✓	✓	?	?	?	?
Nepali	✓	✓	?	%	✓	?
Turkish	✓	✓	?	✓	✓	?
San Lucas	✓	✓	%	X	X	X
Quiaviní Zapotec						
Korean	✓	✓	%	?	?	✓
Japanese	✓	✓	%	✓	%	✓
Chagga	✓	✓	✓	?	?	?
Texmelucan	✓	✓	✓	?	?	✓
Zapotec						
French	✓	✓	✓	?	?	✓
Croatian	✓	✓	✓	?	✓	?
Tamil	✓	✓	✓	?	✓	?
German	✓	✓	✓	✓	✓	?
English	✓	✓	✓	✓	✓	✓

✓ = acceptance; X = rejection; % = variable acceptance; ? = missing data

Because of the frequency of speaker-anchoring, anchoring to a perspective other than that of the speaker is sometimes referred to as **perspective shift**. However, I will use this term to refer to any change in perspective, whether or not it is a transition from the speaker's perspective.

While the cross-linguistic data in Figure 2.1 is incomplete, it seems to present an implicational hierarchy: if a language allows any non-speaker perspective holder, it allows the listener. There is no attested language in which the attitude holder, but not the listener, is a valid perspective holder.

American English allows a relatively large set of perspective holders, making it a useful language in which to explore the semantics of perspectival motion verbs. It allows both listener and attitude holder anchoring of *come*, as shown in (12) and (13).

#### 12. Listener-anchoring:

*Hercules Poirot is talking on the telephone to Colonel Hastings.*

Poirot: You say there is a corpse in the library? I will come at once!

#### 13. Attitude holder-anchoring:

Miss Lydgate showed no signs of being ashamed of Miss Vane. On the contrary, she greeted her warmly, begged her to come and see her on Sunday morning, [...] and commended her for keeping up a scholarly standard of English, even in mystery fiction.  
(Sayers, 1935)

The set of perspective holders for *come* in English is shown in (14).

#### 14. Set of perspective holders for English *come*:

{ speaker, listener, attitude holder }

This is not, however, an exhaustive set of perspective holders. Any sufficiently prominent perspective-holding<sup>5</sup> individual can serve as the perspectival anchor for *come* in English. In narrative text, for instance, it is common for the protagonist of the story to serve as the perspectival center, as in (15), where *come* describes motion towards Peter Wimsey, the protagonist of the novel.

15. Sir Impey Biggs walked up and down for some half-hour, smoking. Then he **came** across with determination, brutally switched on a reading-lamp right into Peter's face,

---

<sup>5</sup>I mean here to exclude non-sentient individuals, since they cannot self-ascribe locations.

sat down opposite to him, and said: “Now, Wimsey, I want to know all you know.”  
(Sayers, 1926)

There are also special kinds of narrative text in which the identity of the perspective holder established by convention. In Free Indirect Discourse, a discourse style that mixes the narrator’s perspective with that of the main character or **protagonist**, the protagonist serves as the default perspective holder for perspectival motion verbs (rather than the narrator, who represents the speaker and therefore might ordinarily be the most likely perspective holder) (Banfield, 1982).

16. Mrs. Dalloway said she would buy the flowers herself. For Lucy had her work cut out for her. The doors would be taken off their hinges; Rumpelmayer’s men were **coming**. And then, thought Clarissa Dalloway, what a morning— fresh as if issued to children on a beach. (Woolf, 1925)

In (16), Clarissa Dalloway, the protagonist of the Free Indirect Discourse passage, serves as the anchor for *come*.

For this reason, (14) is merely meant to represent the set of individuals whose perspectives tend to be prominent enough to serve as perspectival anchors for *come*.

### 2.2.2 Anchoring times

In many languages, including American English, the speaker’s perspective at either utterance time or at the time of the motion event can anchor *come*. In (17a), the speaker’s perspective is evaluated at utterance time: although Harriet was not in her room at the time of the motion event, *come* can be used to describe motion to her room since she is currently located there. This is an **utterance-time anchoring** of *come*.

17. *Harriet Vane is surveying the floor of her room, which is covered in shards of glass.*
  - (a) **Utterance time:** Harriet: While I was in the library, someone came in here and smashed my chess set.
  - (b) **Event time:** Harriet: When I was in the library, Miss Lydgate came to retrieve a book.

In (17b), Harriet is no longer located in the library, but she can describe motion to the library using *come* if it occurred during the time that she was in the library. This is an **event-time anchoring** of *come*. For present-tense event descriptions, of course, utterance time and event time generally coincide.

(18) shows the set of perspective holders for *come* in American English, updated with time indexing.

#### 18. Set of perspective holders for English *come*:

{ speaker @ UT, speaker @ ET, listener @ UT, listener @ ET, attitude holder }

While all languages allow speaker-anchoring of *come*, not all languages allow event time anchoring, as shown in Table 2.1. Again, there seems to be an implicational hierarchy: no language allows event time licensing without allowing utterance time licensing.

#### 2.2.3 Anchoring relations

So far I have spoken of the perspectival implication of *come* as requiring the perspective holder to be located at the destination of motion (at either utterance time or event time). I have not spelled out very precisely what is required to satisfy this condition. For instance, this could mean that the goal of the motion must be the location of the perspective holder, or merely that the perspective holder is located in roughly the same location at the destination, or, even more loosely, that the direction of motion is such that the mover ends up closer to the perspective holder.

Barlew (2015) explores the spatial licensing conditions for *come* more precisely. He shows that *come* can be used to describe motion in the direction of the perspective holder, even if the intended destination is not the perspective holder's location. In (19a), Tom's destination is not Salt Lake City, but because his motion was in the direction of Salt Lake City, where the perspective holder is located, *come* is licensed.

#### 19. *The speaker and listener are in Salt Lake City.*

- (a) Tom came through Salt Lake City today, so we had lunch while he was here.  
(Barlew, 2015)

- (b) Tom came through Denver on his way here. (Barlew, 2015)

Moreover, Barlew (2015)'s example (19b) shows that the anchoring relation is not specified by the prepositional modifier of *come*, since in (19b), the destination specified by the prepositional phrase is not the location of the perspective holder. Barlew ultimately posits that what is required is a motion path with a subpath that ends at the location of the perspective holder, but begins elsewhere. This subpath may or may not be part of a larger motion path with a different destination.

The spatial extent of the location of the perspective holder for anchoring purposes also seems to be contextually determined: for instance, in the context shown in (20), the relevance of the nationality of the perspective holder licenses *come* when describing motion to any US city, even ones that the perspective holder has not visited.

20. *Carolyn and Paty are talking in San Lucas Quiaviní, Mexico. Carolyn is from Massachusetts and has never been to Texas. She is speaking with Paty, who is from San Lucas and has never been to the US.*

Carolyn: Felipe came to Texas last winter.

This context sensitivity seems to be true in other languages as well; this example originally came from an experience that I had in Oaxaca, Mexico, and was originally elicited in San Lucas Quiaviní Zapotec, where the same judgments were given.

The exact spatial configuration that licenses *come* is not my focus. For the most part, I will treat these licensing conditions loosely, and continue to speak as if they are satisfied when the perspective holder is located at the destination of motion.

In addition to the utterance time or event time presence of the perspective holder at the destination of motion, there are two other anchoring relations that many languages, including English, allow. First, *come* may be used to describe motion to the homebase of the perspective holder, even if that individual is not located there at the time of motion (21).

21. **Homebase anchoring:** *Mary Morstan, Sherlock Holmes and Doctor Watson are in a tea parlor. Mary shows them a pearl and tells them how it came into her possession.*

Mary: It was delivered to my flat in an unmarked envelope

Watson: Did you see who delivered it?

Mary: No, I wasn't at home when it came.

Second, English has a deictic accompaniment construction where *come* can describe motion alongside the perspective holder, as shown in (22).

22. **Accompaniment anchoring:** *Sherlock Holmes is about to set off to Dartmoor to investigate a mysterious hound.*

Holmes: Come with me to Dartmoor, Watson!

Multiple anchoring relations are sometimes used with the same perspective holder in close proximity. For instance, in (23), the speaker is the perspective holder for both instances of *come*. The first is licensed by her utterance time presence in L.A., while the second is licensed by the fact that London is her homebase.

23. Daisy Ridley: I love to come to L.A. to visit, and then I like to come to rainy old London because it's home.<sup>6</sup>

The set of anchoring relations for *come* in English is shown in (24).

24. **Set of anchoring relations for English *come*:**

$$\left\{ \begin{array}{l} \text{perspective holder @ destination,} \\ \text{perspective holder accompaniment,} \\ \text{destination = perspective holder's homebase} \end{array} \right\}$$

Unlike the set of perspective holders in (14), the set of anchoring relations shown in (24) is exhaustive, as far as I know.

#### 2.2.4 Summary

As we have seen, there is considerable cross-linguistic variation in the set of perspective holders, the anchoring relations, and the evaluation time of anchoring for the cross-linguistic

---

<sup>6</sup>Siegel, Tatiana. 2015. "Next Gen 2015: How Unknown Daisy Ridley's "Weird Feeling" Helped Her Land 'Star Wars' Role." *The Hollywood Reporter*. [www.hollywoodreporter.com/news/next-gen-2015-how-unknown-836637](http://www.hollywoodreporter.com/news/next-gen-2015-how-unknown-836637).

equivalents of *come*. In addition to the fact that not all languages allow all perspective holders and anchoring relations, some languages other than English lexicalize these differences.

In some languages, there are separate perspectival motion verbs for different perspective holders. For instance, Palauan is reported by Josephs (1975) to make a three way distinction between *me* ‘come towards speaker’, *eko* ‘come towards listener’, and *mo* ‘go’, as shown in (25) and (26).

25. Ng səbəch-em əl me ər a bli-k ər a klukuk?  
it ability-your to come to house-my on tomorrow

‘Can you come to my house tomorrow?’ (Josephs, 1975, glossing taken from (Nakazawa, 2007))

26. Chochoi. Ng səbəch-ek əl eko ər a bli-k ər a klukuk.  
yes it ability-my to come to house-your on tomorrow

‘Yes. I can come to your house tomorrow.’ (Josephs, 1975, glossing taken from (Nakazawa, 2007))

In other languages, it is the anchoring relation that is lexically specified. In San Lucas Quiaviní Zapotec, for instance, there are two verbs that correspond to American English *come*: *ried*, which describes motion towards the location of the perspective holder, and *ria*<sup>7</sup> which describes motion towards a homebase.

27. b-ya’=a’ liaz=a  
PERF-go.home=1s house=1s

‘I came home.’ (Munro et al., 1999)

28. zhyúahn zh:ih=ihzy b-yáa’ll=a’  
? days=only PERF-come=1s

‘I came for only a few days.’<sup>8</sup>

---

<sup>7</sup>This verb is homophonous with *ria* ‘go’ in the habitual 3rd person singular form, the form in which Zapotec verbs are conventionally cited, but not in other forms.

<sup>8</sup>This example is from my own fieldwork.

In (27), the verb that describes motion towards the homebase is used, while in (28), the verb that describes motion towards the location of a perspective holder is used.

In this section, I have provided only a brief survey of the cross-linguistic variation in the semantics of *come* and *go*, but there is a wealth of research on the topic. A comprehensive discussion of the issues around basic motion verbs as a linguistic universal can be found in Wilkins and Hill (1995), while information about cross-linguistic variation in the allowed perspective holders and anchoring relations can be found in Gathercole (1987); Nakazawa (2007, 2009).

### 2.3 Properties of perspectival motion verbs

In the previous section, I have sketched the licensing conditions of *come* and *go*, including the licit perspective holders and anchoring relations in American English. In this section I discuss some general properties of perspectival motion verbs.

#### 2.3.1 Non-perspectival uses

In addition to the perspectival uses of *come* and *go* that I have been discussing, it is important to note that each verb also has some non-perspectival uses. There is a large set of expressions with *come* or *go* as one of their components that no longer have any perspectival implication.

These non-perspectival uses are relatively easy to distinguish when they do not describe actual motion. For example, *came out* is often used to describe the metaphorical emergence or appearance of its subject, as in (29), rather than physical motion. In this example, it is used to describe the appearance of the lists, without any anchoring relation to the perspective holder.

29. Mary, said all her friends, was marked for a First; only the dim, inscrutable dons had not been surprised when the lists came out with Harriet's name in the First Class and Mary's in the Second. (Sayers, 1935)

It is harder to distinguish non-perspectival from perspectival uses when they describe actual motion. In (30), for example, there is a physical motion event. However, this is not a

perspectival use of *come*; *came into sight* can be used regardless of whose perspective is adopted, without the anchoring relation to the perspective holder usually required by *come*.

30. The procession came into sight, beneath the archway; a small crocodile-walk of elderly people, dressed with the incongruous brilliance of a more sumptuous era, and moving with the slovenly dignity characteristic of university functions in England. (Sayers, 1935)

One diagnostic for identifying non-perspectival uses is to replace *come* with *go* (or vice versa) and try to devise a context in which the modified expression is felicitous; if no such context exists, then the motion verb is probably being used as part of a lexicalized expression rather than perspectivally. This diagnostic correctly identifies both (29) and (30) as non-perspectival uses. Example (31) is only felicitous under the interpretation that the lists are actually being mailed out (i.e., there is physical motion), unlike (29), where the lists may simply be posted. Example (32) is infelicitous.

31. ?? Mary, said all her friends, was marked for a First; only the dim, inscrutable dons had not been surprised when the lists went out with Harriet's name in the First Class and Mary's in the Second. (Sayers, 1935)

32. # The procession went into sight, beneath the archway; a small crocodile-walk of elderly people, dressed with the incongruous brilliance of a more sumptuous era, and moving with the slovenly dignity characteristic of university functions in England. (Sayers, 1935)

In addition to the plethora of lexicalized non-perspectival constructions in which *come* and *go* appear, *go* is also used as a near-future auxiliary in English, as in (33).

33. Inspector Alleyne: Inspector Fox is going to come stay with us for the weekend.

Cross-linguistically, it is extremely common for perspectival motion verbs to take part in lexicalized non-perspectival constructions and to grammaticalize into other functional categories, such as passive markers, intensifiers, and aspect markers (Cook, 1994; Bilmes, 1995; Hassler, 1999; Hooper, 2002; Carlson, 2014; Dragomirescu and Nicolae, 2014).

Many languages also contain complex motion verb constructions involving perspectival motion verbs, which may or may not have lost their perspectival flavor (Cardinaletti and Giusti, 2001; de Vos, 2005; Mauri and Sansó, 2014). English contains two such constructions. One remains perspectival and requires actual motion, as shown in (34).

34. **‘Go get’ construction (Pullum, 1990):** Flavia will come get her specimen kit before investigating.

35. **Adversative pseudo-coordination:** *Nancy, Bess, and George are getting ready for a dance. George accidentally knocks over her tea, and Bess complains to Nancy:*  
Bess: George went and spilled her tea all over my gown.

The other is not perspectival; it does not require real motion, but conveys that the event was unfortunate according to the speaker, as shown in (35), where George does not go anywhere.

For the remainder of this work I will focus specifically on the perspectival uses of *come* and *go*. However, it is useful to understand the kinds of non-perspectival uses these verbs have in order to be able to distinguish the examples of interest.

### 2.3.2 Embedded anchoring

As I discussed in Section 2.2.1, attitude holders are one common kind of perspective holder. This arises most frequently when a perspectival motion verb is embedded under an attitude verb; in these contexts, the perspective of the subject of the attitude verb is usually prominent. In (36), for instance, neither the narrator, nor Jenny (her listener) are located at Vera’s house, which is the destination of the motion event described by *come*. The only possible anchor for *come* is Vera, the subject of the attitude verb *think*.

36. When I have Jenny sitting in front of coffee, pecking at watery scrambled eggs, I ask how Vera ended up in the hospital. “Who found her?” “The nurse told me Mamma called EMS at four o’clock this morning. Said she wasn’t feeling good and **she thought they better come and get her.**” (Davies, 2008)

This instance of *come* is in fact embedded multiple times over. First, *come* is embedded in the thought report, of which Vera is the subject. This thought report is embedded in Vera's speech report, of which she is also the subject. This speech report is actually communicated to Jenny via the nurse at the hospital, who is the subject of *told*. The nurse's speech report is in turn reported by Jenny using a direct speech report indicated by quotation marks.

Attitude holders can serve as perspective holders in contexts where they are not introduced as subjects of attitude verbs. In the synopsis of the novel *Small Island* in (37), *come* describes motion to England. Since Gilbert, but not Hortense, is located in England at that point in the discourse, he must be the perspective holder. Although he is not the subject of an attitude verb, he is an attitude holder: it is the content of his dream that is being described.

37. Gilbert Joseph was one of the several thousand Jamaican men who joined the RAF to fight against Hitler. Returning to England after the war he finds himself treated very differently now that he is no longer in a blue uniform. It is desperation that makes him remember a wartime friendship with Queenie and knock at her door. **Hortense shared Gilbert's dream of leaving Jamaica and coming to England to start a better life-** that's why she married him. But when she at last joins her husband, she is shocked by London's shabbiness and horrified at the way the English live. (Baroni et al., 2009)

Thus, although attitude holder perspectives are often introduced by attitude verbs, they can also be introduced by other means.

In the cases of attitude holder anchoring that I have presented so far, the perspective holder for *come* is always the nearest attitude holder. Is this always the case? I do not think so. For one thing, it is possible for *come* to be anchored to the speaker even when the motion verb is embedded under an attitude predicate. I judge *come* to be felicitous in (38), and given that Saint-George is not located at Shrewsbury, *come* must be anchored to the perspective of the speaker.

38. *Harriet Vane is talking to the Dean of Shrewsbury College about a possible visit by Lord Peter Wimsey. Her information comes via Lord Peter's nephew, Lord Saint-George Wimsey, who is a student at Christ Church College*

Harriet: Lord Saint-George says that his uncle is planning to come to Shrewsbury for the Opening of the Library.

In addition, *come* seems to allow readings where it is anchored to an attitude holder that is not the subject of the nearest attitude verb. I judge *come* to be felicitous in (39), although it is anchored to the Harriet's perspective rather than that of Mr. Padgett, the nearest attitude holder.

39. *Lord Peter Wimsey is in his flat in London, talking to his manservant Bunter about a letter he has received from Harriet Vane, who is at Shrewsbury College.*

Lord Peter: Harriet says that Mr. Padgett knew that Miss de Vine was coming to see her.

Attitude holder anchorings of *come* do not, therefore, seem to be obligatorily determined by the scope of the nearest attitude verb. This conclusion is further supported by the fact that the subject of an attitude verb can also serve as an anchor for perspectival motion verbs that occur outside of its scope. In (40), for instance, the first speaker starts an attitude report, introducing the perspective of her friend, the subject of the attitude verb, but then interrupts the attitude report with a long appositive. In this appositive, she uses *come* anchored to the perspective of her friend. The attitude report is picked back up again and finished in her second utterance, after prompting by the second speaker.

40. Ms. Watson: It never crossed my mind. I had been traveling the way everybody else did, by air and - or car, and I was frustrated with that. **And a friend of mine a while back said - when I was going to come and visit her in New Mexico,** I was angry because I couldn't get there fast, and couldn't get there easily. And finally, I thought maybe I could drive, but gee, that's several days.

Conan: A long way, yeah.

Ms. Watson: She said, why don't you take the train? (Conan, 2008, found in (Davies, 2008))

This suggests that the use of an attitude verb makes the perspective of its subject available not just for perspectival expressions within its scope, but also for perspectival expressions within a certain discourse proximity.<sup>9</sup>

Another kind of attitude holder anchoring arises in Free Indirect Discourse, a special split-perspective discourse style in which some classes of context-sensitive expressions are interpreted relative to the perspective of the protagonist, while others are interpreted relative to the narrator's perspective (Banfield, 1982).<sup>10</sup>

Perspectival expressions are generally claimed to be obligatorily protagonist-oriented in Free Indirect Discourse (Eckardt, 2014). However, the anchoring of *come* is not strictly determined in Free Indirect Discourse environments. The protagonist of Free Indirect Discourse is frequently the perspective holder for *come*, as in (41), where the Free Indirect Discourse protagonist is Sally Seton.

41. All these years the Dalloways had never been once. Time after time they had asked them. **Clarissa (for it was Clarissa of course) would not come.** For, said Sally, Clarissa was at heart a snob— one had to admit it, a snob. And it was that that was between them, she was convinced. Clarissa thought she had married beneath her, her husband being— she was proud of it— a miner's son. (Woolf, 1925)

However, it is also possible for other characters to be the perspective holder for *come*, as in (42). In this example, Lady Bruton is the Free Indirect Discourse protagonist, but since she is the subject of *come*, she cannot be its perspective holder. Instead, her listener, Richard Dalloway, is the perspective holder: the destination of motion is his homebase.

---

<sup>9</sup> A point about *come* that I will return to in Chapter 3.

<sup>10</sup> Note that the term *protagonist* is used in the Free Indirect Discourse literature to refer to the non-narrator perspective holder, since there are two perspective holders active in Free Indirect Discourse contexts. However, the protagonist of a Free Indirect Discourse passage does not have to be the protagonist of the narrative in a more general sense: in (41), the Free Indirect Discourse protagonist is Sally Seton, but the main character or protagonist of the novel is Clarissa Dalloway.

42. Richard turned to Lady Bruton, with his hat in his hand, and said, “We shall see you at our party to-night?” whereupon Lady Bruton resumed the magnificence which letter-writing had shattered. **She might come; or she might not come.** Clarissa had wonderful energy. Parties terrified Lady Bruton. But then, she was getting old. So she intimated, standing at her doorway. (Woolf, 1925)

These examples indicate that the perspective holder of *come* is not strictly determined in Free Indirect Discourse. However, I do think that Free Indirect Discourse imposes some restrictions on who the perspective holder of *come* can be. My sense is that the perspective holder for *come* must be someone whose perspective is accessible by the Free Indirect Discourse protagonist, for example, in (42), the Free Indirect Discourse protagonist’s listener. This follows from Maier (2015) claim that Free Indirect Discourse passages must generally be verbatim faithful to the thoughts of the protagonist.

Taken together, these facts suggest that the anchoring of *come*, while not completely unconstrained, is not strictly determined by the syntactic domain of attitude reports or by pragmatic environments like Free Indirect Discourse.<sup>11</sup>

### 2.3.3 Obligatory *de se* interpretation

So far, I have spoken of the most common anchoring relation for perspectival motion verbs as requiring the presence of the perspective holder at the destination of motion. However, the truth conditions are subtly more complex: what matters is not the perspective holder’s actual location, but their self-ascribed location.

---

<sup>11</sup>Of course, under many accounts, Free Indirect Discourse environments are semantically and syntactically similar to attitude reports. For instance, Maier (2018) proposes that there is an implicit speech report scoping over Free Indirect Discourse content. Context parameter approaches to Free Indirect Discourse also treat Free Indirect Discourse similarly to attitude reports in the sense that protagonist-oriented interpretations of indexicals arise via similar mechanisms as indexical shift in attitude reports (Schlenker, 2004; Sharvit, 2008; Eckardt, 2014).

Barlew (2017) argues that perspectival motion verbs are obligatorily interpreted *de se*: the perspective holder must self-ascribe the content that is being expressed with the perspectival expression.<sup>12</sup>

43. **Attitude terminology** (Lewis, 1979a):

- (a) *De se*: an attitude that an agent holds about themselves in a first-person way
- (b) *De re*: an attitude that an agent holds, but that is expressed in words that they may not agree with

In order for *come* to be used felicitously, the perspective holder must believe themselves to be located at the destination of motion; if they are mistaken about their location, it is their believed location, not their actual one, that matters. This is shown in (44).

44. *Mark and his friend Sarah are driving down the West Coast. Mark falls asleep shortly after they leave Seattle. Sarah pulls over in Tacoma, wakes him up, and convinces him that they're already in San Francisco. He gets out of the car to take pictures of the Narrows bridge, thinking that it is the Golden Gate bridge. Sarah calls their friend Jeremy to tell him about the prank.*

- (a) ✓*De se*: Sarah: Mark wishes that you had come to San Francisco too—he thinks you would have been really astounded by how beautiful the water is!
- (b) ?*De re*: Sarah: Mark wishes that you had come to Tacoma too—he thinks you would have been really astounded by how beautiful the water is!

In (44), Sarah is able to use *come* even though she is well aware that none of them are actually in San Francisco, since Mark believes that he is in San Francisco. According to Mark's perspective, he is located in San Francisco, so he serves as a perspective holder at the destination of motion to license *come*.

---

<sup>12</sup>Here I am speaking somewhat sloppily, as *de se*-ness is a property of propositional attitudes, not of individual expressions, and perspectival expressions like *come* on their own do not comprise propositions.

The extent to which (44b) is felicitous is due to the fact that Sarah is a potential perspective holder who self-locates in Tacoma; if Jeremy repeats what Sarah has told him, he can only use Mark’s perspective, and the *de se* reading is the only one that is available (45).

45. *Mark and his friend Sarah are driving down the West Coast. Mark falls asleep shortly after they leave Seattle. Sarah pulls over in Tacoma, wakes him up, and convinces him that they’re already in San Francisco. He gets out of the car to take pictures of the Narrows bridge, thinking that it is the Golden Gate bridge. Jeremy, who is in Florida, is telling the story to his partner.*

- (a) ✓ *De se*: Jeremy: Mark wished that I had come to San Francisco too—he thought I would have been really astounded by how beautiful the water is!
- (b) # *De re*: Jeremy: Mark wished that I had come to Tacoma too—he thought I would have been really astounded by how beautiful the water is!

Thus, *come* imposes a *de se* requirement on its perspectival anchor: the perspective holder must believe themselves to be located at the destination of motion.

## 2.4 Summary

In this chapter, I have surveyed some of the key properties of perspectival motion verbs. First, perspectival motion verbs have two components of meaning: a motion description and a perspectival anchoring implication. There is much cross-linguistic variation in the perspectival component of motion verbs: the set of perspective holders, the allowed anchoring relations, and the evaluation times for the anchoring relations vary between languages. Despite this, some cross-linguistic generalizations can be drawn: the self-location of the speaker at the destination of motion at utterance time is the most common licensing condition for the cross-linguistic equivalents of *come*.

Next, I surveyed the licensing conditions for *come* in American English, which, compared to other languages, are fairly broad. All cross-linguistically attested perspective holders, evaluation times, and anchoring relations are allowed in American English.

I also sketched some of the general semantic properties of perspectival motion verbs in American English. I discussed the existence of many non-perspectival constructions involving perspectival motion verbs; the anchoring behavior of embedded perspectival motion verbs; and their obligatory *de se* interpretation.

Throughout this discussion, I have tried to stay narrowly focused on perspectival motion verbs, though some discussion of other perspectival phenomena was necessary. In the next chapter, I turn to a broader discussion of perspective in language.

## CHAPTER 3

### THE SEMANTICS OF PERSPECTIVE

In this chapter, I will lay out some of the key characteristics of perspectival phenomena and key issues in providing a semantic and pragmatic theory of perspective. Many different definitions of perspective have been given, which I will discuss briefly in Section 3.1. I decompose the analysis of perspective into two key questions: (1) how the perspectival component is represented in the semantics of the expression, and (2) how the perspective holder is determined.

In this chapter, and the rest of Part I, I explore the first question, returning to the second in Part II. In Section 3.2, drawing on previous analyses of perspectival expressions, I outline four methods of encoding the perspective holder in the semantics of perspectival expressions. In Section 3.3, I develop a set of diagnostics to sort perspectival expressions into four semantic families. I conclude in Section 3.4 with a case study of how the set of diagnostics can be applied, using perspectival motion verbs in American English as an example.

#### 3.1 Perspectival phenomena

Many heterogeneous phenomena in natural language have been described as perspectival. Barlew (2017) summarizes proposed perspectival phenomena according to five different definitions, reproduced in Table 3.1 below. As this table shows, classifications of perspectival expressions vary widely, and few expressions are recognized as perspectival by all five.<sup>1</sup>

---

<sup>1</sup>One of the motivations for using perspectival motion verbs as an example phenomenon throughout this dissertation is that it is one of the most agreed-upon perspectival expressions.

**Table 3.1.** Candidate perspectival expressions proposed by Bylinina et al. (2015), Fillmore (1966), Mitchell (1986), Speas and Tenny (2003), and Smith (2009), as summarized by Barlew (2017)

Expression type	Example	B. et al.	Fillmore	Mitchell	S & T	Smith
deictic motion verb	<i>come</i>	✓	✓	✓	✓	
spatial descriptions	<i>right</i>	✓	✓	✓	✓	
modals	<i>might</i>	✓	✓		✓	✓
logophoric pronouns	Ewe <i>yè</i>	✓			✓	✓
long distance	Mandarin <i>ziji</i>	✓			✓	✓
reflexives						
evidentials	Japanese <i>rashii</i>	✓			✓	✓
discourse adverbs	<i>frankly</i>		✓		✓	✓
tense			✓		✓	✓
proximity pred.	<i>nearby</i>	✓		✓		
directed paths	<i>across the street</i>	✓		✓		
socio-cultural	<i>foreign</i>	✓		✓		
predicates						
pred. of	<i>tasty</i>	✓		✓		
personal taste						
cognitive pred.	<i>obvious</i>	✓		✓		
person			✓		✓	
spatial indexicals	<i>here</i>		✓		✓	
perception/	<i>loud</i>			✓	✓	
psych pred.						
epithets	<i>the jerk</i>				✓	✓
expressives	<i>damn</i>				✓	✓
FID					✓	✓
mood					✓	✓
ability pred.	<i>easy</i>			✓		
control PRO				✓		
pred. ascriptions	<i>seem</i>			✓		
switch reference					✓	
speech pred.	<i>say</i>					✓
attitude pred.	<i>think</i>					✓
perception verbs	<i>see</i>					✓
possessive pronouns	<i>your</i>					✓

### 3.1.1 Proposed diagnostics

Various diagnostics have been proposed for categorizing perspectival phenomena. In this section, I review some of the previously proposed approaches.

First, many authors make a distinction between spatial perspective and mental perspective (Mitchell, 1986). Spatial perspective phenomena depend on the visual point-of-view and/or spatial location of their perspective holder. For example, the meaning of spatial prepositions like *in front of* depends on the visual perspective on a scene that is taken, as well as a reference point (Figure 3.1, left).

Mental perspective has to do with the attitudes and beliefs on the perspective holder. For instance, predicates of personal taste, such as *tasty*, depend on the preferences of a perspective holder. While two individuals with identical spatial perspectives will agree on whether or not a dessert is in front of them, they will not necessarily agree about whether the cake is tasty (Figure 3.1, right).

**Figure 3.1.** Disagreements in spatial perspective (left) and mental perspective (right)



However, the distinction between spatial and mental perspective is not straightforward. For instance, the perspectival motion verb *come* obviously involves spatial perspective, since its perspectival component references the spatial location of the perspective holder. Yet as we saw in Chapter 2, it also involves mental perspective, because it is obligatorily interpreted *de se*: it is not the perspective holder's actual spatial location that matters, but their perceived spatial location.

A second line of work uses structural diagnostics to classify perspectival phenomena on the basis of particular syntactic and semantic analyses. For instance, Speas and Tenny (2003) propose that there is a point-of-view projection in the left periphery with a **seat of knowledge** argument (roughly corresponding to the perspective holder). Because they assume that there can only be one seat of knowledge node, certain diagnostics for perspectival phenomena fall out of this account, for instance, that all perspectival expressions in the same sentence will have the same perspective holder.

However, adopting an analysis in order to generate a definition of the phenomena seems like putting the cart before the horse. I prefer to start from some account-independent generalizations about perspective. One recent proposal for a theory-neutral definition of perspectival expressions comes from Bylinina et al. (2015), who posit that perspectival expressions have three fundamental characteristics: shiftability, default orientation to the speaker, and Shift Together effects.

### 3.1.1.1 Diagnostic 1: shiftability

One of the most notable properties of perspectival phenomena is their ability to be interpreted relative to more than one individual. Mitchell (1986) observed that certain expressions can either receive a speaker-oriented interpretation, or, when embedded under an attitude predicate such as *think*, an attitude holder-oriented interpretation. For instance, in (46), *around the corner* may either be interpreted relative to Miss Marple's perspective, or relative to that of the speaker.

46. Miss Marple thinks that that the restaurant is around the corner.

Bylinina et al. (2015) generalize Mitchell (1986)'s observation to a shiftability diagnostic: all perspectival expressions must be shiftable in certain grammatical constructions. These constructions include, but are not limited to attitude predicates. Bylinina et al. (2015) do not require ambiguity in these environments; some perspectival expressions might be obligatorily shifted, while others might be compatible with multiple interpretations.

This diagnostic leads to a fairly broad notion of perspective that will encompass indexical shift (at least in languages where indexical shift is non-obligatory), and, as Barlew (2017) points out, any expression that is subject to a *de re/de dicto* ambiguity.

### 3.1.1.2 Diagnostic 2: default speaker interpretations

Bylinina et al. (2015) propose that perspectival expressions are, by default, anchored to the speaker's perspective. They do not spell out exactly what this means, however. In one sense, a speaker default may simply be the observation that the majority of uses of perspectival expressions are anchored to the speaker's perspective. If this is what is meant by a speaker default, it is not really about the semantics of the expressions themselves, but rather about general patterns of their usage.

On the other hand, this diagnostic could be spelled out semantically, by assuming that all perspectival expressions are semantically fixed to the speaker's perspective unless some operation intervenes. That is to say, in (47) below, *come* is semantically fixed to the speaker's perspective, but something induces perspective shift, leading *come* to be anchored to the listener's perspective.

47. *Miss Marple is on the phone to vicar's wife, Griselda Clement.*

Marple: I'll come there at once, dear.

The assumption that perspectival expressions are frequently anchored to the speaker's perspective is fairly common. However, this does not seem like a very informative diagnostic, given that context-sensitive linguistic context is generally interpreted relative to the speaker, or at least, relative to the utterance context, which generally amounts to the same thing.

### 3.1.1.3 Diagnostic 3: Shift Together effects

The third diagnostic that Bylinina et al. (2015) propose is the existence of Shift Together constraints for perspectival expressions. Shift Together effects, initially discussed in the context of indexical shift (Anand and Nevins, 2004), arise when all context-sensitive expressions in the same domain receive a uniform interpretation. In the context of perspective, Shift Together means that all perspectival expressions are anchored to the same perspective.

For instance, if predicates of personal taste are subject to Shift Together effects, then both *spicy* and *tasty* in (48) should be interpreted relative to the same point-of-view. Either Hastings thinks the soup is both tasty and spicy, or Poirot does; if there were a Shift Together constraint, (48) could not be used in a situation where Poirot thinks the soup is spicy, but not tasty, and Hastings thinks the soup is tasty, but not spicy.

48. Poirot: Hastings thinks that he could eat this tasty, spicy soup every night of his life.

It is necessary to distinguish Shift Together effects within perspectival phenomena from Shift Together effects between perspectival phenomena. If all classes of perspectival phenomena spring from a unified underlying semantics, then we might find Shift Together effects between different classes of perspectival expressions. However, it is also possible that there are various ways of encoding perspective in language, and that different classes of expressions do not interact with each other.

The obligatoriness of Shift Together effects does not meet my criteria of a theory-neutral diagnostic, since such effects are predicted only by accounts of perspective in which the perspective holder is governed syntactically, such as Speas and Tenny (2003). Shift Together effects are essentially a kind of locality constraint on expressions that are in the scope of an operator.

In what follows, I would like to avoid assuming that there is such an operator for all perspectival expressions, at least for the moment. Therefore, I will not use Shift Together effects to define perspectival expression, though I will discuss them again later in this chapter as a diagnostic for certain analyses of perspectival expressions.

### 3.1.2 A definition of perspectival expressions

It is very difficult to devise diagnostics for perspective that capture the range of proposed phenomena without presupposing a particular syntactic or semantic account. Although aspects of Bylinina et al. (2015)'s proposed definition are compelling, their proposed third property is problematic for this reason.

I will instead adopt a very broad definition in the interest of presupposing as little as possible about the implementation of perspective and encompassing as many phenomena as possible. I propose that a perspectival expression is one whose meaning depends on the location, perception, or beliefs of a contextually prominent individual.

49. **Definition of perspectival expression (preliminary):** a perspectival expression is an expression whose meaning depends on the location, perception, or beliefs of a prominent individual.

By a prominent individual, I do not mean to presuppose that the individual is contextually supplied rather than syntactically represented. I merely mean that the perspective holder cannot be just any individual; they must, at least, be discourse-given.

The reliance on an individual is therefore a critical component of perspective sensitivity. However, some expressions meet this criterion but do not meet the full definition because they do not involve reference to the individual's location, perception, or beliefs. For instance, person pronouns refer relative to an individual, but they do not have the subjective quality of perspectival expressions: they do not depend in any way on the individual's location, perception, or beliefs.

I have purposefully stated a definition that is broad enough to encompass both spatial and mental perspectival expressions. By including expressions that are sensitive to the location of a perspective holder, I have included spatial prepositions, which refer relative to the location and/or orientation of an individual. For instance, *left* and *right* are interpreted relative to the orientation of a perspective holder, as shown in (50).

50. *The speaker and listener are facing each other. There is a red lamp on one side of the listener and a green lamp on the other.*

The lamp on the right (of you/of me) is green/red.

By including perception and beliefs, I have also included predicates of personal taste, like *tasty*. In (51), whether or not the hot chocolate is tasty depends on the personal preferences of each of the characters: Poirot thinks it is, but Hastings thinks it is not.

51. Poirot: This chocolate is very tasty!

Hastings: No it isn't. Drinking hot chocolate for breakfast is a revolting habit!

Because of this subjective quality, some analyses of predicates of personal taste propose that their meaning depends on a **judge** (Lasersohn, 2005; Stephenson, 2007). However, I will avoid using this term in favor of the more general term *perspective holder*, since I do not wish to argue for these analyses over other approaches to predicates of personal taste, like the judge-free approach proposed by Pearson (2013).

Although (51) made it clear whose preferences were being described by *tasty*, it is not always easy to identify the perspective holder of a perspectival expression.

52. Hastings said that drinking delicious hot chocolate for breakfast is a revolting habit.

While predicates of personal taste embedded under attitude predicates are often anchored to the perspective of the subject of the attitude verb, as with *revolting* in (52), they can also be speaker-oriented, as with *delicious* in the same example.<sup>2</sup>

Because my definition is very broad, however, we might ask what kinds of context-sensitive expressions it does exclude. Pure temporal anaphora like *the next minute* will not meet the definition unless their referent is mediated through an individual. In (53), the interpretation of *the next minute* does not depend in any way on the evaluation of an individual: it refers relative to the time mentioned in the preceding sentence.

53. The clock struck nine. The next minute, a bolt of lightning lit up the sky.

According to the definition in (49), an expression that refers relative to a prominent time is not perspectival, but one that refers to the temporal location of a prominent individual is.

However, the definition of perspectival expressions that I have given is still very broad. At the moment, the definition encompasses spatial and temporal indexicals like *here*. As (54) shows, the referent of *here* is dependent on the speaker's location.

---

<sup>2</sup>Incidentally, (52) is evidence against taking Shift Together constraints as a necessary condition for perspectival expressions, since predicates of personal taste are one of Bylinina et al. (2015)'s example perspectival phenomena and yet, Shift Together constraints do not hold.

54. (a) *Poirot is talking to Colonel Hastings in his London flat.*

I live here.

(b) *Miss Marple is talking to Raymond West in her cottage.*

I live here.

Since the identity of the speaker is context-dependent, indexicals meet the definition of perspectival expressions proposed in (49). However, the meaning of indexicals is more fixed than the expressions that I am interested in exploring: I want to distinguish perspectival expressions who allow some optionality in the choice of perspective holder, like *come*. Therefore, I propose the modified definition of perspectival expressions in (55).

55. **Definition of perspectival expression (revised):** a perspectival expression is an expression whose meaning depends on the location, perception, or beliefs of a prominent individual chosen by the speaker.

The meaning of a perspectival expression must depend on the subjective assessment of some contextually prominent individual, and that individual is selected by the speaker. This revision serves to exclude classes of context-sensitive expressions like indexicals, whose meaning relies on the location of an individual but whose reference point is fixed, without excluding spatial prepositions like *right*.<sup>3</sup> The intent is to adopt Bylinina et al. (2015)'s shiftability diagnostic without adopting their speaker default diagnostic.

### 3.1.3 Perspective shift environments

The linguistic phenomena that have been described as perspectival comprise a diverse group. While many of the candidate perspective-sensitive phenomena shown in Table 3.1 meet the broad definition that I have proposed for perspectival expressions, some are not really expressions, but environments. I will use the term **perspective shift environment** to categorize linguistic environments that affect the interpretation of perspectival expressions, but do not qualify as perspectival expressions in their own right.

---

<sup>3</sup>Indexicals could be excluded by appealing to their non-subjectivity, but this would exclude spatial prepositions as well, since they depend only on spatial perspective and not mental perspective.

Certain narrative styles affect the interpretation of perspectival expressions in their scope. The best known of these environments is **Free Indirect Discourse**, a narrative style that is used to simultaneously narrate events and report a protagonist's thoughts or comments on them. In Free Indirect Discourse, some context-sensitive expressions refer relative to the narrator's perspective, while others refer relative to the narrator's perspective (Banfield, 1982; Doron, 1991; Schlenker, 2004; Sharvit, 2008; Eckardt, 2014; Maier, 2015). For instance, tense and the person features on pronouns are interpreted relative to the narrator's perspective, while indexicals, adverbials, and the gender features of pronouns are interpreted relative to the protagonist's point-of-view.

In Free Indirect Discourse, perspectival expressions are interpreted relative to the protagonist's perspective (Eckardt, 2014). For instance, in the following passage from *Mrs. Dalloway*, *come* must be interpreted relative to the perspective of the protagonist, Clarissa Dalloway, not the narrator.<sup>4</sup>

56. Mrs. Dalloway said she would buy the flowers herself. For Lucy had her work cut out for her. The doors would be taken off their hinges; Rumpelmayer's men were coming. And then, thought Clarissa Dalloway, what a morning— fresh as if issued to children on a beach. (Woolf, 1925)

Another kind of perspective shift environment is **Protagonist Projection** (Stokke, 2013; Abrusán, 2018, 2019), also known as **Viewpoint Shift** (Hinterwimmer, 2017b).<sup>5</sup> Protagonist Projection is used to report the pre-verbal perceptions, feelings, and beliefs of the protagonist, rather than their internal thoughts or speech (Abrusán, 2019). In (57), for example, the protagonist's sensations as he is falling asleep are described, but there is no conscious thought of his that is being reported.

---

<sup>4</sup>As I discussed in Chapter 2, however, *come* is not always anchored to the protagonist's perspective (exemplified in (42)). However, I am not aware of any instances of Free Indirect Discourse where *come* is anchored to the narrator's perspective.

<sup>5</sup>The extent to which Protagonist Projection and Viewpoint Shift differ is a topic of ongoing debate (Abrusán, 2019).

57. A great brush swept smooth across his mind, sweeping across it moving branches, children's voices, the shuffle of feet, and people passing, and humming traffic, rising and falling traffic. Down he sank into the plumes and feathers of sleep, sank, sank and was muffled over. (Woolf, 1925, cited in Abrusán (2019))

Both Free Indirect Discourse and Protagonist Projection mix the perspectives of the narrator and the protagonist, but in different ways. Free Indirect Discourse reports conscious thought or speech of the protagonist,<sup>6</sup> while Protagonist Project reports sensations or perceptions that do not have to be consciously realized or expressed. Because Protagonist Projection does not convey already formed speech or thought, it is not required to be faithful to the protagonist's style, as Free Indirect Discourse is. The expressions that shift in each environment are also different: in Protagonist Projection, indexicals are interpreted relative to the narrator's perspective Abrusán (2019).

I do not classify either Free Indirect Discourse or Protagonist Projection as perspectival expressions, because both environments affect more than just perspectival expressions. They impact the interpretation of non-perspectival pronouns and adverbials as well as perspectival ones. Thus, I think it is more useful to think of these discourse styles as **perspective shift environments**: environments that affect the interpretation of perspectival expressions within their scope.

Another environment that affects the interpretation of perspectival expressions is quotation. In quotation, the words of the quotee are repeated exactly as they were originally used. Quoted perspectival expressions are interpreted relative to the perspective that the quotee originally used; they cannot be interpreted relative to the quoter's perspective. For example, (58) would not be felicitous if the speaker, rather than Sally Ride, had been at NASA. When it is embedded in a quote, *come* cannot be interpreted as referring relative to the perspective of the quoter.

58. *A professor at Swarthmore College is giving a speech.*

Sally Ride said, "I did not come to NASA to make history," because she wanted to be

---

<sup>6</sup>Though see Fludernik (1995) for a dissenting view.

seen first and foremost as an astronaut. Nevertheless, as the first American woman in space, she became an icon and a role model for many women.

Again, although quotation can be used to convey the point-of-view of the quotee, quotes should not be considered perspectival expressions, since quotation affects the interpretation of non-perspectival expressions as well. Rather, I classify quotation as a perspective shift environment, since it affects how perspectival expressions are interpreted.

Lastly, we might also consider attitude verbs as a kind of perspective shift environment. Although attitude, speech, and perception verbs are included in Table 3.1 as candidate perspectival expressions, they do not meet my criteria, because their meaning does not depend on a contextually prominent individual: it depends on their subject.

59. Poirot: Hastings thinks that John Cavendish murdered Emily Inglethorp.

There is no context in which (59) conveys that Poirot, rather than Hastings, believes that Cavendish is the murderer, because *think* must be interpreted as describing its subject's beliefs. Thus, attitude verbs (as well as speech and perception verbs) are not perspectival expressions. They are perspective shift environments, however, because they generally make the perspective of their subjects available to perspectival expressions in their scope.

I will not posit an explanation of the mechanism by which these verbs affect the interpretation of perspectival expressions at this time. There are multiple possible ways for perspective shift environments to affect the interpretation of perspectival expressions: they may introduce new perspective holders into the discourse context; they may affect the prominence of discourse-given perspectives; or they may bind perspectival variables. The availability of some or all of these strategies depends in part on the analysis posited for the perspectival expressions whose interpretation they affect. I leave further exploration of the semantics of perspective shift environments for future work in order to focus on the semantics of perspectival expressions.

### 3.1.4 Summary

Many kinds of linguistic phenomena have been described as perspectival. However, it is far from clear that these phenomena form a natural class, or even have something in common beyond the label perspectival. I have proposed that perspectival expressions are expressions whose meanings convey the perceptions, beliefs, experiences, or other subjective evaluation of a contextually prominent individual selected by the speaker.

Having established a somewhat broad working definition of perspectival expressions, I turn to various ways of analyzing their perspectival content. In particular, I focus on the issue of perspective encoding: how the perspective holder is incorporated into the semantics of perspectival expression. In the next section, I discuss five different approaches to this question.

Because perspectival motion verbs are a class of expressions considered perspectival under most proposed diagnostics, I will use them as a case study in the following section and throughout this dissertation. Although I focus on this particular class of expressions, the issues raised and the analyses discussed are relevant for any phenomenon that appears to be sensitive to perspective.

## 3.2 Possible semantics for perspectival expressions

Because perspectival expressions refer relative to a perspective holder, a critical question for their analysis is how the perspective holder is encoded in their semantics. In order to help generalize over perspectival expressions, it is useful to think of the semantics of perspectival expressions as consisting of two components: a component that varies based on the kind of expression, and a perspectival component that encodes whose perspective is being used. For instance, in Chapter 2, I decomposed the meaning of *come* into a motion event description and a perspectival anchoring condition. For *come*, this perspectival implication requires that a perspective holder is located at the destination of the motion event.

60. *Poirot is in London.*

- (a) Poirot: Come to London at once, Hastings!

- (b) Poirot: # Come to Suffolk at once, Hastings!

We can similarly decompose the meaning of other perspectival expressions into these two components. For instance, Potts (2005) proposed separating the contribution of epithets like *goose* into two components: a descriptive component (picking out an individual) and a conventional implicature about the emotional stance of a perspective holder (in (61), that the speaker holds a dismissive attitude towards the individual picked out by *goose*).

61. Flavia: Ophelia is so self-absorbed. The vain goose didn't even notice when I stole her lipstick.

Although I will not adopt the two-dimensional treatment of conversational implicature proposed in Potts (2005), I will adopt the distinction between the expressive-specific content and perspectival content of such expressions, since it is useful in distinguishing the component of interest.

Similarly, predicates of personal taste have been analyzed as an expression-specific component and a perspectival component that conveys whose opinion is being expressed (Laser-sohn, 2005; Stephenson, 2007; Moltmann, 2010; Keshet, 2005; Pearson, 2013).

62. **Predicate of personal taste (attitude holder-oriented)**: Hastings: Although I find the stuff disgusting, Poirot said that he would like another cup of the delicious hot cocoa.

One of the most important questions about perspectival expressions, therefore, is the nature of this perspectival component. How is perspective encoded in the grammar of these expressions?

In this section, I discuss several possible answers to this question. As Section 3.1 showed, the landscape of perspective is rich, and a correspondingly diverse set of analyses for the semantics of perspectival expressions have been proposed. I divide the proposed accounts into five categories. First, there are accounts that encode the perspective holder directly in the lexical semantics of the expressions. Second, there are indexical approaches that rely on the context parameter to encode perspective. Third, there are logophoric accounts that

govern perspective via semantic composition and binding mechanisms. Fourth, there are anaphoric accounts that treat the perspective like a pronoun. Last, there are proposals that treat perspectival phenomena as a special case of quotation.

My goal in this section is not to adjudicate between these accounts but to illustrate each of them. Although the approaches have been proposed for various perspectival phenomena, for explanatory purposes, I will use perspectival motion verbs as an example phenomenon when sketching out these accounts.

### 3.2.1 A lexical approach

One approach to the perspectival component is to directly encode it in the truth conditions of the expression. For instance, in the case of *come*, we could propose that its semantics stipulate that the destination of motion is the speaker’s location:

#### 63. Lexically stipulated semantics for *come*:

$$[[\text{come}]]^{C,g} = \lambda x. \exists e. \text{MOVE}(e) \wedge \text{DEST}(e, x) \wedge x = \text{LOC}(\text{speaker})$$

This approach will under-generate for *come* in American English, since there are multiple licit perspective holders, including the speaker, listener, and subjects of attitude verbs. It might work, however, for Shibe *come*, which Nakazawa (2007) reports can only be anchored to the utterance time perspective of the speaker.

Goddard (1997) proposes a different kind of lexical approach to perspective. He proposes that *come* that does not rigidly encode the perspective holder, but merely requires that there is someone at the destination of motion who can think, of the subject X, “X is in the same place as me.”<sup>7</sup>

#### 64. Lexically stipulated semantics for *come*:

$$[[\text{come}]]^{C,g} = \lambda x. \exists e. \exists z. \text{MOVE}(e) \wedge \text{DEST}(e, x) \wedge x = \text{LOC}(z), \text{ where } z \text{ is a sentient individual}$$

---

<sup>7</sup>Interestingly, this does capture the *de se* property of *come* discussed in Section 3.3.2; as far as I am aware, it is the earliest work to discuss this property of *come*.

However, this approach overgeneralizes for most perspectival expressions since it allows any sentient individual to serve as the perspective holder. For instance, this approach predicts that *come* should be able to describe movement towards any location where a person is located. As (65) shows, however, the fact that at least one person besides the speaker was located in New York and aware that the speaker was also there does not license *come*: the desk clerk's perspective is not available even though a desk clerk must have been present (as well as a thief).

65. *Captain Hastings and Poirot are in London.*

Hastings: # Last year I came to New York. While I was checking into my hotel, someone ran off with my bags.

This kind of account essentially reduces to an animacy (or perhaps sentience) restriction: a perspectival expression with these semantics is predicted to be licit in any situation where a sentient individual is present. For perspectival expressions that allow some but not all individuals mentioned in a situation to be perspective holders, the lexical approach is either too restrictive or too permissive.

These perspectival expressions require a way of imposing restrictions on the identity of the perspective holder without stipulating their exact identity. This entails two things: first, that the perspective holder must be represented with some kind of variable (since their identity cannot be directly stipulated in the lexical semantics), and second, that the variable must be governed in some way other than existential quantification (since not just any individual can be the perspective holder in any situation).

This leads me to conclude that for perspectival expressions with a restricted, but non-singleton set of perspective holders, the perspective holder must be represented by a context-sensitive variable in the semantics of the perspectival component. In the following sections, I describe several proposals that posit a perspectival variable in the semantics of a perspectival expression. Each involves a different way of governing the variable.

### 3.2.2 An indexical approach

One way of deriving context-sensitive selection of the perspective holder is to treat perspectival expressions as indexicals. This approach has been pursued for *come* (Taylor, 1988; Oshima, 2006a,b; Sudo, 2018) and other perspectival expressions (Korotkova, 2016).

In an indexical analysis, the semantics of perspectival expressions contain a perspectival variable whose value is determined by the context parameter. This allows the perspective holder to covary with the discourse context. For instance, for *come*, one possible indexical semantics is shown in (66).

#### 66. Indexical semantics for *come*:

$$[[\text{come}]]^{C,g} = \lambda x. \exists e. \text{MOVE}(e) \wedge \text{DEST}(e, x) \wedge x = \text{LOC}(C_{\text{perspective}})$$

One issue for indexical approaches is how to incorporate perspective into the context parameter. In (66), I have represented the perspectival component as a field in the context parameter. Oshima (2006a) proposes augmenting the context parameter with a ranked set of perspective holders. Similarly, Korotkova (2016)'s analysis of evidentials proposes enriching the context parameter with an *Origo* field tracking the person whose information state is being conveyed, and Potts (2007b)'s analysis of predicates of personal taste proposes enriching the context parameter with a judge field.

Another possibility is to make use of the fact that the speaker and listener are already encoded in the context parameter, as in the denotation given in (67).

#### 67. Indexical semantics for *come*, revised:

$$[[\text{come}]]^{C,g} = \lambda x. \exists e. \text{MOVE}(e) \wedge \text{DEST}(e, x) \wedge (x = \text{LOC}(C_{\text{speaker}}) \vee x = \text{LOC}(C_{\text{listener}}))$$

This would allow speaker- or listener-anchored uses of perspectival motion verbs. In order to derive attitude holder anchorings without augmenting the context parameter, we must appeal to **indexical shift**. This is the approach taken by Sudo (2018).

Indexical shift is a phenomenon where indexicals in speech or attitude reports are interpreted relative to the embedded context of utterance or belief, rather than the matrix utterance

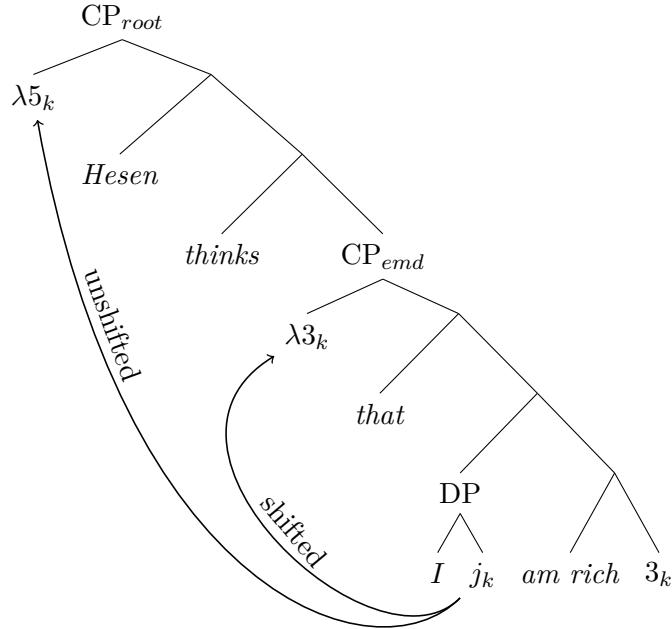
context.<sup>8</sup> For instance, in the Zazaki example shown in (68),  $\varepsilon z$  ‘I’ can refer to the subject of the attitude verb, Hesen, or to the speaker of the sentence.

68. Hesen-i va ke εz dəwletia  
 Hesen said that I rich.be-PRES

‘Hesen<sub>i</sub> said that I<sub>i/I<sub>speaker</sub></sub> am rich.’ (Anand and Nevins, 2004)

Several approaches to indexical shift have been proposed. Schlenker (2003) proposed that in cases of indexical shift, the attitude verb is a universal quantifier over intensional contexts. Some indexicals cannot be bound: these **pure indexicals** refer rigidly to the matrix utterance context. Others can be bound: shifted indexicals are bound by the attitude verb and interpreted relative to the quantified-over contexts.

**Figure 3.2.** Quantifier approach to indexical shift (Schlenker, 2003, tree from Sundaresan (2020a))



By contrast, Anand and Nevins (2004) and Deal (2014) propose that indexical shift is caused by context shift operators projected in finite CPs. These operators overwrite the context

---

<sup>8</sup>Indexical shift has been attested in a large number of languages; see Deal (2017) and Sundaresan (2020a) for more comprehensive discussions of the phenomenon.

parameter of the matrix utterance context (or some of its fields) with that of the embedded utterance context. There is nothing particularly special about the semantics of the shifted indexicals themselves: they refer relative to the context parameter, as pure indexicals.

#### 69. Context shift operator approach (Anand and Nevins, 2004):

Hesen<sub>i</sub> said that I<sub>i</sub> am rich.

- (a)  $[[I]]^{C,g} = C_{speaker}$
- (b)  $[[OP_{cs} \alpha]]^{C,i,g} = [[\alpha]]^{i,i,g}$ , where  $\alpha$  is an attitude report and  $i$  is an intensional context
- (c)  $[[\text{Hesen}_i \text{ said that } I_i \text{ am rich}]] = [[\text{say}]]^{c,i,g}(\lambda i'.[[OP_{cs}[I \text{ am rich}]]])^{c,i',g}([[Hesen]]^{c,i,g}) = 1 \text{ iff, } \lambda i' \text{ compatible with what Hesen said in } i \text{ that } [CPAUTHOR}(i') \text{ is rich}]^{i',i',g}$

A third proposal charts a middle ground between these two approaches: Sundaresan (2020a) proposes that context shift operators are intensional complementizers that quantify over a set of contexts. As in the quantifier approach, indexical shift comes about via quantification over contexts, but, as in the context-overwriting approach, the operator is located in the complementizer rather than the verb.

In addition, Maier (2016) proposes that certain kinds of indexical shift are best analyzed as quotation: I will discuss this proposal in Section 3.2.5, since adopting this proposal for perspectival expressions is a quotational approach.

However indexical shift is analyzed, perspectival expressions with an indexical semantics are predicted to allow attitude holders as perspective holders or non-utterance time anchoring evaluation under the scope of attitude verbs, since this is the environment in which context shift operators are found. In all other circumstances, the perspective holder is predicted to be either the listener or speaker at utterance time.

#### 3.2.3 A logophoric approach

Another way of deriving context-sensitive valuing of the perspective holder is to assume that the perspective holder is represented by a variable that is bound by a logophoric operator. Logophoric binding approaches have been developed for a number of proposed perspectival

expressions: Japanese reflexives (Nishigauchi, 2014); Tamil exempt anaphors (Sundaresan, 2018); and French exempt anaphors (Charnavel, 2019).

Exempt anaphors are anaphors that are exempt from Condition A of Binding Theory. As many authors have noted (Clements, 1975; Sells, 1987), such exemption occurs for anaphors in domains whose content is expressed according to the point-of-view of a perspective holder. One diagnostic for exemption is whether the anaphor occurs in a syntactic environment that inanimate anaphors cannot appear in, since perspective holders must be animate.<sup>9</sup> Charnavel (2019) provides the contrasting examples in (70) to illustrate this requirement in French. Although *son* can ordinarily have either an animate or inanimate referent, when it is exempt, its referent must be animate.

70. (a) \*[Cette auberge]<sub>i</sub> bénéficie du fait que les touristes préfèrent son<sub>i</sub> propre jardin à ceux des auberges voisines.

Intended: ‘This inn<sub>i</sub> benefits from the fact that the tourists prefer its<sub>i</sub> own garden to that of the neighboring inns.’ (Charnavel, 2019)

- (b) Marie<sub>i</sub> bénéficie du fait que les touristes préfèrent son<sub>i</sub> propre hôtel à ceux de ses concurrents.

‘Mary<sub>i</sub> benefits from the fact that the tourists prefer her<sub>i</sub> own hotel to those of the competitors.’ (Charnavel, 2019)

Charnavel (2019) proposes that exempt anaphors are only licensed in the domain of a logophoric operator, where they are bound by a silent logophoric pronoun. This logophoric operator may be projected in any spellout domain (TP, vP, DP, and any other XP that has a subject). The proposed configuration is shown in (71).

#### 71. Syntactic configuration for logophoric binding (Charnavel, 2019):

[XP	[YP [LogP prolog <sub>i</sub> OPlog [α ... exempt anaphor <sub>i</sub> ... ]]]]
phase edge	spellout domain

The logophoric operator takes a logophoric pronoun as its subject and requires that the content of its complement is interpreted from the perspective of its subject.

---

<sup>9</sup>Or rather, sentient, which generally correlates with grammatical animacy.

72. Semantics of logophoric operator (Charnavel, 2019):

$$[[\text{OP}_{\log}]] = \lambda\alpha.\lambda x.\alpha \text{ from } x\text{'s first-person perspective}$$

Although proposed in order to explain exempt anaphors, the logophoric binding approach that Charnavel takes can be applied to other perspectival expressions as well. For *come*, the logophoric binding approach posits a semantics like that shown in (73): the perspectival component contains a variable that is bound by the subject of a logophoric operator.

73. Logophoric semantics for *come*:

$$[[\text{come}_i]]^{C,g} = \lambda x.\exists e.\text{MOVE}(e) \wedge \text{DEST}(e,x) \wedge x = \text{LOC}(l_i), \text{ where } l_i \text{ is bound by the subject of a logophoric operator } \text{OP}_{\text{LOG}}.$$

One of the motivations for Charnavel's proposal is that anaphors must generally be obligatorily bound. By adopting a logophoric binding approach, Charnavel is able to account for the apparent exempt behavior of perspectival anaphora without actually violating Condition A of Binding Theory, since even exempt anaphors are, under her account, bound.

However, for other perspectival expressions, like *come*, we do not have evidence that the perspectival variable shown in (73) is obligatorily bound. There are therefore two different ways of implementing the general logophoric operator approach for perspectival expressions like *come*.

### 3.2.3.1 Obligatory binding approach

One possibility is to assume that the perspectival variable of *come* is obligatorily bound, as in the case of exempt anaphors. In order to derive speaker-oriented readings, we would need to posit a top-level speech act projection that behaves like a logophoric projection whose subject is the speaker (or, optionally, the listener) (Speas and Tenny, 2003). In this case, *come*'s perspectival variable is obligatorily bound either to the subject of this top-level projection, in which case the perspective holder is the speaker, or to the subject of a lower logophoric operator, such as one introduced by an attitude verb.

This works out to be similar to the Pearson (2013)'s proposal for predicates of personal taste. She implements their perspectival component via an individual-type variable that

is obligatorily bound by the most local abstraction operator. In order to derive speaker-oriented readings of predicates of personal taste in matrix clauses, she posits a silent operator in the left periphery that binds individual variables to the speaker.

### 3.2.3.2 Binding by capture approach

An alternative is to assume that the perspectival variable of expressions like *come* can be free, but that logophoric operators obligatorily capture any such free variables. This is the direction that Charnavel (2018) pursues in her analysis of *venir* ‘come’ in French. She proposes that *venir* has an implicit logophoric variable that can be free, but is obligatorily bound by any logophoric operator in whose scope it occurs. If it remains free, its referent is determined pragmatically (i.e., through anaphora resolution).

This means that the optional binding approach requires a mechanism for resolving the referent of the perspectival variable, unlike the obligatory binding approach. However, since both approaches require some kind of explanation of how the speaker decides whether to project a logophoric operator, this is not a reason to prefer the obligatory binding approach over the optional one.

Whether or not the perspectival variable in *come* is obligatorily bound, the logophoric binding approach provides a way of governing the perspective holder as a context-sensitive perspectival variable. As in the indexical shift approach, under the logophoric binding analysis, perspective is governed syntactically, and the identity of the perspective holder is restricted by syntactico-semantic factors.

### 3.2.4 An anaphoric approach

Another approach to the semantics of perspectival expressions is to treat the perspective holder anaphorically. I argued in Section 3.2.1 that the perspective holder cannot be completely unrestricted (i.e., existentially bound). In the two alternative approaches presented so far, the perspective holder is represented by a variable that is governed syntactico-semantically, either indexically or via binding. Another possibility, however, is that this perspective variable is governed pragmatically.

An anaphoric approach to perspectival expressions treats the perspective holder as a free variable whose value is determined directly by the discourse context instead of by binding or by the context parameter. One existing anaphoric proposal is Barlew (2017)'s analysis of American English *come*, which treats *come* as anaphoric to a prominent perspective in the discourse context.

#### 74. Anaphoric semantics for *come*:<sup>10</sup>

$[[\text{come}]]^{C,g} = \lambda x. \exists e. \text{MOVE}(e) \wedge \text{DEST}(e, x) \wedge x = \text{LOC}(p)$ , where  $p$  is a prominent perspective holder in the Common Ground.

In Barlew (2017)'s approach, the Common Ground contains a set of familiar perspectives, perhaps in the form of a ranked list. The perspective holder of *come* must hold a discourse familiar perspective. The perspectives of conversation participants always satisfy this constraint on discourse familiarity: their perspectives are automatically entered into the Common Ground. Thus, this approach easily allows speakers and listeners to serve as perspective holders for *come*.

It also allows other kinds of perspective holders. As the discourse context evolves, new perspectives can be added and perspectives can rise and fall in prominence. When new characters in a narrative are introduced, their perspectives become familiar. Certain constructions may also introduce or make prominent other perspectives: for instance, attitude predicates make the perspectives of their subjects prominent.

Because the anaphoric account relies on prominence in the discourse context to govern the interpretation of the perspectival expressions, it requires an explanation of how perspectives are tracked in the Common Ground. Barlew (2017) proposes a ranking of perspectives that is separate from the Common Ground set of individuals. He treats perspectives as sets of **centered worlds**.

---

<sup>10</sup>Barlew (2017)'s dynamic semantics are adapted here into an event semantics.

75. **Centered world** (Stalnaker, 2008): a centered world is pair consisting of a world and an individual in the world at a particular time and place, which is the individual that the perspective holder believes themselves to be.

The intuition is that instead of viewing an individual's beliefs as a set of propositions they hold to be true, it is also possible to view them as a set of possible worlds in which they believe they may be located. If Hercule Poirot believes that Jacqueline de Bellefort is the murderer, he believes himself to be located in one of the worlds in which this proposition is true. Viewing beliefs through the lens of self-location is particularly helpful for perspective, because it unifies the spatial and mental aspects of perspective. From the set of locations in possible worlds in which an individual self-locates (their centered worlds), we can recover both their spatial perspective and their mental perspective.

Barlew (2017)'s proposal to add a ranking of perspectives, in the form of sets of centered worlds, enriches the kinds of objects tracked by the Common Ground. It also involves an ontological enrichment, since perspectival expressions like *come* now involve variables that resolve to perspectives, which, in his proposal, have a complex type.

Alternatively, it may be possible to derive perspectival prominence from the combined prominence of an individual and a time. The prominence of times is important because of how *come* allows different evaluation times of the anchoring relation. The perspective holder can be located at the destination of motion at either event time or utterance time.

76. (a) **Utterance time:** Colonel Hastings is coming here now.  
(b) **Event time:** When I was in the library, Miss Lydgate came to retrieve a book.

In order to explain these facts, perspectives must either be time-indexed,<sup>11</sup> or represented in a way that allows a time index to be retrieved. The centered worlds approach that Barlew (2017) proposes is sufficient to distinguish among different time-indexed perspectives, since an individual's self-ascribed location includes both their temporal and spatial location. However, rather than enriching the set of objects tracked by the Common Ground, as

---

<sup>11</sup>Or event-indexed or situation-indexed.

Barlew proposes, it might also be possible to track perspective prominence with a set of prominent individuals and a set of prominent times. A set of prominent perspectives could then be derived by combining the two.

In addition to these two approaches, there may be other feasible systems for tracking perspective prominence. Regardless of the method of tracking perspective prominence that is adopted, the anaphoric approach is strictly more powerful than the other context-sensitive accounts proposed so far. Any configuration of perspectival expressions that is possible according to the indexical or logophoric binding account is also possible according to the anaphoric account. However, because perspectival expressions shift more freely in the anaphoric account, developing an anaphoric analysis requires a specification of the conditions under which perspectives enter into and become prominent in the Common Ground.

### 3.2.5 A quotational approach

The last kind of analysis for perspectival expressions that I consider is a quotational account. In this account, perspectival expressions like *come* are treated as a kind of mixed quotation.

**Mixed quotation** is the use of someone else's speech directly in the context of an indirect speech report (Davidson, 1979; Maier, 2017a). In mixed quotation, the quoted words are both used and mentioned. (77) illustrates the two components of mixed quotation: (1) the content of an indirect speech report (that Harriet Vane appeared at a party), and (2) a meta-linguistic contribution that those actual words were used by the source of the speech report (that the newspaper used the phrase *a well-known detective authoress*).

77. “Three days later, while reading in the morning paper that among the guests at a literary luncheon-party had been seen ‘Miss Harriet Vane, the well-known detective authoress,’ she was interrupted by the telephone.” (Sayers, 1935)

Like pure quotation, mixed quotation makes a meta-linguistic contribution; unlike pure quotation, however, it is grammatically incorporated into the rest of the sentence. Because mixed quotation has this meta-linguistic component, it necessarily conveys the perspective of the person being quoted. That is, the appositive in (77) must be interpreted with respect to the newspaper writer’s perspective, not the speaker of the matrix utterance. Interpreting

mixed quotation therefore involves accessing a non-speaker perspective. In this sense, all quotation is perspectival, but for our purposes, the relevant question is whether perspectival phenomena like *come* can be explained using the same mechanisms as quotation.

Maier has contributed several closely related accounts of quotation.<sup>12</sup> Maier (2015, 2016) outlines a two-dimensional account of quotation building on Geurts and Maier (2005) and Potts (2007a). The use component is a property variable  $P$ , ranging over semantic objects of the type matching the syntactic category of the quoted phrase. The mention component states that a contextually salient individual used the material within the quotation marks to refer to  $P$ .

78. Maier (2015) quotation:

$[[“A”]] = \text{whatever a contextually salient individual } x \text{ used } A \text{ to mean.}$

79. Ann said that this music was “not mah cup ’o tea.”

Ann used ‘not mah cup ’o tea’ to refer to the property  $P \wedge$  she said that this music was  $P$ .

The contextually salient individual is governed by pragmatic factors; the property is interpreted with respect to their perspective, since it is interpreted as whatever they intended the linguistic material to mean.

This account could be adapted for perspective as follows. First, assume that quotation can apply at a subword level, an assumption that Maier (2015) makes to explain the split interpretation of pronouns in Free Indirect Discourse.<sup>13</sup> Second, for *come*, we posit that the destination argument of *come* alone is quoted. Since the destination is often implicit, there is no observable stylistic faithfulness, the hallmark of quotation. We cannot assume the entire motion verb is quoted, since in this case, we would expect the form of the motion

---

<sup>12</sup>Here I focus on Maier’s earlier anaphoric approach to mixed quotation, rather than his more recent event modification approach (Maier, 2017b, 2018), as the latter posits both a syntactically represented quotee and a quoted proposition, making it less compatible with perspectival expressions.

<sup>13</sup>He assumes that the gender features of pronouns are quoted, but the person features are unquoted.

verb to be faithful to the style of the perspective holder (reflecting, for instance, any stylistic traits of their dialect).

#### 80. Quotational semantics for *come*:<sup>14</sup>

$$[[\text{come}]]^{C,g} = \lambda x. \exists e. \text{MOVE}(e) \wedge \text{DEST}(e, x) \wedge x = \text{"my location"}$$

$$[[\text{come}]]^{C,g} = \lambda x. \exists e. \text{MOVE}(e) \wedge$$

$\text{DEST}(e, x) \wedge x = \text{whatever a contextually salient individual } y \text{ could use } my\ location \text{ to mean}$ )

The key question is how to treat the quotee. If we propose that the perspectival component of *come* is quoted, we must also provide a way of determining who the perspective holder, or quotee, is. In Maier (2015, 2016), the quotee is essentially determined anaphorically. This means that an account of perspective based on Maier (2015, 2016) is essentially anaphoric.

A quotational approach to perspective shift following Maier (2015) therefore turns out identically to an anaphoric approach. For the remainder of this chapter, I will omit direct discussion of the quotational approach, since my discussion of the predictions of the anaphoric approaches applies to the quotational approach as well.

### 3.3 A set of diagnostics for perspectival encoding

The four approaches to perspective sketched above propose different methods of encoding the perspective holder in the semantics of perspectival expressions. They make correspondingly different predictions about the behavior of perspectival expressions in a number of environments. In this section, I develop a set of diagnostics for probing the representation of perspective based on these predictions. In many places, I will again use *come* as a case study to illustrate the different approaches; however, the diagnostics are general and could be applied to any perspectival expression for which the right environment can be constructed.

My goal in laying out this set of diagnostics is to guide exploration of the landscape of perspectival expressions. In the previous section, I discussed four ways that the perspective holder can be encoded in the semantics of perspectival expressions, using, for the most

---

<sup>14</sup>Barlew (2017)'s dynamic semantics are adapted here into an event semantics.

part, independently motivated mechanisms. Within each of these four approaches, there may be many possible variants. My descriptions of the predictions of each account should be taken as a description of the predictions of the simplest or most standard version of each approach. Not all variants of a particular approach will necessarily share these predictions, but it is helpful to understand which environments we would expect to adjudicate among these four approaches.

### 3.3.1 Cardinality of the perspective set

I argued in Section 3.2.1 that the lexical approach can only be applied to a narrow class of perspectival expressions: those with either an unrestricted set of licit perspective holders, or a singleton set. This is because the lexical account must stipulate the identity of the perspective holder lexically. If there is a single licit perspective holder, this is simple, because their identity can be lexically specified. On the other hand, if there are no constraints on who the perspective holder can be, this is also simple, because existential binding over individuals can be used.

Most of the expressions that have been called perspectival do not fall into either of these classes. In order to be noticeably perspectival, expressions generally allow some but not all individuals to serve as perspective holders.

This is predicted by the remaining three accounts. Under the anaphoric account, the dis-course prominence of the individual determines whether they are a licit perspective holder. Under the logophoric binding account, the perspective holder is governed by the logophoric operator. In the indexical account, the set of perspective holders is either encoded in the context parameter, or contains only the individuals already tracked in the context parameter (the speaker and listener); other perspective holders become available when context shift operators manipulate the contents of those fields in the context parameter.

**Table 3.2.** Perspective set cardinality predictions by account

	Lexical	Indexical	Logophoric	Anaphoric
Singleton perspective set	✓	X	X	X
Restricted perspective set	X	✓	✓	✓

### 3.3.2 *De se* effects

In Chapter 2, I presented data showing that the anchoring relation of *come* is obligatorily *de se* interpreted: the perspective holder must self-locate at the destination of motion. This is a commonly posited property of perspectival expressions: the perspective holder must self-ascribe the content of the perspectival expression.

81. **Attitude terminology** (Lewis, 1979a):

- (a) *De se*: an attitude that an agent holds about themselves in a first-person way
- (b) *De re*: an attitude that an agent holds, but that is expressed in words that they may not agree with

Indeed, it is hard to imagine what it means for an expression to be perspectival if it is not interpreted *de se*, since what it means for an expression to convey an individual's point-of-view is that it is consistent with their beliefs, including their beliefs about themselves. One desirable property, therefore, for any semantic analysis of a perspectival expression is that it derives obligatory *de se* interpretation.

The indexical shift account predicts obligatory *de se* interpretation. Both the cross-linguistic data on indexical shift and the predictions of the context shift account support an obligatory *de se* interpretation of shifty indexicals.<sup>15</sup> In a context shift account, the location and time fields of the utterance context will be replaced with those of the embedded speech context. In theory, this only guarantees that the destination is the *de se* location of the perspective holder if the perspective holder believes themselves to be located at the location of the embedded speech context. However, most approaches to context shift assume contexts of thought, as well as contexts of speech, in which case, the embedded context parameter will encode the perspective holder's self-ascribed location.

The logophoric binding account also predicts that perspectival expressions obligatorily receive *de se* interpretations. In this account, the perspectival variable is bound by the

---

<sup>15</sup>A *de se* requirement has been widely reported for shifty indexicals in a variety of languages: see Deal (2017) for a more extensive discussion.

argument of a logophoric operator, which is a logophoric center. The logophoric center represents who the individual believes themselves to be, including their self-ascribed spatial and temporal location (Charnavel, 2019). Therefore, the individual must believe themselves to be located at the destination of motion.

The predictions of the anaphoric account rest on how perspective is formalized. The anaphoric account proposes that the perspectival variable refers to a perspective holder in the Common Ground. If the Common Ground representation of a perspective is a set of centered worlds, as Barlew (2017) proposes, the obligatory *de se* property follows from the fact that a perspective encodes the perspective holder’s self-ascribed location. However, if some other representation of perspective is used (for instance, if the perspectival variable merely refers to an individual in the Common Ground), then some additional presupposition might be necessary in order to capture the *de se* facts.

The lexical account makes no particular predictions about the *de se* interpretation of perspectival expressions. In order to derive obligatory *de se* readings, they would have to be stipulated in the lexical semantics of the expression.

**Table 3.3.** *De se* predictions by account

	Lexical	Indexical	Logophoric	Anaphoric
Obligatory <i>de se</i> anchoring	◊	✓	✓	◊

### 3.3.3 Syntactic domain of perspective shift

The logophoric and indexical accounts posit that perspective is governed by syntactic operators. Both accounts therefore predict that the interpretation of a perspectival expression depends in part on its syntactic environment. However, the two accounts make different predictions about the syntactic domains in which the perspective-governing operators occur.

On the indexical view, the perspectival variable is predicted to refer to attitude holders only when *come* is in the scope of a context shift operator. In Deal (2017)’s and Sundaresan (2020a)’s treatments of indexical shift, context shift operators can only be projected in finite CPs. This makes a strong prediction about the syntactic environment in which attitude holder-anchored perspectival expressions should occur.

## 82. Environment: Perspective shift within finite CP

[<sub>IP</sub> Mary<sub>i</sub> said [<sub>CP</sub> that ... *persp. expression<sub>i</sub>*]]

The logophoric account, by contrast, is more permissive, at least as formulated by Charnavel (2018): she proposes that logophoric operators can be projected in any spellout domain (TP, vP, DP, or any other XP with a subject). Thus, attitude holder-anchored perspectival expressions should also be able to occur in any such domain.

## 83. Environment: Perspective shift in XPs with subjects

[<sub>XP</sub> Meg<sub>i</sub> ... *persp. expression<sub>i</sub>* ... ]

The anaphoric account, by contrast to the others, does not impose restrictions on the syntactic domain in which non-speaker-anchored interpretations of perspectival expressions should arise. The lexical approach also does not predict that the syntactic environment affects the interpretation of perspectival expressions: either there is only a single perspective holder, or the perspective is existentially bound.

**Table 3.4.** Syntactic domain for perspective shift predictions by account

	Lexical	Indexical	Logophoric	Anaphoric
Shift outside finite CP	✓	X	✓	✓
Shift outside XP with subj.	✓	X	X	✓

### 3.3.4 Shift Together effects

Shift Together effects occur when all context-sensitive expressions in a domain must shift together if any shift. While originally proposed for indexical shift, Bylinina et al. (2015) proposes that perspectival expressions are also subject to Shift Together constraints.

This is expected under an indexical approach, since Shift Together effects are robustly reported for indexical shift in most languages.<sup>16</sup> In the Slave example in (84), for instance, we see that the embedded first persons must either both refer to the attitude holder or both refer to the speaker; shifty indexicals in the same syntactic domain cannot shift independently of one another.

---

<sup>16</sup>There is ongoing debate about whether Shift Together effects hold for all cases of indexical shift; see Sundaresan (2020a) for a cross-linguistic survey of the evidence.

84. Sehlégé segħa gonīħkie rárulu yudeli.  
1s.friend 1s.for slippers 3s.will.sew 3s.want.4s

- (a) ✓‘She<sub>i</sub> wants her<sub>i</sub> friend to sew slippers for her<sub>i</sub>.’
- (b) ✓‘She<sub>i</sub> wants my friend to sew slippers for me.’
- (c) X ‘She<sub>i</sub> wants my friend to sew slippers for her<sub>i</sub>.’
- (d) X ‘She<sub>i</sub> wants her<sub>i</sub> friend to sew slippers for me.’

(Rice, 1986; Anand and Nevins, 2004; Deal, 2017)

Context overwriting accounts of indexical shift (Anand and Nevins, 2004; Deal, 2014) predict that Shift Together effects are obligatory, since a context, once overwritten, cannot be recovered. However, Anand and Nevins (2004) posit that context shift operators can overwrite some but not all of the fields of the context parameter, which predicts Shift Together effects only for indexicals of the same class (person indexicals will affect each other but not temporal or locative indexicals).

Schlenker (2003)’s context quantification approach only predicts Shift Together effects for obligatorily shifted indexicals in the same domain. These will shift together since they are bound by the same quantifier. Indexicals that optionally shift, on the other hand, will not be subject to Shift Together effects.

Sundaresan (2020a) points to this as a weakness of the Schlenker (2003) approach, given the cross-linguistic frequency of Shift Together effects. She proposes a locality rule to derive Shift Together effects: the silent context pronoun associated with a shiftable indexical must be co-indexed with its locally c-commanding operator. This blocks any non-local shifting (such as in the case of multiply-embedded indexicals).

Although the context quantification theory of indexical shift does not derive Shift Together effects, Shift Together effects are widely attested cross-linguistically. Under an indexical account of perspectival expressions, therefore, we would expect Shift Together effects for multiple instances of the same perspectival indexical within the same finite CP.

## 85. Environment: Shift Together within finite CP

- (a)  $[_{IP} \text{Susan}_j \text{ said } [_{CP} \text{that } \text{Mary}_i \text{ said } [_{CP} \text{that } \dots \text{ persp. } \text{expr.}_i \dots \text{ persp. } \text{expr.}_{i/*j}]]]$
- (b)  $[_{IP} \text{Susan}_j \text{ said } [_{CP} \text{that } \text{Mary}_i \text{ said } [_{CP} \text{that } \dots \text{ persp. } \text{expr.}_j \dots \text{ persp. } \text{expr.}_{j/*i}]]]$

However, it is important to note that the predicted Shift Together effects for perspectival indexicals are different from those of shifty (non-perspectival) indexicals. For non-perspectival indexicals, Shift Together effects have been observed only among indexicals that refer relative to the same field in the context parameter. For instance, person and locative indexicals in Nez Perce can shift independently of each other (Deal, 2014). Unlike the shifty indexicals that Deal (2014) discusses, however, all perspectival indexicals refer relative to the same field in the context parameter.<sup>17</sup> This means that all perspectival indexicals should be subject to obligatory Shift Together effects.

What about the other accounts?

The logophoric account also relies on syntactic operators and therefore also predicts Shift Together effects among perspectival expressions in the same spellout domain. The predicted Shift Together effects apply not just to multiple instances of the same perspectival expression, but to any expressions bound by logophoric operators in the same domain.

## 86. Environment: Shift Together within XPs with subjects

- (a)  $[_{XP} \text{Sue}_i \dots [_{XP} \text{Meg}_j \dots \text{persp. } \text{expr.}_i \dots \text{persp. } \text{expr.}_{i/*j}]]$
- (b)  $[_{XP} \text{Sue}_i \dots [_{XP} \text{Meg}_j \dots \text{persp. } \text{expr.}_j \dots \text{persp. } \text{expr.}_{j/*i}]]$

The anaphoric account does not predict obligatory Shift Together behavior, since each perspective variable's referent is resolved pragmatically. However, the discourse effects that govern perspective prominence may effectively impose Shift Together constraints, since the prominence of perspective holders is unlikely to change rapidly between two local perspectival expressions. These indirect Shift Together effects will not be syntactic in nature: the binding domains in which the perspectival expressions occur will not impact their behavior.

---

<sup>17</sup>In Section 3.2.2, I outlined two variants of a perspectival indexical account. In one, the context parameter contains a field that tracks a set of perspectives. In the other, perspectival expressions refer relative to the speaker's perspective, tracked by the speaker field. In either variant, perspectival indexicals take their reference from the same field of the context parameter.

The lexical approach predicts no Shift Together effects at all. If there is only one possible perspective holder, then Shift Together effects are impossible to observe. If the perspective is existentially bound, then there should be no interactions between perspectival expressions.

**Table 3.5.** Shift Together predictions by account

	Lexical	Indexical	Logophoric	Anaphoric
Shift Together effects	X	✓	✓	◊

### 3.3.5 Utterance boundaries

The accounts sketched above also differ with respect to whether perspectival licensing is possible across utterance boundaries. The indexical and logophoric accounts do not predict this to occur, because they govern the perspectival variable syntactically. Thus, a perspective introduced in one utterance should not be accessible in the next utterance.

#### 87. Environment: Licensing across utterance boundaries

[<sub>IP</sub> Mary<sub>i</sub> said [<sub>CP</sub> that the pheasant tasted off.]] [<sub>IP</sub> ... *perspectival expression<sub>i</sub>* ... ]

The indexical account predicts that the configuration shown in (87) is not licit. Because the indexical account relies on context shift operators to introduce the attitude holder's perspective via the context parameter, it predicts that the perspective of the attitude holder should not be available outside of the scope of the attitude predicate (since the context parameter will only be shifted within the scope of the attitude verb). Thus, the indexical analysis does not provide a mechanism for cross-utterance perspective licensing.

Similarly, the logophoric binding account predicts that this configuration should not be possible. If there is a logophoric binding operator whose subject is a logophoric pronoun with the same index as the attitude holder, that operator cannot bind a perspectival expression in the second utterance, since its binding domain does not extend across the utterance boundary.

By contrast, the anaphoric account predicts that cross-sentential perspectival anchoring should be possible. Under the anaphoric account, the attitude verb *thought* introduces Captain Hasting's perspective into the set of discourse-prominent perspectives in the Common Ground. So long as it remains prominent, perspective variables in the remainder of

the discourse can continue to refer to it, even if they are not in the scope of the attitude verb that introduced it.<sup>18</sup>

The lexical approach does not predict licensing across utterance boundaries, since it does not really include a notion of licensing the perspective.

**Table 3.6.** Cross-utterance anchoring predictions by account

	Lexical	Indexical	Logophoric	Anaphoric
Anchoring across utterances	X	X	X	✓

### 3.3.6 Quantificational binding

In quantificational binding contexts, pure indexicals do not covary with the quantifier, since their value is determined solely by the utterance context. For instance, in (88), *here*, a pure indexical, does not covary with the quantifier. Even though there is a prominent location for each instantiation of the quantifier, *here* picks out the location of utterance, leading to a contradictory interpretation (unless there is some way for the speaker to try food at a distance).

88. # Every time I travel somewhere new, I always try the food here.

Covariation with quantifiers is also unattested for shifty indexicals (Deal, 2017). For instance, in the Korean example in (89), the shifty indexical *yeki* ‘here’ cannot covary with the quantifier, but must be interpreted as strictly relative to utterance time. By contrast, the locative adverbial *palhwa cangso* ‘the speech location’ can, as shown in (90).

89. \*Obama-ka malhal ttyay.mata manhun salamtul-i **yeKi**-ey issta.  
Obama-NOM speaks whenever many people-NOM here-at be

Intended: ‘Whenever Obama speaks, many people are here.’ (Park, 2014a)

---

<sup>18</sup>The strength of the evidence derived from this diagnostic may vary based on the direction of the argument. If licensing across utterance boundaries is not observed, then it is a strong indication for a logophoric or indexical approach over an anaphoric approach. However, if licensing across utterance boundaries is observed, the argument for the anaphoric approach is slightly weaker, due to the possibility of enriching the logophoric or indexical account in order to capture licensing across utterance boundaries in a different way (i.e., through some kind of sentential binding). In general, it is possible to enrich any of these accounts and lead to different predictions than the ones I have described; my goal is to illustrate the predictions of each account given a minimal set of assumptions.

90. Obama-ka malhal ttyay.mata manhun salamtul-i **palhwa**  
 Obama-NOM speaks whenever many people-NOM speech  
**cangso-ey** issta.  
 location-at be

'Whenever Obama speaks, many people are at the speech location.' (Park, 2014a)

This lack of covariation with quantification is predicted by the dominant theories of indexical shift. Since Deal (2014)'s and Sundaresan (2020a)'s accounts only posit context shift operators in finite CPs of attitude verbs, (89) is not predicted to allow indexical shift. The fact that the indexical in (89) is not in the scope of an attitude verb also blocks indexical shift in the quantification over contexts approach taken by Schlenker (2003).

On the other hand, the logophoric binding account does not predict infelicity in quantificational binding contexts. If we assume that *her* introduces a logophoric operator that governs the judge of *wayward*, a predicate of personal taste, the value of the logophoric *pro* will covary with the quantifier since *her* does. Thus, the logophoric approach predicts that the perspective variable should be able to covary with the quantifier so long as the logophoric pronoun does.

The anaphoric account also predicts that perspectival expressions should be able to covary with the quantifier in quantificational binding contexts. Expressions that are anaphoric to other objects, such as the time-anaphoric expression *the next day*, covary (91).

91. Bunter: Every time Lord Peter catches a murderer, he falls into a depression the next day.

Perspective anaphoric expressions are predicted to covary so long as the context makes a different perspective prominent for each instantiation of the quantifier.

The lexical approach does not predict covariation. If the perspective is lexically stipulated, it never varies; if it is existentially bound, there is no reason that it should vary.

**Table 3.7.** Quantificational binding predictions by account

	Lexical	Indexical	Logophoric	Anaphoric
Covariation in quant. contexts	X	X	✓	✓

### 3.3.7 A diagnostic toolkit for perspectival expressions

I have discussed four methods of encoding the perspective holder in the semantics of perspectival expressions. The lexical approach is only viable for perspectival expressions that either have a single licit perspective holder whose identity can be lexically stipulated, or expressions that allow any individual to serve as the perspective holder. The quotational analysis reduces to an anaphoric approach, and therefore makes identical predictions to it.

The three remaining accounts (indexical, logophoric, anaphoric) are the most interesting, because they can be applied to perspectival expressions that allow a restricted, but non-singleton set of perspective holders. These accounts make different predictions about the behavior of perspective expressions in a number of environments, summarized in Table 3.8.

**Table 3.8.** Predictions by account

	Lexical	Indexical	Logophoric	Anaphoric
Singleton perspective set	✓	X	X	X
Restricted perspective set	X	✓	✓	✓
Obligatory <i>de se</i> anchoring	◊	✓	✓	◊
Shift outside finite CP	✓	X	✓	✓
Shift outside XP with subj.	✓	X	X	✓
Shift Together effects	X	✓	✓	◊
Anchoring across utterances	X	X	X	✓
Covariation in quant. contexts	X	X	✓	✓

Within each of these four broad categories of approaches, there may be many variants. Some of these many not make exactly the same predictions as the standard version that I have presented. Each individual diagnostic should therefore be seen not as conclusive evidence for or against a particular approach, but as a guide for identifying a critical environment in which to test the behavior of perspectival expressions. If a perspectival expression behaves contrary to the predictions of one account, as I sketch them, it is suggestive (but not conclusive) evidence that it might belong to a different class of perspectival expressions.

Taken together, the predictions described in Table 3.8 constitute a set of diagnostics for fine-grained analysis of the semantics of perspectival expressions. The behavior of a given

perspectival expression can be tested in each of these environments in order to determine whether it should be analyzed lexically, indexically, logophorically, or anaphorically.

### 3.4 Applying perspectival diagnostics to *come*

In this section, I demonstrate how the set of diagnostics developed in Section 3.3 can be used to explore which analysis is most appropriate for a given perspectival expression. I use a well-studied canonical perspectival expression, American English *come*, as my test case. Although proposals have been made for perspectival motion verbs in all four families of analyses (Taylor, 1988; Goddard, 1997; Oshima, 2006a,b; Barlew, 2017; Charnavel, 2018; Sudo, 2018), I show that the diagnostics suggest that its behavior is most consistent with the predictions of the anaphoric account.

#### 3.4.1 Diagnostic 1: cardinality of perspective set

The lexical approach stands out from the rest in that it assumes that the perspective holder is either existentially bound or hardcoded into the lexical semantics. This predicts either a singleton perspective set or an unrestricted set of perspective holders.

As I have argued in Chapter 2 and Section 3.2.1, the set of licit perspective holders for American English *come* is neither completely unrestricted nor a singleton set. As (65) (repeated as (92) below) shows, the fact that there is an individual at the destination of motion is not enough to license *come*, which shows that the perspectival anchor of *come* is not existentially quantified.

92. *Captain Hastings and Poirot are in London.*

Hastings: # Last year I came to New York. While I was checking into my hotel, someone ran off with my bags.

We have also seen several examples of many kinds of perspective holders for *come*: the speaker, the listener, and, as in (93), the subject of an attitude verb.

93. *Miss Marple is in her cottage, reflecting on the murder of Agnes, the maid at the Symmingtons' house.*

Miss Marple: Agnes must have realized that the letter had arrived before the postman came to the house.

This is inconsistent with a unique lexically stipulated perspective holder, unless we want to posit a separate lexical entry for every possible perspective holder.

Thus, the facts about the cardinality of the set of licit perspective holders for *come* rule out a lexical approach. They are consistent with an indexical, logophoric, or anaphoric approach.

### 3.4.2 Diagnostic 2: quantificational binding

The indexical approach differs from the logophoric binding and anaphoric approaches in its predictions about quantificational binding contexts. According to the indexical account, the perspective holder of *come* is fixed by the context parameter. Because the context parameter does not vary in quantificational binding contexts, the perspective holder of *come* should not either.

This prediction does not seem to be supported by the data. As Barlew (2017) points out, the perspective holder for *come* can covary with the quantifier in quantificational binding cases. Barlew (2017) presents (94) as evidence for an anaphoric approach to *come* and against an indexical approach.

94. Every mother was grateful that her wayward child came home for Christmas.

However, it seems possible to analyze this as a case of indexical shift. (94) contains an attitude verb (*was grateful*) and a complement of the right height for indexical shift in a Deal (2014)-style analysis. It could be the case that for each instantiation of the quantifier (each mother), a different context shift operator is introduced to manipulate the context parameter.<sup>19</sup> If this were the case, the perspective holder for *come* would also be expected to covary with the quantifier.

---

<sup>19</sup> As far as I am aware, such environments have not been discussed in the shifty indexical literature, so I cannot draw upon evidence from languages that have shifty indexicals.

If we turn to other quantificational binding examples, however, the facts become clearer. (95) is a near paraphrase of (94) that does not contain an attitude verb, and therefore cannot be explained by indexical shift.

95. Every mother bought a special present for her wayward child who was coming home for Christmas.

Contrary to the predictions of an indexical account of *come*, this sentence is felicitous. Quantificational binding contexts therefore pose a serious challenge for indexical approaches to American English *come*. The ability of the perspective holder of *come* to covary with the quantifier provides evidence in favor of either the logophoric binding or anaphoric account.

### 3.4.3 Diagnostic 3: perspective shift outside finite CPs

The indexical shift account predicts that perspective shift should only occur when perspectival expressions are inside finite CPs. Under Deal (2017) and Sundaresan (2020a)'s analyses of indexical shift, context shift operators may only be projected in finite CPs. If *come* is an indexical, it should receive speaker-oriented interpretations in all other environments.

This prediction does not seem to hold. We have already seen one example that contradicts it: in (37), repeated below as (96), *come* is interpreted relative to an attitude holder who is introduced by an dream report that does not contain a CP.

96. Gilbert Joseph was one of the several thousand Jamaican men who joined the RAF to fight against Hitler. Returning to England after the war he finds himself treated very differently now that he is no longer in a blue uniform. It is desperation that makes him remember a wartime friendship with Queenie and knock at her door. **Hortense shared Gilbert's dream of leaving Jamaica and coming to England to start a better life-** that's why she married him. But when she at last joins her husband, she is shocked by London's shabbiness and horrified at the way the English live. (Baroni et al., 2009)

There are also many corpus examples containing attitude holder-oriented uses of *come* inside non-finite CPs, as in (97). In (97), the perspective holder for *come* is the security worker, but *come* appears in a non-finite clause.

97. In an email to Kotaku, Taylor explained that the venue gave him his money back, and that even after he was escorted out, he asked if he could stay. “When I was escorted out to the lobby,” Taylor said, “no one put their hands on me or anything. A security worker politely asked me to come with him and I walked out under my own power.”  
(Davies, 2008)

These examples provide evidence against the indexical account for *come* and in support of either the logophoric binding or the anaphoric account.

#### 3.4.4 Diagnostic 4: perspective shift outside XPs with subjects

Although the logophoric binding account does not predict that perspective shift should be constrained to finite CPs, it does posit some limits on perspective shift. Because the logophoric approach uses logophoric operators to govern the perspectival variable of *come*, perspective shift is predicted to occur only when *come* is in the domain of a logophoric operator. Charnavel (2019) posits that logophoric operators may be projected in any XP with a subject.

The anaphoric account, by contrast, predicts that perspective shift should also happen outside of XPs with subjects, since the value of the perspectival variable is determined by the discourse context.

This environment is particularly challenging to test. Because it is a verb and requires a subject, *come* must occur in an XP with a subject. However, there do seem to be examples where the perspective holder for *come* is introduced in a separate syntactic domain from *come*. For instance, in (40), repeated below as (98), the perspectival anchor for *come* is an attitude holder introduced as the subject of the attitude verb *said*. However, the attitude report is interrupted with a long appositive, in which *come* is used.

98. Ms. Watson: It never crossed my mind. I had been traveling the way everybody else did, by air and - or car, and I was frustrated with that. **And a friend of mine a while back said - when I was going to come and visit her in New Mexico**, I was angry because I couldn't get there fast, and couldn't get there easily. And finally, I thought maybe I could drive, but gee, that's several days.

Conan: A long way, yeah.

Ms. Watson: She said, why don't you take the train? (Conan, 2008, found in (Davies, 2008))

If there were a logophoric operator introduced in the CP of the attitude report, it would not scope over the content of the appositive. However, this rests on an analysis of appositives as outside the syntactic scope of the rest of the utterance.

Another environment to test is when the attitude holder is introduced in a syntactic domain that is usually considered to have a limited scope. For instance, in (99), the attitude holder Daisy is introduced in an adverbial. Any logophoric operator whose subject is bound by Daisy should not extend over the rest of the utterance.

99. *Nick is complaining to a friend about how out-of-touch his friend Daisy is about money.*

Nick: Despite Daisy's unrelenting guilt trips, no one came to her wedding. It turned out that the cost of travel to the International Space Station was prohibitive for most of her guests.

As (99) shows, however, instances of *come* in the main body of the utterance can still be interpreted relative to the attitude holder. This suggests that perspective shift is bound by the syntactic domain of logophoric operators introduced in XPs with subjects, as predicted by the logophoric account.

### 3.4.5 Diagnostic 5: Shift Together effects

Both indexical and logophoric accounts predict Shift Together effects for *come*. This is because syntactic operators govern the perspective holder of *come* under either analysis. Multiple instances of *come* under the same operator should receive the same interpretation.

The anaphoric account does not predict obligatory Shift Together effects, though weaker Shift Together effects might be expected, depending on how the anaphora resolution is posited to proceed. If the anaphora resolution involves deterministic perspective selection, like choosing the single most prominent perspective from the discourse context, then multiple nearby perspectival expressions would likely receive the same interpretation, since the ranking of prominent perspectives is unlikely to change significantly between each instance of *come*.

The indexical account predicts that Shift Together effects for perspectival expressions should be observed within finite CPs, since this is the domain of context shift operators. We can test this by looking at multiple instances of *come* in the scope of a single attitude holder, such as in (100).

100. *Nick and Carolyn are siblings. Nick lives in Texas, Carolyn lives in Massachusetts, and their parents live in Washington. Nick is in Texas talking to his friend.*

Nick: Carolyn is annoyed that our mom is coming to see me instead of coming to see her.

The indexical account predicts this to be infelicitous, since the two instances of *come* in the finite CP have different destinations and therefore different perspective holders. This example is particularly problematic for an indexical account, since the first instance of *come* is speaker-oriented. This should block any lower perspectival expressions from receiving attitude holder-oriented interpretations, but it does not. Thus, the obligatory Shift Together effects within finite CPs predicted by the indexical account do not seem to exist.

The logophoric account predicts Shift Together effects in a narrower syntactic domain: within XPs with subjects. It is more challenging to construct examples testing this prediction, since it requires two instances of *come* to co-occur in a fairly small domain. The simplest environment that satisfies the conditions is VP coordination, as in (101).

101. *Nick and Carolyn are siblings. Nick lives in Texas, Carolyn lives in Massachusetts, and their parents live in Washington. Nick is in Texas talking to his friend.*

Nick: Carolyn says that our mom will come to Texas during bluebonnet season and come to Northampton during asparagus season.

In this example, the subject of both moving events is the mother, but the destination of motion is different for each instance of *come*. In the first, it is the location of the speaker, while in the second, it is the location of the attitude holder. If there is a logophoric projection, it must be in an XP with a subject, and the syntactically lowest subject is the subject of the coordinated VP. Since this means that there cannot be a logophoric projection within a branch of the coordinated VP projection, any logophoric operator must scope over both instances of *come*.

Therefore (101) provides evidence either that attitude holder-oriented readings can arise without logophoric operators, or that the perspectival variable of *come* is not obligatorily bound when it is in the scope of a logophoric operator. This data is hard to reconcile with a logophoric view. To maintain a logophoric analysis, it would be necessary to either add a mechanism besides logophoric binding that licenses attitude holder-oriented readings, or to propose a variant in which the perspectival variable of *come* is only optionally bound by the logophoric operator.

By contrast, the behavior of *come* in (101) is exactly as predicted by the anaphoric approach, which neither predicts nor rules out Shift Together Effects.

### 3.4.6 Diagnostic 6: anchoring across utterance boundaries

Under both the logophoric and indexical approaches, perspectives other than those of the speaker and listener are not predicted to be accessible outside of the utterance in which they are introduced. This is because both accounts employ syntactic operators whose scope determines the interpretation of the perspective variable. The anaphoric account, by contrast, predicts that once a perspective is introduced into the discourse context, it should be accessible in later utterances, so long as it remains prominent.

In (102) below, the attitude verb in the first sentence may introduce a context shift operator or logophoric operator that licenses the attitude holder as a perspective holder for *come*. However, the syntactic scope of this operator will not extend to the second utterance.

Therefore, the logophoric and indexical approaches predict that perspectival expressions in the second utterance should not be able to be interpreted relative to the perspective of the attitude holder (Captain Hastings).

102. *Hercules Poirot, Captain Hastings, and Miss Lemon are in Poirot's London flat, discussing a recent murder at the country estate of Hunterbury. While Hastings and Poirot were at Hunterbury investigating the first murder, another murder occurred. The murderer had placed poison in her own tea cup while everyone was present. After creating a diversion to lure everyone out of the room, the murderer went back into the drawing room to switch her cup with that of her victim. On Poirot's instructions, Captain Hastings hid in the curtains of a room that he thought was the drawing room in order to prevent the murder. Because Hastings was unfamiliar with the large house, he accidentally hid in the parlor instead of the drawing room, and the murderer's plot succeeded. Poirot is relating this to Miss Lemon.*

Poirot: Captain Hastings<sub>i</sub> thought that he<sub>i</sub> was in the drawing room<sub>j</sub>, but he<sub>i</sub> was actually in the parlor. When the murderer<sub>k</sub> came to switch the tea cups, there was no one there<sub>j</sub> to stop her<sub>k</sub>.

This is not the case, however. The destination of motion in the second sentence of (102) is where Captain Hastings imagined himself to be (the drawing room), not his actual location (the parlor) or his current location (Poirot's flat in London). This shows that perspectival expressions can be anchored to perspectives introduced in previous utterances.

This falls out naturally under the perspective-anaphoric account, since once perspectives are introduced into the discourse context, they continue to be accessible. It is much harder to explain in the indexical or logophoric approaches, which predict that the accessibility of perspectives is governed by the syntactic scope of operators. The ability of *come* to refer relative to a perspective introduced in a preceding utterance is therefore evidence in support of an anaphoric analysis of *come*.

### 3.4.7 The hypothesis space for American English *come*

In Section 3.3, I laid out a set of contrasting predictions made by four approaches to the semantics of perspectival predictions. In this section, I applied these diagnostics to one perspectival expression, American English *come*. The evidence aligns most closely with the predictions of the anaphoric account.

The fact that *come* allows a restricted but non-singleton set of perspective holders is challenging for the two variants of a lexical encoding approach that I have sketched, since it means the perspective holder can be neither existentially bound nor directly stipulated in lexical semantics.

In order to pursue an indexical approach to *come*, there are several challenges that would need to be overcome. One of these is the fact that *come* allows its perspective holder to covary with quantifiers, contrary to the attested non-shiftability of indexicals in such environments. Another significant challenge is the fact that *come* receives non-speaker-oriented readings outside of finite CPs. The absence of Shift Together behavior within finite CPs is also unexpected under an indexical analysis.

The logophoric binding approach also faces challenges. To pursue a logophoric binding analysis of *come*, it is necessary to account for the fact that *come* allows perspective shift outside of XPs with subjects and across utterance boundaries.

These findings are summarized in Table 3.9.

**Table 3.9.** Predictions by account compared to *come* data

	Lex.	Index.	Log.	Anaph.	<i>come</i>
Singleton perspective set	✓	X	X	X	X
Restricted perspective set	X	✓	✓	✓	✓
Obligatory <i>de se</i> anchoring	◊	✓	✓	◊	✓
Shift outside finite CP	✓	X	✓	✓	✓
Shift outside XP with subj.	✓	X	X	✓	✓
Shift Together effects	X	✓	✓	◊	X
Anchoring across utterances	X	X	X	✓	✓
Covariation in quant. contexts	X	X	✓	✓	✓

### 3.4.8 Cross-linguistic variation in *come*, revisited

Although I have argued that the behavior of *come* in American English is most consistent with a perspective-anaphoric analysis, the behavior of *come* equivalents in other languages may differ. My goal is not to argue that perspective-anaphoricity is the only mechanism that natural language provides for encoding perspective, or even the only mechanism used by perspectival motion verbs.

In fact, there is already some evidence that perspectival motion verbs in other languages encode the perspective holder differently. Charnavel (2018) argues that the perspectival motion verb *venir* ‘come’ is sensitive to logophoric binding. Her argument takes a different form from the one that I have presented for American English *come*, resting on Shift Together interactions between *venir* and exempt anaphors.

She presents evidence that while *venir* ‘come’ can ordinarily refer relative to either the speaker or an attitude holder, as shown in (103), if there is an exempt anaphor present, it must be anchored to the same perspective as the exempt anaphor, as shown by the infelicity of (104a).

#### 103. No exempt anaphor:

- (a) *The speaker is in Lyon. Claire’s son is in Paris.*

[Le fils de Claire]<sub>i</sub> craint que la pluie n’empêche son<sub>i</sub> fils de venir à Lyon.

‘[Claire’s son]<sub>i</sub> is afraid that the rain prevents his son from coming to Lyon.’

- (b) *The speaker is in Paris. Claire’s son is in Lyon.*

[Le fils de Claire]<sub>i</sub> craint que la pluie n’empêche son<sub>i</sub> fils de venir à Lyon.

‘[Claire’s son]<sub>i</sub> is afraid that the rain prevents his son from coming to Lyon.’

#### 104. Exempt anaphor:

- (a) *The speaker is in Lyon. Claire’s son is in Paris.*

# [Le fils de Claire]<sub>i</sub> craint que la pluie n’empêche son<sub>i</sub> propre fils de venir à Lyon.

‘[Claire’s son]<sub>i</sub> is afraid that the rain prevents his own son from coming to Lyon.’

- (b) *The speaker is in Paris. Claire's son is in Lyon.*

[Le fils de Claire]<sub>i</sub> craint que la pluie n'empêche son<sub>i</sub> propre fils de venir à Lyon.

'[Claire's son]<sub>i</sub> is afraid that the rain prevents his own son from coming to Lyon.'

(Charnavel, 2018)

This is unexpected in the Barlew (2017)-style perspective-anaphoric account that I have sketched, which does not predict obligatory Shift Together effects between perspectival expressions.

This suggests that there are at least two strategies for encoding the perspective holder of perspectival motion verbs that are used cross-linguistically. Further work may also reveal cases of lexically and indexically encoded perspectival motion verbs.

A related question of interest is whether there is a relation between the narrowness of a perspectival motion verb's licensing conditions and the encoding of its perspectival component. In Chapter 2, I discussed cross-linguistic variation in three aspects of the semantics of perspectival motion verbs: the set of perspective holders, the set of anchoring relations, and the evaluation time. American English *come* was among the least restrictive in all three of these dimensions, allowing a large set of perspective holders, multiple anchoring relations, and two evaluation times.

Is the fact that American English *come* is comparatively less restricted related to the finding that it is best analyzed as perspective-anaphoric? On the one hand, this seems possible: it is easy to introduce new perspective holders in an anaphoric approach, since any discourse-given perspective is available. This makes it simple to capture the fact that prominent individuals in a discourse can serve as perspective holders, even if they are not conversation participants.<sup>20</sup>

By contrast, the simplest indexical approach would allow only the speaker and listener, since those are the two individuals tracked in the context parameter. To allow other perspective holders, an indexical approach must either enrich the context parameter with new fields or

---

<sup>20</sup>This is also true of a logophoric binding approach, since the value of the logophoric pronoun that binds the perspective variable is determined by the discourse context.

employ context shift. This is not an insurmountable difficulty, since either of these options is available, but it suggests that all else being equal, we might expect a smaller set of perspective holders for indexical perspectival expressions.<sup>21</sup>

We can also approach the question from the other direction. As discussed in Chapter 2, perspectival motion verbs in some languages allow only the speaker to serve as the perspective holder. Does this indicate that these languages are more likely to be lexical or indexical, rather than logophoric or anaphoric?

This turns out to be a very tricky question to answer, for this reason: the more restricted the licensing of the motion verb is, the more difficult it becomes to probe its behavior using the proposed set of diagnostics. To see why this is the case, consider the diagnostic related to the syntactic domain of perspective shift. If we wanted to test whether the equivalent of *come* in a language like Shibe, which allows only the speaker's utterance time perspective, allows perspective shift outside of finite CP, we would need set up an example with two perspectives. But we already know that Shibe allows only one perspective; therefore, we know that this example will be ungrammatical.

Can we conclude from this that Shibe *come* is therefore not anaphoric or logophoric? Not necessarily, since we have no way of finding the source of the ungrammaticality. Perhaps Shibe *come* only allows the speaker's perspective because the lexical semantics of the verb directly stipulate the perspective holder. Or perhaps Shibe *come* only allows the speaker's perspective because the perspectival component is logophorically bound (say, by a top-level speaker-anchored speech act projection, as in the obligatory logophoric binding variant I sketched), but the language severely restricts the projection of logophoric operators. It may be impossible to tell.

To conclude, the existence of such well-attested cross-linguistic variation in the semantics of perspectival motion verbs suggests that the argument that I have made for American English *come* may not generalize across languages. I hope that the set of diagnostics that I have proposed will be useful in future cross-linguistic work exploring the semantics of

---

<sup>21</sup>Especially since not all languages allow (non-perspectival) indexical shift.

perspectival motion verbs whose licensing conditions differ from those of American English *come*. However, I acknowledge that in cases where the set of perspective holders is very restricted, it becomes difficult to apply the diagnostics that probe perspective shift.

### 3.4.9 Diagnostic challenges

As discussed above, one challenge scenario for diagnosing perspective is when the set of perspective holders is very small. However, this exploration of the behavior of American English *come* illustrates some of other challenges inherent in identifying the right analysis of a novel perspective-sensitive expression.

One finding that has emerged from the discussion of the different ways of encoding perspective is that the licensing environments for the various approaches overlap to a great extent. The behavior of indexical and logophoric expressions is predicted to be identical except in two environments: XPs with subjects that are not finite CPs, and quantificational binding environments. Similarly, the behavior of logophoric and perspective-anaphoric expressions is predicted to be identical except when the discourse context makes a perspective prominent enough that it can license perspectival expressions across syntactic domain boundaries. Without applying diagnostics that probe these relatively rare environments, it is very hard to distinguish among these classes of context sensitivity.

The critical examples can be challenging to construct, particularly the ones needed to test Shift Together effects and the availability of perspectives outside XPs with subjects. For certain categories of perspectival expressions, it may be very difficult to design examples in which multiple instances co-occur in the same syntactic domain; or examples that occur outside of XPs with subjects, making it challenging to pinpoint a unique correct analysis.

In many ways, *come* is one of the easiest expressions to test, in that it occurs with high frequency in a variety of environments, and judgments about its felicity tend to be robust. In the next chapter, I discuss a perspectival expression that is more challenging because there is significant interspeaker variation in judgments about its felicity in certain contexts.

### 3.5 Summary

In this chapter, I have provided a working definition of perspectival expressions and delved into one of the crucial questions about their semantics: how the perspective holder is encoded. I have outlined a number of ways of representing the perspectival component of expressions like *come* using, for the most part, mechanisms that are independently motivated. While the quotational account reduces to the anaphoric account, each of the four other accounts defines a possible family of perspectival expressions: lexically-specified perspectival expressions, perspectival indexicals, logophoric perspectival expressions, and perspective-anaphoric expressions.

As I have outlined in Section 3.3, these accounts make different predictions about the behavior of perspectival expressions in various environments. While all accounts are capable of deriving the obligatory *de se* interpretation of perspectival expressions, they differ in their predictions about the ability of the perspective holder to covary in quantificational binding contexts; the availability of perspective shift outside finite CPs, outside spellout domains, and across utterance boundaries; and the obligatoriness of Shift Together effects. These diagnostics can be applied in order to determine which of the three ways of encoding perspective is most appropriate for any given perspectival expression.

In Section 3.4, I applied the set of diagnostics to one canonical perspectival expression: American English *come*. I concluded that *come* was best analyzed as perspective-anaphoric, following Barlew (2017)'s previous proposal. These findings are not necessarily expected to hold across languages, given the well-documented cross-linguistic variation in the semantics of perspectival motion verbs. However, by providing a standard set of diagnostics for perspective encoding, I hope that the findings of this chapter will be useful in further cross-linguistic exploration of perspectival motion verbs, as well as other classes of perspectival expressions.

In the next chapter, I turn to a second case study: non-utterance time readings of *tomorrow* in American English. Unlike *come*, this case study focuses on a context-sensitive expression that has not previously been analyzed as perspectival: American English *tomorrow*. Using the set of perspectival diagnostics as well as diagnostics for other forms of context sensitivity,

I develop an analysis of a perspectival variant of *tomorrow* that is available for some (but most likely, not all) American English speakers.

## CHAPTER 4

### EXPLORING TEMPORAL PERSPECTIVE

#### 4.1 Introduction

In Chapter 3, I proposed a set of diagnostics for probing how the perspective holder is encoded in the semantics of perspectival expressions. I demonstrated the application of these diagnostics using a canonical perspectival expression: American English *come*. However, when analyzing a context-sensitive expression, it is often not known whether the expression is perspectival.

In this chapter, I present a second case study that focuses on an expression not usually thought to be perspectival. I explore the context sensitivity of American English *tomorrow*, a temporal adverbial long analyzed as a pure indexical. I show that for one group of speakers, it behaves like a pure indexical, as it has been analyzed (Kamp, 1971; Kaplan, 1989). However, I also show that for a second group, *tomorrow* behaves like a perspectival expression, an observation likely overlooked both because of the interspeaker variation and because of the significant overlap in licensing conditions between pure indexicals and perspectival expressions.

Unlike in the case study of *come*, judgments about the behavior of *tomorrow* are not clear-cut or well-studied. In order to support my claim that *tomorrow* is sometimes perspectival, I present experimental evidence from environments where the prominent perspective is not an utterance-time perspective, and show that for some speakers, *tomorrow*'s behavior is not consistent with pure indexicality. I then turn to probing how the perspectival component of *tomorrow* should be analyzed, using the four perspective encoding approaches laid out in Chapter 3. I show that *tomorrow* belongs in the logophoric or anaphoric families of perspectival expressions and present a perspective-anaphoric semantics for *tomorrow*. I conclude with a discussion of the challenges of applying the remaining perspectival diagnostics.

The data I consider in this chapter are non-utterance time readings of *tomorrow*, as in (105). This sentence does not make sense if *tomorrow* is interpreted as the day after utterance time; if it is judged felicitous, *tomorrow* must refer to the day after the saying event.

105. % Last week, Jane said that she would order the cake tomorrow, but she didn't.

Non-utterance time readings of *tomorrow* like (105) receive mixed judgments from American English speakers. Some speakers judge them perfectly acceptable, while others reject them outright. For this reason, I explore the behavior of *tomorrow* experimentally. I build up an argument that *tomorrow* has a perspectival semantics through a series of experimental studies that narrow the hypothesis space for the semantics of *tomorrow*.

In the first half of this chapter I present experimental evidence in support of the claim that for some American English speakers, *tomorrow* is not a pure indexical. In Experiment 1, I show that non-utterance time readings are accepted by many, though not all, American English speakers. I consider two main possibilities: (1) that these readings arise through context shift and *tomorrow* refers relative to the context parameter; and (2) that they are anaphoric uses of *tomorrow*. In Experiment 2, I show that such readings arise outside of quoted contexts and cannot be analyzed as partial quotation.<sup>1</sup> In Experiments 3a, 3b, and 4, I show that such readings arise outside of two environments for which context shift has been proposed: Free Indirect Discourse and attitude reports.

In the second half of this chapter, I develop the case for a perspectival view of *tomorrow*. I present two quantificational binding tasks that show that *tomorrow* is not anaphoric to just any discourse-given time, and that the acceptability of *tomorrow* in binding environments improves when the context allows perspectives to covary with the quantifier. On the basis of this data, I argue that for some American English speakers, the temporal reference of *tomorrow* is mediated by a perspective, and that this behavior has been overlooked because the temporal location of the perspective holder is almost always identical to the time index of the context parameter.

---

<sup>1</sup>This experiment concerns the possibility of a quotational analysis of *tomorrow*, not the quotational approach to perspectival expressions that I sketched in Chapter 3.

The experimental data is consistent with two of treatments of perspectival expressions proposed in Chapter 3: the logophoric approach and the perspective-anaphoric approach. Having narrowed the hypothesis space to these two families of approaches, I turn to deciding between them. I present some judgments that tentatively favor the perspective-anaphoric approach, and develop a perspective-anaphoric analysis of *tomorrow* that draws on Barlew (2017)'s treatment of perspectival motion verbs. However, given the attested interspeaker variation in judgments of *tomorrow*, the judgments that I present should be seen as tentative pending future experimental testing. I argue that regardless of whether the logophoric or perspective-anaphoric approach is adopted, the semantics of *tomorrow* (for speakers who accept the non-UT readings) must involve a perspective variable that is not bound by the context parameter.

Exploring the semantics of *tomorrow* at a fine-grained level is useful as a case study of the challenges involved in distinguishing among classes of context sensitivity. Moreover, the work presented in this chapter highlights the importance of careful diagnostics for context sensitivity and perspective sensitivity. Because the licensing conditions for different classes of context-sensitive expressions overlap in all but a few environments, perspective sensitivity may be hiding in plain sight among even the best-studied classes of context-sensitive expressions.

## 4.2 The semantics of *tomorrow*

In most contexts, *tomorrow* picks out the day after the day on which it is spoken. In (106), for example, *tomorrow* refers to Wednesday, the day after the conversation.

106. *It is Tuesday. Miss Marple is speaking to Cherry Baker.*

Marple: My nephew Raymond is arriving tomorrow at 3pm.

Such readings are called **utterance time readings** because they refer relative to the time of utterance. All American English speakers agree that utterance time readings of *tomorrow* are acceptable.

It is also possible to design contexts in which *tomorrow* refers relative to a time other than utterance time. An example **non-utterance time reading** is shown in (107), where *tomorrow* must refer relative to the time of the embedded speech event, rather than the current conversation.

107. Last week, Jane said that she would order the cake tomorrow, but she didn't.

Non-utterance time readings of *tomorrow* like (107) receive mixed judgments. While some speakers judge (107) to be perfectly felicitous, other speakers reject it as infelicitous, because they interpret *tomorrow* as referring to the day after the matrix sentence is uttered (in which case, it is not yet known whether Jane will order the cake).

#### 4.2.1 Naturally-occurring non-utterance time readings of *tomorrow*

Despite the variable judgments of (107), non-utterance time readings are well-attested in American English corpus data. For instance, in (108), *tomorrow* is interpreted relative to the previous day (the day of Bush's speech) and picks out the day of the conversation between Krause and Wiethop (rather than the following day).

108. *The previous day, President Bush had given a speech in Fulton, but Wiethop decided not to cover it on the front page. In the morning, however, the president declared war on Iraq, making his speech the previous day more newsworthy.*

Krause: Also listening last night, Dave Wiethop, news editor at the Fulton Sun. He listened closely, but decided not to lead this morning's paper with the President's speech. Wiethop: I played Bush pretty low because like I said, we didn't hear anything that we hadn't heard before. If he said we were going to bomb Baghdad tomorrow morning, I think that might have been my lead story there. (Davies, 2008)

Because of the tense/aspect that Wiethop uses in his counterfactual speech report, it is clear that it is an indirect speech report rather than (imagined) quotation. Therefore *tomorrow* is not directly quoted.

Shipping complaint forums are particularly rich sources of these uses of *tomorrow*, since the timeline is usually described carefully. (109) comes from a dissatisfied UPS customer.

109. He then said they would reattempt delivery on Monday. Monday came and I tracked the package and same message as before, “The receiving business was closed. / Your delivery has been rescheduled for the next business day.” Again, my business was not closed.

I called UPS and after being on hold for almost an hour a supervisor told me he couldn’t get a hold of the local service center but would call me back in an hour. Almost 4 hours later, I receive a call from the local center that the package was never attempted to be delivered, and that this looked strange. The agent said he would locate it and make sure it went out tomorrow (Tuesday).

Tuesday morning I track the package and now it says that I am going to pick it up, some 30+ miles from my home.

In (109), *tomorrow* refers relative to the day of the speaker’s conversation with the agent at the local call center (Monday), not relative to the current day (Tuesday or later). As in the previous example, *tomorrow* cannot be interpreted as a part of a quotation, since the pronouns are inconsistent with a direct report of the UPS agent’s speech.

Thus, although non-utterance time readings of *tomorrow* are judged unacceptable by many American English speakers, they are also judged acceptable by many, and occur fairly frequently.<sup>2</sup> This is puzzling, since *tomorrow* has long been thought to be a pure indexical: an expression whose meaning is determined by the context of utterance alone.

#### 4.2.2 Context-sensitive temporal modifiers

There are two ways that temporal modifiers can be sensitive to context. Temporal modifiers that are sensitive to the context of utterance refer **indexically**: their reference time comes from the context parameter. In his foundational work on indexicality, Kaplan (1989) includes three English temporal modifiers in his list of indexicals: *tomorrow*, *now*, and *yes-*

---

<sup>2</sup>Note also that they are not limited to oral contexts, where they might be interpreted as speech errors; (109) is from a written complaint, whose author is very deliberate in trying to communicate the timeline of events clearly. Moreover, the author’s parenthesizing of the day of the week suggests that they were aware of a potential ambiguity in the interpretation of *tomorrow*.

*terday*. He points out that the meaning of (110a) is different depending on the day on which it is said, because *yesterday* will pick out a different date depending on when it is spoken.

110. (a) I was insulted yesterday.

(b) I was insulted on March 30th, 2020.

This is in contrast to temporal modifiers that are not context-sensitive, such as actual dates; the date of the insult described by (110b) is the same no matter when it is spoken.

Temporal modifiers can also be sensitive to the discourse context in which they occur. Temporal modifiers that refer **anaphorically** select a discourse-given time as their reference time. For instance, *the next day* refers to the day after a prominent time in the discourse context, which does not need to be the utterance time. In (111a), *the next day* refers relative to the time of the shopping event. By contrast, (111b), where *the next day* is meant to refer relative to utterance time, is marked.

111. (a) Jane is about to go shopping, and she will cook the next day.

(b) ?Jane will cook the next day.

Expressions may have both indexical and anaphoric uses. Expressions that have only indexical uses are known as **pure indexicals**. These expressions refer relative to times other than utterance time only if the context parameter is manipulated in some way (such as in cases of indexical shift).

#### 112. Kaplanian definition of (pure) indexicality:

An indexical is an expression whose interpretation relies on the context of utterance and not on the circumstances of evaluation. (Kaplan, 1989)

I use the term *pure indexical* to refer to indexicals that do not have anaphoric uses and the term *pure anaphoric expressions* to refer to expressions that have only anaphoric uses.<sup>3</sup> There are also expressions that can have both indexical and anaphoric uses (Altshuler, 2016). I refer to these expressions simply as *indexicals*.

---

<sup>3</sup>Such as *the next day*, as demonstrated by the contrast in (111).

113. Temporal modifier terminology:

- (a) **Pure indexical**: an expression that must refer relative to the utterance context
- (b) **Indexical expression**: an expression that can refer relative to the utterance context
- (c) **Pure anaphoric expression**: an expression that must refer relative to a discourse-given reference point

Since Kaplan (1989), *tomorrow* has been treated as a pure indexical and thought to refer strictly relative to the temporal field of the context parameter. This makes the existence of non-utterance time uses of *tomorrow* surprising. However, there are a handful of environments in which pure indexicals receive non-utterance time readings, so the existence of non-utterance time readings is not necessarily conclusive evidence against a pure indexical view. In the next section, I lay out the space of analyses for *tomorrow* for speakers who do accept non-utterance time readings.

#### 4.2.3 The space of analyses for non-utterance time *tomorrow*

Although non-utterance time readings are unexpected if *tomorrow* is a pure indexical, there are some circumstances in which pure indexicals receive shifted readings. First, in some languages, indexicals may receive shifted readings when embedded in speech or attitude reports. These languages have been proposed to contain context shift operators that can manipulate the context parameter. Although indexical shift has not been attested in American English, perhaps it is available for the speakers who accept non-utterance time readings of *tomorrow*.

Second, indexicals in quotation are always interpreted relative to the original context of utterance. Although the corpus examples discussed above do not appear at first glance to be quoted, perhaps there is a way of analyzing non-utterance time readings of *tomorrow* as instances of quotation. This is appealing because the mechanisms used in quotation should be available for American English speakers; the interspeaker variability may be in accessing a quoted reading of *tomorrow*.

Third, indexicals are also known to receive shifted interpretations in certain perspective shift environments, most notably, in Free Indirect Discourse. If non-utterance time readings of *tomorrow* are limited to these environments, then *tomorrow* may be a pure indexical even for speakers who allow such readings. Free Indirect Discourse interpretations are predicted to be available for all American English speakers; however, there could be variability in how likely speakers are to interpret a given piece of discourse as Free Indirect Discourse.

A fourth possibility is that *tomorrow* is not, in fact, a pure indexical (for the speakers who accept non-utterance time readings). Perhaps for these speakers, *tomorrow* belongs to another class of context-sensitive expressions, such as temporal anaphoric expressions. This involves positing a novel lexical entry for *tomorrow*.

In this section, I will sketch out the different analyses of non-utterance time *tomorrow* readings, highlighting the contrasting predictions that they make.

#### **4.2.3.1 Analysis 1: non-utterance time readings occur due to context shift in Free Indirect Discourse**

Free Indirect Discourse is a style of discourse that makes two perspectives prominent at once: the perspective of the narrator and the perspective of the protagonist. In Chapter 3, I discussed Free Indirect Discourse as an example of a perspective shift environment.<sup>4</sup> However, Free Indirect Discourse also affects non-perspectival context-sensitive expressions, such temporal indexicals.

In Free Indirect Discourse, tense and person pronouns are used relative to the narrator's perspective, while temporal and locative indexicals, expressives, and perspectival items are used relative to that of the protagonist (Banfield, 1982). This means that temporal indexicals generally receive non-utterance time readings in Free Indirect Discourse contexts. For instance, in (114), *tomorrow* is used relative to the perspective of the protagonist, and therefore refers to a past time.

---

<sup>4</sup>As discussed in Chapter 3, there are other similar perspective shift environments, such as Protagonist Projection. However, in Protagonist Projection, temporal indexicals are fixed to the narrator's perspective, so it would not give rise to non-utterance time readings of *tomorrow* (Stokke, 2013; Abrusán, 2018).

114. Tomorrow was Monday, Monday, the beginning of another school week! (Lawrence, 1920, cited in Banfield (1982))

A common analysis of Free Indirect Discourse is that it involves two different context parameters: an external context parameter  $C$ , representing the narrator's utterance situation, and an internal context parameter  $c$ , representing the protagonist's thought situation.<sup>5</sup> While in direct speech, only the external context is available, Free Indirect Discourse environments introduce an internal context. When there are two context parameters, shiftable indexicals (temporal and locative) refer relative to the internal context, while rigid indexicals (person) remain fixed to the external context.

Under this account, pure indexicals have their usual semantics, except that they refer relative to the internal context parameter rather than the external one, and therefore receive shifted interpretations: they are interpreted relative to the protagonist's speech or thought context rather than the matrix utterance context.

One possible explanation of non-utterance time readings of *tomorrow* is that they arise only in Free Indirect Discourse environments. Assuming a double context parameter analysis of Free Indirect Discourse like the one sketched above, non-utterance time readings would arise when *tomorrow* is interpreted relative to an internal context, and the semantics of *tomorrow* would be as in (115).

115. FID derivation of non-utterance time *tomorrow*:

$$[[\text{tomorrow}]]^{C,c,g} = \lambda Q. \lambda e. Q(e) \wedge \tau(e) \subset \iota t. \text{DAY-AFTER}(t, c_t)$$

where  $C$  is the external context parameter,  $c$  is the internal context parameter, and  $c_t$  is the temporal index of  $c$

This account is consistent with pure indexicality, since *tomorrow* would refer to a context parameter even when receiving protagonist-oriented interpretation.

116. Tomorrow was Monday.

---

<sup>5</sup>This analysis was proposed by Doron (1991) and adopted by Sharvit (2008) and Eckardt (2014), but see Schlenker (2004) and Maier (2015, 2017b) for other analyses.

- (a)  $[[\text{tomorrow}]]^{C,c,g} = \lambda Q. \lambda e. Q(e) \wedge \tau(e) \subset \iota t. \text{DAY-AFTER}(t, c_t)$
- (b)  $[[\text{Sarah was arriving}]]^{C,c,g} = \lambda e. \text{ARRIVE}(e) \wedge \text{AGENT}(e, s)$
- (c)  $[[\text{Sarah was arriving tomorrow}]]^{C,c,g} = \lambda e. \text{ARRIVE}(e) \wedge \text{AGENT}(e, s) \wedge \tau(e) \subset \iota t. \text{DAY-AFTER}(t, c_t)$

In order to test this hypothesis, we need to be able to distinguish Free Indirect Discourse from other environments, since this account predicts that non-utterance time readings of *tomorrow* should not occur outside of Free Indirect Discourse. The literature on Free Indirect Discourse is broad, and there is some disagreement on the scope of the phenomenon, but the following characteristics of Free Indirect Discourse have been proposed.

**Tense and aspect:** Generally relative to the narrator’s viewpoint. Past under past is used to refer to the protagonist’s temporal location (Sharvit, 2008); past perfect is used for the protagonist’s past (Eckardt, 2014).

**Embedding:** Although Free Indirect Discourse reports the thoughts of the protagonist, it is characteristically unembedded. Schlenker (2004) claims that Free Indirect Discourse is never preceded by a complementizer. Hinterwimmer (2017a) claims that Free Indirect Discourse is blocked in the complements of attitude verbs.

**Syntactic scope:** Hinterwimmer (2017a) provides evidence that Free Indirect Discourse always applies to a whole proposition.

**Person pronouns:** Person pronouns refer relative to the narrator’s point-of-view. For this reason, first-person pronouns do not typically appear in Free Indirect Discourse; the exception is when the protagonist is a past self of the narrator.

**Definite descriptions:** Definite descriptions are generally interpreted from the point-of-view of the subject in Free Indirect Discourse (Sharvit, 2008), which may be related to the fact that Free Indirect Discourse is obligatorily faithful to the style as well as the content of the protagonist’s thoughts. However, there is not uniform agreement on this point: while Schlenker (2004) claims that everything in Free Indirect Discourse is interpreted *de dicto* rather than *de re*, Doron (1991) claims that “the referential use of definite descriptions is

connected to the speaker, while the attributive use is connected to the subject” (Doron, 1991, p. 54).

A Free Indirect Discourse context-shift account of non-utterance time *tomorrow* makes several predictions. First, if Free Indirect Discourse has proposition-wide scope, *tomorrow* should pattern along with other expressions that shift in Free Indirect Discourse environments: if *tomorrow* is non-utterance time-interpreted, then expressives and definite descriptions should also be interpreted relative to the protagonist’s point-of-view. Second, non-utterance time interpretations of *tomorrow* should only occur in clauses where the tense and aspect use supports a Free Indirect Discourse interpretation.

#### 4.2.3.2 Analysis 2: non-utterance time readings occur due to context shift in attitude contexts

Context shift has also been proposed to occur under speech and attitude verb embedding. In a number of languages, there are so-called **shifty indexicals**, indexicals whose interpretation changes when they are embedded under speech or attitude verbs. In embedded contexts, these shifty indexicals can be interpreted relative to the embedded context rather than the matrix utterance context.

Although person indexicals are the most cross-linguistically common kind of shifty indexical (Deal, 2017), shifty temporal indexicals have also been documented. In the Korean example in (117), *nayil* ‘tomorrow’ can refer either to the day after utterance time (matrix interpretation), or the day after Mary’s speech act (shifted interpretation).

117. *It is January 8th.*

cinan	cwu-ey	Mary-ka	nwuka	nayil	ttenanta-ko	malhayss-ni?
last	week-in	Mary-NOM	who-NOM	tomorrow	leave-COMP	said-Q

‘Who did Mary say a week ago would leave on January 2nd/9th?’ (Park, 2014a)

As discussed in Chapter 3, one main analysis of indexical shift posits a covert syntactic operator that shifts the context parameter in attitude reports (Anand and Nevins, 2004; Deal, 2014, 2017). Other accounts propose that shifted environments involve quantification over intensional contexts (Schlenker, 2003; Sundaresan, 2020a). Under any of the proposed

analyses, shifty indexicals are still considered pure indexicals, because they are evaluated relative to a context parameter.

118. Indexical shift derivation of non-utterance time *tomorrow*:

$$[[\text{Mary said OP}_t \text{ John would come tomorrow}]]^{c,g} = \exists e.\text{SAY}(e) \wedge \\ \text{AGENT}(e, \text{Mary}) \wedge \text{THEME}(e, [[\text{John would come tomorrow}]]^{c[\text{TIME}(e) \rightarrow \text{TIME}], g})$$

There are two diagnostics for indexical shift that arise from the context shift operator view of the phenomenon: the existence of Shift Together effects for shifty indexicals; and the unacceptability of shifty indexicals in quantificational binding contexts.

Because the shift operator overwrites the matrix context parameter, if one indexical shifts, all others in the same syntactic domain ought to shift as well,<sup>6</sup> since the matrix context parameter is no longer accessible.<sup>7</sup> This springs from a key property of pure indexicals: because they all refer relative to the context parameter, they receive uniform interpretations.

In addition, because shifty indexicals are still pure indexicals, they are not licit in quantificational binding contexts. In (119), for example, the indexical *cikum* ‘now’, despite being shiftable, is infelicitous because of the quantification (Park, 2014a).<sup>8</sup>

119. \*Obama-ka malhal ttyaymyun manhun salamtul-i cikum  
 Obama-NOM speaks when many people-NOM now  
 pakswuchinta  
 clap

Intended: ‘When Obama speaks, many people clap now.’ (Park, 2014a)

<sup>6</sup>At least, if they are the same type; Anand and Nevins (2004) proposes separate shift operators for time and person parameters.

<sup>7</sup>The data on Shift Together effects in some languages is contested; see Deal (2017) and Sundaresan (2020a) for more discussion.

<sup>8</sup>The indexical status of *cikum* ‘now’ is controversial; although Park (2014a) treats it as a shifty indexical, Lee and Choi (2009) analyzes it as anaphoric. However, Park (2014a) reports the same quantificational binding facts for other temporal indexicals in Korean as well, including *ece* ‘yesterday’, *onul* ‘today’, and *nayil* ‘tomorrow’, though she does not include examples.

Similar quantificational binding data have been reported for shifty indexicals in other languages as well (Deal, 2017). Thus, shifty indexicals behave like pure indexicals except when they are embedded under an attitude predicate.

One possible explanation for non-utterance time interpretations of *tomorrow* in English is that *tomorrow* is a shifty indexical and the readings arise from context shift. This account, like the Free Indirect Discourse account, is consistent with pure indexicality, given that the dominant analyses of indexical shift posit that the original context parameter is overwritten by a context parameter representing the embedded context.

An indexical shift account makes three predictions. First, non-utterance time interpretations should only arise when *tomorrow* is embedded under a speech or attitude verb. Second, like other indexicals, *tomorrow* should be infelicitous in quantificational binding environments. Lastly, we expect uniform behavior across temporal indexicals in the same clause (with the caveat that some indexicals seem to be rigidly referent to the matrix context); we would expect other temporal adverbials, like *yesterday*, to behave similarly to *tomorrow*.

#### 4.2.3.3 Analysis 3: non-utterance time readings occur due to (partial) quotation

Another environment in which pure indexicals receive non-utterance time interpretations is when they are quoted. When an entire utterance is quoted, the indexicals are interpreted relative to the original context of saying, rather than the matrix utterance context. Example (107), repeated as (120) below, is not an instance of full quotation, since the quotee is referred to in the third person. However, it might be an instance of mixed quotation: perhaps *tomorrow* alone is quoted.

120. Last week, Jane said that she would order the cake tomorrow, but she didn't.

**Mixed quotation** is the use of someone else's speech as part of an indirect speech report (Davidson, 1979; Maier, 2017a). For instance, the description of Harriet Vane in (121) is quoted as part of a larger indirect speech report.

121. Three days later, while reading in the morning paper that among the guests at a literary luncheon-party had been seen ‘Miss Harriet Vane, the well-known detective authoress,’ she was interrupted by the telephone. (Sayers, 1935)

Mixed quotation is interesting because the quoted content is incorporated grammatically into the broader syntactico-semantic context of the utterance. This means that the semantic contribution of the quote must be incorporated compositionally into the rest of the sentence.

As I discussed in Chapter 3, Maier (2015) proposes a two-dimensional analysis of mixed quotation where the quoted material has both a use and a mention component. The use component is a property variable  $P$ , ranging over semantic objects of the type that matches the syntactic category of the quoted phrase. The mention component states that some contextually determined individual used the linguistic material within the quotation marks to refer to  $P$ .

122. Maier (2015) quotation:

$[[“A”]] = \text{whatever a contextually salient individual } x \text{ used } A \text{ to mean.}$

123. Ann said that this music was “not mah cup ’o tea.”

Ann used ‘not mah cup ’o tea’ to refer to the property  $P \wedge$  she said that this music was  $P$ .

Because the identity of individual is contextually determined, the quoted content is essentially anaphoric to the discourse context.

A quotational approach to the non-utterance time readings of *tomorrow* would posit that *tomorrow* only receives non-utterance time readings when it is quoted.

124. Quoted derivation of non-utterance time *tomorrow*:

$[[“tomorrow”]]^{C,g} = \text{whatever a contextually salient individual } x \text{ used } tomorrow \text{ to mean.}$

An immediate objection is that this semantics does not derive the fact that *tomorrow* is used to refer to the day following some other day. However, if most individuals have a

standard pure indexical semantics for (unquoted) *tomorrow*, then they will use it to refer to the day following the day on which they are speaking.

The hallmark of quotation is that quotes are faithful to the exact words of the quotee, including any stylistic features of the speaker (as in (122)). Therefore a quoted account of non-utterance time *tomorrow* predicts that non-utterance time readings should only arise in contexts where the quotee actually uses the word *tomorrow*. If the quotee refers to the same day, but uses a different expression, such as *in the morning* or *in 24 hours*, *tomorrow* should not receive a non-utterance time reading in an indirect speech report describing the quotee's utterance.

#### **4.2.3.4 Analysis 4: non-utterance time readings occur because *tomorrow* is anaphoric**

The previous three accounts are all consistent with a pure indexical semantics for *tomorrow*. They posit that non-utterance time readings arise due to something special about the environment in which *tomorrow* occurs. Another possibility, however, is that non-utterance time readings arise because *tomorrow* itself is special. Perhaps *tomorrow* is not a pure indexical all of the time, for all speakers.

What kind of lexical semantics for *tomorrow* would predict non-utterance time readings? One possibility is that *tomorrow* can refer anaphorically as well as indexically. Under this account, non-utterance time readings would arise when *tomorrow* is anaphoric to a time other than utterance time.

125. Anaphoric semantics for *tomorrow*:

$$[[\text{tomorrow}]]^{c,g} = \lambda Q. \lambda e. Q(e) \wedge \tau(e) \subset \iota t. \text{DAY-AFTER}(t, t')$$

where  $t'$  is a prominent time in the Common Ground

Under an anaphoric account of *tomorrow*, we would expect *tomorrow* to be licit in quantificational binding contexts. Although pure indexicals resist quantificational binding, the referent for anaphoric expressions can covary with the quantifier in contexts that involve quantification over times. Because *here* is a pure indexical, (126) can only mean that for each place that the speaker visits, it rains at the utterance location, which is nonsensical.

By contrast, in (127), the referent for the anaphoric expression *the next day* can covary with the quantification over times.<sup>9</sup>

126. # Everywhere I visit, it rains here.

127. Whenever I wash my car, it rains the next day. (Deal, 2014)

In order to be able to apply the quantificational binding diagnostic, however, we have to understand the referent of the expression of interest. As work on *now* highlights, temporal adverbials with anaphoric uses are not always anaphoric to any prominent time in the discourse context; they may be anaphoric to something more complex, such as a result state (Altshuler and Stojnić, 2015; Altshuler, 2016, 2017).

When this is the case, anaphoric expressions may appear indexical because their referent does not covary with the quantifier in the particular quantificational binding environments tested. For instance, *now* is infelicitous in (128) but felicitous in (129), where for each artist that instantiates the quantifier, there is a different prominent result state.

128. #Whenever I am in North Hadley, I'm happy now. (Altshuler, 2017)

129. Every artist reaches that point, where he is now ready for his masterpiece. (Altshuler and Stojnić, 2015)

Saying that non-utterance time interpretations of *tomorrow* are anaphoric therefore is only a partial account, since in order to understand such readings, we also need to know what kind of object *tomorrow* can reference anaphorically. For now, I set aside this question and sketch out the predictions that all anaphoric accounts make.

First, all anaphoric accounts predict that non-utterance time readings could arise even when there is no manipulation of the context parameter. Because of this, an embedding attitude predicate is not predicted to be necessary.

Second, under an anaphoric account, *tomorrow* will not necessarily pattern together with other context-sensitive expressions in the same clause, since anaphoric reference is not as

---

<sup>9</sup>This could be viewed as quantification over situations rather than times; nothing in my analysis rests on this distinction.

constrained as indexical reference. We might expect *tomorrow* to behave similarly to *the next day*, since *the next day* is also anaphoric; however, if *tomorrow* is anaphoric to a different kind of referent than *the next day*, they may behave differently.

Lastly, under the other three accounts, we would expect to find non-utterance time readings of *yesterday* as well as *tomorrow*. This is because those accounts do not posit a modified semantics for *tomorrow*; they posit that non-utterance time readings of *tomorrow* arise in context shift environments of one kind or another. Each of the proposed environments (Free Indirect Discourse, indexical shift, and quotation) should affect the interpretation of all temporal indexicals in their scope, including *yesterday*. By contrast, under the anaphoric approach, it is the semantics of *tomorrow* that give rise to non-utterance time readings, rather than the environments in which it occurs. Therefore we would not necessarily expect other temporal indexicals to behave similarly.

#### 4.2.3.5 Evaluating analyses of non-utterance time *tomorrow*

I have outlined four analyses for non-utterance time *tomorrow*: three that are indexical and one that is anaphoric. These accounts make different predictions about the availability of non-utterance time readings, summarized in Figure 4.1.

Using experimental methods, we can evaluate these accounts by testing the behavior of *tomorrow* in the environments for which their predictions differ. This would be challenging to do if we relied on corpus data or native speaker intuitions, since judgments about the felicity of non-utterance time readings varies among speakers. Experimental methods provide a way of quantifying both the variability in judgments and the relative acceptability of non-utterance time readings in different environments.

**Figure 4.1.** Predicted availability of non-UT readings of *tomorrow* by analysis

Hypothesis	Quant. binding	UT epithet	Non-verbatim report	Unembedded	Attitude report
FID	!	!	!	★	★
Shifty indexical	!	★	★	!	★
Quotation	★	★	!	★	★
Anaphoric	★	★	★	★	★

★ = predicted felicity, ! = predicted infelicity

In Sections 4.3 - 4.5, I present a series of experiments evaluating the context shift accounts. Experiment 1 establishes a baseline, and Experiments 2, 3a, 3b, and 4 test the predictions of the pure indexical-consistent accounts by manipulating the environment in which *tomorrow* occurs. Experiment 2 considers the quotation proposal; Experiments 3a and 3b explore the Free Indirect Discourse account; and Experiment 4 tests the indexical shift approach.<sup>10</sup>

### 4.3 Experiment 1: establishing a baseline

In Experiment 1,<sup>11</sup> I establish a baseline of acceptability for non-utterance time interpretations of *tomorrow*. I measure their acceptability with a task that avoids explicit grammaticality judgments, since non-utterance time readings may violate the usage rules for *tomorrow* that speakers have been taught. Instead of asking participants to rate the grammaticality of sentences, I ask them to rate how well a sentence fits a context depicted by a comic strip. All sentences presented to participants (including filler items) are grammatical, but not all of them are felicitous in the given context.

Experiment 1 uses an environment that is predicted to allow non-utterance time readings under all of the accounts discussed above: embedding under a speech verb.

#### 4.3.1 Method

Data on the acceptability of non-utterance time interpretations of *tomorrow* in American English was collected through a comic-captioning task where participants rated captions for comic strips on a 7-point Likert scale (where 7 indicates high naturalness). Ratings for *tomorrow* were compared against the anaphoric expression *the next day* and a factually correct and factually incorrect baseline; captions that are accurate according to the comic strip depiction are expected to receive higher ratings than ones which are false. If speakers of American English accept non-utterance time readings of *tomorrow*, then participants should rate the *tomorrow* items higher than the false baseline and close to *the next day*.

---

<sup>10</sup>All materials are publicly available through the Open Science Foundation: <https://osf.io/8x9w2/>.

<sup>11</sup>The experiments are presented in the order that makes for the clearest argument; their chronological order was as follows: Experiment 1, Experiment 3b, Experiment 4, Experiment 2, Experiment 3a.

**Table 4.1.** Experiment 1 predictions

Condition	Truth	Predicted ratings
False control	False	Low
True control	True	High
<i>the next day tomorrow</i>	True True if participant allows non-UT reading False otherwise	High High Low

#### 4.3.1.1 Participants

For Experiment 1, 126 participants were recruited through Amazon’s Mechanical Turk platform. 4 participants were excluded because English was not the language of their childhood household; 50 participants were removed because their mean ratings for the good baseline condition were not at least 1 point higher than for the bad baseline.<sup>12</sup> This left 72 monolingual English participants residing in the US. These exclusion criteria were preregistered through the Open Science Foundation.<sup>13</sup>

#### 4.3.1.2 Materials

20 critical items were developed and distributed across four Latin square lists. Each list was combined with the same set of 10 fillers. Each item included a three-panel comic strip and a sentence below it. The same comic strips were used for all experiments, though the text was edited between experiments.<sup>14</sup>

In the first panel of each comic, two characters are shown, and one of them says that they will do something the following day. Nothing happens in the second panel, indicating that they did not follow through. In the third panel, the other character expresses frustration with the first character’s lack of action (Figure 4.2).

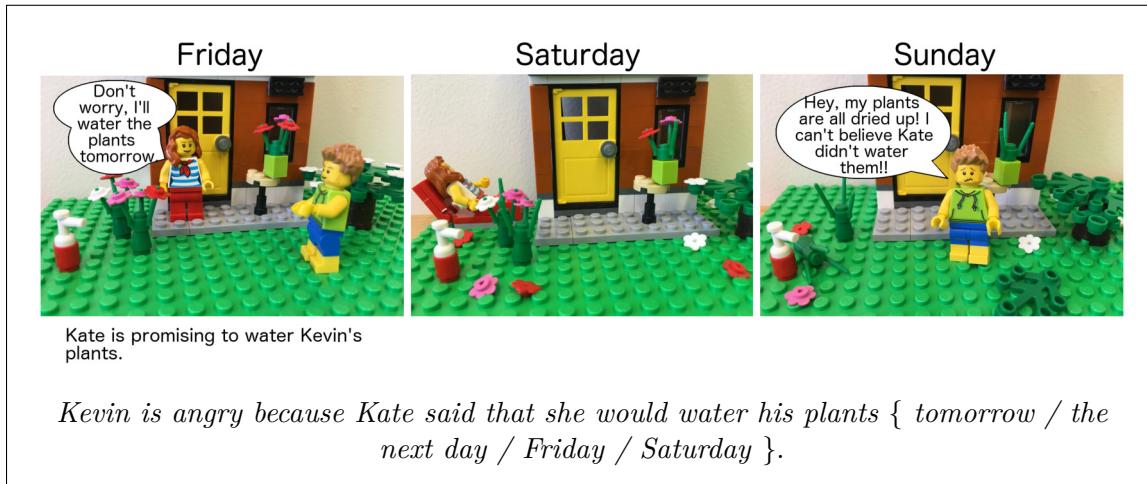
---

<sup>12</sup>The high rate of failure of the baseline measure is likely due to the fact that none of the training items required participants to read the day-of-week labels for the comics. A training item highlighting the day-of-week labels was added in subsequent experiments, reducing the participant removal rate.

<sup>13</sup>The experimental designs, participant exclusion criteria, and planned analyses for all experiments in this dissertation were preregistered, except where post-hoc analyses were motivated by the unpredicted findings. The post-hoc nature of these analyses is noted in their discussion.

<sup>14</sup>The items used in the experiments reported in this Chapter can be found in Appendices F-L.

**Figure 4.2.** Experiment 1 example stimulus

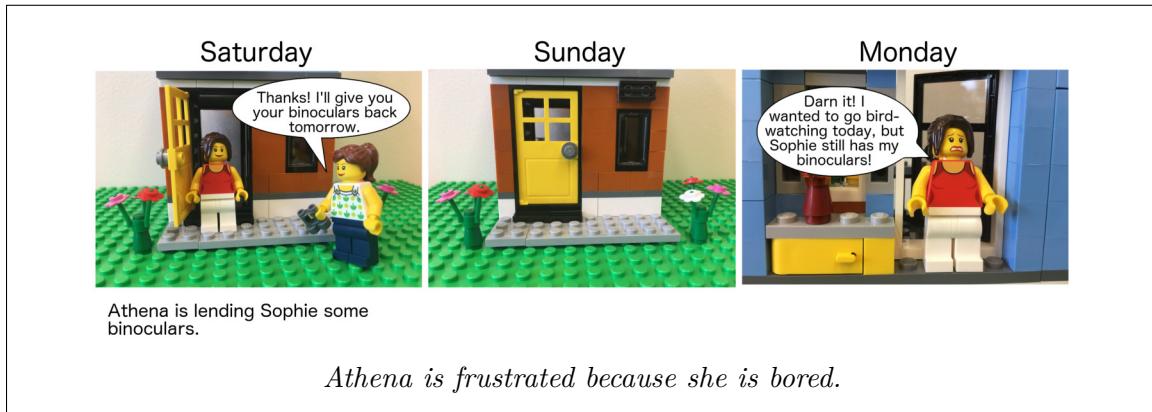


The target sentence was shown below the comic strip, and participants were asked to judge the sentence as a caption for the third panel. The target adverbials were embedded under the speech verb *say*, followed by the overt complementizer *that* to block a full quotation interpretation; the verb forms in the embedded clause are also inconsistent with quotation.

Four conditions were created by manipulating the temporal expression in the caption: *tomorrow*, the critical condition; *the next day*, the anaphoric condition; the day-of-week name of the first panel, a factually incorrect baseline; and the second day-of-week name, a factually correct baseline. Whether or not speakers allow non-utterance time readings of *tomorrow*, they are expected to interpret *tomorrow* in the first panel as referring to the second day. The *tomorrow* captions are felicitous only under a non-utterance time reading of *tomorrow*, since the first character promises to act on the second day; the other character has no grounds for anger on a utterance time reading.

Three kinds of fillers were used: one group which had captions that were obviously incorrect, which I will refer to as the bad fillers; one group which had correct captions, which I refer to as the good fillers; and pragmatically subpar fillers, which were factually correct, but under- or over-informative, which I refer to as the medium fillers. Figure 4.3 shows an example medium filler: the participant must draw the inference that Athena is bored because she would have gone birdwatching had Sophie returned her binoculars.

**Figure 4.3.** Example medium filler



#### 4.3.1.3 Procedure

Stimuli were displayed and responses collected using the Ibex Farm platform for web-based experiments (Drummond, 2019). Each experimental session began with an informed consent form and a demographic survey.<sup>15</sup> Next, participants were given instructions and practiced on 3 training items: a true item, which they were told most people would rate at 7; a false item, which they were told most people would rate at 1; and a medium filler, which they were told most people would rate at 4. In subsequent experiments, an additional training item calling attention to the day-of-week labels was added.

#### 4.3.2 Regression analysis

An analysis using paired t-tests was preregistered, but after discussion with colleagues, I decided to use a mixed effects ordinal regression model.<sup>16</sup> Because the response data was ordinal, I used a cumulative link model, which is appropriate for ordered non-continuous data like Likert scale responses. The maximal random effects structure was used for all models: random intercepts and slopes were included for all fixed-effects predictors, for participants, and for items. All models were fitted using the ordinal package in R (Christensen, 2019).

---

<sup>15</sup>Demographic information was collected in order to explore whether there was any significant sociolinguistic variation by age or geographic region, but these factors were not found to be informative. These results can be found in Appendix A.

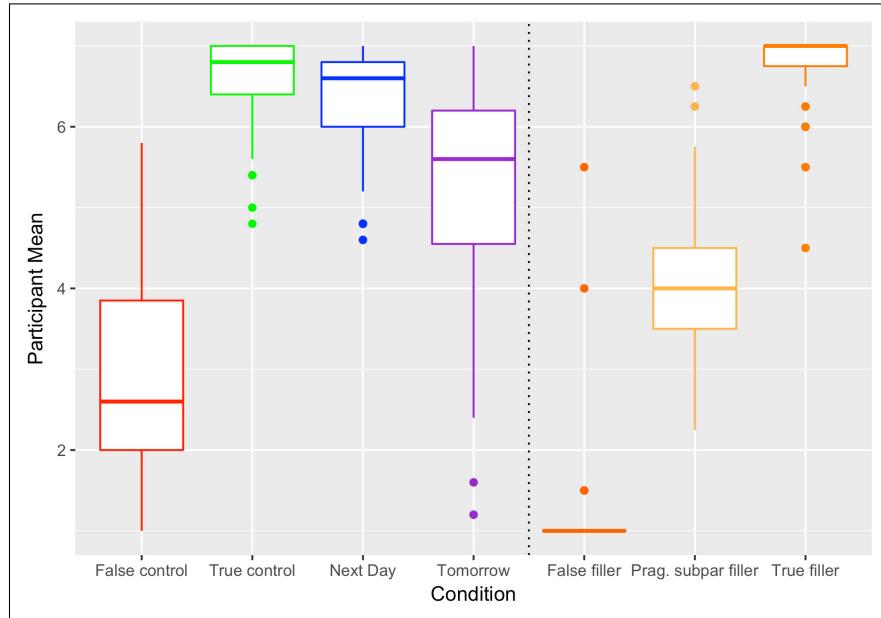
<sup>16</sup>The comparisons of interest were the same under both analyses.

Treatment coding was used, treating *tomorrow* as the baseline condition. This resulted in the following fixed-effects contrasts: *the next day*, 1 for *the next day* and 0 otherwise; false control, 1 for the false control and 0 otherwise; and true control, 1 for the true control and 0 otherwise. Treating *tomorrow* as the baseline allows us to interpret the intercepts as indicating how far the mean scores in other conditions differ from those in the *tomorrow* condition. For example, finding a significant effect of the *the next day* predictor would demonstrate that the responses in the *the next day* condition were reliably different from those in the *tomorrow* condition. A model that included the medium fillers as a fixed-effect predictor was also run, so that a comparison could be made between the pragmatically suboptimal items and the items in the *tomorrow* condition.<sup>17</sup>

#### 4.3.3 Experiment 1 results

The results showed that participants rated the *tomorrow* items much higher than the false baseline items, but somewhat lower than the true baseline and *the next day*.

**Figure 4.4.** Experiment 1 participant means by condition



<sup>17</sup>The comparison between *tomorrow* and the medium fillers was not preregistered.

**Table 4.2.** Experiment 1 results

Condition	Mean ratings	95%CI for part. means
False control	2.9	[2.6-3.2]
<i>Tomorrow</i>	5.3	[5.0-5.6]
<i>The next day</i>	6.4	[6.2-6.5]
True control	6.6	[6.5-6.7]
Bad fillers	1.1	[0.9-1.3]
Medium fillers	4.0	[3.8-4.2]
Good fillers	6.8	[6.7-6.7]

In the regression analysis, all three coefficients were reliable effects at  $p < 0.0001$ . This shows that the ratings in the *tomorrow* condition were significantly different from the false baseline, but also from *the next day*. Participants accepted non-utterance time readings of *tomorrow*, but found them worse than *the next day*.

**Table 4.3.** Experiment 1 mixed effects regression analysis, fixed effects (N=1440)

	$\hat{\beta}$	<i>z</i>	<i>p</i>
False control	-3.36(+/- 0.36)	-9.3	< <b>0.0001</b>
True control	2.47(+/- 0.38)	6.45	< <b>0.0001</b>
<i>next day</i>	1.53(+/- 0.29)	5.35	< <b>0.0001</b>

Because the *tomorrow* items were rated lower than *the next day*, it could be the case that they are not truly felicitous, but that speakers can interpret them as true through semantic coercion. However, the ratings of the pragmatically sub-optimal fillers argue against this interpretation of the data: these items require only a small amount of accommodation to fit the context, yet participants rate the *tomorrow* items more highly than them.

A second regression model that included the pragmatically subpar fillers was run. The coefficient for the medium filler condition was significant, indicating that *tomorrow* was rated significantly higher than the medium fillers. This strengthens the claim that non-utterance time readings of *tomorrow* are accepted as grammatical.

**Table 4.4.** Experiment 1 mixed effects regression analysis including medium fillers, fixed effects (N=1728)

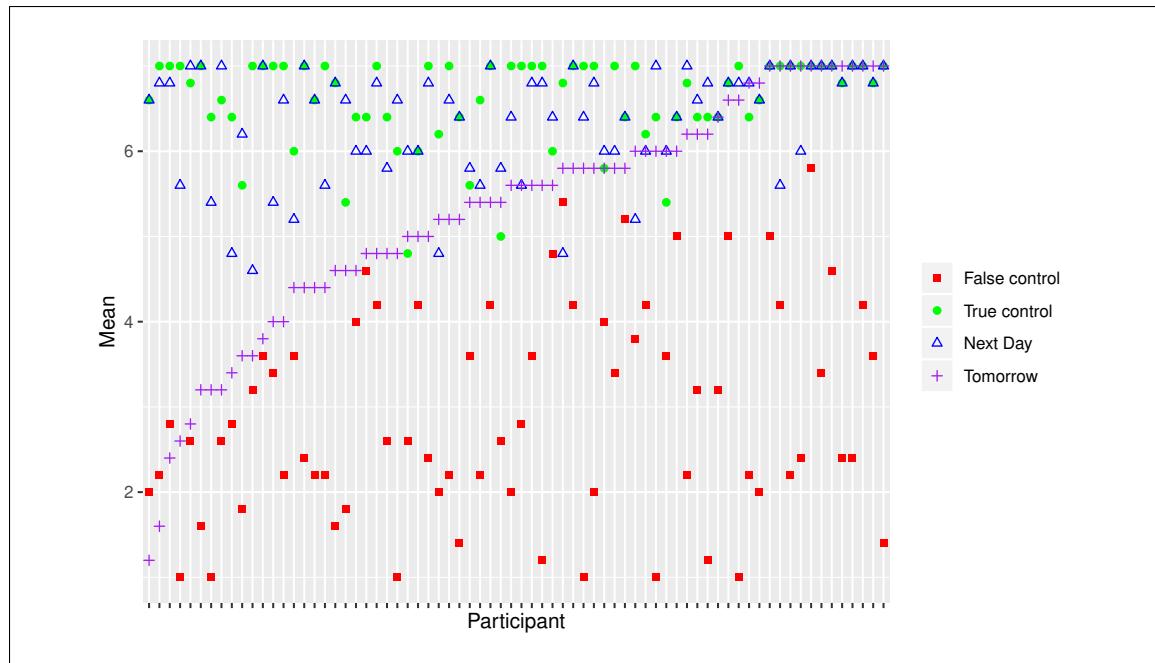
	$\hat{\beta}$	<i>z</i>	<i>p</i>
False control	-2.78(+/- 0.29)	-9.5	< <b>0.0001</b>
True control	2.25(+/- 0.33)	6.74	< <b>0.0001</b>
<i>next day</i>	1.34(+/- 0.25)	5.45	< <b>0.0001</b>
Medium fillers	-1.53(+/-0.27)	-5.59	< <b>0.0001</b>

#### 4.3.3.1 Interspeaker variation

There was interspeaker variation in the acceptability of non-utterance time *tomorrow*. Figure 4.5 shows participant means in each condition in order of increasing *tomorrow* means. While the *tomorrow* means for most participants are close to their *the next day* means, the *tomorrow* means of some participants are just as low as their bad baseline means.

This suggests that there is a small group of participants who do not accept non-utterance time uses of *tomorrow*, whose low ratings for *tomorrow* items depress the overall mean for the condition.

**Figure 4.5.** Experiment 1 means by participant



#### 4.3.4 Discussion

Experiment 1 shows that participants rate non-utterance time interpretations of *tomorrow* lower than *the next day*, but well above the false control items and the pragmatically subpar fillers, establishing that non-utterance time readings of *tomorrow* are accepted by a large group of American English speakers.

Having established this baseline, I turn to evaluating the analyses outlined in Section 4.2 (Figure 4.6). Experiment 2 explores whether non-utterance time interpretations of *tomorrow* can be explained as instances of mixed quotation.

**Figure 4.6.** Attested versus predicted availability of non-UT readings of *tomorrow* by analysis

Hypothesis	Quant. binding	UT epithet	Non-verbatim report	Unembedded	Attitude report
FID	!	!	!	★	★
Shifty indexical	!	★	★	!	★
Quotation	★	★	!	★	★
Anaphoric	★	★	★	★	★
Actual	?	?	?	?	✓

★ = predicted felicity, ! = predicted infelicity, ✓ = attested felicity, X = attested infelicity

### 4.4 Experiment 2: testing the quotational account

Experiment 2 tests the possibility of a quoted interpretation of *tomorrow*. The same comic strip items were used, but the character's speech in Panel 1 was paraphrased so that it did not contain *tomorrow*. This blocks the possibility that the *tomorrow* in the third panel is interpreted as a quotation of the *tomorrow* that was spoken in the first panel.

#### 4.4.1 Method

Experiment 2 explores the possibility of interpreting the critical *tomorrow* captions as mixed quotation (Maier, 2015). In the previous experiments, *tomorrow* appeared in the first panel as well as the third panel. This gives rise to the possibility that when the speaker in the third panel uses *tomorrow*, they could be quoting the other character's utterance in the first panel.<sup>18</sup> Although the tense and pronoun use are not consistent with a direct speech

---

<sup>18</sup>Thank you to Kristen Syrett for pointing out this possibility.

report, they are consistent with mixed quotation, a phenomenon in which some words in a speech report are quoted while others are unquoted. Removing *tomorrow* from the first panel makes it impossible to interpret the *tomorrow* in the third quote as a direct quote, because it is no longer a verbatim report of what was spoken on Day 1.

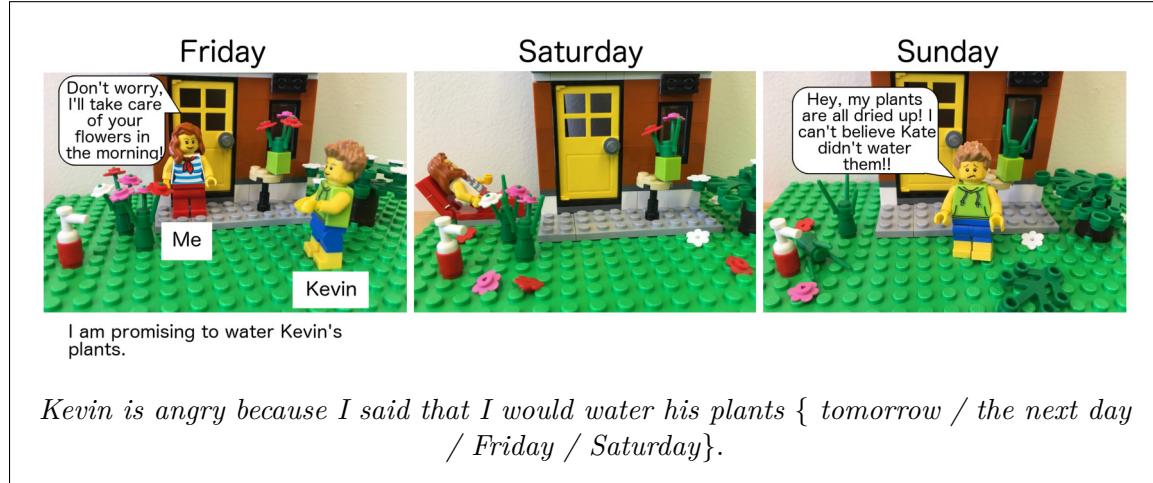
#### 4.4.1.1 Participants

93 participants were recruited through Amazon's Mechanical Turk platform. 13 participants were excluded for failing the baseline criterion. The remaining 80 participants were evenly balanced across experimental lists.

#### 4.4.1.2 Materials

The same comic strips were used as in Experiment 1. The speech bubbles were edited so that *tomorrow* did not appear in the comic strip, and the predicates were reworded so that they were not the same as the predicates in the critical caption (Figure 4.7).

**Figure 4.7.** Experiment 2 example stimulus



#### 4.4.1.3 Procedure

The same experimental methods were used as in Experiment 1. However, an additional training item calling attention to the day-of-week labels was added in order to reduce the number of participants rejected by the baseline criterion.

#### 4.4.2 Regression analysis

The same ordinal regression analyses were run as in Experiment 1.

#### 4.4.3 Experiment 2 results

Experiment 2 replicated the findings of Experiment 1. The mean for the *tomorrow* condition was slightly lower, but still significantly above that of the bad baseline and medium filler conditions. Ratings in all of the main conditions were lower in Experiment 2 compared to Experiment 1.

**Table 4.5.** Experiment 2 results

Condition	Mean	95%CI
false control	2.1	[1.9-2.3]
<i>Tomorrow</i>	4.9	[4.5-5.2]
<i>The next day</i>	6.1	[5.9-6.3]
True control	6.4	[6.3-6.6]
Bad fillers	1.3	[1.1-1.5]
Medium fillers	4.2	[4.0-4.3]
Good fillers	6.7	[6.5-6.8]

**Table 4.6.** Experiment 1 and 2 mean scores

Condition	Exp. 1	Exp. 2
False control	2.9	2.1
<i>Tomorrow</i>	5.3	4.9
<i>The next day</i>	6.4	6.1
True control	6.6	6.4
Bad fillers	1.1	1.3
Medium fillers	4.0	4.2
Good fillers	6.8	6.7

All significant effects reported in Experiment 1 were also found in Experiment 2. The only significant difference between ratings in each condition in Experiments 1 and 2 was for the true control condition, which received lower ratings in the Experiment 2.

**Table 4.7.** T-tests between main conditions in Experiments 1 and 2 using z-transformed means

Condition	Result
Exp. 1 <i>tomorrow</i> v. Exp. 2 <i>tomorrow</i>	$1.3(148.7) = 0.19$ , $p > 0.0125$ <sup>19</sup>
Exp. 1 <i>the next day</i> v. Exp. 2 <i>the next day</i>	$0.8(146.2) = 0.44$ , $p > 0.0125$
Exp. 1 true control v. Exp. 2 true control	$4.4(136.4) = 2.2\text{e-}05$ , $p < 0.0001$
Exp. 1 false control v. Exp. 2 false control	$-0.04(149.1) = 0.96$ , $p > 0.0125$

#### 4.4.4 Discussion

Since Experiment 2 replicated the findings of Experiment 1, the ratings for the non-utterance time readings of *tomorrow* reported in Experiment 1 are not due to mixed quotation. This

---

<sup>19</sup>p-values are reported using Bonferroni-corrected thresholds to account for multiple comparisons.

provides evidence against a mixed quotation account of non-utterance time readings of *tomorrow*, since such readings can occur even when *tomorrow* is not used to report the speaker's verbatim utterance.

This is not, however, evidence against a Free Indirect Discourse account; although Free Indirect Discourse also requires verbatim faithfulness, Free Indirect Discourse can report a character's thoughts, rather than their speech, a possibility that was not excluded by this design.

**Figure 4.8.** Attested versus predicted availability of non-UT readings of *tomorrow* by analysis

Hypothesis	Quant. binding	UT epithet	Non-verbatim thought report	Non-verbatim speech report	Unembed.	Att. report
FID	!	!	!	★	★	★
Shifty indexical	!	★	★	★	!	★
Quotation	★	★	!	!	★	★
Anaphoric	★	★	★	★	★	★
Actual	?	?	?	✓	?	✓

★ = predicted felicity, ! = predicted infelicity, ✓ = attested felicity, X = attested infelicity

## 4.5 Experiments 3a and 3b: testing the Free Indirect Discourse account

One hypothesis about non-utterance time readings of *tomorrow* is that they arise from Free Indirect Discourse effects. Experiments 3a and 3b assess whether participants accept non-utterance time readings of *tomorrow* outside of Free Indirect Discourse environments. One possibility for the stimuli used in 1 and 2 is that participants might be interpreting them as instances of Free Indirect Discourse, such as (130).

130. She hoped he would not bang his books on the floor above their heads, she thought, still thinking how annoying Charles Tansley was. For neither of them slept well; they were excitable children, and since he said things like that about the Lighthouse, it seemed to her likely that he would knock a pile of books over, just as they were going to sleep, clumsily sweeping them off the table with his elbow. For she supposed that he had gone upstairs to work. Yet he looked so desolate; yet she would feel relieved when he went; yet she would see that he was better treated tomorrow. (Woolf, 1927)

This is possible because the critical Day 3 caption reports the thoughts and feelings of a character, which can be done through Free Indirect Discourse. If the caption is an instance of Free Indirect Discourse that reports the Day 1 thoughts of the character, then *tomorrow* would be expected to receive a non-utterance time reading, since temporal adverbials in Free Indirect Discourse are interpreted relative to the protagonist's perspective.

In some ways this is not a likely hypothesis from the outset. The entire caption cannot be a Free Indirect Discourse report of the Day 1 perspective of the character, because the emotions reported are not experienced until Day 3. It is possible that the embedded clause containing *tomorrow* is an instance of Free Indirect Discourse, though this goes against Hinterwimmer (2017b)'s claim that Free Indirect Discourse is blocked in the complement of attitude verbs and Schlenker (2004)'s claim that Free Indirect Discourse cannot be preceded by an overt complementizer.

Nonetheless, in Experiments 3a and 3b, I test the possibility of an Free Indirect Discourse interpretation using two manipulations. Experiment 3a manipulates pronoun use to control for Free Indirect Discourse effects, under the assumption that first-person pronouns block Free Indirect Discourse (Banfield, 1982). While it is true that in Free Indirect Discourse, the protagonist is usually referred to in the third-person, the exception is when the protagonist's perspective is that of the narrator at some earlier time, as in (131).<sup>20</sup>

131. *The narrator, Modiano, knows that he will not succeed in finding Jacqueline. However, his past self, the protagonist, thinks that he will.*

J'ai sorti de ma poche le “compte-rendu” que j'avais signé. Elle habitait donc square de l'Alboni. Je connaissais cet endroit pour être souvent descendu à la station de métro toute proche. Aucune importance si le numéro manquait. Avec le nom: Jacqueline Beausergent, je me débrouillerais.

'I drew out of my pocket the "report" I had signed. So she was living in the square de l'Alboni. I knew that place because I had often got down at the nearest underground

---

<sup>20</sup>Thank you to Amy Rose Deal and Vincent Homer for discussion on this point.

station. No problem that the number was missing. With the name: Jacqueline Beausérgent, I would manage.' (Modiano, 2005, cited in Reboul et al. (2016))

Since this is an interpretative possibility for the comic strip items, Experiment 3b uses narrator-oriented epithets to block Free Indirect Discourse readings. In Free Indirect Discourse environments, expressives and epithets are interpreted according to the protagonist's perspective (Eckardt, 2014). For example, in (132), because Megan is not aware at this point in the story that her stepfather is the murderer, the use of *dastardly* is inconsistent with a Free Indirect Discourse reading of the passage.

132. *Megan Hunter is locked up in a closet, having been taken unawares by a mysterious prowler. Unbeknownst to her, the person was her stepfather, who had poisoned her mother.*

# Megan's heart was racing. How her hands ached! But surely her dastardly stepfather would call the police when he realized that she was missing. She could not be trapped in here for long.

Including narrator-oriented epithets in the critical captions should therefore block Free Indirect Discourse readings. If the Experiment 3 *tomorrow* ratings are similar to those in Experiments 1 and 2, this is evidence against the view that non-utterance time interpretations of *tomorrow* arise due to context shift in Free Indirect Discourse environments.

**Table 4.8.** Experiment 3 predictions

Condition	Truth	Predicted ratings
<i>tomorrow</i>	True if non-UT <i>tomorrow</i> allowed outside of FID False otherwise	High Low

#### 4.5.1 Method

The methods used were the same as in Experiments 1 and 2, with the exception of the critical manipulations described below and the sample size.

##### 4.5.1.1 Participants

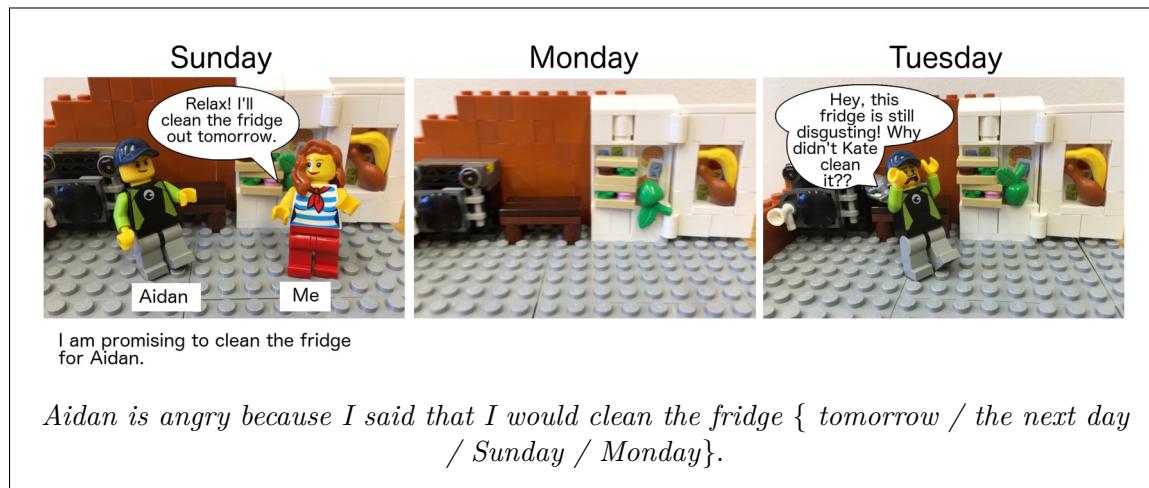
In Experiment 3a, 52 participants were recruited on Amazon's Mechanical Turk platform. 4 participants were excluded based on the baseline criterion, leaving 48 participants balanced

across experimental lists.<sup>21</sup> In Experiment 3b, 56 participants were recruited on the Prolific platform. 8 were excluded based on the baseline criterion, leaving 48 participants balanced across experimental lists.

#### 4.5.1.2 Materials

In Experiment 3a, the stimuli from Experiment 1 were modified to use first-person narration. The captions were changed to include first-person subjects, and the promise-maker was labeled as the narrator (“Me”) (Figure 4.9). The experimental procedure was as in Experiments 1 and 2, except that participants were instructed that the captions represented diary entries written by the “Me” character on the day they describe.

**Figure 4.9.** Experiment 3a example stimulus



In Experiment 3b, the stimuli were modified to include a narrator-oriented epithet. The character who was promised something was labeled as the narrator, and a narrator-oriented epithet was added to the critical captions. For example, in Figure 4.10, the epithet “that liar” cannot be interpreted as reporting Kevin’s Day 1 thoughts, since he has no grounds for thinking that the other character is a liar until Day 3.

<sup>21</sup> Analysis of the effect size in Experiment 1 suggested that a smaller number of participants could be used. Large effect sizes were found for the difference between *tomorrow* and the medium fillers (Cohen’s  $d = 0.82$ ), and between *tomorrow* and *the next day* (Cohen’s  $d = 0.81$ ). If the observed effect sizes were reliable indicators of the population effect sizes, power of 0.9 could be achieved with 20 participants. Given that post-hoc effect size analyses overestimate, subsequent experiments used 48 participants.

**Figure 4.10.** Experiment 3b example stimulus



#### 4.5.1.3 Procedure

The same experimental methods were used as in Experiment 2.

#### 4.5.2 Regression analysis

The same ordinal regression analyses were run as in Experiments 1 and 2.

#### 4.5.3 Experiment 3a results

The results of Experiment 3a were similar to Experiment 1. Participants rated the *tomorrow* items higher than the medium fillers and false items, but lower than *the next day* (Table 4.9).

**Table 4.9.** Experiment 3a mean ratings by condition

Condition	Mean ratings	95%CI for part. means
False control	2.2	[1.9-2.4]
<i>Tomorrow</i>	5.6	[5.2-6.0]
<i>The next day</i>	6.5	[6.3-6.7]
True control	6.6	[6.4-6.7]
False fillers	1.2	[1.0-1.5]
Medium fillers	3.9	[3.6-4.2]
True fillers	6.7	[6.5-6.9]

In the mixed-effects ordinal regression model, all three coefficients were reliable effects at  $p < 0.001$ . Thus, ratings for the *tomorrow* condition were significantly different from both the false control condition and *the next day* condition.

**Table 4.10.** Experiment 3a mixed effects regression analysis, fixed effects (N=960)

	$\hat{\beta}$	<i>z</i>	<i>p</i>
False control	-5.48 (+/- 0.56)	-9.8	< <b>0.0001</b>
True control	1.78 (+/- 0.48)	3.7	< <b>0.001</b>
<i>next day</i>	1.56 (+/- 0.45)	3.5	< <b>0.001</b>

As in previous experiments, there was a small group of participants who gave *tomorrow* items low ratings.

A mixed-effects ordinal regression model was run to compare Experiments 1 and 3a. The Experiment 3a coefficient was not significant, indicating that the *tomorrow* scores did not differ significantly between experiments.

**Table 4.11.** Experiment 1 and Experiment 3a comparison mixed-effects regression analysis, fixed effects and interactions (N=2400)

Condition	$\hat{\beta}$	<i>z</i>	<i>p</i>
False control	-3.48(+/-0.36)	-9.63	< <b>0.0001</b>
True control	2.45(+/-0.36)	6.78	< <b>0.0001</b>
<i>the next day</i>	1.58(+/-0.30)	5.22	< <b>0.0001</b>
Exp. 3a	0.58(+/-0.47)	1.24	0.21
False control * Exp. 3a	-1.51(+/-0.53)	-2.83	<b>0.005</b>
True control * Exp. 3a	-0.71(+/-0.52)	-1.37	0.17
<i>the next day</i> * Exp. 3a	-0.18(+/-0.44)	-0.42	0.67

The false control scores were lower in Experiment 3a, as reflected by the significant coefficient for the interaction between experiment and the false control condition.

#### 4.5.4 Experiment 3b results

Experiment 3b replicated Experiment 3a using epithets rather than first-person pronouns to control for Free Indirect Discourse effects. The results of Experiment 3b were similar to previous experiments. Participants rated the *tomorrow* items lower than *the next day*, but above the pragmatically suboptimal fillers and the false control items (Table 4.12).

**Table 4.12.** Experiment 3b results

Condition	Mean ratings	95% CI for part. means
False control	2.5	[2.2;2.8]
<i>Tomorrow</i>	5.2	[4.7;5.7]
<i>The next day</i>	6.5	[6.3;6.7]
True control	6.7	[6.6;6.8]
Bad fillers	1.2	[1.0;1.3]
Medium fillers	4.2	[3.9;4.5]
Good fillers	6.5	[6.3;6.8]

In the mixed-effects ordinal regression model, all three coefficients were reliable effects at  $p < 0.001$ . Thus, ratings for the *tomorrow* condition were significantly different from both the false control condition and *the next day* condition (Table 4.13).

The coefficient for the medium filler condition was also significant in the regression model that included the medium fillers (Table 4.14).

**Table 4.13.** Experiment 3b mixed-effects analysis, fixed effects ( $N=960$ )

	$\hat{\beta}$	$z$	$p$
False control	-3.9	-9.4	<b>&lt; 0.0001</b>
True control	4.1	145.5	<b>&lt;0.0001</b>
<i>next day</i>	3.4	5.8	<b>&lt;0.0001</b>

**Table 4.14.** Experiment 3b mixed-effects analysis with medium fillers, fixed effects ( $N=960$ )

	$\hat{\beta}$	$z$	$p$
False control	-3.4	-8.0	<b>&lt; 0.0001</b>
True control	2.8	4.7	<b>&lt;0.0001</b>
<i>next day</i>	2.5	4.5	<b>&lt;0.0001</b>
Medium filler	-1.4	-3.2	<b>&lt;0.01</b>

Mean participant ratings in each condition were similar in Experiment 1 and Experiment 3b. A mixed-effects ordinal regression model was run to compare the results of Experiments 1 and 3b. The Experiment 3b coefficient was not significant, indicating that the *tomorrow* scores did not differ significantly between experiments.

**Table 4.15.** Experiment 1 and Experiment 3b comparison mixed-effects regression analysis, fixed effects and interactions (N=2400)

Condition	$\hat{\beta}$	z	p
False control	-3.41(+/-0.37)	-9.22	< 0.0001
True control	2.47(+/-0.41)	6.00	< 0.0001
<i>the next day</i>	1.59(+/-0.32)	4.83	< 0.0001
Exp. 3b	-0.12(+/-0.48)	-0.25	0.81
False control * Exp. 3b	-0.52(+/-0.53)	-0.99	0.33
True control * Exp. 3b	0.51(+/-0.60)	-0.85	0.40
<i>the next day</i> * Exp. 3b	0.75(+/-0.50)	1.48	0.14

#### 4.5.5 Discussion

Experiments 3a and 3b replicated the results of Experiment 1. There was no significant difference between the *tomorrow* ratings in either Experiment 3a or Experiment 3b when compared to Experiment 3a. Contrary to the predictions of the Free Indirect Discourse account, participants rated the *tomorrow* items with narrator-oriented epithets similarly to the Experiment 1 items. This suggests that non-utterance time interpretations of *tomorrow* are not eliminated when Free Indirect Discourse interpretations are blocked.

**Figure 4.11.** Attested versus predicted availability of non-UT readings of *tomorrow* by analysis

Hypothesis	Quant. binding	UT epithet	Non-verbatim thought	Non-verbatim speech	Unembed.	Attitude report
FID	!	!	!	★	★	★
Shifty indexical	!	★	★	★	!	★
Quotation	★	★	!	!	★	★
Anaphoric	★	★	★	★	★	★
Actual	?	✓	?	✓	?	✓

★ = predicted felicity, ! = predicted infelicity, ✓ = attested felicity, X = attested infelicity

The experiments so far have not provided evidence for or against the other two hypotheses: the indexical shift and the anaphoric accounts. Experiment 4 tests the predictions of the indexical shift account.

#### 4.6 Experiment 4: testing the indexical shift account

The previous experiments have established that non-utterance time readings of *tomorrow* are possible under speech-verb embedding. Experiment 4 tests whether such embedding is necessary. Shifty indexicals are only predicted to have non-utterance readings when embedded under attitude predicates, while anaphoric expressions pick up their referent from the discourse context. The indexical shift account therefore predicts that non-utterance time readings should not arise outside of embedded contexts, while anaphoric accounts predict that non-utterance time readings should be available whenever there is a prominent referent.

One way to adjudicate between these two accounts is to create an unembedded context in which a non-utterance time reference time is prominent. Adding emotive content is often helpful in order to increase the prominence of the past time.<sup>22</sup> In the context in (133), for instance, both perspective-anaphoric and time-anaphoric expressions are able to pick up referents from the discourse context introduced in the previous sentences.

133. *Carolyn and Jyoti are chatting in their shared office.*

Carolyn: Last week I had a really frustrating experience at a restaurant. I showed up around 7pm.

- (a) **Time-anaphoric:** A few minutes later the power went out and everybody had to leave.
- (b) **Perspective-anaphoric:** By the time Beth **came**, it was too crowded to get a table.
- (c) **Indexical:** # By the time Beth got **here**, it was too crowded to get a table.
- (d) **Shifty indexical (predicted):** # By the time Beth got **here<sub>shifty</sub>**, it was too crowded to get a table.

By contrast, indexicals, whether shifty or rigid, should refer relative to the utterance context, since there is no attitude predicate that might introduce a context shift operator.

---

<sup>22</sup>Possibly by centering the Question Under Discussion around the past experience.

In this experiment, I test whether shifted readings of *tomorrow* arise in environments where there is emotive content heightening the prominence of a past time, but no embedding attitude predicate. If the results show that non-utterance time readings of *tomorrow* do not arise, it will be strong evidence in favor of an indexical shift account. On the other hand, if the results show that such readings arise in unembedded contexts, the anaphoric account will be the most promising, since this would be evidence that non-utterance time readings arise outside of the environments in which context shift has been proposed to happen.

#### 4.6.1 Method

The methods for Experiment 4 were the same as in previous experiments, except for the manipulation of the critical caption described below.

##### 4.6.1.1 Participants

53 participants were recruited on Amazon Mechanical Turk. 5 failed the baseline criterion and were removed, leaving 48 balanced across experimental lists.

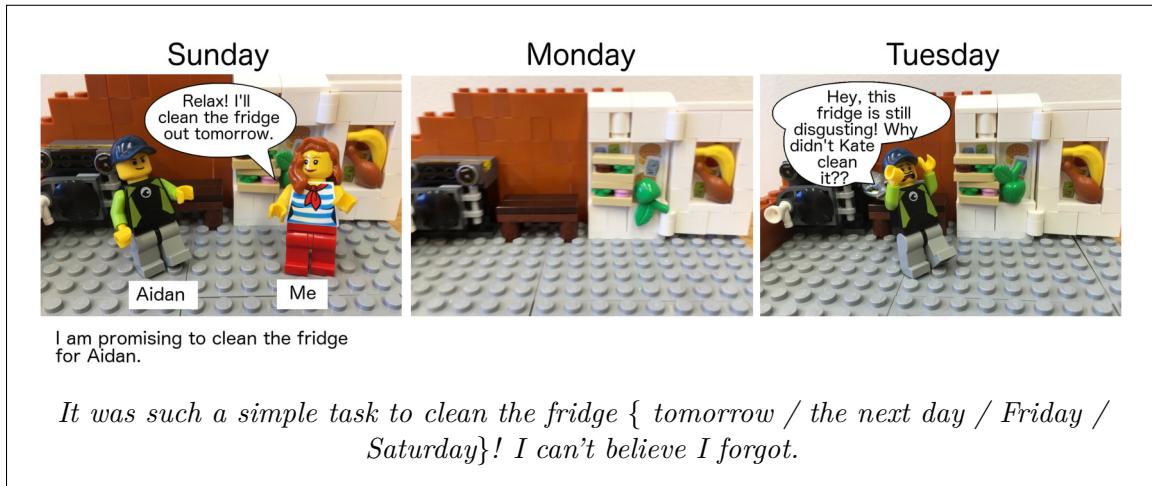
##### 4.6.1.2 Materials

In Experiment 4, the embedding speech verbs were removed from the captions. To provide a prominent previous time for anaphoric reference, the conversation between the characters is mentioned in the caption with heightened emotive content (Figure 4.12).<sup>23</sup>

---

<sup>23</sup>Following the strategy of eliminating a single hypothesis at a time, the grammatical environment tested is compatible with a logophoric binding account as well as a time-anaphoric and perspective-anaphoric account.

**Figure 4.12.** Experiment 4 example stimulus



Under an anaphoric account, we expect similar ratings for *tomorrow* as in previous experiments, since the mention of the Day 1 conversation should provide a prominent time different than utterance time. Under an indexical shift account, however, we expect unavailability of non-utterance time readings of *tomorrow*. Thus, if the *tomorrow* ratings are unaffected by removing the attitude predicates, it is evidence that non-utterance time interpretations of *tomorrow* do not arise solely from context shift in attitude reports.

**Table 4.16.** Experiment 4 predictions

Condition	Truth	Predicted ratings
<i>tomorrow</i>	False if speech embedding needed True otherwise	Low High

#### 4.6.1.3 Procedure

The experimental methods were the same as in Experiments 2-3b.

#### 4.6.2 Regression analysis

The same ordinal regression analyses were run as in Experiments 1-3b.

#### 4.6.3 Results

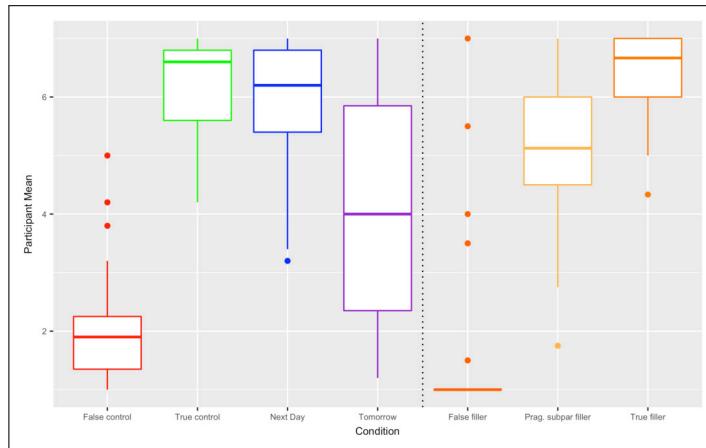
The mean ratings for *tomorrow* items were lower than in previous experiments, but still significantly higher than the false control items and the false fillers. The true controls,

the false controls, the true fillers, *the next day*, and *tomorrow* were all rated lower than in previous experiments, while the medium fillers and the false fillers were rated higher than in previous experiments.

**Table 4.17.** Comparison of mean ratings across experiments

Condition	Exp. 1	Exp. 2	Exp. 3a	Exp. 3b	Exp. 4
False control	2.9	2.1	2.2	2.5	2.1
<b>Tomorrow</b>	<b>5.3</b>	<b>4.9</b>	<b>5.6</b>	<b>5.2</b>	<b>4.1</b>
<i>The next day</i>	6.4	6.1	6.5	6.5	5.9
True control	6.6	6.4	6.6	6.7	6.3
False fillers	1.1	1.3	1.2	1.2	1.4
Medium fillers	4.0	4.2	3.9	4.2	5.0
True fillers	6.8	6.7	6.7	6.5	6.4

**Figure 4.13.** Experiment 4 main condition ratings



In the mixed-effects ordinal regression model, all three coefficients were reliable effects at  $p < 0.0001$ . Thus, despite the lower *tomorrow* scores in this experiment, the *tomorrow* condition was still significantly different than the false control condition.

**Table 4.18.** Experiment 4 mixed effects regression analysis, fixed effects (N=960)

	$\hat{\beta}$	$z$	$p$
False control	-3.41(+/- 0.50)	-6.81	< 0.0001
True control	3.20(+/- 0.48)	6.63	<0.0001
next day	2.54(+/- 0.50)	5.08	<0.0001

In the model that included the medium fillers, however, the coefficient for the medium fillers was not significant at  $p = 0.0125$ , indicating that the difference between the *tomorrow* condition and the pragmatically subpar fillers was not statistically significant.

**Table 4.19.** Experiment 4 mixed effects regression analysis with medium fillers, fixed effects (N=960)

	$\hat{\beta}$	<i>z</i>	<i>p</i>
False control	-3.16(+/- 0.46)	-6.84	< <b>0.0001</b>
True control	3.03(+/- 0.45)	6.69	< <b>0.0001</b>
<i>next day</i>	2.39(+/- 0.47)	5.08	< <b>0.0001</b>
Medium fillers	0.98(+/-0.45)	2.16	0.03

#### 4.6.3.1 Comparison with previous experiments

Although the main comparisons in Experiment 4 were similar to Experiment 1, a mixed-effects ordinal regression model comparing Experiment 1 and 4 finds significant interactions between experiment and *tomorrow*, indicating that the differences between the Experiment 1 and Experiment 4 ratings for the *tomorrow*. This suggests that the manipulation in Experiment 4 did affect participants' acceptance of the non-utterance time *tomorrow* items.<sup>24</sup>

**Table 4.20.** Experiment 1 and Experiment 4 comparison mixed-effects regression analysis, fixed effects and interactions (N=2400)

Condition	$\hat{\beta}$	<i>z</i>	<i>p</i>
False control	-3.44(+/-0.36)	-9.44	< <b>0.0001</b>
True control	2.31(+/-0.36)	6.47	< <b>0.0001</b>
<i>the next day</i>	1.50(+/-0.33)	4.52	< <b>0.0001</b>
Exp. 4	-1.72(+/-0.47)	-3.6	< <b>0.001</b>
False control * Exp. 4	0.35(+/-0.53)	0.65	0.52
True control * Exp. 4	0.87(+/-0.52)	1.68	0.09
<i>the next day</i> * Exp. 4	0.97(+/-0.47)	2.06	0.039

In addition, the 95% confidence intervals for participant means were wider than in previous experiments. Although the confidence interval for the *tomorrow* condition is widest in all experiments, in Experiment 4 it was wider than 1 Likert scale point.

---

<sup>24</sup>Or that this experiment sampled more heavily from participants who do not accept non-utterance time readings of *tomorrow*.

**Table 4.21.** Experiment 4 results

Condition	Mean rating	95%CI for part. means
False control	2.1	1.8-2.3
<i>Tomorrow</i>	4.1	3.5-4.6
<i>The next day</i>	5.9	5.6-6.2

#### 4.6.3.2 Interspeaker variation

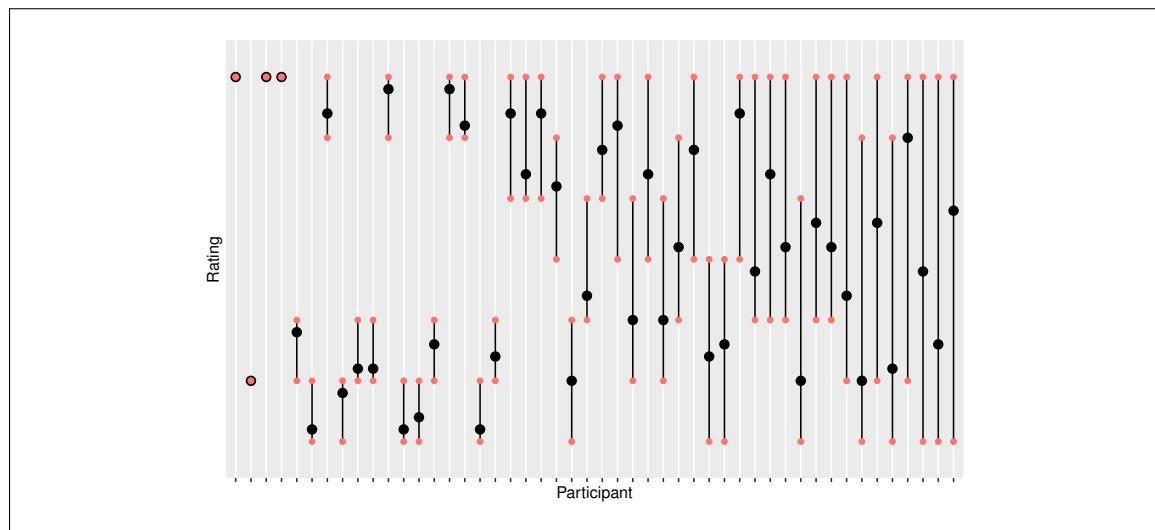
There was a higher amount of interspeaker variation in *tomorrow* ratings in this experiment.

While in all experiments there was a group of participants who gave *tomorrow* items consistently low ratings, the distribution of *tomorrow* scores in Experiment 4 is almost bimodal.

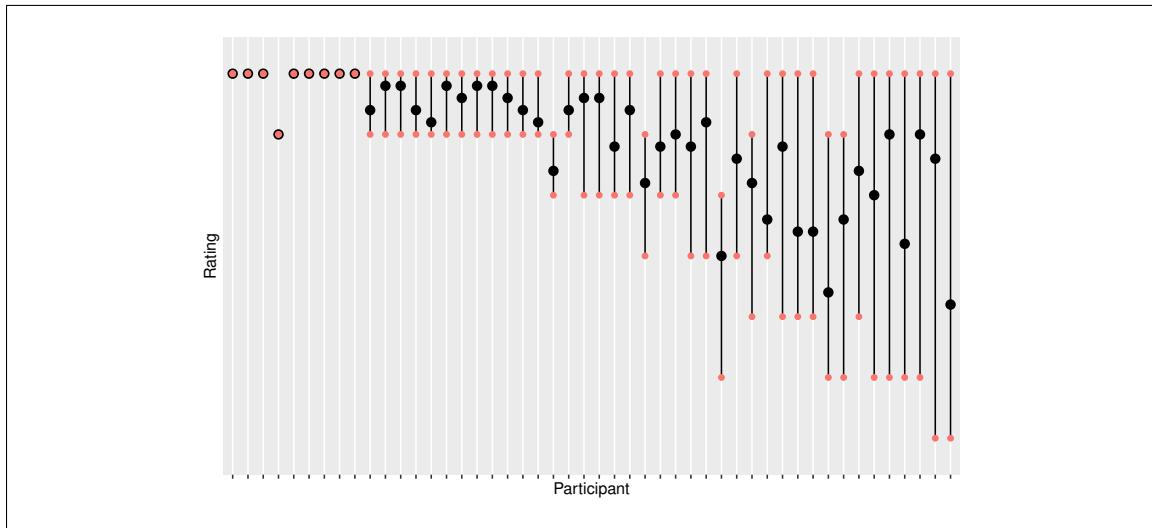
To see this, it is helpful to look at the range of responses given by each participant.

Figure 4.14 plots the mean *tomorrow* rating for each participant, with their highest and lowest ratings indicated in red. On the left side of the figure, we see two distinct groups: one which consistently rates *tomorrow* near *the next day*, and one which rates it near the bad baseline. On the right side of the figure lie participants who produced at least one rating on either side of the scale (indicated by a long line connecting the participant's lowest and highest ratings).

**Figure 4.14.** Experiment 4 *tomorrow* participant means in order of difference between highest and lowest rating (minimum and maximum ratings shown in red)



**Figure 4.15.** Experiment 4 *the next day* participant means in order of difference between highest and lowest rating (minimum and maximum ratings shown in red)



By contrast, Figure 4.15 plots the mean *the next day* ratings for each participant along with their highest and lowest ratings. On the left side of this figure, there is only one kind of participant: participants who rate every *the next day* item highly. There are no participants who give uniformly low ratings. The right side of the figure shows about the same number of participants who give at least one rating on either end of the scale.

The question is whether the participants who give *tomorrow* low ratings in this experiment are the same as those in previous experiments. It may be that Experiment 4 sampled more heavily from this population; or it may be that removing the embedding verb has a real effect on the acceptability, and that these participants would have accepted the *tomorrow* items in the other experiments.

#### 4.6.4 Discussion

Ratings for *tomorrow* items were lower in Experiment 4 than in previous experiments. They were not significantly different than the ratings for the medium fillers, though they were significantly above the bad baseline.

**Figure 4.16.** Attested versus predicted availability of non-UT readings of *tomorrow* by analysis

Hypothesis	Quant. binding	UT epithet	Non-verbatim thought	Non-verbatim speech	Unembed.	Attitude report
FID	!	!	!	★	★	★
Shifty indexical	!	★	★	★	!	★
Quotation	★	★	!	!	★	★
Anaphoric	★	★	★	★	★	★
Actual	?	✓	?	✓	✓	✓

★ = predicted felicity, ! = predicted infelicity, ✓ = attested felicity, X = attested infelicity

In addition, there was more interspeaker variation in *tomorrow* ratings than in previous experiments. Although a larger group of participants rejected non-utterance time *tomorrow* in Experiment 4 compared to other experiments, another substantial group gave the *tomorrow* items high ratings. This interspeaker variability poses an interpretative dilemma.

One possibility is to assume that there are two underlying populations of speakers: a population that accepts the non-utterance time readings and a population that does not. This would be a more comfortable assumption to make if there were a known sociolinguistic variable that correlated with the two groups of participants. However, none of the demographic variables measured (age, location, gender) correlated with the participants' ratings of the *tomorrow* items.<sup>25</sup>

Another possibility is to assume that there is a uniform population of speakers. If we make this assumption, we must make a decision about the grammaticality of non-utterance time readings of *tomorrow*. Are the ratings in Experiment 4 high enough to indicate that *tomorrow* is grammatical? Concluding that they are not grammatical seems unwarranted, given that many participants consistently rate them as highly as the items in the good baseline and the *next day* conditions. This is unpredicted on the view that *tomorrow* is a pure indexical; shifted readings should be entirely inaccessible in this context. However, concluding that they are grammatical is also problematic, given that there are participants who consistently rate the *tomorrow* items as low as the truth-conditionally false items.

---

<sup>25</sup>See Appendix A.

I will pursue the first interpretative approach, despite the fact that there is no known sociolinguistic correlate for the population split. I propose a split analysis for American English *tomorrow*: for one group of speakers, *tomorrow* has a standard pure indexical semantics. For another group of speakers, however, *tomorrow* is not a pure indexical, since the experimental data presented show that for these speakers, *tomorrow* can receive non-utterance time readings outside of the environments that allow pure indexicals to shift (Free Indirect Discourse, quotation, and attitude reports).

Although I will discuss two different semantics for *tomorrow* based on these two groups of speakers, I do not want to commit to the difference among speakers being a grammatical difference. It is also possible that the difference lies in processing. To preview my analysis of the non-utterance time accepting speakers, I will ultimately propose that *tomorrow* is perspective-sensitive. There is some work that reveals processing differences related to the accessibility of different perspectives both in terms of individual differences (Brown-Schmidt, 2009; Duff, 2018a) and at a population level (Köder and Maier, 2016; Mizuno et al., 2011; Long et al., 2018). One possibility is that the participants who do not accept shifted readings of *tomorrow* find it much more difficult to access these readings, even though their grammar theoretically allows it.

Setting aside the interesting questions around the source of interspeaker variation for the moment, for the rest of the paper, I explore the semantics of *tomorrow* with the goal of explaining how non-utterance time readings could arise outside of indexical shift environments. Since the standard pure indexical account is already well developed, I focus on proposing an analysis of *tomorrow* for the group of speakers who accept *tomorrow* in Experiment 4.

#### 4.7 Developing a non-indexical account of *tomorrow*

Experiments 2, 3a, 3b, and 4 test the predictions of three accounts that are consistent with a pure indexical semantics for *tomorrow*: the quotation account, the Free Indirect Discourse account, and the indexical shift account. Experiment 2 showed that contrary to the quotation account, non-utterance time readings of *tomorrow* can occur even in contexts that are not faithful to the original speaker's words. Experiment 3b showed that contrary to

the Free Indirect Discourse account, non-utterance time interpretations of *tomorrow* occur alongside narrator-oriented epithets. Experiment 4 showed that unlike shifted interpretations of shifty indexicals, non-utterance time readings of *tomorrow* are available outside of embedded contexts (for one group of speakers). Taken together, these experiments establish that for one group of American English speakers, non-utterance time readings of *tomorrow* are available outside the environments in which pure indexicals receive shifted readings.

This leads us to the view that for this group of speakers, *tomorrow* is not a pure indexical. The next most intuitive account is that *tomorrow* can also be anaphoric to a discourse-given time. However, there is some preliminary evidence against this approach, since *tomorrow* receives lower ratings than the temporal adverbial *the next day* in all of the experiments.

Instead, I will argue that *tomorrow* is, for at least some American English speakers, perspective-sensitive. I propose that the reference time of *tomorrow* is taken from the temporal field of a discourse-given perspective. This may not be as intuitive as a time-anaphoric account, but previous work on temporal adverbials has revealed unexpected complexities in deriving the reference times of seemingly time-anaphoric expressions (Stojnić and Altshuler, 2019). In addition, a perspectival view can explain the graded acceptability of *tomorrow*: non-utterance time readings of *tomorrow* are acceptable to the degree that a non-utterance time perspective is accessible for the participant.

My proposal does not rest on these points, however: in Section 4.7.2, I present evidence from quantificational binding that *tomorrow* cannot be anaphoric to just any prominent time; and that perspectival contexts heighten the acceptability of *tomorrow* in quantificational binding contexts.

#### **4.7.1 Perspectival versus time-anaphoric reference**

In Chapter 3, I laid out four families of perspectival expressions. Having argued against a pure indexical view of *tomorrow*, there are two families of perspectival analyses that I

consider, in addition to the possibility that *tomorrow* is time-anaphoric: the logophoric binding view and the perspective-anaphoric view.<sup>26</sup>

The perspective-anaphoric analysis builds upon Barlew (2017)'s analysis of *come* as anaphoric to a discourse-given perspective. One of the most convincing pieces of evidence that Barlew (2017) presents in favor of his account and against an indexical analysis of *come* (Oshima, 2006b; Sudo, 2018) is the behavior of *come* in quantificational binding contexts.

As discussed in Section 4.2.3.4, anaphoric expressions, unlike pure indexicals, can covary with quantifiers in quantificational binding contexts. Although the pure indexical *I* cannot covary in (134), since it always refers to the speaker, the anaphoric expression *the next day* can covary in (135).

134. # Whenever someone reads this sentence out loud, I am speaking.

135. [[Whenever I wash my car, it rains the next day]]<sup>*C,g*</sup> =

$$\begin{aligned} \forall t \text{ s.t. } \exists e. \text{WASH}(e) \wedge \text{AGENT}(e, C_{\text{speaker}}) \wedge \text{THEME}(e, \text{my car}) \wedge \tau(e) = t \rightarrow \\ \exists e'. \text{RAIN}(e') \wedge \tau(e') \subset \iota t'. \text{DAY-AFTER}(t, t') \end{aligned}$$

If *tomorrow* is time-anaphoric, then it should covary with the quantifier in similar quantificational binding contexts: contexts where there is a prominent time for each instantiation of the quantifier, like (135). Thus, the time-anaphoric account of *tomorrow* predicts felicity for (136).

136. [[Whenever I wash my car, it rains tomorrow]]<sup>*C,g*</sup> =

$$\begin{aligned} \forall t \text{ s.t. } \exists e. \text{WASH}(e) \wedge \text{AGENT}(e, C_{\text{speaker}}) \wedge \text{THEME}(e, \text{my car}) \wedge \tau(e) = t \rightarrow \\ \exists e'. \text{RAIN}(e') \wedge \tau(e') \subset \iota t'. \text{DAY-AFTER}(C_t, t') \end{aligned}$$

Prediction: infelicitous if *tomorrow* is a pure indexical; felicitous if *tomorrow* is time-anaphoric

On the other hand, if *tomorrow* had a pure indexical semantics, then (136) would mean that whenever the speaker washes their car, it rains the day after the sentence is spoken,

---

<sup>26</sup>The perspectival indexical family of analyses is ruled out by the same evidence that I used to argue against the pure indexical view of *tomorrow*.

since the context parameter does not covary with the quantifier in quantificational binding environments.<sup>27</sup>

Covariation in quantificational binding contexts is also one of the perspectival diagnostics that I proposed in Chapter 3. As Barlew (2017) points out, perspective-anaphoric expressions like *come* can covary along with the quantifier so long as there is a perspective holder who covaries with the quantifier. In (137), each instantiation of the woman provides a prominent perspective as an anchor for *come*; *was glad* heightens the prominence of each woman’s perspective.

137. Every woman gave her wayward child an extra serving of ham when he came to Christmas dinner.<sup>28</sup>

The same behavior is predicted for logophoric perspectival expressions, so long as the subject of the logophoric operator can covary with the quantifier.

Quantificational binding therefore provides a diagnostic for the semantics of non-utterance time *tomorrow* with three possible outcomes. If *tomorrow* is a pure indexical, its meaning should not covary with the quantifier, since the context parameter is not affected by quantification. If *tomorrow* is time-anaphoric, then it should covary freely in contexts where the quantification is over times. If *tomorrow* is perspectival (whether logophoric or perspective-anaphoric), then it should only covary in contexts where the perspective holder covaries with the quantifier.

#### 4.7.2 Quantificational binding behavior of *tomorrow*

We can apply the quantificational binding diagnostic in order to probe the semantics of *tomorrow*. There are three accounts that make different predictions. First, the pure indexical account, which I have already presented evidence against, predicts that the meaning

---

<sup>27</sup>With the possible exception of quantification over attitude report contexts, which might actually covary over context parameters if the quantifier scopes over a context shift operator.

<sup>28</sup>Reworded from Barlew (2017)’s original example, *Every woman was glad that her wayward child came to Christmas dinner* to avoid the confound of the attitude verb, as suggested by Josh Dever.

of *tomorrow* should be invariant under quantificational binding. We expect to observe this behavior of any speakers who do not accept shifted readings of *tomorrow* in Experiment 4.

Second, the time-anaphoric account of *tomorrow* predicts that *tomorrow* should covary freely in contexts involving quantification over times.<sup>29</sup>

Lastly, if *tomorrow* is perspectival, it is predicted to covary with the quantifier in some, but not all quantificational binding contexts: only those in which the perspective holder also covaries with the quantifier.

Thus the task is to set up environments in which the perspective holder can covary with the quantifier. One natural environment to test is contexts that involve quantification over individuals, since all perspective holders are individuals. We can, for instance, adapt Barlew (2017)'s *come* example, repeated as (138) below, for *tomorrow*, as in (139).

138. Every woman gave her wayward child an extra serving of ham when he came home to Christmas dinner. (Barlew, 2017)

139. Every little girl stays awake for hours the night before her birthday wondering what awesome new toys she will unwrap tomorrow morning.

Prediction: felicitous if *tomorrow* is perspective-anaphoric

If *tomorrow* is perspectival, (139) is predicted to be felicitous, since each girl provides a perspective for *tomorrow*. The prominence of their perspectives is heightened by *wonder*.

Another quantificational binding context in which perspectives are predicted to covary is when the quantification is over times and each time provides a prominent event time perspective. In (140) and (141), the expressive content should increase the prominence of the event time perspectives of the listener and speaker, respectively.

140. Every time you have to kick a drunk idiot out of the bar, you get to gloat about how hungover the jerk will be tomorrow.

Prediction: felicitous if *tomorrow* is perspective-anaphoric

---

<sup>29</sup> And perhaps even more widely: quantification over situations or events might also provide a discourse-given time that covaries for *tomorrow* to refer relative to.

**Figure 4.17.** Predicted covariation of *tomorrow* in quant. binding environments

Hypothesis	Temporal quantification	Perspective quantification
Pure indexical	!	!
Perspective anaphoric	!	★
Time-anaphoric	★	★

★ = predicted felicity, ! = predicted infelicity, ✓ = attested felicity, X = attested infelicity

141. My coworker is such a brat. Every time the jerk thinks it'll be sunny tomorrow, he calls in ‘sick’ and I have to cover his shift.

Prediction: felicitous if *tomorrow* is perspective-anaphoric

These are examples of temporal quantification<sup>30</sup> where each instantiated time provides a distinct perspective (made prominent by the expressive content). I have not given a good diagnostic for environments in which perspectives can covary versus environments that do not support perspective covariance; I suspect this is a gradient phenomenon. There is a clear prediction, however, that the acceptability of examples like (140) and (141) should increase when the expressive content is added, since this should increase the accessibility of the time-indexed perspectives.

The behavior of *tomorrow* in quantificational binding contexts was tested in two small experiments.

#### 4.7.3 Quantificational Binding Task 1: time and speech context quantification

Following Experiment 3a, participants ( $n=72$ ) completed a short sentence acceptability task exploring the grammaticality of *tomorrow* in quantificational binding contexts.<sup>31</sup>

##### 4.7.3.1 Participants

Participants were selected as part of Experiment 3a.

---

<sup>30</sup>Or quantification over situations, where the situation time is available for anaphoric reference.

<sup>31</sup>A less extensive task quantificational binding task was included after Experiment 1. The results were similar and are therefore omitted here for brevity.

#### 4.7.3.2 Materials

The sentence acceptability task asked participants to rate sentences on a 7-point Likert scale where 1 indicated complete unnaturalness and 7 indicated complete naturalness. The task probed the acceptability of *tomorrow* and *the next day* in two quantificational contexts: quantification over times and over speech contexts (Figure 4.18). Participants rated 2 time quantification and 4 speech-context quantification items for each adverbial, for a total of 12 items, presented in a Latin square design.

**Figure 4.18.** Quant. Binding Task 1 example stimuli

**Time quantification:** Every time I wash my car, it rains { tomorrow / the next day}.

**Speech context quantification:** Every time the UPS person says that the package has been delivered, it doesn't show up until {tomorrow / the next day }.

#### 4.7.3.3 Procedure

Participants took part in the quantificational binding task after completing the main study in Experiment 3a.

#### 4.7.3.4 Results

Quantificationally bound *tomorrow* items received low ratings in all conditions. An ANOVA of the *tomorrow* items showed no significant difference by condition ( $1.77(1,382) = 0.18$ ;  $p > 0.05$ ).

**Table 4.22.** Quant. Binding Task 1 results

Adverbial	Time quant. mean	95% CI	Speech quant. mean	95% CI
<i>tomorrow</i>	2.4	[2.0-2.9]	2.7	[2.4-3.0]
<i>the next day</i>	6.2	[5.9-6.5]	5.8	[5.6-6.1]

In addition, the ratings for *tomorrow* in Quantificational Binding Task 1 do not appear correlated with the differences between *the next day* and *tomorrow* ratings in the main task in Experiment 3a ( $\rho = -0.10$  for speech quantification;  $\rho = -0.15$  for time quantification). There was no observable linear relationship between giving high ratings to *tomorrow* in quantificational binding environments and rating *tomorrow* highly in the main task.

**Figure 4.19.** Predicted versus attested covariation of *tomorrow* in quant. binding environments

Hypothesis	Temporal quantification	Perspective quantification
Pure indexical	!	!
Perspective anaphoric	!	★
Time-anaphoric	★	★
Attested	X	?

★ = predicted felicity, ! = predicted infelicity, ✓ = attested felicity, X = attested infelicity

#### 4.7.3.5 Discussion

The fact that *tomorrow* receives very low ratings in quantificational binding contexts where there is a prominent time available at each instantiation of the quantifier suggests that *tomorrow*, unlike *the next day*, is not anaphoric to just any prominent time in the preceding discourse.

#### 4.7.4 Quantificational Binding Task 2: perspectival contexts

A second quantificational binding task was designed to test the quantificational binding cases that are predicted to be felicitous if *tomorrow* is perspectival. This task was included after the main task in Experiment 3b.

##### 4.7.4.1 Participants

Participants were selected as part of Experiment 3b.

##### 4.7.4.2 Materials

4 kinds of quantificational binding items were included: 3 *tomorrow* and 3 *the next day* items from Quantificational Binding Task 1; 3 naturally-occurring *tomorrow* items collected from corpora (Figure 4.20); and 3 examples constructed to test the predictions of a perspectival account of *tomorrow* (Figure 4.21).

**Figure 4.20.** Quantificational Binding Task 2 naturally occurring *tomorrow* items

1. I try to start winding down by 9pm so that I have time to tidy up the kitchen, think through what I'm going to wear tomorrow, and generally get my things together. That way I can sleep in as long as possible and still get to work on time!
2. One of my professors wears a different hat to work each day. He even has a website where you can pick a hat for him to wear tomorrow.
3. I set up an app on my phone that notifies me every time tomorrow's forecast calls for snow.

The temporal quantification examples from Quantificational Binding Task 1 were predicted to receive low ratings. The naturally occurring and perspectival examples were predicted to receive high ratings.

**Figure 4.21.** Quantificational Binding Task 2 perspectival *tomorrow* items

1. On Christmas Eve, every little girl stays awake for hours wondering what she will find under the Christmas tree tomorrow morning.
2. Every time you have to kick a drunk idiot out of the bar, you get to gloat about how hungover the jerk will be tomorrow.
3. My coworker is such a brat. Every time the jerk thinks it'll be sunny tomorrow, he calls in "sick" and I have to cover his shift.

#### 4.7.4.3 Procedure

Participants rated these 12 items after the main task in Experiment 3b.

#### 4.7.4.4 Quantificational Binding Task 2 results

The results of this task differ from those of Quantificational Binding Task 1. In this task, the bound *tomorrow* items did not uniformly receive low ratings. Although the *tomorrow* items presented in the previous task, labeled 'previously used *tomorrow*' in Figure 4.23, again received low ratings, the naturally-occurring corpus examples of bound *tomorrow* and the constructed perspectival items received higher ratings.

**Table 4.23.** Quantificational Binding Task 2 results

Condition	Mean rating	95%CI for part. means
Previously used <i>the next day</i>	5.6	[5.2;6.0]
Previously used <i>tomorrow</i>	3.7	[3.2;4.1]
Perspectival <i>tomorrow</i>	4.8	[4.4;5.3]
Naturally-occurring <i>tomorrow</i>	5.2	[4.8;5.6]

The ratings for individual items are shown in Figure 4.22.

Differences among ratings for quantificational binding items were investigated using a cumulative link mixed effect model. The *the next day* items were treated as the baseline, and three fixed effects were included: tomorrow, coded as 1 if the adverbial used was *tomorrow* and 0 otherwise; perspectival, coded as 1 if the item was a perspectival item and 0 otherwise; and natural, coded as 1 if the item was a naturally-occurring quantificational binding instance and 0 otherwise.

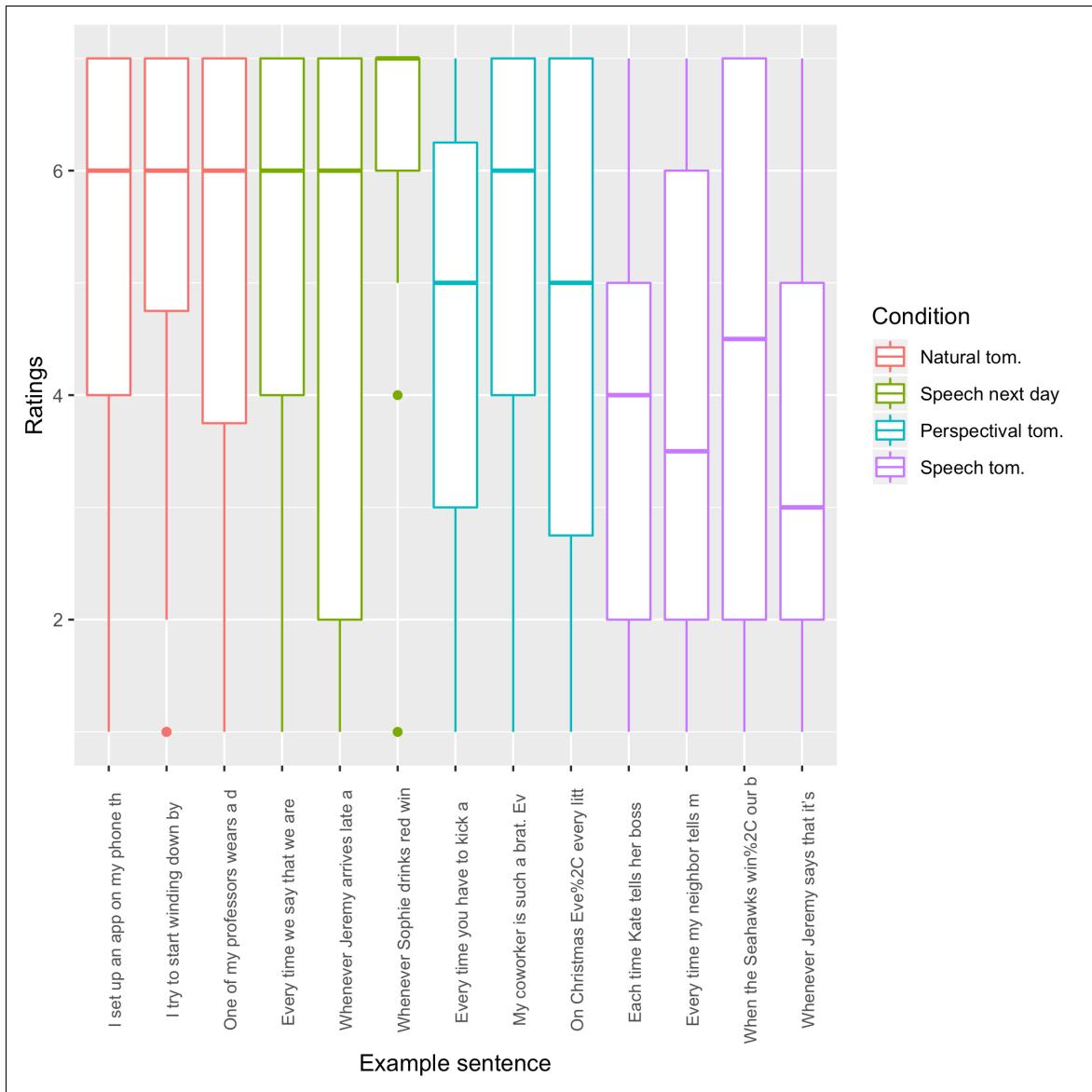
All three fixed effects were reliable at  $p < 0.0001$  (Table 4.7.4.4), indicating that while the use of *tomorrow* decreases participants' ratings, the ratings for the perspectival and natural conditions have a significant positive effect on participants' ratings compared to the *tomorrow* items from Quantificational Binding Task 1.

**Table 4.24.** Quant. Binding Task 2 mixed effects regression analysis, fixed effects (N=576)

	$\hat{\beta}$	$z$	$p$
Perspectival	1.2(+/- 0.25)	5.02	< <b>0.0001</b>
Natural	1.7(+/- 0.37)	6.79	< <b>0.0001</b>
Tomorrow	-2.3(+/- 0.26)	-6.05	< <b>0.0001</b>

In addition, there was a moderate negative correlation between participants' perspectival *tomorrow* scores and the difference between their mean *tomorrow* and *the next day* scores in the main task ( $\rho=-0.45$ ). The closer that participants rated *tomorrow* and *the next day*, the higher they rated perspectival *tomorrow*; this is expected if acceptance of *tomorrow* in the perspectival quantificational binding task is contingent on acceptance of non-utterance time readings of *tomorrow* (Figure 4.23)

**Figure 4.22.** By-item ratings for items in Quant. Bind Task 2

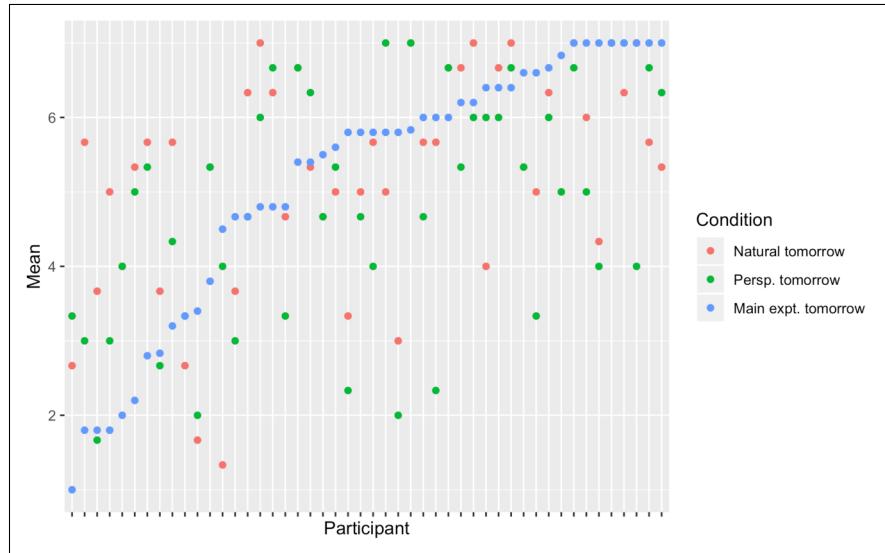


**Figure 4.24.** Predicted versus attested covariation of *tomorrow* in quant. binding environments

Hypothesis	Temporal quantification	Perspective quantification
Pure indexical	!	!
Perspective anaphoric	!	★
Time-anaphoric	★	★
Attested	X	✓

★ = predicted felicity, ! = predicted infelicity, ✓ = attested felicity, X = attested infelicity

**Figure 4.23.** By-participant means for *tomorrow* ratings in Expt. 3b and in Quant. Bind Task 2



The higher ratings for the *tomorrow* items in the perspectival condition support a perspectival view of *tomorrow*. Instances of bound *tomorrow* in which there are perspectives that covary with the quantifier, whether the perspective holder varies, as in the Christmas example, or the time-index on the perspective varies, as in the expressive examples, are rated higher than instances of bound *tomorrow* under ordinary temporal quantification. The results provide evidence against a view of *tomorrow* as anaphoric to a discourse-given time, as well as further evidence against a pure indexical view of *tomorrow* (for this subgroup of participants).

The higher ratings of the naturally-occurring *tomorrow* items are more difficult to explain, since two of the three seem to involve quantification over times similar to the items from Quantificational Binding Task 1, and the other seems to involve quantification over thought

situations. This example also has a habitual flavor, and it may be that modality is playing a role in its relatively high ratings.

#### 4.8 Towards a perspectival account of *tomorrow*

I have presented evidence that *tomorrow* (1) is not a pure indexical for one group of American English speakers; (2) is not anaphoric to discourse-given times; and (3) shows covariation with quantifiers in environments where perspectival expressions are predicted to covary. In this section, I develop a perspectival semantics for *tomorrow*. Again, I am not proposing that this is the only semantics that *tomorrow* has: for the group of American English speakers who reject non-utterance time readings of *tomorrow*, it is most natural to suppose that *tomorrow* is a pure indexical. I am proposing a semantics for *tomorrow* for the group of speakers who do accept such readings.

In Chapter 3, I described four different ways of encoding the perspective holder in the semantics of perspectival expressions: the lexically, indexically, logophorically, and anaphorically. I proposed a set of diagnostics for classifying perspectival expressions into these families, summarized in Table 4.25 below.

**Table 4.25.** Predictions by account compared to *tomorrow* data

	Lex.	Index.	Log.	Anaph.	<i>tomorrow</i>
Singleton perspective set	✓	X	X	X	X
Restricted perspective set	X	✓	✓	✓	✓
Obligatory <i>de se</i> anchoring	◊	✓	✓	◊	?
Perspective shift outside finite CP	✓	X	✓	✓	✓
Perspective shift outside XP with subj.	✓	X	X	✓	?
Shift Together effects	X	✓	✓	◊	?
Anchoring across utterances	X	X	X	✓	?
Covariation in quant. contexts	X	X	✓	✓	✓

The experimental work that I have presented in this chapter overlaps partially with the diagnostics that I proposed in Chapter 3. For instance, Experiment 4 provided evidence that non-utterance time readings of *tomorrow* can occur outside of finite CPs under attitude verbs. This provides evidence against both the traditional pure indexical analysis of *tomor-*

*row*, and against an analysis of *tomorrow* as a perspectival expression whose perspective holder is encoded indexically.<sup>32</sup>

The experimental findings narrow the space of possible perspectival analyses of *tomorrow*. Because *tomorrow* has a restricted but non-singleton set of perspective holders, the lexical approach is not viable. The experimental evidence that *tomorrow* can shift outside of finite CPs and can covary in quantificational binding contexts are not compatible with the indexical approach. However, there are two accounts that remain viable: the logophoric binding account and the perspective-anaphoric account.

I have not tested the environments that adjudicate between these two accounts. There are three diagnostics that are relevant for deciding between a logophoric and a perspective-anaphoric approach: the existence of Shift Together effects in logophoric binding environments; the existence of non-utterance time readings outside of logophoric binding environments; and the existence of non-utterance time readings licensed across utterance boundaries.

Testing the existence of Shift Together effects in logophoric binding environments involves setting up examples with two instances of *tomorrow* in the same spellout domain in order to observe whether they obligatorily receive the same interpretation. There are two reasons this is difficult. First, it is generally easier to set up examples with two instances in the same spellout domain if there is more than one member of the perspectival class being examined. This is not the case with *tomorrow*, since I have not explored whether non-utterance time readings are available for similar temporal adverbials like *yesterday*. Second, temporal adverbials as a class are challenging because they rarely co-occur in the same spellout domain.

---

<sup>32</sup>An indexical perspectival account of *tomorrow* is in theory different from the standard pure indexical analysis of *tomorrow*: in a perspectival account, *tomorrow* would take its reference time from the time index of a perspective tracked by the context parameter, rather than from the time index of the context parameter directly. Of course, the contexts that adjudicate between these two approaches are rare: the two analyses would only differ in contexts where the perspective holder tracked in the context parameter is mistaken about their own temporal location.

Testing the existence of non-utterance time readings outside of XPs with subjects is also difficult. As I argued in Section 3.4, this is one of the more challenging diagnostics to apply in general, and it is particularly challenging for temporal modifiers, since it can be difficult to diagnose the attachment height of temporal adverbials.<sup>33</sup> If American English contained a known logophoric element, its infelicity could be used to establish that there is no logophoric operator present in a particular environment; the felicity of *tomorrow* could then be tested. However, I do not know of any such expression. American English reflexives, for instance, do not display the same logophoric sensitivity as the French reflexives discussed by Charnavel (2019).

The third diagnostic is whether non-utterance time readings can be licensed across utterance boundaries. If the perspectival variable receives its value from the discourse context, then it should be able to refer to any sufficiently prominent perspective in the Common Ground, even if it was introduced in a previous sentence. This is the easiest diagnostic to test.

In order to set up the right environment, there needs to be a prominent perspective in one utterance that is used as the anchor for a perspectival expression in the following utterance. One naturally occurring candidate environment is shown in (142). In this example, the speaker's perspective of the night before is made prominent by the discussion of their restless mental state. The first instance of *tomorrow* is clearly from this perspective. Then the next utterance starts *It's tomorrow now*, which indicates a mixed temporal reporting: *tomorrow* is from the previous perspective, the speaker's perspective of the night before, while *now* is evaluated with respect to a new time, utterance time.

142. “I owe you one,” Dilly’s mother said, and hung up the phone. I sat at my desk for half an hour, not knowing what to think or what to do. There was a meeting of the committee that night, and I called my friend Laura and said I was under the weather. Laura was upset: “You can’t afford to be under the weather right now. Every day

---

<sup>33</sup>To take an item from Experiment 1 as an example, in *Kevin is angry because Kate said that she would water his plants tomorrow*, *tomorrow* could in theory attach high and modify either the time of Kevin’s state of anger, or the time of Kate’s speech event. In this example, of course, both the verb tenses and the discourse context provided by the comic strip rule out these possible attachment heights.

counts.” I said I’d be back on board tomorrow. **It’s tomorrow now; tomorrow is today and Dilly is home from Orlando.** What will I say to her when I see her? All I can think of are the things I can’t say. (Davies, 2008)

I have also constructed a few candidate examples that I judge natural. In (143), *tomorrow* refers to the day after the event time of Ariana’s promise; either it is evaluated relative to Ariana’s event time perspective or Laurel’s. In (144), *tomorrow* refers relative to Sam’s perspective on the 24th, when they spoke with UPS the first time. In (145), *tomorrow* refers relative to Jeremy’s perspective on the previous day, when he promised to finish the project by the next day.

143. *It is the day before prom. Her friend Ariana has been making some alterations to Laurel’s prom dress, but promised that it would be ready the day before prom. Laurel is debating with herself whether or not to stay home and wait for it so she can try it on right away.*

Ariana said that my dress would be ready tomorrow. But what time tomorrow did she mean? Should I stay in this morning in case it’s done early, or should I go out to distract myself?

144. *Sam ordered centerpieces for their wedding online. They paid for express shipping so that they would be delivered on June 25th, the day before their wedding. By the end of the day, the centerpieces still hadn’t showed up. When they called UPS, the employee promised them that the shipment would arrive early the next morning. However, the centerpieces did not arrive before the wedding. Sam has called to complain.*

UPS employee: Our records show that you spoke with a representative on the 24th. You indicated that you were satisfied with the service you received.

Sam: Yes, but the shipment didn’t arrive tomorrow, like the guy promised. It came five days later!

145. *Jeremy has accepted a last minute video editing job. He promised a quick turnaround, but the footage was poorly shot and required more work than expected. The next morning, his customer arrived to find him asleep at his desk, still working on the project.*

Customer: You promised that the video would be ready within 24 hours!

Jeremy: Yeah, well, I stayed up all night, but tomorrow came too soon. I'll give you a refund and finish the video as soon as I can.

These examples suggest that the perspectival variable of *tomorrow* is governed anaphorically rather than through logophoric binding (Table 4.26).

**Table 4.26.** Predictions by account compared to *tomorrow* data

	Lex.	Index.	Log.	Anaph.	<i>tomorrow</i>
Singleton perspective set	✓	X	X	X	X
Restricted perspective set	X	✓	✓	✓	✓
Obligatory <i>de se</i> anchoring	◊	✓	✓	◊	✓*
Shift outside finite CP	✓	X	✓	✓	✓
Shift outside XP with subj.	✓	X	X	✓	?
Shift Together effects	X	✓	✓	◊	?
Anchoring across utterances	X	X	X	✓	✓?
Covariation in quant. contexts	X	X	✓	✓	✓

\* Not confirmed experimentally. See Appendix C for judgments and discussion.

In the remainder of this chapter, I adopt this approach, and present a perspective-anaphoric semantics for *tomorrow*. However, given the interspeaker variation in judgments about non-utterance time *tomorrow*, this should be seen as a tentative conclusion pending future experimental evidence about the felicity of these examples. In Appendix B, I describe three versions of a logophoric analysis that could be pursued if future experimental evidence supports the logophoric approach instead of the anaphoric approach.

#### 4.8.1 A perspective-anaphoric semantics for *tomorrow*

I propose a perspective-anaphoric treatment for *tomorrow* similar to the analysis of American English *come* proposed by Barlew (2017) that I argued for in Chapter 3: the perspectival component is represented by a free perspective variable. As shown in (146), *tomorrow* takes its reference time from a perspective holder.

##### 146. Perspective anaphoric semantics for *tomorrow*:

$$[[\text{tomorrow}_{ana}]]^{C,g} = \lambda Q_{<v,t>}.\lambda e_v.Q(e) \wedge \tau(e) \subset \iota t_i.\text{DAY-AFTER}(t, \text{TIME}(a)), \text{ where } \text{TIME}(a) \text{ returns the temporal location of the holder of perspective } a$$

The perspectival variable in the semantics of *tomorrow* is like a perspective pronoun: it is resolved to a discourse-given perspective in the Common Ground via an anaphora resolution process. Following Barlew (2017), I assume that the Common Ground contains a set of discourse-given perspectives that are ranked by prominence.

I am not committed to a particular theory of anaphora resolution. For illustrative convenience, I will assume that perspectival variables denote indices mapped to perspectives in the Common Ground via the assignment function  $g$ . I will not put forward a proposal for how a particular perspective is selected from among the set of discourse-given perspectives here. In Part II of this dissertation, I take up the questions of perspective selection and identification in greater detail.

#### 4.8.1.1 Representing perspectives

My account posits a perspectival variable in the semantics of *tomorrow*. I will introduce an atomic perspective type for this variable as shown in Table 4.27. I use  $u$  to denote this atomic perspective type<sup>34</sup> and  $a$  as the canonical variable for perspectives.<sup>35</sup>

**Table 4.27.** Semantic ontology

Category	Type
Truth values	$t$
Entity	$e$
Time	$i$
Event	$v$
World	$w$
Perspective	$u$

What kinds of objects are perspectives? One way of representing perspectives is as a set of **centered worlds** (Stalnaker, 2008; Roberts, 2015; Barlew, 2017). Each centered world is a world-individual pair: a world that the perspective holder believes to be possible, along with the individual that the perspective holder believes themself to be in that world.

---

<sup>34</sup> $u$  for unique, since everyone holds a unique perspective.

<sup>35</sup> $a$  for anchor.

**147. Perspective (preliminary):** An individual  $x$  holds perspective  $a$  if  $a$  is a tuple  $\langle c_1, \dots, c_n \rangle$ , where  $c_n$  is a centered world  $\langle w_n, i_n \rangle$  such that  $x$  self-identifies as individual  $i_n$  in world  $w_n$ .

Given this representation of a perspective, however, it is not obvious how to derive the truth conditions of various kinds of perspectival expressions. To take an example familiar from Chapters 2 and 3, the licensing conditions for *come* require the location of the perspective holder to be the destination of motion, as shown in (148).

**148. Anaphoric semantics for *come* (preliminary):**

$$[[\text{come}]]^{C,g} = \lambda x_e. \lambda e_v. \text{MOVE}(e) \wedge \text{DEST}(e, x) \wedge \text{LOC}(x, a), \text{ where } a \text{ is a perspective}$$

In order to satisfy the truth conditions, the perspectival variable must be able to enter into the LOC relationship with a destination of motion. Otherwise, there would be no way to check whether the perspective holder self-locates at the destination of motion. One necessary property of a perspective object, therefore, is the ability to extract the spatial location of the perspective holder. How does this work?

If a perspective is represented as a set of individual-world pairs representing the individuals with which the perspective holder self-identifies, then each of those individuals have a particular spatial location. One way of deriving the perspective holder's self-ascribed location, therefore, is to iterate through all worlds in the set of centered worlds and collect the locations of the individuals with whom the perspective holder self-identifies. This seems difficult if not all of those self-identified individuals are located at the same destination, so I will assume for the moment that they are.<sup>36</sup> For simplicity, I propose that each perspective has a spatial location field: a spatiotemporal perspective is a tuple consisting of an individual, a spatiotemporal location, and a set of centered worlds.

<sup>36</sup>If they are not, there are a couple of solutions. First, we could place an existential operator within the locational part of the semantics, so that it is true if there is at least one individual with whom the perspective holder self-identifies who is located at the destination of motion. Or perhaps we can find the nearest belief world to the world against which the sentence is being evaluated, and check if the individual who the perspective holder self-identifies as in that world is located at the destination of motion.

149. **Perspective (revised):** An individual  $x$  holds perspective  $a$  if  $a$  is a tuple

$\langle \{c_1, \dots, c_n\}, l_x \rangle$  where:

$c_n$  is a centered world  $\langle w_n, i_n \rangle$  such that  $x$  self-identifies as individual  $i_n$  in world  $w_n$  and  $l_x$  is the location at which  $x$  self-locates.

This spatial field makes it easier to represent the truth conditions of *come*. The perspectival variable  $a$  tracks the perspective holder's self-ascribed location along with the set of their belief worlds. We can now define the LOC function trivially as a function that accesses the spatial field of the perspective, which must be identical to the destination of motion.<sup>37</sup>

150. **Anaphoric semantics for *come* (revised):**

$[[\text{come}]]^{C,g} = \lambda e_e. \lambda e_v. \text{MOVE}(e) \wedge \text{DEST}(e, x) \wedge x = \text{LOC}(a)$ , where  $\text{LOC}(a)$  returns the spatial field of perspective  $a$

To aid in the analysis of temporal perspectival expressions like *tomorrow*, we can treat temporal perspective in a similar way. As well as being located in space, individuals are located in time, in both their belief worlds and the real world. The self-ascribed temporal location of an individual can also be derived from the set of centered worlds representing their beliefs. For ease of reference, I will add a temporal location field to the perspective object. Thus, I will treat a perspective as a tuple containing a set of centered worlds, a spatial location index, and a temporal location index, as shown in (151).

151. **Perspective (final):** An individual  $x$  holds perspective  $a$  if  $a$  is a tuple

$\langle \{c_1, \dots, c_n\}, t_x, l_x \rangle$  where:

$c_n$  is a centered world  $\langle w_n, i_n \rangle$  such that  $x$  self-identifies as individual  $i_n$  in world  $w_n$ ,

$t_x$  is the time at which  $x$  self-locates, and  $l_x$  is the location at which  $x$  self-locates.

I adopt this notation for perspectives in what follows; however, my proposal does not rest on this way of formalizing perspective, so long as there is a linguistic object representing

---

<sup>37</sup>I omit the careful consideration of the motion path of *come* developed by Barlew (2017).

a perspective from which the perspective holder's self-ascribed temporal location can be extracted.

#### 4.8.1.2 Deriving utterance time and non-utterance time readings of *tomorrow*

In my analysis, *tomorrow* refers relative to the temporal field of a perspective, as shown in (152). The perspective variable is free and receives its value from the discourse context.

152. Perspective-anaphoric semantics for *tomorrow*:

$$[[\text{tomorrow}]]^{C,g} = \lambda Q_{<v,t>}. \lambda e_v. Q(e) \wedge \tau(e) \subset \iota t_i.\text{DAY-AFTER}(t, \text{TIME}(a)), \text{ where } \text{TIME}(a) \\ \text{returns the temporal field of perspective } a^{38}$$

Non-utterance time readings of *tomorrow* arise when the value of the perspective variable is resolved to a perspective whose holder does not self-locate at utterance time.

Example (153) illustrates the derivation of a non-utterance time reading in an attitude context. The perspective of the attitude holder, Mary, is available in the Common Ground. In this example, the perspectival variable is resolved to Mary's perspective by the assignment function, which results in an event time reading of *tomorrow*.

153. Mary said that John sings tomorrow.

- (a) Perspective set= {  $a_{\text{speaker}} : <\{w_1, \dots\}, t_{\text{speaker}}, l_{\text{speaker}} >$   
 $a_{\text{Mary}} : <\{w_1, \dots\}, \text{TIME}(e_{\text{say}}), \text{LOC}(e_{\text{say}}) >$  }
- (b) Assignment function:  $g(a) = a_{\text{Mary}}$
- (c)  $[[\text{tomorrow}_{ana}]]^{C,g} = \lambda Q_{<v,t>}. \lambda e_v. Q(e) \wedge \tau(e) \subset \iota t_i.\text{DAY-AFTER}(t, \text{TIME}(a))$
- (d)  $[[\text{John sings}]]^{C,g} = \lambda e. \text{SING}(e) \wedge \text{AGENT}(e, j)$
- (e)  $[[\text{John sings tomorrow}]]^{C,g}$   
 $= \lambda e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, \text{TIME}(a)) \wedge \text{SING}(e) \wedge \text{AGENT}(e, j)$  By Function Application

---

<sup>38</sup>The functions  $\tau$  and  $\text{TIME}$  are similar: both are functions that take a single argument and return a time interval. However, because the type of their argument is different ( $\tau$  takes an event and returns its duration, while  $\text{TIME}$  takes a perspective and returns the value of its temporal field), I represent them differently.

$= \exists e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, \text{TIME}(a)) \wedge \text{SING}(e) \wedge \text{AGENT}(e, j)$  By Existential Closure

(f)  $[[\text{said John sings tomorrow}]]^{C,g}$   
 $= \lambda e'. \text{SAY}(e') \wedge \text{THEME}(e', \exists e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, \text{TIME}(a)) \wedge \text{SINGS}(e) \wedge \text{AGENT}(e, j))$   
 By Function Application

(g)  $[[v \text{ said John sings tomorrow}]]^{C,g}$   
 $= \lambda x. \lambda e'. \text{SAY}(e') \wedge \text{AGENT}(e, x) \wedge \text{THEME}(e', \exists e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, \text{TIME}(a)) \wedge \text{SINGS}(e) \wedge \text{AGENT}(e, j))$  by Function Application

(h)  $[[\text{Mary said John sings tomorrow}]]^{C,g}$   
 $= \lambda e'. \text{SAY}(e') \wedge \text{AGENT}(e, m) \wedge \text{THEME}(e', \lambda e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, \text{TIME}(a)) \wedge \text{SINGS}(e) \wedge \text{AGENT}(e, j))$  By Function Application  
 $= \exists e'. \text{SAY}(e') \wedge \text{AGENT}(e, m) \wedge \text{THEME}(e', \lambda e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, \text{TIME}(a)) \wedge \text{SINGS}(e) \wedge \text{AGENT}(e, j))$  By Existential Closure  
 $= \exists e'. \text{SAY}(e') \wedge \text{AGENT}(e, m) \wedge \text{THEME}(e', \lambda e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, a_{m_t}) \wedge \text{SING}(e) \wedge \text{AGENT}(e, j))$  By Assignment Function  
 $= \exists e'. \text{SAY}(e') \wedge \text{AGENT}(e, m) \wedge \text{THEME}(e', \lambda e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, \tau(e_{say})) \wedge \text{SING}(e) \wedge \text{AGENT}(e, j))$  By evaluation of the temporal field of Mary's perspective

However, there is nothing in the compositional semantics shown in (153) that stipulates a non-utterance time reading of *tomorrow*. The assignment function could map the perspectival variable to the speaker's perspective instead, which would result in a utterance time interpretation of *tomorrow*. Whether a particular instance of *tomorrow* is interpreted as shifted or not depends on the pronoun resolution algorithm and the ranking of perspectives in the Common Ground.

The derivation of an utterance time reading of *tomorrow* would be nearly identical. Only the last step, the resolution of the perspectival variable, would be different. This is shown

in (154). The assignment function now maps  $a$  to the perspective of the speaker, rather than the perspective of the attitude holder.

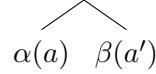
154. Mary said that John sings tomorrow.

- (a) Perspective set = {  $a_{\text{speaker}} : < \{w_1, \dots\}, t_{\text{speaker}}, l_{\text{speaker}} >$   
 $a_{\text{Mary}} : < \{w_1, \dots\}, \text{TIME}(e_{\text{say}}), \text{LOC}(e_{\text{say}}) >$  }
- (b) Assignment function:  $g(a) = a_{\text{speaker}}$
- (c)  $[[\text{Mary said John would come to the bank tomorrow}]]^{C,g}$   
 $= \exists e'. \text{SAY}(e') \wedge \text{AGENT}(e, m) \wedge \text{THEME}(e', \lambda e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, \text{TIME}(a)) \wedge$   
 $\text{SINGS}(e) \wedge$   
 $\text{AGENT}(e, j))$   
 $= \exists e'. \text{SAY}(e') \wedge \text{AGENT}(e, m) \wedge \text{THEME}(e', \lambda e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, a_{\text{speaker}_t}) \wedge$   
 $\text{SING}(e) \wedge$   
 $\text{AGENT}(e, j)) \quad \text{By Assignment Function}$   
 $= \exists e'. \text{SAY}(e') \wedge \text{AGENT}(e, m) \wedge \text{THEME}(e', \lambda e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, \tau(\text{speaker})) \wedge$   
 $\text{SINGS}(e) \wedge$   
 $\text{AGENT}(e, j)) \quad \text{By evaluation of the temporal field of the speaker's perspective}$

Because each perspectival variable can be resolved independently, the perspective-anaphoric account does not predict grammatical Shift Together effects.<sup>39</sup> As Figure 4.25 shows schematically, when multiple perspective-anaphoric expressions appear in a derivation, their perspectival variables are not identified, since they are free variables.

**Figure 4.25.** No Shift Together effects for perspective-anaphoric *tomorrow*

$\alpha(a) \wedge \beta(a')$  by Predicate Modification

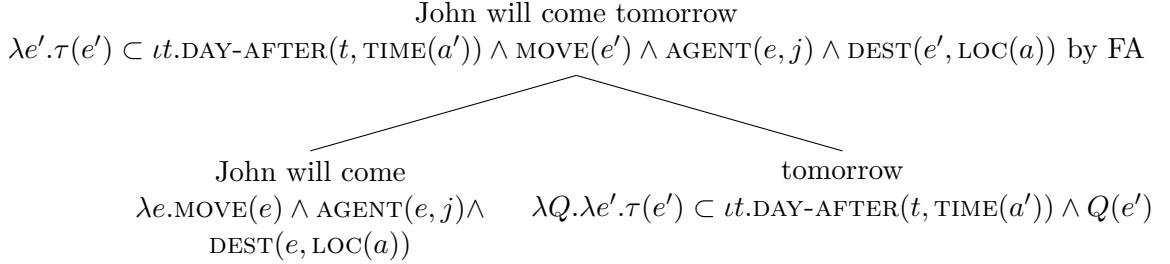



---

<sup>39</sup>Depending on the theory of anaphora resolution adopted, however, Shift Together effects could still arise. For instance, if the Common Ground contains a ranked list of perspectives that is fixed for the duration of a sentence and the topmost perspective is always selected, then all perspectival variables will refer relative to that perspective. This will result in the emergence of pragmatic Shift Together effects. However, if the ranking of perspectives varies dynamically during the evaluation of the sentence, or if perspectives are sampled from the ranking, then Shift Together effects will not emerge.

The perspective-anaphoric account also does not predict Shift Together effects for multiple perspective-anaphoric expressions in the same sentence. For instance, if *tomorrow* and *come* co-occur, there is no need for the perspectival variables of the two expressions to co-refer, since they are valued independently by the assignment function. This is shown schematically in Figure 4.26.

**Figure 4.26.** No Shift Together effects for perspective-anaphoric *tomorrow* and *come*



Furthermore, since the interpretation of the perspective pronoun is determined by the discourse context, the perspective-anaphoric account predicts that non-utterance time readings are not restricted to any particular syntactic domain. Once the perspective of the attitude holder is introduced to the Common Ground, it may remain accessible regardless of the syntactic scope of the expression that introduced it. This predicts that attitude holder perspectives may be available outside of the syntactic scope of attitude reports and even across utterance boundaries.

However, these predictions depend in part on the theory of perspectival anaphora resolution that is adopted. For instance, Shift Together effects might arise pragmatically via constraints on anaphora resolution, even though they are not predicted by the structures that I have proposed. I will not propose a specific theory of anaphora resolution, but I return to some of these issues in Part II, where I address questions of perspective selection and identification.

## 4.9 Summary

I have argued that for a substantial group of speakers, *tomorrow* is not a pure indexical. In a series of experiments, I have shown that this group of speakers accepts non-utterance time readings of *tomorrow* in environments beyond those in which a pure indexical would

shift, such as quotation, Free Indirect Discourse, and attitude reports. I have also shown that *tomorrow* can appear in some, but not all, quantificational binding environments, and used this evidence to argue against a time-anaphoric view of *tomorrow*.

I have proposed that for the American English speakers who accept these readings, *tomorrow* refers relative to the time index of a perspective, and that non-utterance time readings arise whenever there is a mismatch between utterance time and this time index. This proposal can be spelled out using either the logophoric or the perspective-anaphoric approach introduced in Chapter 3.

Both the logophoric and perspective-anaphoric approaches are consistent with the experimental evidence, since both predict shifted readings outside of attitude reports and in quantificational binding environments when there are perspectives that covary with the quantifier. However, I also provided some tentative evidence of perspectival licensing across utterance boundaries. This supports the perspective-anaphoric approach over the logophoric approach, although, given the interspeaker variability of judgments of *tomorrow*, experimental evidence comparing these two accounts should be collected.

On the basis of these judgments, I presented a perspective-anaphoric analysis of *tomorrow*, along the lines of Barlew (2017)'s treatment of *come*. In this account, *tomorrow* takes its reference time from the temporal field of a prominent perspective whose value is determined by the discourse context. Three alternative logophoric approaches are presented in Appendix B.

Although I have presented experimental evidence in favor of a perspectival view of *tomorrow*, there are several questions that remain unanswered. The most critical is the nature of the interspeaker variation observed across experiments. For one thing, if both *tomorrow* and *come* are anaphoric to perspectives, why are judgments about non-utterance time readings of *tomorrow* so much more varied than judgments about non-speaker oriented readings of *come*? Although there is some interspeaker variation in how frequent non-speaker-oriented

uses of *come* are, they are extremely common in corpus data, and shifted readings are well-accepted by most speakers.<sup>40</sup>

What makes *tomorrow* so different? Perhaps there truly is a population-level split in the semantics of *tomorrow*, such that some speakers can never shift *tomorrow* while others can. Perhaps temporal perspective is, for some reason, more challenging to shift than locative perspective. Or perhaps there is interspeaker variation in both the grammar of *tomorrow* and the accessibility of non-utterance time perspectives. These are interesting questions for future work.

One of the contributions of this chapter has been to explore the perspective sensitivity of temporal expressions in depth. Although individuals are situated in both space and time, and hold beliefs about both dimensions of their location, temporal perspective is less frequently discussed than spatial perspective. Temporal perspective-sensitive expressions also seem to be rarer, judging from the list of candidate perspectival expressions in Table 3.1, which contains only one temporal expression type (tense) and five spatial expression types. A more speculative hypothesis about the variability in *tomorrow* judgments is that temporal perspective is a particularly challenging kind of perspective for speakers to track or access. Perhaps the apparent rarity of perspective-sensitive temporal expressions and the interspeaker variability uncovered in this investigation are connected.

The conclusions reached in this chapter might be strengthened by work on other related expressions. In particular, it is an open question is whether non-utterance time readings can be found for other temporal indexicals, such as *yesterday*. Examples involving *yesterday* are more difficult to collect from corpus data, since the collocation of past tense cannot be used as an indicator of a shifted reading. Nonetheless, we might expect that speakers who have a perspectival semantics for *tomorrow* also have one for *yesterday*. Having multiple documented expressions in this class of perspectival expressions would make it easier to test some of the diagnostics that I did not apply, such as the existence of Shift Together effects.

---

<sup>40</sup>I omit quantifying the variability of *come* here; however, some relevant data is presented in Part II, in particular, the results of the comprehension studies in Chapter 7.

This chapter does not just contribute a case study of a particularly challenging kind of perspectival expression; it also highlights the difficulty in differentiating among classes of context sensitivity. I hope that it motivates more experimental work in this area, since experimental methods are particularly valuable in this domain. The environments in which perspectival expressions receive non-utterance-context interpretations overlap to a great extent with those in which indexical shift occurs. When indexicals participate in context shift, particular care must be taken to test the environments in which perspective shift, but not context shift, is predicted.

It seems likely that the perspective sensitivity of *tomorrow* has been overlooked for so long both because of the variability of speaker judgments and also because the licensing conditions of perspective-sensitive expressions overlap so closely with those of pure indexicals. The environments that decide among classes of context sensitivity are so rare that even well-studied expressions like American English *tomorrow* can turn out to have surprising complications. Investigating the seemingly exceptional uses of well-studied context-sensitive expressions can shed light on the landscape of context-sensitive reference in the spatiotemporal domain.

## CHAPTER 5

### INTERIM CONCLUSION

Part I has focused on the question of how the perspective holder is encoded in the semantics of perspectival expressions. In Chapter 3, I outlined four ways of encoding the perspectival component: a lexical stipulation approach; an indexical approach; a logophoric binding approach; and an anaphoric approach. I then presented a set of diagnostics for classifying perspectival expressions into these families of analyses.

In the remainder of Part I, I presented two case studies demonstrating how the set of perspectival diagnostics can be applied. In Chapter 3, I gave a case study focusing on a well-studied perspectival expression, the motion verb *come*, and concluded that it was best analyzed as an anaphoric expression, as proposed by Barlew (2017). In Chapter 4, I turned to a less well-studied phenomenon: the existence of non-utterance time readings of *tomorrow* in American English. I argued that *tomorrow* is perspectival, at least for some American English speakers, and tentatively suggested that it is best analyzed as perspective-anaphoric, though a logophoric binding approach may also be possible.

In this chapter, I conclude my discussion of the issue of perspective encoding by reviewing what we have learned about the landscape of perspective from the approaches that I have explored and from the two case studies I have presented.

#### 5.1 The theoretical landscape of perspectival expressions

Throughout Part I, I have argued that there are multiple ways that the perspective holder could be encoded in the semantics of perspectival expressions. In Chapter 3, I outlined four main approaches to perspective encoding and presented a set of diagnostics for distinguishing among them, presented again in Table 5.1.

**Table 5.1.** Predictions by account

	Lexical	Indexical	Logophoric	Anaphoric
Singleton perspective set	✓	X	X	X
Restricted perspective set	X	✓	✓	✓
Obligatory <i>de se</i> anchoring	◊	✓	✓	◊
Shift outside finite CP	✓	X	✓	✓
Shift outside XP with subj.	✓	X	X	✓
Shift Together effects	X	✓	✓	◊
Anchoring across utterances	X	X	X	✓
Covariation in quant. contexts	X	X	✓	✓

The set of perspectival diagnostics that I have proposed can be used to guide the analysis of a candidate perspectival expression by identifying a likely subdivision of the hypothesis space. I hope that it will be valuable for fieldwork on understudied perspectival expressions, as well as in refining our understanding of well-studied context-sensitive expressions. This is important because, as Chapter 4 illustrated, many classes of context-sensitive expressions appear identical outside of a few key environments.

As the proposed set of diagnostics illustrates, the licensing environments for lexical, indexical, logophoric, and anaphoric perspectival expressions display a subset-superset relationship. In every environment in which perspectival indexicals can shift, logophoric perspectival expressions can also shift, as noted by Sundaresan (2020b). In every environment in which logophoric perspectival expressions are licensed to shift, perspective-anaphoric expressions can also shift. To distinguish among classes of perspectival expressions, therefore, it is necessary to test the handful of environments that do not fall in the intersection of these different analyses.

### 5.1.1 A revised diagnostic toolkit for perspectival expressions

In Chapter 4, I narrowed the hypothesis space for a perspectival variant of *tomorrow* through a series of experimental studies, concluding that either a logophoric binding or perspective-anaphoric approach was possible. I then tentatively argued in favor of a perspective-anaphoric approach on the basis of a handful of examples showing licensing of non-utterance time *tomorrow* across utterance boundaries. I proposed a Barlew (2017)-style perspective-

anaphoric approach, but left open the possibility of a logophoric approach if future experimental work supported it.

It is important to note that within each of the four broad categories of perspectival analyses that I have proposed, there may be many possible variations. For example, in Appendix B, I sketched three variants on a logophoric approach in Appendix B, which, together with the perspective-anaphoric approach, form a complete set of logical outcomes for two of the perspectival diagnostics: whether Shift Together effects are obligatory in XPs with subjects and whether shifted readings happen outside of logophoric environments.

An updated set of diagnostics for perspective encoding that distinguishes among these approaches is illustrated in Figure 5.2. This revised toolkit distinguishes between obligatory Shift Together effects in the environments in which indexicals shift can shift (finite CPs), and obligatory Shift Together effects in the environments in which logophoric binding occurs (XPs with subjects).

**Table 5.2.** Predictions by account

	Lex.	Ind.	Ind.+Log.	Obl.	Log.	Opt.	Log.	Ana.
Singleton perspective set	✓	X	X		X	X		X
Restricted perspective set	X	✓	✓		✓	✓		✓
Obligatory <i>de se</i> anchoring	◊	✓	✓		✓	✓		◊
Shift outside finite CP	✓	X	✓		✓	✓		✓
Shift outside XP with subj.	✓	X	X		X	✓		✓
Shift Together in finite CP	X	✓	X		X	X		◊
Shift Together in XP with subj.	X	✓	X		✓	✓		◊
Anchoring across utterances	X	X	X		X	✓		✓
Covariation in quant. contexts	X	X	✓		✓	✓		✓

It would also perhaps be possible to further distinguish among anaphoric approaches, based on the method adopted for representing perspectives. In Chapter 3, I sketched two possibilities: Barlew (2017)'s approach, which treats perspectives as sets of centered worlds, and an approach that instead tries to derive perspectives from combinations of prominent individuals and times. Although I adopted a Barlew (2017)-style approach in Chapter 4, it requires enriching the ontology with perspective objects and enriching the Common Ground with

a mechanism for keeping track of perspective prominence. The latter approach, although fully worked out, might prove a more parsimonious alternative if the empirical evidence supports the reduction of prominent perspectives to prominent individuals and times. The choice between the two approaches impacts the issue of obligatory *de se* anchoring. Barlew (2017)'s centered worlds approach predicts obligatory *de se* anchoring, but the approach of deriving a perspective via prominent individuals and times does not.

### 5.1.2 perspective sensitivity versus context sensitivity

The issue of overlapping licensing conditions extends beyond perspectival expressions and into the broader landscape of context sensitivity. All perspective-sensitive expressions with restricted, non-singleton sets of perspective holders are context-sensitive, but not all context-sensitive expressions are perspective-sensitive. One observation about the landscape of perspective that emerges from Part I is that for each class of context sensitivity, there is a corresponding possible class of perspective sensitivity. For expressions that are perspectival but not context-sensitive, there is the lexical account. For perspectival expressions that are sensitive to the local grammatical context, there is the logophoric binding account.<sup>1</sup> For perspectival expressions that are utterance context-sensitive, there is the indexical account, with or without context shift operators. Lastly, for perspectival expressions that are discourse context-sensitive, there is the anaphoric account.

When analyzing a candidate perspectival expression, it may be useful to first establish the class of context sensitivity displayed by the expression before trying to distinguish among the different treatments of the perspectival component. This is the approach that I took in Chapter 4: I presented experimental evidence against viewing *tomorrow* as either a pure indexical or a time-anaphoric expression, before developing a perspectival analysis.

One drawback of this approach is that it is difficult to eliminate local context sensitivity without prior beliefs about what kind of operator might affect the meaning of the expression. The meaning of the expression may make some kinds of operators more plausible candidates

---

<sup>1</sup>Logophoric perspectival expressions are also sensitive to the discourse context, since in the logophoric analyses considered, there is an implicit logophoric pronoun. Whether there are perspectival expressions that are sensitive only to the local grammatical context is an interesting question for future research.

than others. For instance, a temporal expression like *tomorrow* might be expected to be governed by operators that bind times.

This approach may also lead to some diagnostic redundancy, since some diagnostics may need to be reapplied in order to test the perspectival variant of each kind of context sensitivity. For instance, in the work on *tomorrow*, we saw that *tomorrow* does not covary with temporal quantifiers ordinarily. However, it would be a mistake to conclude from this that *tomorrow* is indexical: as we saw, *tomorrow* can covary with quantifiers if they provide prominent perspectives that also covary. Thus, if the expression does turn out to be perspectival, it may be necessary to revisit the categorization of its context sensitivity.

Another kind of approach is to first establish the perspective sensitivity of a suspected perspectival expression. This is the approach that I followed for *come* in Chapter 3, motivated by previous literature establishing *come* as a perspectival expression.<sup>2</sup> If the perspectival nature of the expression is not established, there are a couple of ways the argument for perspective sensitivity can be made. If the expression is obligatorily *de se*-interpreted, this can motivate a perspectival account, since this is one of the hallmarks of perspectival expressions. If the interpretation of the expression shifts in the same environments in which the interpretation of other perspectival expressions whose behavior is well-understood, then it is likely to be perspectival.<sup>3</sup> Once the case that the expression is perspectival has been made, then the diagnostic toolkit presented in Chapter 3 can be followed.

### 5.1.3 Summary

Distinguishing among classes of context sensitivity can be challenging, given the overlap in licensing conditions among various classes of expressions. Introducing perspective sensitivity into the mix complicates the task further. Perspectival expressions can be analyzed as a subclass within any of the three existing classes of context sensitivity with minimal

---

<sup>2</sup>For instance, Barlew (2017)'s argument that *come* is obligatorily *de se*-interpreted.

<sup>3</sup>However, this approach assumes that the set of perspective holders is the same for all perspectival expressions (in the same language), which is an untested assumption.

additional assumptions. This means that there are a plethora of possible analyses for any candidate perspectival expression.

I have proposed a set of diagnostics to guide the analysis of perspectival expressions. However, as the exploration of *tomorrow* revealed, there are many possible variants within each of the four families of perspectival expressions that I have proposed. For certain expressions, the challenge of applying some of the diagnostics may make it infeasible to pinpoint a unique analysis. Instead, the diagnostics should be seen as a way of organizing the space of attested and theoretical perspectival expressions.

## 5.2 The attested landscape of perspective sensitivity

Part I has focused for the most part on the theoretical landscape of perspectival expressions: given the set of independently motivated semantic mechanisms, what are the ways in which it would be possible to encode the perspective holder? I have laid out four broad classes of perspectival expressions that are predicted to be possible: the lexical stipulation family, the indexical family, the logophoric binding family, and the anaphoric family.

A separate question is what classes of perspectival expressions actually exist. I have not gone very far towards answering this question. In Part I, I presented two case studies: a study of American English *come* and a study of American English *tomorrow*. In both cases, I ultimately argued against lexical and indexical approaches. In the case of *come*, I argued specifically in favor of a perspective-anaphoric approach, while taking Charnavel (2018)'s case for a logophoric treatment of French *venir* 'come' as evidence of cross-linguistic variation. In the case of *tomorrow*, I argued more tentatively in favor of a perspective-anaphoric approach, while leaving open the possibility of a logophoric approach.

Although I do not present any cases where an indexical or lexical approach is supported, I do not wish to argue against their existence. I have looked at only two kinds of perspectival expressions in a single language. As we have seen in Chapter 3, many kinds of expressions have been claimed to be sensitive to perspective. Much more work is needed before we can draw any firm conclusions about the actual, versus the theoretical, landscape of perspectival expressions.

My goal in outlining the four broad families of approaches to perspective has been to guide future work on the cross-linguistic typology of perspective. The licensing conditions predicted by these families can be visualized as a set of concentric circles: the licensing conditions of the anaphoric approach are a superset of those of the logophoric approach, which in turn are a superset of the indexical approach. This overlapping property makes it critical to pinpoint the right environments to test. I hope that the set of diagnostics that I have provided will make it easier to adjudicate among these families of approaches when conducting cross-linguistic work on perspective.

### 5.3 Open questions

Setting aside the question of the cross-linguistic landscape of perspective, there are a number of additional open questions about perspectival expressions raised by the findings in Part I. I survey them briefly in this section.

#### 5.3.1 Interactions among perspectival classes

In the set of diagnostics that I have proposed, the existence of Shift Together effects plays an important role in distinguishing among anaphoric and logophoric binding analyses. The Shift Together effects that I have discussed concern multiple instances of the same expression: given a perspectival expression A, a Shift Together effect holds if two instances of A within the same domain obligatorily receive the same interpretation. For instance, the critical example for *come* was (101), repeated as (155) below.

155. *Nick and Carolyn are siblings. Nick lives in Texas, Carolyn lives in Massachusetts, and their parents live in Washington. Nick is in Texas talking to his friend.*

Nick: Carolyn says that our mom will come to Texas during bluebonnet season and come to Northampton during asparagus season.

There may also be Shift Together effects among expressions from different perspectival classes. For instance, Sundaresan (2020b) claims that in Tamil, which has both shifty indexicals (in the form of shifty verb agreement) and perspectival anaphors, the presence of indexical shift constrains the interpretation of the anaphor. In a case of multiple embedding

like (156), where the verb agreement is unshifted, the perspectival anaphor *taan* can refer either to the nearest attitude holder (Seetha), or the higher attitude holder (Maya).

156. [Seetha<sub>i</sub> [taan<sub>i/j</sub> ḍjej-čč-aaŋ-nnū] nene-čč-aaŋ-ünnū] Mia<sub>j</sub>  
 Seetha ANAPH win-PST-3fs-COMP think-PST-3fs-COMP Maya  
 so-nn-aaŋ  
 say-PST-3fs

‘Mia<sub>j</sub> said [that Seetha<sub>i</sub> thought [that she<sub>i/j</sub> had won the contest]].’ (Sundaresan, 2020b)

However, if the agreement is shifted, the anaphor can only refer to the closest attitude holder, as shown in (157).<sup>4</sup>

157. [Seetha<sub>i</sub> [taan<sub>i/\*j</sub> ḍjej-čč-een-nnū] so-nn-aaŋ-ünnū] Mia<sub>j</sub>  
 Seetha ANAPH win-PST-1s-COMP say-PST-3fs-COMP Maya  
 so-nn-aaŋ  
 say-PST-3fs

‘Mia<sub>j</sub> said [that Seetha<sub>i</sub> said [that she<sub>i/\*j</sub> had won the contest]].’ (Sundaresan, 2020b)

Park (2014b, 2017) reports that a similar Shift Together constraint holds in Korean between shifty indexicals and long-distance reflexives.

Shift Together constraints between logophoric and anaphoric perspectival expression have also been reported. Charnavel (2018) reports a Shift Together constraint between exempt anaphors and *venir* ‘come’ in French. Although *venir* can usually receive either a speaker-oriented or an attitude holder-oriented interpretation, as shown in (158a) and (159a), when it co-occurs with an exempt anaphor, its perspective holder must be the same as the referent of the anaphor, as shown by the infelicity of (158b).

158. **Speaker-anchoring:** *The speaker is in Lyon. Claire’s son is in Paris.*

---

<sup>4</sup>Note that this effect does not follow from the mechanisms for indexical shift and logophoric binding proposed in Chapter 3 alone, since a manipulated context parameter should not block binding by a logophoric operator. Sundaresan (2020b) proposes a one-way structural restriction between logophoric operators and context shift operators to derive this Shift Together interaction.

- (a) [Le fils de Claire]<sub>i</sub> craint que la pluie n'empêche son<sub>i</sub> fils de venir à Lyon.  
 '[Claire's son]<sub>i</sub> is afraid that the rain prevents his son from coming to Lyon.'
- (b) # [Le fils de Claire]<sub>i</sub> craint que la pluie n'empêche son<sub>i</sub> propre fils de venir à Lyon.  
 Intended: '[Claire's son]<sub>i</sub> is afraid that the rain prevents his own son from coming to Lyon.'
159. **Attitude holder-anchoring:** *The speaker is in Paris. Claire's son is in Lyon.*
- (a) [Le fils de Claire]<sub>i</sub> craint que la pluie n'empêche son<sub>i</sub> fils de venir à Lyon.  
 '[Claire's son]<sub>i</sub> is afraid that the rain prevents his son from coming to Lyon.'
- (b) [Le fils de Claire]<sub>i</sub> craint que la pluie n'empêche son<sub>i</sub> propre fils de venir à Lyon.  
 '[Claire's son]<sub>i</sub> is afraid that the rain prevents his own son from coming to Lyon.'
- (Charnavel, 2018)

Charnavel (2018) reports that the same Shift Together constraint holds for the Mandarin exempt anaphor *ziji* and *lai* 'come'. This is shown in (160)-(161), where *lai* must be interpreted according to Lisi's perspective when the exempt anaphor is used, but can otherwise receive a speaker-oriented reading.

160. **Speaker-anchoring:** *The speaker is in Beijing. Lisi is in Tianjin.*
- (a) Ta<sub>i</sub> de haizi bu neng lai Beijing de xiaoxi shi Lisi hen shangxin.  
 'The news that his<sub>i</sub> child cannot come to Beijing makes Lisi<sub>i</sub> very sad.'
- (b) # Ziji<sub>i</sub> de haizi bu neng lai Beijing de xiaoxi shi Lisi hen shangxin.  
 Intended: 'The news that his<sub>i</sub> child cannot come to Beijing makes Lisi<sub>i</sub> very sad.'

161. **Attitude holder-anchoring:** *The speaker is in Tianjin. Lisi is in Beijing.*
- (a) Ta<sub>i</sub> de haizi bu neng lai Beijing de xiaoxi shi Lisi hen shangxin.  
 'The news that his<sub>i</sub> child cannot come to Beijing makes Lisi<sub>i</sub> very sad.'

(b)  $Ziji_i$  de haizi bu neng lai Beijing de xiaoxi shi Lisi hen shangxin.

'The news that his<sub>i</sub> child cannot come to Beijing makes Lisi<sub>i</sub> very sad.' (Charnavel, 2018)

Testing the existence of Shift Together effects between different kinds of perspectival expressions can be used as a diagnostic when probing the semantic encoding of perspective. For instance, Charnavel (2018) uses the data above to argue for an optional logophoric binding analysis of *venir* and *lai*: they contain perspectival variables that can be free, but are obligatorily bound by any logophoric operator in whose scope they occur.

However, testing the existence of Shift Together effects across classes of perspectival expressions is less useful as a diagnostic because it requires both a theory of how classes of perspectival expressions interact and an accepted analysis for at least one of the expressions tested. Charnavel (2018), for instance, first establishes that exempt anaphors are logophorically bound, and then uses Shift Together effects to motivate her analysis of Mandarin and French perspectival motion verbs. The strength of her analysis of these verbs rests on both her analysis of exempt anaphors and the judgments shown in (158)-(161).

In order to draw a conclusion about the perspectival encoding of one kind of perspectival expression from Shift Together behavior with another kind of perspectival expression, it is necessary to establish both the existence or absence of Shift Together effects and the semantic encoding of one of the perspectival expressions. This makes an argument from Shift Together effects among classes of perspectival expressions more complicated than an argument based on applying the perspectival encoding diagnostics directly.

### 5.3.2 Consistency of the perspective set

Another open question is whether the set of perspective holders is consistent across perspectival expressions in the same language. If *come* allows speaker, listener, and attitude holder perspectives, should we expect all other perspectival expressions in English to allow them as well?

In the literature on many perspectival expressions, the listener perspective is relatively little discussed. Although it is commonly discussed for perspectival expressions that have

a spatial component, like *come*, listener-anchoring is rarely explored in work on mental perspectival expressions like epithets and expressives.<sup>5</sup> Is this an accidental gap, or is it a reflection of the set of perspectives available for such expressions?

Of course, this question intersects with the question of perspective encoding: all else being equal, we might expect that within a language, perspectival expressions that encode perspective in the same way will allow the same set of perspective holders. For instance, if two expressions are both perspective-anaphoric, the values of their perspectival variables are both determined by the Common Ground, so we would expect them to share the same candidate perspective holders.<sup>6</sup>

The question of how consistent the perspective set is within a language is an interesting direction for future experimental work. Perspectival expressions could be probed in the same set of environments to test how consistently their perspectival anchorings are interpreted.

### 5.3.3 Variation

One of the most intriguing findings in Chapter 4 was the high degree of interspeaker variation in ratings of non-utterance time readings of *tomorrow*. Throughout the chapter, I focused mainly on providing an analysis of *tomorrow* for the American English speakers who do accept *tomorrow*. However, the source of the variability is in itself an interesting topic.

There are a number of possible sources. One possibility is that there is a difference in the grammar of speakers who do accept non-utterance time readings and those who do not. For some speakers, *tomorrow* is a pure indexical, while for others, it is perspectival.<sup>7</sup>

---

<sup>5</sup>An exception to this is discussion of listener-anchoring in questions, where perspectival expressions are commonly interpreted relative to the listener's perspective. There is a fairly extensive amount of work on this phenomenon under the label **interrogative flip** (Faller, 2002).

<sup>6</sup>However, it would also be possible for a perspectival expression to impose a lexical restriction on its perspective holder that is not shared by other expressions in the same class: we could imagine, for instance, a distal version of *come* that imposes a constraint that its perspective holder be distant; if this were the case, other perspective-anaphoric expressions would not have exactly the same set of perspective holders.

<sup>7</sup>Or, by the lexical ambiguity variant of the logophoric analysis sketched in Appendix B, for some speakers, there is a perspectival version of *tomorrow* available as well as the indexical one.

Another possibility is that the variability springs (in part or in whole) from variability in the prominence of different perspectives. In a perspective-anaphoric analysis of *tomorrow*, the perspectival variable is resolved via an anaphora resolution algorithm to one of the discourse-given perspectives in the Common Ground. Perhaps different participants calculate the prominence of perspectives differently. In this case, some speakers may not be able to access a perspective that licenses the non-utterance time reading.

There are several populations known to have more difficulty with perspective shift: young children (Loveland, 1984; Ricard et al., 1999; Köder and Maier, 2016), who struggle with shifts between first- and second-person pronouns; autistic individuals (Mizuno et al., 2011);<sup>8</sup> and older adults, who are reported to rely less on perspective shift in spatial relation tasks (Long et al., 2018). In the case of *tomorrow*, however, there is no evidence that the inter-speaker variability corresponds to any of these known population differences.<sup>9</sup> There are also well-documented individual differences in ability to perform perspective shift (Brown-Schmidt, 2009). The sample sizes of the Chapter 4 experiments are not large enough to explore individual differences among participants in detail; however, this is an interesting area for future work.

Another open question is whether the interspeaker variability found with *tomorrow* exists for other perspectival expressions. In experimental work on epithets, Duff (2018a) found that around half of the participants always chose speaker-oriented interpretations of the epithets when asked to judge items like the one in Figure 5.1. He failed to find a demographic variable that was a good predictor of this variability.

This suggests that interspeaker variability in the accessibility of different perspectives may be found across classes of perspectival expressions. This is a fruitful area for future work.

These questions relate directly to the question of how perspectives are identified, which will be addressed in Part II. However, they are also related to the question of how perspective is

---

<sup>8</sup>While I don't want to entirely omit mention of the extensive literature on autism and perspective-taking, I also want to acknowledge that many strands of this research are problematic and have been weaponized against members of the autistic community (Gernsbacher, 2007; Yergeau, 2013).

<sup>9</sup>See Appendix A for a discussion of the impact of various demographic features on *tomorrow* ratings.

**Figure 5.1.** Experimental item from Duff (2018a)

*Miranda knew that her supervisor gave her a negative performance review. The dirtbag is only nice to his tennis buddies.*

***Miranda's supervisor is a “dirtbag” according to who?***

- a. *Miranda*      b. *the narrator*

encoded. Whether or not variability in the accessibility of perspectives is expected to extend across classes of perspectival expressions depends on the theory of perspective encoding. Given the assumption that perspectives are tracked in the Common Ground, for instance, interspeaker variability observed for one perspective-anaphoric expression it should extend to others. This is not necessarily the case in other accounts of perspectival encoding: for instance, the lexical account relies on lexical ambiguity, which is not expected to vary uniformly across different perspectival expressions.

The question of variability in judgments related to perspectival expressions therefore spans across the questions addressed in both halves of this dissertation.

#### 5.4 Taking stock

In Part I, I have explored how perspective is encoded in the semantics of perspectival expressions. I have described four different ways of encoding perspective: the lexical approach, the indexical approach, the logophoric binding approach, and the anaphoric approach. I have also proposed a set of diagnostics to aid in classifying perspectival expressions into these families.

I presented two case studies to demonstrate the use of the proposed set of perspectival diagnostics. First, I applied the set of diagnostics to a well-known and well-studied perspectival expression: the perspectival motion verb *come*. I found that its behavior is most consistent with the perspective anaphoric-approach. Second, I explored novel uses of an expression not previously considered perspectival: American English *tomorrow*. I presented a series of experiments demonstrating that many speakers accept non-utterance time readings of *tomorrow* in contexts where pure indexicals are not predicted to shift. I argued that for these

speakers, *tomorrow* has a perspectival component, and should be analyzed as a member either of the logophoric or perspective-anaphoric families of perspectival expressions.

A number of interesting issues emerged from these case studies. First, some of the perspective proposed diagnostics were difficult to apply to *tomorrow*. In particular, testing Shift Together effects is challenging when the class of expressions that does not have many members. Second, the *tomorrow* case study revealed a high degree of interspeaker variation in the acceptability of shifted readings. This variation remains unexplained; it could arise from population-level variation in the semantics of *tomorrow*, or from individual variability in the accessibility of different perspectives.

Before concluding Part I, I want to reiterate one of the findings from the exploration of *tomorrow* in Chapter 4: even canonical instances of non-perspectival classes of context sensitivity may have overlooked perspectival properties. There is a large overlap in the licensing conditions not just among families of perspectival expressions, but also among context-sensitive expressions more broadly. The fine-grained diagnostics applied to *tomorrow* might also be applied fruitfully to other expressions that we think we understand well, because in order to confirm the classification of a context-sensitive expression, its behavior must be tested in exactly the right environments.

Throughout Part I, I have focused on the question of perspective encoding. In order to explore the semantics of perspectival expressions, I relied on the environments that set up contrasts between the predictions of different accounts: one analysis predicts felicity, while the other predict infelicity. However, as we saw in Chapter 2, there are a wealth of environments in which perspectival expressions can be interpreted in multiple ways. Exploring these environments will enable us to understand not just whether a perspectival expression can be used, but also, how it will be interpreted when it is used.

In Part II, I turn to a fresh set of questions about **perspective identification** and *perspective selection*. Given the availability of multiple perspective holders, how do listeners interpret perspectival expressions? How do speakers decide which perspective to adopt? These questions are relevant regardless of the theory of perspective encoding that is adopted, al-

though the form of the question changes slightly. For the lexical approach, it is a question of lexical ambiguity. For the indexical and logophoric approaches, it is a question of whether or not a silent syntactic operator is projected. For the logophoric and anaphoric approaches, it is a question of anaphora resolution. Regardless of how perspective is encoded semantically, the identity of the perspective holder is frequently underspecified. In Part II, I seek to develop an explanation of how the perspective holder's identity is determined by both speakers and listeners.

## **PART II: TAKING PERSPECTIVES**

## CHAPTER 6

### PERSPECTIVE SELECTION AND IDENTIFICATION

In Part I, I focused on the question of how the perspective holder is encoded in the semantics of perspectival expressions. In Part II, I focus on a fresh set of questions about perspectival expressions: how do speakers select a perspective to take, and how do listeners identify the perspective being used?

When a speaker uses a perspectival expression, they must decide whose perspective to adopt. This decision is necessary for any perspectival expression that allows multiple perspective holders. A theory of perspective encoding does not provide an answer to the problem of perspective selection, since every account of perspective encoding described in Part I involves some kind of underspecification or ambiguity in order to allow possible multiple perspective holders. For the lexical account, it is a question of lexical selection. For the indexical and logophoric binding accounts, it is a question of whether or not a certain kind of covert operator is projected. For the logophoric binding and anaphoric accounts, it is a question of pronominalization.

The listener faces an inverse decision when they interpret a perspectival expression: in order to understand the speaker's meaning, they must identify the perspective that the speaker has adopted. Again, the question of perspective identification arises regardless of how the perspectival component is encoded. When the listener interprets a lexical perspectival expression, they must reason about lexical ambiguity. When they interpret an indexical or logophoric perspectival expression, they must reason about the presence or absence of a covert operator. When they interpret a logophoric perspectival expression or a perspective-anaphoric expression, they must use an anaphora resolution strategy.

In this half of the dissertation, I address the question of perspective identification and selection. In this chapter, I give an overview of the issues involved and of previous approaches that have been taken. In Sections 6.1 and 6.2, I review previously proposed models of perspective identification and selection. I focus in particular on a proposal by Harris (2012), who probes the extent to which conversation participants reason about perspectives. In his proposal, processing perspectival expressions involves a two part system: a simple set of heuristics for perspective identification and a more costly perspective reasoning process.

Then I turn to some of the factors that have been proposed to play a role in these processes. In Sections 6.3 and 6.4, I review the empirical evidence in support of two cognitive biases: a bias towards our own perspectives, and a bias against perspective shift. In Section 6.5, I review some of the discourse factors that have been proposed to affect the perspective identification and selection processes.

In the following two chapters, I develop and implement a perspective reasoning model. In Chapter 7, I propose a model of the listener's perspective identification process based in the Rational Speech Acts framework (Bergen et al., 2012; Frank and Goodman, 2012). In this model, listeners jointly reason about the speaker's adopted perspective and intended meaning using a mental model of the speaker's production process. Using perspectival motion verbs as a case study, I show that this rational approach to perspective identification captures several key properties of the interpretation of perspectival expressions. I compare the predictions of the model with a simpler heuristic-based process in two comprehension studies. I show that the experimental evidence provides some support for the claim that listeners reason simultaneously over multiple possible perspectives.

In Chapter 8, I turn to the question of how speakers select which perspective to use. I present a rational production model mirroring the comprehension model presented in Chapter 7. I compare its predictions to those of a heuristic-based selection process in a production study and find that the predictions of the rational perspective production model are not supported by the production data. I also explore several modifications to the proposed production and comprehension models, but conclude that none of the theoretically motivated modifications can capture the observed production behavior.

## 6.1 Approaches to perspective identification

Throughout this dissertation, I have focused on perspectival expressions: expressions that reference, but do not introduce, a perspective. The perspectival expressions that I am interested in allow multiple perspective holders, which poses an interpretative problem for the listener: understanding the speaker's meaning requires understanding whose perspective the speaker has adopted.

Consider (162): if Poirot, the speaker, is using his own perspective, then the murderer's destination is Poirot's flat, but if he is adopting the perspective of the attitude holder, Japp, the murderer's destination is Scotland Yard.

162. *In his London flat, Poirot is telling Miss Lemon about a call he received from Chief Inspector Japp, at Scotland Yard.*

Poirot: Chief Inspector Japp thinks that the murderer will come to confess.

In order to figure out Poirot's intended meaning, the listener, Miss Lemon must identify the perspective that Poirot is using. How does she do this?

One possibility is that Miss Lemon has a small set of rules that she uses. A rule-based system would make processing perspectival expressions easy, provided that the rules are simple to apply. The simplest rule-based system would simply stipulate a default perspective holder. Since utterances are usually taken to express commitments of the speaker (Gunlogson, 2008), let us assume that the default perspective is the speaker's. A minimal rule-based system of perspective identification, therefore, is the **Speaker Default model**.

163. **Speaker Default model of comprehension (initial version):** when interpreting a perspectival expression, assume that the perspective holder is the speaker.

This minimal system has many drawbacks, however. An immediate one is that there are some circumstances in which the speaker simply cannot be the perspective holder. For instance, when the speaker describes their own motion using *come*, as in (164), they cannot be using their own perspective.

164. *From a telephone box, Poirot calls up Inspector Japp at Scotland Yard.*

Poirot: I have figured out who the murderer is. I will come to you immediately!

In this case, the truth conditions of the sentence do not allow a speaker-oriented interpretation of the motion verb. We can revise the definition of the Speaker Default model to take this possibility into account: if the speaker's perspective is ruled out by the truth conditions of the sentence, then the listener selects another perspective and reinterprets the sentence.

165. **Speaker Default model of comprehension (revised):** when interpreting a perspectival expression, assume that the perspective holder is the speaker. If the sentence is false under this interpretation, reinterpret it according to a different perspective.

Already our model has become underspecified: what perspective should we use when the speaker's is unavailable? One proposed solution is to use a ranked list of perspective holders.

### 6.1.1 Ranked list approaches

In work on long distance reflexives, Anand and Hsieh (2005) propose a set of rules for perspective identification. The speaker is the default perspective holder, but if their perspective is unavailable, the perspective of another conversation participant is used. If no conversation participant is available, third persons are considered.

166. **Anand and Hsieh (2005)'s ranked list approach:**

- (a) In unmarked contexts, the perspective center is the speaker.
- (b) When a conversation participant is the matrix subject, the perspective center is that conversation participant.
- (c) The perspective center can be a non-speech act participant in marked contexts (like narratives), where the 3rd person is established by discourse to be the perspective holder.

This specific set of rules does not generalize well to other classes of perspectival expressions; the second rule would not apply to perspectival motion verbs, since the subject of a motion verb cannot be its perspectival anchor. However, the idea of using the first available

perspective holder in a ranked list is an appealing approach to perspective identification. A more general ranked list model is described in (167).

167. **Ranked List model:** when interpreting a perspectival expression, use the highest ranked perspective holder. If the sentence is false under this interpretation, select the next highest ranked perspective holder and reanalyze it.

**Perspective holder ranking:**

- (1) Speaker
- (2) Listener
- (3) Subject of an attitude verb
- (4) Named third person

The Ranked List model of perspective identification works well for (164): since the speaker's perspective is eliminated by the truth conditions of the sentence, Inspector Japp will reanalyze the sentence according to his own perspective (the listener). For (162), the Ranked List model works less well. Since the speaker's perspective is available, the Ranked List model predicts an unambiguous speaker-oriented interpretation. The possibility of the attitude holder-oriented interpretation is never considered.

### 6.1.2 Harris (2012): a two part system

The Ranked List model predicts that non-speaker-oriented readings of perspectival expressions should not arise unless the speaker's perspective is inconsistent with the truth conditions of the sentence. Harris (2012) presents a series of experiments showing that this is not the case: comprehenders sometimes interpret epithets and appositives as non-speaker-oriented even when the speaker's perspective is available.

One of his studies is a comprehension task probing the availability of non-speaker-oriented readings of appositives.<sup>1</sup> The truth conditions of the critical sentences are consistent with either the speaker's perspective or that of the subject of an attitude verb (an example stimulus is shown in Figure 6.1). Despite the availability of the speaker's perspective,

---

<sup>1</sup>Originally reported in Harris and Potts (2009) .

**Figure 6.1.** Example appositive stimuli from Harris and Potts (2009)

<b>Experiment 1 example stimulus:</b>			
<b>Context:</b> I am increasingly worried about my roommate. She seems to be growing paranoid.			
<b>Embedded condition:</b> The other day, she told me that we need to watch out for the mailman, <b>a possible government spy</b> .			
<b>Unembedded condition:</b> The other day, she refused to talk with the mailman, <b>a possible government spy</b> .			
<b>Prompt:</b> Whose view is it that the mailman might be a government spy?			
<b>Response:</b>			
(a) Mine		(b) My roommate's	(c) Mine and my roommate's

participants selected the non-speaker-oriented reading in a majority of cases (Table 6.1). This shows that shifted readings of appositives do arise in contexts where the speaker's perspective is not ruled out by the truth conditions.

**Table 6.1.** Results from Harris and Potts (2009)'s Experiment 1

Condition	Subject	Speaker	Both
Embedded	110 (86%)	9 (7%)	9 (7%)
Unembedded	87 (68%)	32 (25%)	9 (7%)
Total	197 (77%)	41 (16%)	18 (7%)

Why do most participants report that the appositive conveys the attitude holder's perspective? Although the speaker-oriented interpretation does not falsify the sentence, it is surprising given the discourse context. To take the stimulus shown in Figure 6.1 as an example, the speaker states that they are worried about their roommate and describes them as 'paranoid'. This sets up an expectation that the speaker and their roommate have differing beliefs about some topic of significance. Moreover, the discourse structure is such that the critical sentence is most naturally interpreted as a justification for the speaker's statement that their roommate is growing paranoid.

Harris (2012) recognizes that the discourse context provides many cues about the perspective in use. Although comprehenders seem to use such cues, neither the Speaker Default nor the Ranked List model of perspective identification are sensitive to the discourse context.

Harris (2012) proposes an alternative model of perspective identification with two components. First, there is an inference system that is able to use all of the available evidence to reason about the perspective holder: the truth conditions of the sentence, the discourse context, and any other information available to the listener.<sup>2</sup> The second component is a simpler heuristic-based model. Because it is expensive for comprehenders to reason about the perspective holder, they rely on this second, cheaper system whenever possible. Harris (2012) writes:

The central proposal is that, in the normal case, the language processor avoids a costly and highly fallible abductive reasoning process by employing a presumptive pragmatic default favoring speaker-orientation. This default is overturned when the result is not coherent, provided that there are enough cues in the input to signal that the speaker intends a non-speaker oriented interpretation. (Harris, 2012, p. 55)

The two-part model of perspective identification is attractive because it balances processing ease with the ability to reason over a variety of discourse cues. However, there are a number of aspects of the proposal that remain to be worked out.

One key issue is the question of how the switch to the costlier inference system is triggered. As I have argued above, it is not enough to invoke perspective shift in situations where the truth conditions of the sentence rule out the speaker's perspective. But if the switch is not triggered by falsification, what triggers it? Harris (2012) writes as if evidence of a perspective shift is accumulated gradually and eventually becomes significant enough to trigger the perspective inference process. However, it is unclear to me how the listener realizes that enough evidence of a non-speaker perspective has accumulated without some sort of continuous reasoning over the active perspective.

The second issue is the nature of the two components of the model. Harris (2012) is most explicit about the simpler heuristic system. He proposes that it consists of two rules. The first is a speaker default rule, as articulated in the Speaker Default model:

---

<sup>2</sup>Harris (2012) mentions many kinds of cues that might signal a perspective shift, including quotation marks, gestural markers, prosody and intonation, attitude predicates, aspect, and mood.

168. **Speaker Default rule of comprehension:** when interpreting a perspectival expression, assume that the perspective holder is the speaker.

The second is a preference for maintaining perspective. Harris (2012) posits that a single active perspective is tracked in the conversational scoreboard and that the comprehender should assume that this perspective is the anchor for any perspectival expressions used by the speaker.

169. **Maintain Perspective rule of comprehension:** when interpreting a perspectival expression, assume that the perspective conveyed is the same as the last perspective conveyed.

Whenever possible, comprehenders use these rules to interpret perspectival expressions. If the resulting interpretation is sufficiently incoherent, however, the comprehender can fall back on the perspective reasoning system to recalculate the active perspective.

The Speaker Default rule may seem very limited if Maintain Perspective is also at play, since it will only affect the interpretation of the very first perspectival expression in a discourse. However, Harris (2012) suggests that because maintaining non-speaker perspectives involves reasoning about the beliefs of others, it is cognitively costly, and the active perspective may drift back to the speaker when the cost of perspective maintenance becomes too high.<sup>3</sup>

Harris (2012) is less explicit about the second component of the model. He posits that it is an abductive inference process that takes into account the discourse context and what is

---

<sup>3</sup>Harris casts this as cooperative behavior on the part of the speaker, writing:

Unlike simple reference points, maintaining a non-speaker perspectival center requires that the processor reason about the beliefs of others, which may be independently taxing. As such, I proposed that shifted contexts are constrained by the cognitive resources it takes to model and assess third-party beliefs, and that the processor may revert back to a speaker-oriented context, unless additional cues indicate otherwise. A cooperative speaker will recognize that maintaining a non-speaker perspectival center requires resources, and provide sufficient cues to guide her interlocutors. (Harris, 2012, p. 162)

However, it's unclear to me why it is easier for the listener to maintain a representation of the speaker's beliefs than to maintain a representation of the beliefs of a third party. It seems more natural to me to reformulate this in terms of the listener knowing that maintaining a non-speaker perspective might be cognitively difficult for the speaker.

known about the candidate perspective holders. He leaves an explicit model of this system for future work.

In Chapter 7, I propose one way such a reasoning system might be implemented. The system that I describe is based in the Rational Speech Acts framework, and posits that listeners reason simultaneously about the speaker’s intended meaning and their adopted perspective using a mental model of the speaker’s production process. Instead of modeling a single active perspective in the conversational scoreboard, I posit that the conversational scoreboard tracks probability distributions over possible worlds and over prominent perspectives. The perspective with the highest probability in my system is analogous to the active perspective in Harris (2012)’s system.

## 6.2 Approaches to perspective selection

What about perspective selection? Perspective selection is more challenging because the speaker’s production process is less constrained. When the listener is attempting to identify the perspective being used, they have access to knowledge about the discourse context as well as knowledge of the truth conditions of the speaker’s utterance. Taken together, the listener’s search space can be narrowed by the incompatibility of some perspectives with the truth conditions of the sentence.

The speaker, on the other hand, has access to the discourse context and some meaning that they wish to convey (in the form of a proposition or an observed possible world or set of possible worlds). However, the set of utterances that they could use to describe this meaning is large. While the listener’s search is reduced by the fact that some perspectives are not compatible with the truth conditions of the sentence, the speaker is not constrained in this way. Consider again the example used to motivate the availability of non-speaker anchoring for *come*, (164), repeated as (170).

170. *From a telephone box, Poirot calls up Inspector Japp at Scotland Yard.*

Poirot: I have figured out who the murderer is. I will come to you immediately!

If the listener tries to adopt the speaker's perspective when interpreting (170), they will arrive at a contradiction, because the speaker cannot both be in motion and located at the destination of motion. The truth conditions of the sentence will guide them towards a listener-oriented interpretation of *come*. The speaker, on the other hand, receives no such cue. If the speaker wants to use their own perspective, they can: they simply have to pick a different verb, as in (171).

171. *From a telephone box, Poirot calls up Inspector Japp at Scotland Yard.*

Poirot: I have figured out who the murderer is. I will go to you immediately!

Because the speaker's search space is less constrained, it is a more challenging task to design a set of rules by which the speaker selects a perspective.

### 6.2.1 Rule-based models of perspective selection

The relative lack of constraint of the speaker's selection process leads a rule-based perspective production model to make very strong predictions. Consider the simplest rule-based model of perspective selection, the Speaker Default model.

172. **Speaker Default model of production:** when using a perspectival expression, use your own perspective. If this results in falsehood, select a different perspective.

As we have just seen, however, the speaker always has a true alternative that they can use (*go*), so the second clause will never come into play. The Speaker Default model of production therefore predicts that the speaker should never shift perspectives. But we have already seen that this is not the case. Speakers can and do shift perspective.

One possibility is to adjust the Speaker Default model by proposing that sentence production consists of two sequential parts: the task of sentence selection and the task of perspective selection. The Speaker Default model is equivalent to selecting the perspective first, and then the sentence: it predicts that the speaker's perspective will always be used. Instead, we could explore a model in which the speaker first selects the utterance, and then the perspective.

173. **Sequential Speaker Default model of production:** first select an utterance. Then select a perspective, starting with your own. If your perspective would make the sentence false, select a different perspective.

But this is counterintuitive. How is the speaker to decide between using *I will come* and *I will go* in this model? Without knowing the perspective, there is nothing to guide the decision between the two verbs.

For this reason, the simplest heuristic systems of perspective selection are implausible.

We might explore a slightly more complex rule-based system of perspective production by assuming that there are heuristics that encode lexical preferences as well as perspective preferences. One rule that is often implicit in theoretical work on the semantics of *come* and *go* is a **Come Default** (Oshima, 2006a; Barlew, 2017): if there is any perspective that would license the use of *come*, use *come* instead of *go*.

174. **Come Default model of production:** try to describe the motion event using *come* anchored to the speaker's perspective. If this would make the sentence false, select a different perspective. If no perspective would make it true, use *go* instead of *come*.

This model predicts that *go* will not be used in a situation where there is any licit perspective holder for *come*. This prediction does not seem borne out by the data; in (175), the attitude holder's perspective is available, yet *go* is used.

175. *Miss Marple is asking Lawrence Redding about his movements on the day Colonel Protheroe was murdered.*

Marple: The vicar's wife said that you went to Vicarage in the early afternoon on the day of the murder.

Moreover, it is not clear how such an approach could be generalized to other classes of perspectival expressions, such as epithets and appositives, which do not come in pairs like *come* and *go*. This suggests that a plausible rule-based model of perspective production would need to incorporate a richer set of factors into its rules.

What about a two-part model of perspective selection? Although Harris (2012) explores the behavior of speakers experimentally, his proposed model is a model of perspective identification, not perspective selection. In the next section, I attempt to outline a version of his system adapted for the task of perspective production.

### **6.2.2 A two-stage model of perspective selection**

Harris (2012) does not explicitly address the question of perspective selection. However, we can imagine a two-stage model of perspective selection that parallels his model of perspective identification. The two-stage model of perspective identification was motivated by concerns about the cost imposed by reasoning. If it is costly for the listener to reason over cues from the discourse context in order to identify the perspective, then it should be similarly costly for the speaker to reason over this evidence in order to select a perspective. As we have seen above, however, the simplest heuristic-based systems do not model perspective production very well. Perhaps speakers, like listeners, rely as much as possible on a simple heuristic-based model, while switching to a more costly reasoning system when necessary.

What kind of heuristic system should the speaker use? All else being equal, a reasonable guess is that the same perspectival pressures govern the selection of perspectives and their identification. We can therefore sketch a heuristic-based system that relies on Harris (2012)'s two rules for perspective identification: Speaker Default and Maintain Perspective.

176. **Speaker Default rule of production:** when selecting a perspectival expression, prefer one that is consistent with the perspective of the speaker.

177. **Maintain Perspective rule of comprehension:** when selecting a perspectival expression, prefer one that is consistent with the last perspective used.

If a non-speaker perspective has already been adopted, Maintain Perspective will predict that this perspective will continue to be used. However, from the production standpoint, perspective shift is still a mystery: given these two rules alone, it is not clear why the perspective would ever shift away from the speaker to start with.

Since we are entertaining a two-part model, the initial perspective shift might happen under the more costly reasoning system. However, this is not much help. One of the most underspecified components of the two-stage model of perspective processing proposed by Harris (2012) is the mechanism for triggering switches between the default and reasoning-based components of the model. In the production model, this is an even bigger mystery, since we cannot appeal to a conflict between the active perspective and the truth conditions of the sentence.<sup>4</sup>

Thus, modeling the perspective selection task required for producing perspectival expressions, whether with a rule-based system or a two-stage model, is more challenging than modeling perspective identification. In Chapter 8 I propose a reasoning-based model of perspective selection, motivated by the difficulty of finding a sufficient set of rules with which to build any simpler model. However, as I will discuss in that chapter, the empirical evidence suggests that some simpler heuristic model might be a better fit for actual speaker behavior. Uncovering a sufficient set of rules to govern perspective selection in production remains a challenging task for future work.

### 6.2.3 Summary

Harris (2012) proposes that at least in some contexts, processing perspectives involves reasoning over various cues. In the following two chapters, I will propose, implement, and test models of perspectival reasoning in comprehension and production. However, I have not yet discussed what kinds of cues this reasoning system operates on.

In the remainder of this chapter, I briefly review factors that have been proposed to influence perspective identification. I start by discussing the empirical support for the two principles invoked in Harris (2012)'s heuristic system: in Section 6.3, I discuss the prominence of the speaker's perspective, and in Section 6.4, I discuss perspective maintenance. I then turn in Section 6.5 to some of the discourse-level factors that have been proposed to affect the interpretation of perspectival expressions. I conclude by previewing the structure of the remaining chapters in Part II.

---

<sup>4</sup>And, as I argued in Section 6.1, this is too restrictive of a trigger even in the comprehension model.

### 6.3 A bias towards the speaker

One of the heuristics that Harris (2012) proposes for guiding perspective identification is the Speaker Default. The argument for this default can be made from general pragmatic principles. Since assertions are interpreted as expressing speaker commitments (Gunlogson, 2008), we might infer that the default interpretation of any expression is speaker-oriented.<sup>5</sup> In addition, as Harris and Potts (2009) points out, perspective shift is a risky conversational move for the speaker because if the listener fails to understand that you have shifted perspective, they will misattribute the beliefs that you are reporting to you:

A speaker who utters an appositive with the intention of having it be understood as non-speaker-oriented has undertaken a risky communicative strategy in the following sense: it runs counter to hearer expectations about how these constructions will be used. Thus, this is a reliable strategy only in contexts that are rich enough to support another perspective in just the right ways. To put it another way: you might always be free to intend your appositive to be understood as non-speaker-oriented, but your audience will often be unable to recover your intentions. (Harris and Potts, 2009)

Because perspective shift runs the risk of being overlooked, it increases the likelihood of your contribution being judged as uncooperative on the grounds of falsehood. A rational speaker will therefore choose to use their own perspective unless the discourse context is rich enough to guarantee that their perspective shift will be tracked by the listener.

Aside from the argument from pragmatic principles, there is also a wealth of empirical evidence in support of a bias towards the speaker's own perspective. Because this is a topic that has been explored in many different research communities, however, the terminology used and the phenomena explored vary from community to community. In this section, I provide a brief overview of the relevant findings from each subfield; however, in many cases, the literature is extensive and there is much interesting work that I will be forced to elide.

In psychology, there is a much-debated and large body of work on **egocentricity** effects: a person's bias towards their own perspective. Much of this work focuses on the distinction between privileged and shared information during conversation. Most of the work in this

---

<sup>5</sup>Except for expressions whose lexical semantics encode a different perspective.

community does not directly address the interpretation of perspectival expressions, focusing instead on the primacy of the speaker as a broader cognitive pressure.

The literature on child language acquisition is more mixed. There is a sizable body of work discussing egocentricity as it relates to the development of **theory of mind**: the ability to maintain representations of other people's cognitive states. The majority of this work focuses on pronoun acquisition. There is a smaller body of work on the acquisition of perspectival motion verbs, and very little work focusing on mental perspectival expressions like epithets, appositives, or predicates of personal taste.<sup>6</sup>

The adult processing literature includes some work focused on the extent to which the speaker's perspective is the preferred interpretation for perspectival expressions. This growing body of work focuses mostly on the interpretation of epithets and appositives in English.

There is also some cross-linguistic evidence in support of a default speaker interpretation of perspectival expressions, although larger cross-linguistic studies of perspectival expressions are needed. The existing research suggests an implicational hierarchy of perspectival anchoring dominated by the speaker.

### 6.3.1 Egocentricity effects in conversation

The existence of an **egocentricity** bias, the bias of an individual towards their own point-of-view, has been well studied in the social psychology community. Egocentricity does not seem to be a language-specific pressure: individuals exhibit egocentric biases in a variety of domains. For instance, individuals tend to overestimate how many hours a week they work (Frazis and Stewart, 2014).

Within the linguistic domain, most work has been on egocentricity effects in two kinds of conversational contexts: contexts that set up a contrast between the visual perspective of the speaker and listener, and contexts that set up a contrast between the knowledge states of the speaker and listener. I will briefly summarize some of the relevant findings below, but

---

<sup>6</sup>There is work on the acquisition of the syntax of these expressions (particularly appositives), but not, as far as I am aware, on the acquisition of shifted interpretations.

Brown-Schmidt and Heller (2018) provide a more comprehensive review of this extensive literature.

Egocentricity effects in these environments have been much debated. Early work on the topic revealed surprisingly common failures to consider the Common Ground, which led to the view that both speakers and listeners are strongly egocentric, at least in automatic processing (Horton and Keysar, 1996; Epley et al., 2004; Keysar and Barr, 2003; Epley et al., 2004; Lin et al., 2010). However, more recent work has called this view into debate. In earlier work, the use of any novel information by a speaker was interpreted as evidence that the speaker failed to consider the listener's perspective. As Brown-Schmidt and Hanna (2011) points out, however, one of the goals of conversation is to exchange information; a speaker who never referenced novel information would likely be considered uncooperative.

More recent findings have supported a less drastically egocentric view of conversation participants. Heller et al. (2008) showed that listeners use information about the speaker's visual perspective when interpreting temporarily ambiguous referring expressions. For instance, in a context where two large objects are visible to the listener, but only one is visible to the speaker, a listener who hears *the big* ... will anticipate the object that is visible to the speaker. Speakers also have been shown to take into account the knowledge state of the listener: when participants are taught names for unknown objects, they rarely use the names when talking to listeners who have not been taught the names (Heller et al., 2012).

The assumption underlying much of this work is that speakers are naturally egocentric; the debate is about the strength of the bias. This means that the experimental paradigms have largely been designed to encourage non-egocentric behavior. Nonetheless, the findings from both visual perspective-taking and mental perspective-taking support the existence of an egocentric bias. Heller et al. (2012) showed that speakers do not always consider the listener's knowledge state,<sup>7</sup> and Ferguson et al. (2017) found in an eye tracking study of visual

---

<sup>7</sup>Heller et al. (2012), for instance, found that speakers use the unfamiliar terms to refer to objects 5% of the time.

perspective that accuracy was reduced and reaction times were slower when participants were asked about a character's perspective instead of their own.

### 6.3.2 Theory of mind and language acquisition

**Theory of mind** is the ability to understand that other people's beliefs may differ from one's own (Brown-Schmidt and Heller, 2018). This is a prerequisite for adult-like processing of perspectival expressions. The development of theory of mind is a critical milestone for children, as it affects linguistic and non-linguistic behavior alike, and therefore has been a topic of much research in language acquisition and development psychology.

There is ongoing debate about how early children acquire theory of mind. Early studies suggested that children are not able to ascribe false beliefs to others until around age 4. However, more recent work using implicit measures of attention like eye-tracking suggests that children develop this ability as young as 13-15 months, but are unable to act based on their understanding until later (de Villiers and de Villiers, 2014).

Much work in this area has focused on the acquisition of first- and second-person pronouns, since children must learn that *you* and *I* change their referents when conversational roles shift. Pronoun reversals have therefore been interpreted as a failure to shift perspective. Pronouns are often mastered at about the same time that children learn to distinguish differing spatial perspectives, leading some researchers to hypothesize a relationship between the two acquisition tasks (Loveland, 1984; Ricard et al., 1999).

In comparison, there has been relatively little work on the acquisition of canonical perspectival expressions. Although children differentiate among conflicting spatial viewpoints around age 2 (Loveland, 1984), they do not use perspectival motion verbs in an adult-like way until much later (5-8 years).<sup>8</sup> Children seem to first learn a simpler, non-perspectival heuristic: they use *come* to describe any motion towards the scene. Because the speaker and listener are frequently located at the scene, this causes their usage appear adult-like

---

<sup>8</sup>There are conflicting findings about the age that children acquire adult-like use of *come* and *go*: Clark and Garnica (1974) report that it is around age 8, while Richards (1976) documents adult-like performance in the same contexts around age 4-5. Winston (1988) concludes that the tasks used in Clark and Garnica (1974) may have required other cognitive skills that the younger children had not yet acquired.

earlier than it actually is. Winston (1988) interprets the data as suggesting that children have adult-like semantics fairly early, but are more willing to shift perspective than adults.

To summarize, the acquisition literature provides mixed support for a speaker default. Young children struggle to use their knowledge about the false beliefs of others, providing some evidence that their own perspective interferes with their ability to consider those of others. However, even young children understand that both mental and spatial perspectives differ, and they do not seem to use a speaker default for perspectival motion verbs.

### **6.3.3 Speaker bias in adult processing of perspectival expressions**

There is a small but growing body of literature on the processing of perspectival expressions that supports the idea of a default speaker orientation. Most of this work focuses on epithets and appositives. Early work on these classes of perspectival expressions assumed that they were obligatorily speaker-oriented (Potts, 2005, 2007b).<sup>9</sup> However, a critique of this view by Amaral et al. (2007) stimulated a line of work evaluating the shiftability of such expressions.

Harris and Potts (2009) present experimental and corpus studies of the shiftability of appositives. In their experiment, participants preferred speaker-oriented readings 32% of the time for unembedded appositives and 9% for appositives embedded under attitude predicates. Subsequent work by Harris (2012) and Kaiser (2015) explored manipulations of various discourse factors. Their results support the view that appositives commonly receive speaker-oriented interpretations, but that shifted readings are fairly easy to induce through a combination of prosodic cues and discourse context factors that indicate multiple distinct perspectives (such as attitude verbs and expressive content). Depending on the environment tested, the rates of speaker-oriented interpretations ranged from 9% to 55% (Harris, 2012).

The shiftiness of epithets has also been explored in a number of experimental studies (Harris and Potts, 2009; Kaiser, 2015; Harris, 2012; Duff, 2018b). Depending on the environment in

---

<sup>9</sup>Despite the clear evidence of shiftability of epithets and appositives presented in Potts' later work with Harris, his earlier claims have persisted in subsequent work on these expressions. In Patel-Grosz (2015)'s extensive work on epithets, for instance, she describes epithets as "generally speaker-oriented (with rare exceptions)", citing both Potts (2007b) and Harris and Potts (2009).

which they occur, the rate of speaker-oriented interpretations of epithets ranges from 24% to 88% (Harris, 2012).

Taken together, these findings support the idea of a default speaker orientation for epithets and appositives. However, it is difficult to estimate the strength the bias from the experimental work. Most of the studies were designed to test the effect of a context factor on the rate of perspective shift. As a result, some use discourse contexts that bias towards a shifted reading (Harris and Potts (2009)'s Experiment 1), while others use contexts that bias towards the speaker's perspective (Harris and Potts (2009)'s Experiment 2).

Harris and Potts (2009) also report a small corpus study of appositives. Of the 34 naturally-occurring appositives for which the annotators agreed there was good textual support for one interpretation or the other, only 6% were non-speaker-oriented. This finding is one of the most directly relevant to the question of a speaker bias.

#### **6.3.4 Cross-linguistic evidence for a speaker preference**

The cross-linguistic landscape of perspectival expressions is not well-studied. From the data that is available, however, a cross-linguistic preference for the speaker's perspective emerges.

First, as discussed in Chapter 2, the cross-linguistic work on the semantics of perspectival motion verbs reveals a preference for speaker-anchoring. In Gathercole (1987)'s and Nakazawa (2007)'s cross-linguistic surveys, all languages allow the speaker as a perspective holder. There is also an implicational hierarchy: if the language allows any event time perspective, it allows the event time perspective of the speaker.<sup>10</sup>

Second, the cross-linguistic work on evidentials in Korotkova (2016) also suggests an implicational hierarchy favoring the speaker. In all of the languages surveyed, evidentials can be speaker-oriented; in most, they shift to the perspective of the listener when used in questions; and in a smaller number of languages, they shift optionally or obligatorily in attitude reports (Table 6.2).

---

<sup>10</sup>See Table 2.1 in Chapter 2.

**Table 6.2.** Cross-linguistic data on the orientation of evidentials from Korotkova (2016)

	Cheyenne	Quechua	Georgian	Tagalog	Tibetan	Turkish
Appear in attitude reports	X	X	✓	✓	✓	✓
Subj.-oriented in attitude reports	N/A	N/A	X	□	□	◊
List.-oriented in questions	✓	✓	N/A	✓	✓	✓
Quotee-oriented in quotes	X	✓	X	✓	X	X

N/A = not available; X = not allowed; ✓ = allowed; □ = must shift; ◊ = may shift  
Data sources: Georgian, Turkish (Korotkova, 2016); Cheyenne (Murray, 2010); Cuzco Quechua (Faller, 2002); Tagalog (Schwager, 2010); Tibetan (Garrett, 2001)

I am not aware of cross-linguistic comparisons of the allowed perspective holders for other perspectival expressions, such as epithets, predicates of personal taste, or expressives.

### 6.3.5 Summary

Research in a variety of subfields has explored the existence of an egocentricity bias. The assumption guiding much of this work is that it is easiest for individuals to access their own perspectives, since every individual has direct access to their own perspective, but must use theory of mind to access those of others. For this reason, many studies have focused on establishing the strength of egocentricity effects, rather than their existence.

Although much of the work that I have discussed focuses on egocentricity as a general cognitive pressure, the processing literature on epithets and appositives provides evidence of a default speaker interpretation for perspectival expressions in particular. However, estimating the strength of this bias is difficult, as it varies according to the discourse context.

One caveat to the converging evidence that I have presented is that the existence of a default speaker interpretation does not necessarily follow from the existence of an egocentricity bias. If egocentricity is a cognitive bias, then it should influence the listener's behavior as well as the speaker's. From a processing standpoint, an egocentricity bias should bias the listener towards their own perspective, not that of the speaker. Therefore, the findings of a default speaker interpretation in the processing of epithets and appositives require further

explanation: the listener's egocentricity bias should not make the speaker's perspective easier to access than that of an attitude holder.

Connecting the frequency of speaker-oriented interpretations of epithets and appositives with a cognitive egocentricity bias requires an additional assumption of some kind. For instance, we might posit that listeners are aware that speakers are egocentric, or that the speaker's perspective is more often similar to the listener's own perspective. Or we might conclude that the default speaker interpretation of epithets and appositives arises from the second processing pressure that Harris (2012) proposes: a penalty for perspective shift.

## 6.4 The cost of perspective shift

The second component of Harris (2012)'s proposed heuristic system is a preference for maintaining the currently active perspective. This is rooted in empirical evidence that perspective shift imposes a cognitive cost. However, because most of this evidence comes from debates about egocentricity, it can be difficult to disentangle the evidence of the costliness of perspective shift from evidence of the costliness of non-speaker perspectives. In the next two sections, I briefly review the relevant literature, doing my best to separate these two factors.

### 6.4.1 The cognitive cost of perspective shift

There is some work in psychology and psycholinguistics on the cognitive effort involved in accessing perspectives. Speakers have been shown to become more egocentric when under time pressure (Horton and Keysar, 1996) or cognitive load (Lin et al., 2010), a finding that has been interpreted as evidence that accessing other people's perspectives is a cognitively difficult task. However, like many studies, these findings do not differentiate the cost of shifting perspective from the cost of accessing non-speaker perspectives.

A handful of studies do distinguish between these two factors. First, Köder et al. (2015) found that comprehenders make more pronoun interpretation errors when processing direct

speech versus indirect speech. This suggests that the switch from the speaker’s perspective to that of the quotee is challenging.<sup>11</sup>

Second, Millis (1995) found that reading times were slower for sentences that introduced a new perspective; accessing the same perspective later in the text, however, was not associated with slower reading times. This suggests that there is a cost for calculating a new perspective, but not for perspective shift to a previously accessed perspective.

In an eye tracking study on visual perspective-taking, on the other hand, Ferguson et al. (2017) finds that reaction times are slower when switching between visual perspectives, regardless of whether the perspective being switched to is familiar (and regardless of whether it is the speaker’s). Of course, visual perspective taking may differ from linguistic perspective taking.<sup>12</sup>

Child et al. (2020) present evidence for a more gradual transition between perspectives. They compare processing times for texts describing subjective narrated in the second person to ones in the third person (Figure 6.2). They find that at the beginning of a text, participants read the third-person narration more slowly than the second-person narration. As the narrative progressed, however, their reading times increased for the third-person pronouns, but not the second-person pronouns. This suggests that the third-person perspective became gradually more prominent over the course of the narrative.

Last, Harris (2012) presents some experimental evidence in favor of perspective maintenance. His Experiment 7 probed whether participants would interpret a sentence following a speech report from the perspective of the attitude holder or the speaker (example stimulus shown in Figure 6.3). Some participants read the sentences, while others listened. Regard-

---

<sup>11</sup>Whether or not this can be interpreted as an egocentricity bias depends on the analysis of pronoun interpretation adopted: if interpreting pronouns involves adopting the speaker’s perspective, an additional penalty for direct speech cannot be explained as an egocentricity bias, since both involve perspective shift on the part of the listener (Wechsler, 2010; Köder and Maier, 2016). However, if first-person pronouns are interpreted as referring to the speaker in the current utterance context, rather than as self-ascription, then the direct speech condition involves perspective shift while the indirect speech condition does not, and the effect can be interpreted as an egocentricity effect.

<sup>12</sup>See Carruthers (2016) for further discussion.

**Figure 6.2.** Example texts from Child et al. (2020)

**Second person narration:** *With a full bag in your hand, you make your way home. It feels quite heavy, but that does not really matter. You had assumed that you would have to spend so much more today. You had been trying to save up for a while, and this was a real bargain. You look at your bag with great satisfaction.*

**Third person narration:** *With a full bag in his hand, Peter makes his way home. It feels quite heavy, but that does not really matter. He had assumed that he would have to spend so much more today. He had been trying to save up for a while, and this was a real bargain. He looks at his bag with great satisfaction.*

**Figure 6.3.** Harris (2012) Experiment 7 stimulus

**Matrix-Past condition:**

There might be a storm today, said Mary. Clouds had been brewing all morning.

**Matrix-Present condition:**

There might be a storm today, said Marry. Clouds have been brewing all morning.

**Subordinate-Present:**

Mary said that there might be a storm today. Clouds have been brewing all morning.

less of the tense and the conversational role of the participant, most interpreted the second sentence from the perspective of the attitude holder (50%-80% of the time).

To summarize, there is some evidence in support of a cognitive cost for perspective shift. However, the existing work makes it difficult to distinguish three hypothetical kinds of perspective shift costs: (1) a one-time cost paid the first time a particular non-speaker perspective is accessed; (2) a cost to accessing non-speaker perspectives that applies every time the perspective is accessed; or (3) a cost to shifting the active perspective, as Harris (2012) proposes. Further work is needed to differentiate among these possibilities.

#### 6.4.2 Individual differences in perspective shift ability

There are also well-documented individual differences in ability to perform perspective shift. Among adults, the ability to distinguish privileged versus shared information has been linked to executive function and working memory (Brown-Schmidt, 2009; Lin et al., 2010; Wardlow, 2013), although other studies have failed to replicate the link, potentially because of the low reliability of the measures of interest (Ryskin et al., 2015).

There are also populations that have been shown to have more difficulty with perspective shift. As mentioned in Section 6.3.2, there is some evidence that children struggle with perspective shifts, whether due to incomplete theory of mind, or due to lack of ability to inhibit their own perspective in tasks measuring perspective shift comprehension (de Villiers and de Villiers, 2014; Köder and Maier, 2016).

There is also evidence that the ability to shift perspective declines with age. In non-linguistic tasks, older adults have shown decreased ability to assess the perspectives of others. A cross-sectional study by Ligneau-Hervé (2005) found that older adults were less likely to take a character's known preferences into account when asked to predict their behavior. A longitudinal study by Pratt et al. (1996) found that participants' perspective taking while discussing moral dilemmas decreased over time. In more language-focused work, Long et al. (2018) found that the performance of adults ages 17-44 and adults ages 66-84 on a spatial referring task were predicted by different individual differences: the performance of the younger group was better predicted by inhibition control, while the performance of the older group was better predicted by attention switching ability.

Autistic individuals are also claimed to have difficulty with perspective shift. There is a large and contentious body of literature on this topic, focusing mostly on theory of mind in the form of pronoun processing (Volden et al., 1997; Mizuno et al., 2011; Pearson et al., 2013; Abbot-Smith et al., 2020).<sup>13</sup>

The individual differences in perspective shift ability provide further evidence that perspective shift is costly. Since not all conversation participants are equally able to track perspective shifts, a rational speaker might avoid shifting perspective when possible.

#### 6.4.3 Summary

The preference to maintain a consistent perspective is motivated by empirical evidence that perspective shift is cognitively costly. This argument is grounded in two kinds of experimental evidence. First, work on adult processing of perspective suggests that there

---

<sup>13</sup>Some strands of this research are deeply problematic; see Gernsbacher (2007) and Yergeau (2013) for a discussion of how research on this topic has been used in harmful and dehumanizing ways.

is an observable cognitive penalty for shifting perspective, in terms of slower processing times, lower task accuracy, and decreased capacity for simultaneous cognitive processes. Second, work on individual differences suggests that individuals differ in their ability to shift perspectives, both within demographic groups and between demographic groups.

There are some major caveats that accompany these findings. First, most evidence comes from work on referring tasks rather than perspectival expressions. While it is intuitively plausible that distinguishing privileged and shared information involves some of the same processes as activating a new perspective, it is also possible that perspective shift in the context of perspectival expressions differs in significant ways.

Second, there are several different forms a perspective shift penalty might take, and they are poorly distinguished in the literature. Many studies do not differentiate between the cost of shifting perspective from the cost of accessing non-speaker perspectives, making it difficult to test whether there is a perspective maintenance effect beyond the egocentricity bias discussed in Section 6.3. A perspective shift cost might also be stated in terms of the cost of calculating a new perspective. This would be a one-time penalty, unlike Harris (2012)'s proposed perspective shift penalty, which applies every time the active perspective is changed. All of these possible perspective shift penalties are intriguing, but more empirical evidence is necessary in order to distinguish among them.

## 6.5 Discourse factors in perspective shift

In addition to the two cognitive pressures that form the basis of Harris (2012) heuristic system, there are a number of discourse factors that are theorized to affect the interpretation of perspectival expressions. I review them briefly in this section. As with the empirical evidence in favor of egocentricity and perspective maintenance pressures, it is often challenging to generalize from the phenomena investigated to other perspectival expressions, since most factors have not been explored systematically across classes of perspectival expressions. Nonetheless, the previous work on discourse factors is rich and opens up many avenues for future exploration.

### 6.5.1 Empathy effects

One discourse factor proposed to affect the interpretation of perspectival expressions is **empathy**: the speaker's sense of identification with entities mentioned in the discourse.

178. **Empathy (Kuno and Kaburaki, 1977)**: the speaker's identification, which may vary in degree, with a person or thing that participates in the event being described.

Kuno and Kaburaki (1977) proposes a ranked list approach to perspective identification using an empathy hierarchy. Conversation participants are generally easier to empathize with; however, depending on the construction used to introduce them, third-persons can also be high in empathy.

179. **Speech-Act Empathy Hierarchy** (Kuno and Kaburaki, 1977):

- (a) The speaker empathizes most strongly with themselves.
- (b) The speaker empathizes more closely with the listener than third persons.

180. **Surface Structure Empathy Hierarchy** (Kuno and Kaburaki, 1977):

- (a) The speaker empathizes most with the subject of the sentence.
- (b) After the subject, the object is the easiest to empathize with.
- (c) Objects of passive by-agentive clauses are the hardest to empathize with.

In his indexical analysis of *come*, Oshima (2006a) proposes that empathy effects determine the ranking of perspective holders. He enriches the context parameter with a set of contextually prominent individuals that always includes the speaker, frequently includes the listener, and occasionally includes other individuals (such as attitude holders). In his view, speaker default effects arise because it is easiest to empathize with the speaker. He also posits that when a conversation participant is a topic, the perspectives of individuals who are not conversation participants are dispreferred.

### 6.5.2 Topicality

Topicality has also been invoked separately from work on empathy. Garrod and Sanford (1988) hypothesize that a single individual serves as the topic or thematic subject in nar-

rative discourse. They tested the interpretation of psychological atmospheric statements using stimuli like (181).

181. *At the restaurant*

Juliet entered the restaurant. There was a table in the corner. The waiter took the order. **Things seemed to go well that night.**

**Prompt:**

- (a) Did things go well for Juliet that night?
- (b) Did things go well for the waiter that night?

Participants interpreted the last sentence as describing the judgment of the first named character (in (181), Juliet) 90% of the time, compared to 50% of the time for the unnamed second character (the waiter). These results suggest that the first character's perspective is more prominent than that of the second character. However, the experimental paradigm makes it difficult to define topicality more precisely, since the characters differ both in definiteness and in when they are introduced.

Hinterwimmer (2017a) also invokes topicality to explain the prominence of perspective holders. He explores pragmatic factors that affect the selection of Free Indirect Discourse protagonists, and posits that both topicality, and subjecthood in the preceding sentence play a role.

Topicality has also been much discussed in the literature on pronominalization; in their probabilistic approach to anaphora resolution, Kehler and Rohde (2013) propose that topicality drives pronominalization. This is may be relevant for perspective-anaphoric expressions, assuming that the perspectival anaphora resolution process is the same or similar to the normal anaphora resolution process.

### 6.5.3 Thematic roles and subjecthood

Subjecthood has been posited to increase the prominence of a perspective holder. Hinterwimmer (2017a) finds that an individual who is the subject of a sentence is more likely to be the protagonist in subsequent Free Indirect Discourse passages (Hinterwimmer, 2017a),

**Figure 6.4.** Example stimulus from Kaiser and Lee (2017b)

<b>Sentence A (Experiencer-Theme condition):</b> Jennifer looked at Amy.	<b>Sentence A (Agent-Patient condition):</b> Jennifer nudged Amy.	
<b>Sentence B (PPT condition):</b> She was irritating.	<b>Sentence B (Multidimensional adj. condition):</b> She was smart.	
<b>Prompt:</b> Whose opinion is it that the other person is {irritating/smart}?		
(a) Jennifer	(b) The narrator	(c) Amy

and Kehler and Rohde (2013) propose that subjecthood affects the resolution of pronouns, which is potentially relevant for perspective-anaphoric expressions.<sup>14</sup>

The impact of subjecthood may also be related to the thematic roles that subjects usually fill. In experimental work, Kaiser and Lee (2017a,b) find that thematic roles differentially affect the interpretation of two classes of subjective adjectives (example stimulus shown in Figure 6.4). The subject of a verb with Agent and Patient thematic roles is more likely to be the perspective holder for both multidimensional adjectives like *healthy* and predicates of personal taste like *tasty* than the object. However, for predicates of personal taste, unlike multidimensional adjectives, the effect of subjecthood is heightened when the verb has Experiencer and Theme roles (Figure 6.5).

Kaiser and Lee interpret these results as evidence that predicates of personal taste involve an experiencer argument (Bylinina, 2014). More generally, these results suggest that the strength of the effect of discourse factors like subjecthood may vary across classes of perspectival expressions.

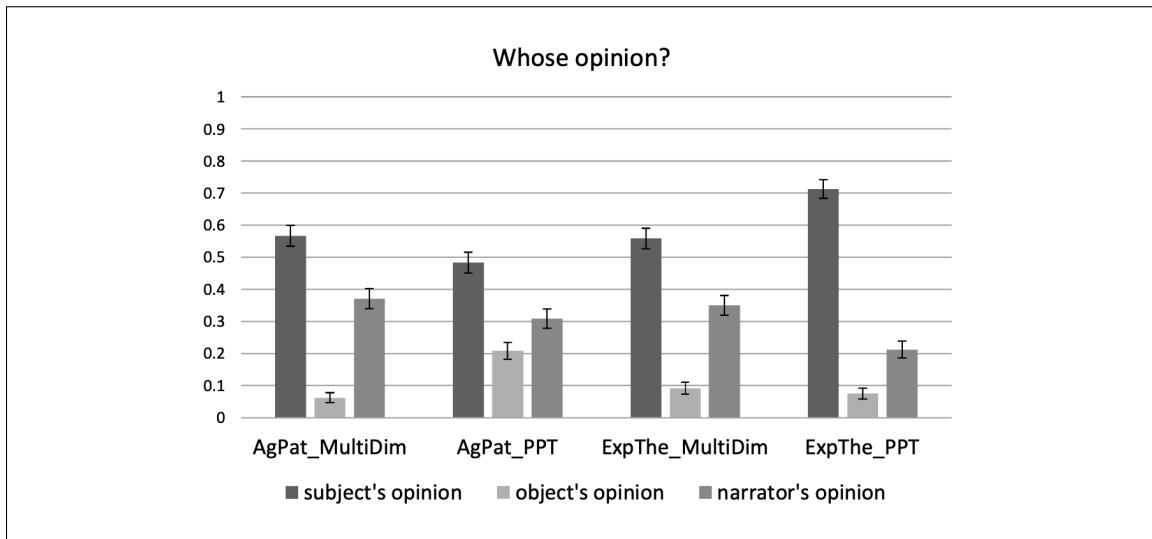
#### 6.5.4 Question Under Discussion

In a well-formed discourse, the contributions of conversation participants are not aimless. One way of thinking about their structure is as a set of issues to be settled called the Question Under Discussion (QUD) (Roberts, 1996).

---

<sup>14</sup>Also, as discussed in Section 6.5.1, Kuno and Kaburaki (1977) incorporate subjecthood into their Surface Structure Empathy Hierarchy.

**Figure 6.5.** Results reported in Kaiser and Lee (2017b)



182. **Question Under Discussion (Roberts, 1996):** the (partially) structured set of questions which discourse participants are mutually committed to resolving at a point in the discourse.

The idea of a QUD is one way of thinking about how the topic of a conversation affects the interpretation of its utterances. As an example, consider the two short dialogues shown in (184) and (183). Both contain the same motion description, *Last weekend Thelma \_\_\_\_ with me to New York*. In an informal Twitter poll that I ran, however, participants' intuitions about which perspectival motion verb was most natural differed between the two discourse contexts: in (184), they preferred *went*, while in (183), they preferred *come* (Figure 6.6).

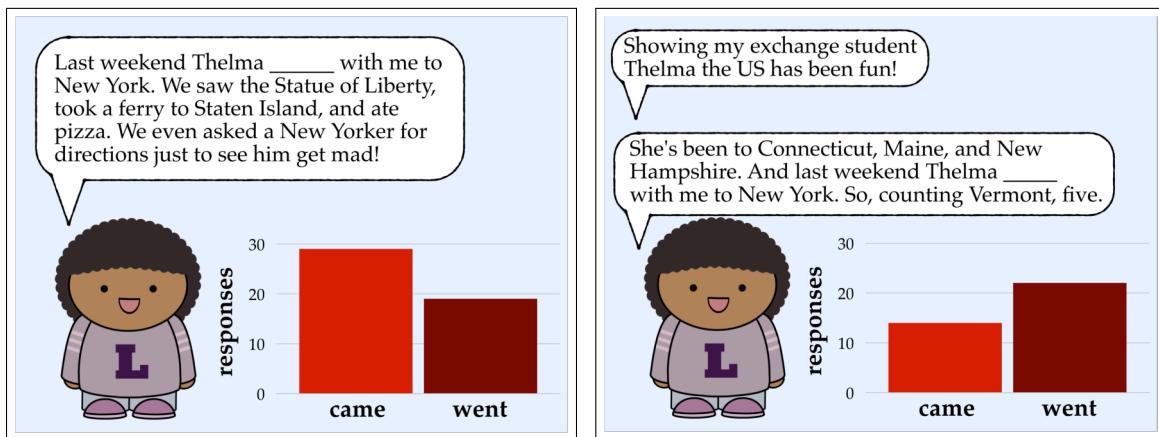
183. Lucy: **Last weekend Thelma \_\_\_\_ with me to New York.** We saw the Statue of Liberty, took a ferry to Staten Island, and ate pizza. We even asked a New Yorker for directions just to see him get mad!

184. Lucy: Showing my exchange student Thelma the US has been fun!

Sam: How many states has she been to so far?

Lucy: She's been to Connecticut, Maine, and New Hampshire. **And last weekend Thelma \_\_\_\_ with me to New York.** So, counting Vermont, five.

**Figure 6.6.** Poll responses for (201) (left) and (202) (right)



One explanation is that the Question Under Discussion differs between the dialogues. In (184), the QUD is about the number of states that Thelma has visited. The speaker’s event time perspective is not particularly relevant to this question. In (183), by contrast, the QUD is about the events of the New York trip, which makes the speaker’s event time perspective relevant. This is preliminary data with an intuitive explanation of how the QUD might affect the interpretation of perspectival expressions, but further work would be necessary to demonstrate how this happens.

### 6.5.5 Discourse coherence

The structure of a discourse is defined not just by the content of its utterance, but also by the relations between those utterances,<sup>15</sup> which are called **coherence relations** (Hobbs, 1979). For instance, the second sentence in (185) is most naturally interpreted as an explanation for the event described in the first sentence. Under this interpretation, the two sentences are related by the Explanation relation.

185. Ariadne Oliver suddenly dropped her sandwich onto her plate. It had a strangely bitter taste.

---

<sup>15</sup>Or propositions; see Gruber and Redeker (2014) for more discussion of the distinction between semantic and pragmatic coherence relations.

186. **Explanation (Kehler, 2002):** Infer that the second sentence describes a cause or reason for the eventuality described in the first sentence.

Discourse coherence theorists posit that discourse comprehension involves reasoning about these relations. During this reasoning process, listeners resolve free variables like pronouns based on the relationships between the utterances that are computed. The coherence relation resolution process might therefore play a particularly important role in the interpretation of anaphoric perspectival expressions. However, since ambiguity resolution is also affected by the coherence relations between utterances, the interpretation of other classes of perspectival expressions might also be affected.

#### 6.5.6 Centering theory

Another theory of how discourses are structured is **centering theory** (Grosz et al., 1995). Centering theory looks at the prominence of individuals over the span of a text. It posits that discourses that refer to the same individual (or set of individuals) are more coherent (Chafe, 1976). In centering theory, the individuals are tracked in a ranked list of **centers**. While there can be multiple forward-looking centers, there is exactly one backward-looking center for each utterance (Grosz et al., 1995). This backward-looking center, which picks out the most prominent individual, can be shifted between utterances, but such shifts affect the coherence of the discourse.<sup>16</sup>

In some ways, the central tenets of centering theory align well with Harris (2012)'s proposal that comprehenders track an active perspective and that they avoid perspective shift. However, centering theory has traditionally been proposed as a model of anaphora resolution: if any center is pronominalized, it is the backward-looking center (Gordon et al., 1993). Thus, while centering theory captures some of the pressures involved in perspectival processing, it is also a more general theory of how discourse coherence involves tracking individuals.

---

<sup>16</sup>See Poesio et al. (2004) for an overview of different variations on centering theory that have been proposed.

### 6.5.7 Attraction and verbatim

Curiously absent from much work on perspective is discussion of the motivations for shifting perspective. In his work on quotation, Maier (2017b) proposes a pragmatic factor that also motivates the use of perspective shift. He proposes that there is an advantage to reproducing other peoples' utterances faithfully, a principle he calls **Verbatim**.

However, this pressure competes with another pragmatic factor, **Attraction**, which is a pressure to refer relative to the point-of-view of the conversation participants.

187. Pragmatic pressures on quotation (Maier, 2017b):

- (a) **Attraction**: the desire to use indexicals that refer directly to the most salient speech act participants.
- (b) **Verbatim**: the desire to reproduce referring expressions directly from the original utterance.

Although Maier (2017b) proposes these pressures to explain the costs and benefits of quotation and unquotation, these factors could be reformulated in terms of perspective. Attraction could be recast as a default towards the speaker's perspective, while Verbatim could motivate the occasional use of perspective shift. In particular, adopting Verbatim as a principle for perspective identification would lead to the prediction that perspective shift is more likely to occur in the context of a speech (or perhaps thought) report, neatly accounting for attitude holder-oriented interpretations of perspectival expressions.

### 6.5.8 Summary

A number of pragmatic factors have been hypothesized to affect perspective prominence (Figure 6.7). Throughout this dissertation, we have seen many cases where the discourse context plays an important role in disambiguating the perspective holder. Exploring the contributions of the discourse factors discussed this section is clearly important. However, it is also challenging. Some factors are difficult to disentangle. For instance, the paradigms used to test the impact of topicality, subjecthood, and thematic roles often overlap.

**Figure 6.7.** Candidate discourse factors in perspective selection

Empathy effects	Oshima (2006a)
Topicality	Hinterwimmer (2017a)
Subjecthood	Hinterwimmer (2017a)
Question Under Discussion	Roberts (2015)
Discourse coherence relations	Hobbs (1979)
Centering theory	Grosz et al. (1995)
Attraction	Maier (2017b)
Verbatim	Maier (2017b)

It is also not yet clear whether the discourse context affects all perspectival expressions in a similar way, since there has been little systematic comparison of these factors across perspectival expressions. Unlike the two perspectival pressures posited by Harris (2012), the factors discussed in this section do not arise from general cognitive pressures, which makes them less likely to apply equally across all classes of perspectival expressions. Kaiser and Lee (2017b,a) provide some preliminary evidence that discourse factors may impact different classes of perspectival expressions in different ways, or at least, to varying degrees. This suggests that the effects of a discourse factor observed for one perspectival expression should not be assumed to transfer to another.

In addition, the method of encoding the perspectival component of various expressions may modulate their sensitivity to the discourse factors discussed. For instance, perspective-anaphoric expressions may be more sensitive to pressures that are involved in other anaphora resolution processes. This is not the case for the bias towards the speaker and the cost of perspective shift, since they reflect general cognitive pressures.

## 6.6 Towards models of perspectival reasoning

In this chapter, I have discussed perspective identification and perspective selection. I have argued that listeners and speakers must engage in these tasks regardless of the semantics of the perspectival expression. I have introduced a number of approaches to modeling these processes. Following Harris (2012), I have argued that relying on a simple set of heuristics cannot satisfactorily capture how perspectival expressions are interpreted in certain contexts. I have also discussed some of the advantages and challenges of Harris (2012)'s

two-stage model of perspective identification, which consists of a costly reasoning system and a cheaper heuristic-based system.

In Chapter 7, I propose a reasoning-based model of perspective identification based in the Rational Speech Acts framework. The Bayesian model that I propose is one possible implementation of the second part of Harris (2012)'s proposed two-stage system: given a representation of the discourse context in the form of a probability distribution over perspectives and a probability distribution over possible worlds, it models how listeners infer the speaker's intended meaning and adopted perspective. I show that this model is capable of capturing the desired perspective identification behavior in a number of small case studies. I also explore one of the model's more surprising predictions experimentally, and find some support for the model's predictions.

The question of perspective selection in production is more challenging, because the task is less constrained than that of perspective identification. In this chapter, I have argued that the most intuitive heuristics cannot explain the pattern of perspective selection on their own (though they may play a role in perspective selection). I have also discussed how Harris (2012)'s system might be adapted to the perspective selection task and reviewed a number of challenges that arise.

In Chapter 8, I propose a reasoning-based model of perspective selection. Although most contemporary work in the Rational Speech Acts framework uses it as a model of comprehension, it was initially proposed as production model as well. I show that although the Bayesian model of perspective production captures some desired properties of perspective selection, it does not correctly predict the behavior of speakers observed in a production study. Thus, perspective selection remains a challenging area for future work.

In this chapter, I have also reviewed some of the pressures hypothesized to influence perspective identification and selection. Harris (2012) codifies two such pressures in his heuristic system: Maintain Perspective and Speaker Default. I have discussed some of the empirical evidence in favor of these two principles. I also discussed some of the discourse factors that may play a role in these processes. However, there are many open questions in this

domain. It is often challenging to compare the impact of various factors across classes of perspectival expressions, since there is not much consistency in the phenomena investigated or the experimental methods used. Larger scale investigations of the impact of discourse factors across classes of perspective would be a valuable direction for future work.

As I have reviewed, there is good evidence that egocentricity biases, perspective shift avoidance, and discourse all play a role in the relative prominence of different perspectives. Each of these is a potential source of evidence for a perspectival reasoning system, such as the models that I will propose in Chapters 7 and 8. However, the models that I will propose implement only the first two factors (speaker bias and perspective shift avoidance).

This is not because I think that the discourse factors discussed in Section 6.5 do not play a role in the processes of perspective identification and selection, but because there are certain advantages to setting aside the more complex discourse factors for the moment. First, simplifying the kind of evidence that the perspectival reasoning models can access makes the reasoning task computationally tractable, which allows me implement working computational models that can generate quantitative predictions about speaker and listener behavior. Second, starting from Harris (2012)'s two basic principles proposes allows us to explore how we can get using a model with minimal inductive biases. Third, given the evidence in Section 6.5 that the impact of discourse factors varies by perspectival expression, and the dearth of quantitative evidence of how the impact of each factor varies by perspectival expression, incorporating these factors would limit the generalizability of the models. I leave the integration of discourse factors as a challenging but important direction for future work.

## CHAPTER 7

### A RATIONAL APPROACH TO PERSPECTIVE INFERENCE

A key property of perspectival phenomena is the availability of multiple perspectives. It is not just the case that a perspectival expression can be anchored to different perspective holders in different contexts; it is often also the case that a perspectival expression has multiple possible anchors within the same context. For instance, in (188), *come* can be interpreted either as describing motion towards the speaker's location (Poirot's flat), or towards the attitude holder's location (Scotland Yard).

188. *In his London flat, Poirot is telling Miss Lemon about a call he received from Chief Inspector Japp, at Scotland Yard.*

Poirot: Chief Inspector Japp thinks that the murderer will come to confess.

This optionality over perspective holders poses an interpretative problem for the listener. In order to infer the intended destination of the subject of *come*, the listener must infer whose perspective the speaker is using.

In this chapter, I use perspectival motion verbs as a case study to explore the question of perspective identification. I propose a model of perspectival reasoning in which listeners jointly infer the speaker's intended meaning and the adopted perspective. This model can be viewed as one way of implementing the more costly reasoning system in Harris (2012)'s two-stage model of perspective processing. The proposed model builds on the Rational Speech Acts framework (Frank and Goodman, 2012), a Bayesian approach to language that has proved useful for capturing many pragmatic phenomena.

Bayesian approaches to language posit that language users are guided by models of the world and of language that they are constantly updating. The model provides a set of prior beliefs, or expectations; the model can be refined by observing how well its predictions match

new data. These approaches make two fundamental assumptions about language users. First, Bayesian approaches to language assume that language users behave **rationally**: they act optimally according to their mental model. For instance, speakers are expected to select the utterance that maximizes their chances of conveying their intended meaning. Second, language users are expected to be constantly updating their models in light of new experiences (evidence). If the speaker observes that their utterance has been misunderstood, they should use this information to adjust their utterance selection process.

In this chapter, I propose a Perspectival Rational Speech Acts model of comprehension that incorporates both of these tenets of Bayesian approaches to language. The core proposal is that listeners consider multiple perspectives simultaneously as they reason about the speaker's intended meaning and the speaker's adopted perspective. I will show that this rational perspective inference model captures several key properties of how listeners interpret perspectival expressions. It also makes a novel prediction about listeners' expectations in one environment. I test this prediction experimentally, comparing the predictions of the model to a simpler heuristic: the **Speaker Default** heuristic.

#### 189. Speaker Default heuristic for perspective identification (Harris, 2012):

When interpreting a perspectival expression, assume that the perspective holder is the speaker.

The Speaker Default heuristic alone cannot capture the perspective identification behavior of listeners, since it fails to predict any ambiguity in cases where multiple perspectives are available, as in (188). Nonetheless, as we have seen in Chapter 6, speakers are biased towards their own perspectives, making this a good baseline for illustrating the kinds of cases of perspective identification where a heuristic approach fails and a reasoning-based system might hope to succeed.

The structure of this chapter is as follows. In Section 7.1, I introduce the Rational Speech Acts framework as a model of comprehension and show how it captures pragmatic reasoning over alternatives. Returning to the question of perspective identification, in Section 7.2, I describe several key properties for a system of perspective inference. In Section 7.3, I present

the Perspectival Rational Speech Acts model of comprehension. In Section 7.4, I illustrate how the proposed model captures the desired properties of perspective identification and discuss a novel prediction that it makes. In Section 7.5, I test this prediction experimentally; I present evidence from two comprehension experiments suggesting that listeners behave as predicted by the proposed perspective inference model. However, the results also suggest that listeners expect strongly egocentric behavior from speakers, which means that the Speaker Default model is also fairly successful at explaining the comprehension data. I conclude with a discussion of open questions about perspective identification.

Throughout this chapter, I use perspectival motion verbs as an example phenomenon. For this reason, I refer to the problem of perspective identification in terms of perspective anaphora resolution, since, in Chapter 3, I argued for a perspective-anaphoric analysis of American English *come*. However, the issue of perspective identification is important regardless of the representation adopted for the perspectival component: anaphoric accounts require reasoning about the value of the perspectival variable; logophoric binding accounts require reasoning about the logophoric variable; and indexical accounts require reasoning about whether a covert context shift operator has been projected. The model of perspectival reasoning that I propose is not phenomena-specific and could be generalized to other perspectival expressions.

## 7.1 Rational approaches to pragmatics

In his foundational work on pragmatics, Lewis (1979b) proposed that conversation is a cooperative game between the participants, where the goal is to determine which world the participants are in. Participants work towards this goal by sharing information, which narrows the set of possible worlds that might be the real world. Information shared between the conversation participants is stored in the Common Ground, which can be viewed as the set of accepted propositions, or as the set of worlds compatible with those propositions.

The Rational Speech Acts (RSA) model is a framework for pragmatic modeling that extends this picture by proposing that the listener uses a Bayesian inference process to determine what meaning the speaker is trying to convey (Bergen et al., 2012; Frank and Goodman,

2012). In this model, the Common Ground contains not just a set of worlds, but also a probability associated with each world, the probability that it is the real world. At each turn in the conversation, the speaker selects a world from the set of worlds, simulating a new piece of information for the speaker to contribute, and chooses an utterance to express it. The speaker strives to select the utterance that is most likely to communicate the observed world to their listener.

Upon hearing the speaker's utterance, the listener must reason about the message that the speaker is trying to convey. The listener assumes that the speaker selects the sentence that maximizes the probability of the observed world. The listener interprets the sentence in order to update the probability distribution over possible worlds in the Common Ground, calculating the likelihood of each world given the sentence selected, according to their model of how the speaker picks sentences.

The RSA model is therefore recursive: the listener's interpretation process involves reasoning about the speaker's utterance selection process, while the speaker's utterance selection process involves reasoning about the listener's interpretation process.

Although initially proposed as both a model of production and comprehension (Bergen et al., 2012; Frank and Goodman, 2012), the RSA model has been explored most thoroughly as a model of comprehension. RSA comprehension models have been applied to a range of phenomena, including hyperbole (Kao et al., 2014); irony (Cohn-Gordon and Bergen, 2019); politeness (Yoon et al., 2016, 2017); projective content (Qing et al., 2016); scalar implicatures (Bergen et al., 2012; Degen et al., 2015; Potts et al., 2016; Bergen et al., 2016; Brochhagen et al., 2016); social meaning (Cohn-Gordon and Qing, 2018; Qing and Cohn-Gordon, 2018); spatial descriptions (Ullman et al., 2016); and word learning (Smith et al., 2013; Frank and Goodman, 2014; Bohn et al., 2019).

### 7.1.1 The standard RSA model

Although a RSA model is potentially infinitely recursive, the first three levels are usually the focal point: the Literal Listener, Pragmatic Speaker, and Pragmatic Listener.

The **Pragmatic Listener** represents the actual listener. Given an utterance, the listener reasons about the speaker's intended meaning (represented by a possible world) using Bayes' rule.<sup>1</sup> They reason about how likely the speaker would have been to select the utterance if they had intended to communicate that meaning, according to a mental model of the speaker's utterance selection process (the Pragmatic Speaker). The Pragmatic Listener calculates the posterior probability of a world given an utterance,  $p(w|u)$ , by calculating the likelihood of the utterance given the world,  $p(u|w)$ , according to the Pragmatic Speaker, discounted by the listener's prior belief in the world,  $p(w)$ . The probabilities over all worlds are then renormalized to form a proper probability distribution using a softmax operation.<sup>2</sup>

The **Pragmatic Speaker** is the listener's mental model of the speaker. The speaker reasons about which utterance to select based on a mental model of how the listener will interpret it. The more likely an utterance is to lead to the listener correctly inferring the speaker's intended meaning (possible world), the higher its utility.<sup>3</sup> However, the Pragmatic Speaker relies on a simpler mental model of how the listener interprets utterances, the Literal Listener, rather than the Pragmatic Listener. After calculating the utility of each utterance in the utterance set, the speaker selects an utterance. The probability of each utterance being selected by the speaker is proportional to its normalized utility.<sup>4</sup>

The **Literal Listener** reasons about the speaker's intended meaning given the utterance. Like the Pragmatic Listener, the Literal Listener calculates the posterior probability of a world given an utterance,  $p(w|u)$ , by taking into account the probability of the utterance given the world,  $p(u|w)$ , and the prior probability of the world,  $p(w)$ . However, unlike the Pragmatic Listener, the Literal Listener does not take into account the speaker's utterance selection process. For the Literal Listener, the probability of an utterance given a world

<sup>1</sup>  $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

<sup>2</sup> I omit the softmax term from the mathematical model in Figure 7.1 and following figures for readability.

<sup>3</sup> Since the speaker uses Bayes' rule to calculate the utility of the utterance, they technically take into consideration the utterance's prior probability. However, I am not aware of any RSA models that use a non-uniform distribution over utterances, so in practice, this is omitted.

<sup>4</sup> Some RSA models select the utterance that is highest in utility rather than sampling proportionally to utility; others set a parameter that controls the rationality of the speaker's selection.

**Figure 7.1.** Vanilla Rational Speech Acts model (Bergen et al., 2012)

<b>Literal Listener</b>		
$p(w u)$	$\propto p(u w) p(w)$	by Bayes' Rule
	$\propto [[u]]^w p(w)$	
<b>Pragmatic Speaker</b>		
$p(u w)$	$\propto p(w u) p(u)$	by Bayes' Rule
	$\propto \text{LitList}(w,u) p(u)$	by definition of Literal Listener
<b>Pragmatic Listener</b>		
$p(w u)$	$\propto p(u w) p(w)$	by Bayes' Rule
	$\propto \text{PragSpeak}(u,w) p(w)$	by definition of Pragmatic Speaker

depends only on whether the utterance is true in the world. The meaning of the utterance is its truth value with respect to the world, as calculated by the interpretation function, which takes a world and an utterance and returns 1 or 0 depending on whether the utterance is true in that world. As in the Pragmatic Listener, the probabilities are then renormalized to produce a well-formed probability distribution.

### 7.1.2 Model components

In order to run an RSA model, there are three components of the model that must be specified: (1) the set of possible worlds and its prior probability distribution; (2) the set of utterances and its prior probability distribution; and (3) the interpretation function that maps each utterance-world pair to a truth value.

Generally, models are run using a uniform prior distribution over possible worlds; this is taken to reflect the empty Common Ground at the onset of a conversation. However, since conversation participants often assume a certain amount of shared world knowledge, it is also fair to start with an initial distribution over worlds that reflects world knowledge.

A uniform distribution over utterances is also generally adopted (and the  $p(u)$  term omitted from the Pragmatic Speaker calculation). This distribution can be manipulated to implement frequency or production complexity effects. However, this is usually done with an utterance cost function (see Section 7.1.5) for more discussion).

**Figure 7.2.** World set, utterance set, and interpretation function for scalar implicature example

	Red apples	Prior		Utterance	Prior		W <sub>0</sub>	W <sub>1</sub>	W <sub>2</sub>	W <sub>3</sub>
W <sub>0</sub>	0	0.25		U <sub>0</sub>	None are red.	0.33	U <sub>0</sub>	1	0	0
W <sub>1</sub>	1	0.25		U <sub>1</sub>	Some are red.	0.33	U <sub>1</sub>	0	1	1
W <sub>2</sub>	2	0.25		U <sub>2</sub>	All are red.	0.33	U <sub>2</sub>	0	0	1
W <sub>3</sub>	3	0.25								

It is important to note that changing the set of utterances will affect the model predictions, since the utterance set defines the set of alternatives in pragmatic competition.

### 7.1.3 Pragmatic competition in the RSA framework

To see how the core RSA model works, we will look at one of the motivating phenomena for the RSA framework: scalar implicature. Consider a context in which Sam Speaker holds a bag of three apples. He opens the bag, observes how many of the apples are red, and describes his observation to Lucy Listener.

**Figure 7.3.** Vanilla Literal Listener calculation for scalar implicature example

<b>Literal Listener: <math>p(w None)</math></b>
$p(w_0 None) = \text{softmax}(p(u_0 w_0)p(w_0)) = \frac{[[u_0]]^{w_0} p(w_0)}{\sum_w [[u_0]]^w p(w)} = \frac{1(0.25)}{\sum_w [[u_0]]^w p(w)} = \frac{0.25}{0.25} = 1$
$p(w_1 None) = \text{softmax}(p(u_0 w_1)p(w_1)) = \frac{[[u_0]]^{w_1} p(w_1)}{\sum_w [[u_0]]^w p(w)} = \frac{0(0.25)}{\sum_w [[u_0]]^w p(w)} = \frac{0}{0.25} = 0$
$p(w_2 None) = \text{softmax}(p(u_0 w_2)p(w_2)) = \frac{[[u_0]]^{w_2} p(w_2)}{\sum_w [[u_0]]^w p(w)} = \frac{0(0.25)}{\sum_w [[u_0]]^w p(w)} = \frac{0}{0.25} = 0$
$p(w_3 None) = \text{softmax}(p(u_0 w_3)p(w_3)) = \frac{[[u_0]]^{w_3} p(w_3)}{\sum_w [[u_0]]^w p(w)} = \frac{0(0.25)}{\sum_w [[u_0]]^w p(w)} = \frac{0}{0.25} = 0$
<b>Literal Listener: <math>p(w Some)</math></b>
$p(w_0 Some) = \text{softmax}(p(u_1 w_0)p(w_0)) = \frac{[[u_1]]^{w_0} p(w_0)}{\sum_w [[u_1]]^w p(w)} = \frac{0(0.25)}{\sum_w [[u_1]]^w p(w)} = \frac{0}{0.75} = 0$
$p(w_1 Some) = \text{softmax}(p(u_1 w_1)p(w_1)) = \frac{[[u_1]]^{w_1} p(w_1)}{\sum_w [[u_1]]^w p(w)} = \frac{1(0.25)}{\sum_w [[u_1]]^w p(w)} = \frac{0.25}{0.75} = 0.333$
$p(w_2 Some) = \text{softmax}(p(u_1 w_2)p(w_2)) = \frac{[[u_1]]^{w_2} p(w_2)}{\sum_w [[u_1]]^w p(w)} = \frac{1(0.25)}{\sum_w [[u_1]]^w p(w)} = \frac{0.25}{0.75} = 0.333$
$p(w_3 Some) = \text{softmax}(p(u_1 w_3)p(w_3)) = \frac{[[u_1]]^{w_3} p(w_3)}{\sum_w [[u_1]]^w p(w)} = \frac{1(0.25)}{\sum_w [[u_1]]^w p(w)} = \frac{0.25}{0.75} = 0.333$
<b>Literal Listener: <math>p(w All)</math></b>
$p(w_0 All) = \text{softmax}(p(u_2 w_0)p(w_0)) = \frac{[[u_2]]^{w_0} p(w_0)}{\sum_w [[u_2]]^w p(w)} = \frac{0(0.25)}{\sum_w [[u_2]]^w p(w)} = \frac{0}{0.25} = 0$
$p(w_1 All) = \text{softmax}(p(u_2 w_1)p(w_1)) = \frac{[[u_2]]^{w_1} p(w_1)}{\sum_w [[u_2]]^w p(w)} = \frac{0(0.25)}{\sum_w [[u_2]]^w p(w)} = \frac{0}{0.25} = 0$
$p(w_2 All) = \text{softmax}(p(u_2 w_2)p(w_2)) = \frac{[[u_2]]^{w_2} p(w_2)}{\sum_w [[u_2]]^w p(w)} = \frac{0(0.25)}{\sum_w [[u_2]]^w p(w)} = \frac{0}{0.25} = 0$
$p(w_3 All) = \text{softmax}(p(u_2 w_3)p(w_3)) = \frac{[[u_2]]^{w_3} p(w_3)}{\sum_w [[u_2]]^w p(w)} = \frac{1(0.25)}{\sum_w [[u_2]]^w p(w)} = \frac{0.25}{0.25} = 1$

There are four equally likely worlds, corresponding to the number of apples that may be red (0-3). The speaker chooses among three equally likely utterances (Figure 7.2).

Intuitively, we expect the listener to interpret *Some are red* as implying that not all of the apples are red, even though the lexical semantics of the utterance do not rule out  $W_3$ , where all apples are red. This is because *All are red* is a more specific alternative for  $W_3$ .

The RSA framework captures this kind of pragmatic competition well. The Literal Listener calculates the likelihood of a particular world given an utterance by looking up the truth of the utterance in that world according to the interpretation function, and then taking into consideration the prior likelihood of the world. As Figure 7.3 shows, *Some are red* is true in Worlds 1-3, while *None are red* is only true in World 0, and *All are red* is only true in World 3. Since the distribution over worlds is uniform, the Literal Listener thinks that the speaker is equally likely to be describing World 1, 2, or 3 when they hear *Some are red*.

**Figure 7.4.** Pragmatic Speaker calculation for scalar implicature example

**Pragmatic Speaker:**  $p(u|w_0)$

$$p(\text{None}|w_0) = \text{softmax}(p(w_0|u_0)p(u_0)) = \frac{p(w_0|u_0)p(u_0)}{\sum_u p(w_0|u)p(u)} = \frac{1(0.33)}{\sum_u p(w_0|u)p(u)} = \frac{0.33}{0.33} = 1$$

$$p(\text{Some}|w_0) = \text{softmax}(p(w_0|u_1)p(u_1)) = \frac{p(w_0|u_1)p(u_1)}{\sum_u p(w_0|u)p(u)} = \frac{0(0.33)}{\sum_u p(w_0|u)p(u)} = \frac{0}{0.33} = 0$$

$$p(\text{All}|w_0) = \text{softmax}(p(w_0|u_2)p(u_2)) = \frac{p(w_0|u_2)p(u_2)}{\sum_u p(w_0|u)p(u)} = \frac{0(0.33)}{\sum_u p(w_0|u)p(u)} = \frac{0}{0.33} = 0$$

**Pragmatic Speaker:**  $p(u|w_1)$

$$p(\text{None}|w_1) = \text{softmax}(p(w_1|u_0)p(u_0)) = \frac{p(w_1|u_0)p(u_0)}{\sum_u p(w_1|u)p(u)} = \frac{0(0.33)}{\sum_u p(w_1|u)p(u)} = \frac{0}{0.11} = 0$$

$$p(\text{Some}|w_1) = \text{softmax}(p(w_1|u_1)p(u_1)) = \frac{p(w_1|u_1)p(u_1)}{\sum_u p(w_1|u)p(u)} = \frac{0.33(0.33)}{\sum_u p(w_1|u)p(u)} = \frac{0.11}{0.11} = 1$$

$$p(\text{All}|w_1) = \text{softmax}(p(w_1|u_2)p(u_2)) = \frac{p(w_1|u_2)p(u_2)}{\sum_u p(w_1|u)p(u)} = \frac{0(0.33)}{\sum_u p(w_1|u)p(u)} = \frac{0}{0.11} = 0$$

**Pragmatic Speaker:**  $p(u|w_2)$

$$p(\text{None}|w_2) = \text{softmax}(p(w_2|u_0)p(u_0)) = \frac{p(w_2|u_0)p(u_0)}{\sum_u p(w_2|u)p(u)} = \frac{0(0.33)}{\sum_u p(w_2|u)p(u)} = \frac{0}{0.11} = 0$$

$$p(\text{Some}|w_2) = \text{softmax}(p(w_2|u_1)p(u_1)) = \frac{p(w_2|u_1)p(u_1)}{\sum_u p(w_2|u)p(u)} = \frac{0.33(0.33)}{\sum_u p(w_2|u)p(u)} = \frac{0.11}{0.11} = 1$$

$$p(\text{All}|w_2) = \text{softmax}(p(w_2|u_2)p(u_2)) = \frac{p(w_2|u_2)p(u_2)}{\sum_u p(w_2|u)p(u)} = \frac{0(0.33)}{\sum_u p(w_2|u)p(u)} = \frac{0}{0.11} = 0$$

**Pragmatic Speaker:**  $p(u|w_3)$

$$p(\text{None}|w_3) = \text{softmax}(p(w_3|u_0)p(u_0)) = \frac{p(w_3|u_0)p(u_0)}{\sum_u p(w_3|u)p(u)} = \frac{0(0.33)}{\sum_u p(w_3|u)p(u)} = \frac{0}{0.44} = 0$$

$$p(\text{Some}|w_3) = \text{softmax}(p(w_3|u_1)p(u_1)) = \frac{p(w_3|u_1)p(u_1)}{\sum_u p(w_3|u)p(u)} = \frac{0.33(0.33)}{\sum_u p(w_3|u)p(u)} = \frac{0.11}{0.44} = 0.25$$

$$p(\text{All}|w_3) = \text{softmax}(p(w_3|u_2)p(u_2)) = \frac{p(w_3|u_2)p(u_2)}{\sum_u p(w_3|u)p(u)} = \frac{1(0.33)}{\sum_u p(w_3|u)p(u)} = \frac{0.33}{0.44} = 0.75$$

**Figure 7.5.** Vanilla RSA calculation for *Some are red*

<b>World 0</b>	$p(w_0 Some) = \frac{\text{softmax}(p(u_1 w_0)p(w_0))}{\sum_w p(u_1 w)p(w)}$	by Pragmatic Listener definition by softmax definition by substitution
	$\frac{p(u_1 w_0)p(w_0)}{\sum_w p(u_1 w)p(w)}$	$0(0.25)$
	$\frac{0}{\sum_w p(u_1 w)p(w)}$	$0$
	$\frac{0}{0.5625}$	$0$
<b>World 1</b>	$p(w_1 Some) = \frac{\text{softmax}(p(u_1 w_1)p(w_1))}{\sum_w p(u_1 w)p(w)}$	by Pragmatic Listener definition by softmax definition by substitution
	$\frac{p(u_1 w_1)p(w_1)}{\sum_w p(u_1 w)p(w)}$	$1(0.25)$
	$\frac{1}{\sum_w p(u_1 w)p(w)}$	$0.25$
	$\frac{0.25}{0.5625}$	$0.44$
<b>World 2</b>	$p(w_2 Some) = \frac{\text{softmax}(p(u_1 w_2)p(w_2))}{\sum_w p(u_1 w)p(w)}$	by Pragmatic Listener definition by softmax definition by substitution
	$\frac{p(u_1 w_2)p(w_2)}{\sum_w p(u_1 w)p(w)}$	$1(0.25)$
	$\frac{1}{\sum_w p(u_1 w)p(w)}$	$0.25$
	$\frac{0.25}{0.5625}$	$0.44$
<b>World 3</b>	$p(w_3 Some) = \frac{\text{softmax}(p(u_1 w_3)p(w_3))}{\sum_w p(u_1 w)p(w)}$	by Pragmatic Listener definition by softmax definition by substitution
	$\frac{p(u_1 w_3)p(w_3)}{\sum_w p(u_1 w)p(w)}$	$0.25(0.25)$
	$\frac{0.25}{\sum_w p(u_1 w)p(w)}$	$0.0625$
	$\frac{0.0625}{0.5625}$	$0.11$

Pragmatic competition comes in at the level of the Pragmatic Listener. The Pragmatic Listener simulates the speaker's reasoning process using the Pragmatic Speaker model. As Figure 7.4 shows, the Pragmatic Speaker is most likely to produce *Some are red* after observing that 1 or 2 of the apples are red, because there is no other true utterance that describes Worlds 1 and 2. The Pragmatic Speaker may also produce *Some are red* in World 3, where all apples are red; however, they are more likely to produce *All are red* in this case.

Because the Pragmatic Speaker is less likely to produce *Some are red* in World 3, where the competitor *All are red* is more likely, than in Worlds 1 and 2, the Pragmatic Listener infers that the most likely interpretation for *Some are red* is the *some but not all* interpretation (World 1 or 2). Figure 7.5 shows the Pragmatic Listener's calculation for *Some are red*.

Thus, the RSA framework is a model of comprehension that naturally incorporates pragmatic competition between alternatives. For this reason, it has been applied with particular success to scalar implicatures, accounting for certain implicatures that are difficult to derive in other systems (Bergen et al., 2012; Degen et al., 2015; Potts et al., 2016; Bergen et al., 2016; Brochhagen et al., 2016).

#### 7.1.4 Contextual reasoning in the RSA framework

The RSA framework incorporates discourse context into the listener's interpretative process in a limited way: through the prior probability distribution over worlds. To see how this works, let us consider a modified version of the scalar implicature example.

Lucy and Sam are engaged in another conversation about the colors of the apples. This time, however, we will assume that the discourse context has already established an expectation that the apples will be all the same color. For instance, perhaps Lucy and Sam are trying to figure out which kind of apple Thelma has purchased: Granny Smiths (green) or Cosmic Crisps (red).

We can model this discourse context by setting higher prior probabilities on the worlds in which all of the apples are the same color: World 0 (where all the apples are green) and World 3 (where all the apples are red). The modified dataset is shown in Figure 7.6.

Because there is now a previously established shared belief that all of the apples are the same color, the Pragmatic Listener's calculations for *Some are red* change, as shown in Figures 7.7-7.9.

**Figure 7.6.** World set, utterance set, and interpretation function for non-uniform scalar implicature example

	Red apples	Prior		Utterance	Prior		W <sub>0</sub>	W <sub>1</sub>	W <sub>2</sub>	W <sub>3</sub>
W <sub>0</sub>	0	0.45		U <sub>0</sub>	None are red.	0.33	1	0	0	0
W <sub>1</sub>	1	0.05		U <sub>1</sub>	Some are red.	0.33	0	1	1	1
W <sub>2</sub>	2	0.05		U <sub>2</sub>	All are red.	0.33	0	0	0	1
W <sub>3</sub>	3	0.45								

**Figure 7.7.** Vanilla Literal Listener calculation for non-uniform scalar implicature example

Literal Listener: $p(w u_0)$	Literal Listener: $p(w u_1)$	Literal Listener: $p(w u_2)$
$p(w_0 u_0) = \frac{0.45}{0.45} = 1$	$p(w_0 u_1) = \frac{0}{0.55} = 0$	$p(w_0 u_2) = \frac{0}{0.45} = 0$
$p(w_1 u_0) = \frac{0}{0.45} = 0$	$p(w_1 u_1) = \frac{0.05}{0.55} = 0.09$	$p(w_1 u_2) = \frac{0}{0.45} = 0$
$p(w_2 u_0) = \frac{0}{0.45} = 0$	$p(w_2 u_1) = \frac{0.05}{0.55} = 0.09$	$p(w_2 u_2) = \frac{0}{0.45} = 0$
$p(w_3 u_0) = \frac{0}{0.45} = 0$	$p(w_3 u_1) = \frac{0.45}{0.55} = 0.82$	$p(w_3 u_2) = \frac{0.45}{0.45} = 1$

**Figure 7.8.** Pragmatic Speaker calculation for non-uniform scalar implicature example

Pragmatic Speaker: $p(u w_0)$	Pragmatic Speaker: $p(u w_2)$
$p(u_0 w_0) = 1$	$p(u_0 w_2) = 0$
$p(u_1 w_0) = 0$	$p(u_1 w_2) = 1$
$p(u_2 w_0) = 0$	$p(u_2 w_2) = 0$
Pragmatic Speaker: $p(u w_1)$	Pragmatic Speaker: $p(u w_3)$
$p(u_0 w_1) = 0$	$p(u_0 w_3) = 0$
$p(u_1 w_1) = 1$	$p(u_1 w_3) = 0.45$
$p(u_2 w_1) = 0$	$p(u_2 w_3) = 0.55$

Because the probability of World 3 is high to begin with, the Pragmatic Listener becomes much more likely to interpret *Some are red* as a description of World 3 than as a description of World 1 or World 2, even though *All are red* uniquely describes World 3.

**Figure 7.9.** Vanilla RSA calculation for *Some are red* in non-uniform model

Pragmatic Listener: $p(w u_1)$
$p(w_0 u_1) = \frac{0}{0.3025} = 0$
$p(w_1 u_1) = \frac{0.05}{0.3025} = 0.17$
$p(w_2 u_1) = \frac{0.05}{0.3025} = 0.17$
$p(w_3 u_1) = \frac{0.2025}{0.3025} = 0.67$

Models in the RSA framework therefore incorporate the discourse context in a limited way. The prior distribution over worlds is one way of incorporating the world set from the Common Ground into a model of comprehension. Because the RSA framework is probabilistic, however, worlds are never completely eliminated from the world set; even when the probability of a particular world is 0, it is possible that a later utterance might

increase that world’s probability again. Thus, RSA models do not guarantee that the set of worlds is monotonically decreasing.

### 7.1.5 Extensions to the vanilla RSA model

There are several common modifications to the vanilla RSA model presented in Section 7.1.1. First, in addition to the basic Bayesian probability calculation, many RSA models also include an utterance cost function in the speaker calculation (Bergen et al., 2012). This represents the trade-off between informativity and sentence complexity: although the speaker could theoretically select an utterance that exactly isolates the observed world, as the size of the set of worlds increases, the complexity of this utterance would also increase. In practice, people often select a simpler, less informative utterance rather than a maximally informative but complex utterance. The RSA model encodes this principle as a cost function that penalizes more complex utterances. A common cost function is the length of the sentence, although syntactic complexity or processing complexity can also be considered.

Another common modification is to add a parameter that controls how rationally the speaker is assumed to behave (Goodman and Stuhlmüller, 2013). The utility parameter  $\beta$  is added to the Pragmatic Speaker to control whether the Pragmatic Speaker always selects the maximally likely utterance. When  $\beta$  is set to 1, the Pragmatic Speaker produces utterances with a frequency proportional to their likelihood; when it is set to 0, the Pragmatic Speaker selects utterances at random (Degen et al., 2019).

Lastly, some RSA models incorporate uncertainty over the lexical semantics of the utterances. In these models, the Pragmatic Speaker reasons jointly over the listener’s lexicon and the utterance to select, and the Pragmatic Listener, over the speaker’s lexicon and intended meaning. This has been useful for modeling scalar implicature (Bergen et al., 2012, 2016; Brochhagen et al., 2016; Potts et al., 2016) and word learning (Smith et al., 2013; Frank and Goodman, 2014).

## 7.2 Key properties of perspectival motion verb comprehension

Having introduced the basic principles of the RSA approach to discourse processing, I would like to lay out some of the key principles that a model of perspective identification should capture. There are four well-established patterns of interpretation for *come* and *go* in English: (1) competition between *go* and *come*; (2) contextual and truth-conditional constraint of perspective; (3) bias towards the speaker’s perspective; and (4) gradient availability of perspectives.

### **Property 1: Competition between *come* and *go***

Although there is disagreement about the lexical semantics of *go*, there is consensus that *go* gives rise to an implication that the perspective holder is not located at the destination of motion. This implication may arise because the lexical semantics of *go* are anti-perspectival: they require that the perspective holder not be located at the destination of motion (Fillmore, 1966; Oshima, 2006a). Or it may arise through pragmatic competition with *come*: the listener may reason that if the perspective holder were located at the destination of motion, the speaker would have used *come* (Wilkins and Hill, 1995; Sudo, 2018).

The RSA framework provides one way to model pragmatic competition, because it is a model of comprehension that involves reasoning about the speaker’s production process. I will show that the core predictions of the RSA model are robust to different lexical semantics of *go*, because the anti-perspectival implication can arise through pragmatic competition. An advantage of the RSA framework that I adopt is that it makes it easy to test out different lexical semantics and compare their impact on the model predictions.

### **Property 2: Context and truth-conditional constraint of perspective**

Relying solely on the Speaker Default heuristic gives rise to a system in which the speaker’s perspective should be assumed to be in use whenever possible. Although a speaker bias is empirically well-founded, it is challenging to define the conditions when the speaker’s perspective should not be assumed. The simplest such condition is when the truth conditions of the sentence are incompatible with a speaker-oriented interpretation. For instance, in (190), regardless of the discourse context, the speaker cannot be the perspective holder

because an individual cannot be simultaneously located at the destination of motion and in motion towards that destination.

190. I am coming to the zoo.

However, I have argued in Chapter 6 that switching from the speaker's perspective only when falsity would arise is too restrictive of a condition. There are also environments in which the speaker's perspective is consistent with the truth conditions of the sentence, but inconsistent with the discourse context, as in (191).

191. *Sam is at the cafe.* Sam: Thelma is coming to the zoo.

Thus, one of the major challenges in Harris (2012)'s two-stage system of perspective identification is defining the conditions for switching between the heuristic-based system and the more costly reasoning system.

The RSA model that I propose, on the other hand, takes into account both the truth-conditional content of utterances and the discourse context (in a limited way). This provides a way of formalizing the conditions under which the speaker's perspective is unavailable, as well as more generally capturing the impact of the discourse context and truth conditions of an utterance on perspective inference.

### **Property 3: Bias towards the speaker's perspective**

We have seen that there is a strong preference for the perspective of the speaker. This preference can be built into a perspective identification using the Speaker Default heuristic. A weakness of rule-based systems, however, is that they predict exactly one interpretation: either the speaker's perspective is available, in which case it is used, or it is unavailable, in which case it is not. Rule-based systems do not predict ambiguity in perspectival anchoring. As we have seen, however, there are environments in which the speaker's perspective is available, but other perspectives may be used, as in (188), repeated as (192).

192. *In his London flat, Poirot is telling Miss Lemon about a call he received from Chief Inspector Japp, at Scotland Yard.*

Poirot: Chief Inspector Japp thinks that the murderer will come to confess.

In this case, *come* may be interpreted relative to the perspective of Poirot, the speaker, or relative to Japp, the attitude holder. This variability is unexpected if listeners are using a Speaker Default heuristic, which posits that if the speaker's perspective is available, then it is obligatorily used.

The model that I propose achieves a bias towards the speaker's perspective by means of a cost function. This imposes a preference for the speaker's perspective that can be overridden when the context and truth-conditional content of the sentence provide enough support for a different perspective. This incorporates the speaker default proposed by Harris (2012) and motivated by empirical evidence discussed in Chapter 6 in a more flexible, graded way.

#### **Property 4: Gradient acceptability of perspectives**

The availability of multiple perspectives in examples like (192) is related to the last property of perspective that I seek to capture: the gradience of judgments about perspectival anchoring. Although *come* and *go* reflect opposite perspectival anchorings (whether lexically encoded or from pragmatic competition), in many contexts, American English speakers accept either verb. This reflects the fact that there are often multiple viable perspective holders within a particular context. Ideally, we would like to model the relative prominence of these perspective holders instead of predicting the single most prominent one.

An advantage of the RSA framework is that its output is probabilistic. Because this makes it easy to map the predictions of the model onto gradient experimental data, an RSA-based model is naturally suited to explore questions about the relative prominence of perspective holders in contexts where multiple perspectives are available. By contrast, the Speaker Default heuristic can only predict whether or not the speaker's perspective is adopted; either the context is consistent with speaker-anchoring, and the speaker's perspective is used, or it is not. The heuristic approach model cannot tell us how *likely* the speaker's perspective is to be used in a particular context.

I have outlined four key properties of perspectival motion verbs that a perspective identification system should capture. In Section 7.3, I present a perspectival identification model in

the Rational Speech Acts framework. I return to the properties discussed above in Section 7.4 and demonstrate how they are captured by the proposed model.

### 7.3 An RSA model of perspectival expressions

So far I have presented the basic RSA framework and shown that it provides a model of sentence comprehension with two desirable properties: (1) reasoning over alternatives and (2) sensitivity to the discourse context. In this section I describe a model of joint perspective inference and sentence comprehension in the RSA framework.

In the vanilla RSA model presented in Section 7.1.1, listeners reason about the speaker’s intended meaning given the utterance that they hear. When a listener hears a perspectival expression, however, they must also figure out whose perspective the speaker has adopted.

In some cases, it is advantageous to reason simultaneously about both these tasks. When inferring the adopted perspective, the truth-conditional content of the utterance can be useful in narrowing the set of possible perspectives. For instance, if the listener hears a sentence like (193), in which the speaker is the subject of *come*, they can infer that the speaker is not using their own perspective, since the speaker cannot simultaneously be in motion and at the destination of motion.

193. Miss Marple: I will come to the Vicarage in 15 minutes.

On the other hand, if the listener knows whose perspective the speaker is using, they may be able to glean more information about the world that the speaker is trying to describe. For instance, if the listener knows that the speaker is using their own perspective and they hear a sentence like (194), where the destination of motion is implicit, they may be able to use their knowledge of the speaker’s location to infer the destination of motion. This will allow them to make a more accurate guess about the speaker’s intended meaning.

194. Holmes: I expect the murderer to come at any moment.

Thus, I propose that comprehension of perspectival expressions should be modeled as a **joint reasoning task**: given an utterance, the listener reasons simultaneously about the speaker’s intended meaning and their adopted perspective.

In the Perspectival Rational Speech Acts model (PRSA), the listener tries to estimate the joint probability of a world/perspective pair, given an utterance selected by the speaker. Formally, the listener's goal is to estimate  $p(w, a|u)$ , where  $a$  represents a perspective,  $w$  a possible world, and  $u$  an utterance.

The Pragmatic Listener reasons about the speaker's adopted perspective and their intended meaning using a mental model of the speaker's production process: the Pragmatic Speaker. In the PRSA model of comprehension, the listener assumes that the speaker is also engaged in a joint reasoning task. The speaker's goal is to select the utterance/perspective pair that is most likely to succeed in communicating their intended meaning to the listener. Formally, the speaker's goal is to calculate  $p(u, a|w)$ ; to do this they again rely on a simplified model of the listener, the Literal Listener.

Since both the Pragmatic Listener and Pragmatic Speaker involve joint reasoning, on the surface their relation to each other is less clear than in the original RSA model: the Pragmatic Speaker's output is  $p(u, a|w)$ , but the Pragmatic Listener's calculation involves  $p(u|w, a)$ . However, we can apply the Chain Rule in reverse, to collapse  $p(u|w, a)p(w, a)$  to the joint probability  $p(u, w, a)$ , and then expand again by the Chain Rule to  $p(u, a|w)p(w)$  to uncover the relationship between the left hand side of the Pragmatic Speaker and the right hand side of the Pragmatic Listener. Figure 7.10 shows a step-by-step derivation.<sup>5</sup>

Thus, in the final model, the Pragmatic Listener reasons about the joint probability of a world-perspective pair, given an utterance, by estimating the probability of the Pragmatic Speaker selecting that utterance to convey that world-probability pair, discounted by the prior probability of the world.

The Pragmatic Speaker selects the best utterance-perspective pair to convey an observed possible world, using the Literal Listener model, discounted by the probability of the utterance and perspective co-occurring. They calculate the joint probability of the utterance and

---

<sup>5</sup>The softmax term is again elided for readability.

**Figure 7.10.** Perspectival Rational Speech Acts model (preliminary)

<b>Literal Listener</b>	
$p(w u,a)$	$\propto p(u,a w) p(w)$ by Bayes' Rule
	$\propto p(u,a,w)$ by definition of joint probability
	$\propto p(u a,w)p(a,w)$ by definition of joint probability
	$\propto p(u a,w)p(a)p(w)$ by assumption of independence
	$\propto [[u]]^{a,w} p(w)p(a)$
<b>Pragmatic Speaker</b>	
$p(u,a w)$	$\propto p(w u,a) p(u,a)$ by Bayes' Rule
	$\propto \text{LitList}(w,u,a) p(u,a)$ by definition of literal listener
	$\propto \text{LitList}(w,u,a) p(u a) p(a)$ by definition of joint probability
<b>Pragmatic Listener</b>	
$p(w,a u)$	$\propto p(u w,a) p(w,a)$ by Bayes' Rule
	$\propto p(w,a,u)$ by Chain Rule
	$\propto p(u,a   w) p(w)$ by Chain Rule
	$\propto \text{PragSpeak}(u,a,w) p(w)$ by definition of pragmatic speaker

perspective by multiplying the prior probability of the perspective by the probability of the message given the perspective, which is found by marginalizing over all possible worlds.<sup>6</sup>

What about the Literal Listener? In the PRSA model, the Pragmatic Speaker makes an additional simplifying assumption about the listener. They assume that the listener has direct access to the perspective that they have adopted. The Literal Listener, therefore is not a joint reasoner; their task is to infer the probability of a world given a perspective-utterance pair,  $p(w|u, a)$ . In practice, the Literal Listener is unlikely to exist, since perspectives are not usually directly observable.

The key insight of the PRSA model of perspectival comprehension is that the listener reasons jointly over the speaker's intended meaning and their adopted perspective. The PRSA model also takes into account the discourse context in a limited form: the prior probability distribution over worlds reflects the shared beliefs of the speaker and the listener, since the same distribution is used in the Pragmatic Listener (the listener's actual comprehension calculation) and the Pragmatic Speaker (the listener's mental model of the speaker's

---

<sup>6</sup>As before, I assume that the speaker's utterances are sampled in proportion to their utility.

production calculation).<sup>7</sup> Thus, the PRSA model fulfills the function proposed by Harris (2012) for the more costly reasoning system: it takes into account information from various sources and reasons over all perspective holders in order to find the optimal fit with the speaker’s selected utterance.

### 7.3.1 Model implementation

Thus far I have presented mathematical models in the RSA framework. However, in order to generate testable predictions, it is useful to implement a computational model that can be run over various datasets. The results that I report are from computational models implemented in the WebPPL programming language (Goodman and Stuhlmüller, 2019). Although theoretically the RSA model is infinitely recursive, I follow common practice by bounding the recursive reasoning process at the first-order Pragmatic Listener calculations.

Running a simulation requires a dataset consisting of a set of utterances, a set of possible worlds, and a set of perspectives, along with their prior probability distributions, and an interpretation function that operates on the lexical semantics of the utterances. Except where noted, I use uniform distributions over perspectives and worlds, to simulate the effect of an empty discourse context. Further details of the model implementation and specification can be found in Appendix D.

## 7.4 Capturing key properties of perspective identification

Having defined the basic PRSA model, let us see how it can be applied to capture the properties of perspective inference outlined in Section 7.2. In this section, I illustrate how the PRSA model works using several small simulations.

### 7.4.1 Property 1: Pragmatic competition between *come* and *go*

One of the benefits of computational modeling in the RSA framework is that it provides a way to test the impact of adopting different lexical semantics. Using the PRSA model, it is easy to compare the two different semantics that have been proposed for *go*. As discussed

---

<sup>7</sup>Equivalent to the set of possible worlds in the Common Ground that are considered candidates for the actual world.

in Section 7.2, American English speakers often judge *go* infelicitous in contexts where the perspective holder is located at the destination of motion, as in (195).

195. *Tommy and Tuppence Beresford are talking in their office at Blunt's Brilliant Detectives Agency in London.* Tuppence: # Albert is going here.

There are two approaches that have been taken to derive this infelicity. Some have posited that the lexical semantics of *go* are similar but opposite to those of *come*: they require that the perspective holder not be located at the destination of motion (Fillmore, 1966; Oshima, 2006a). An alternative view posits that the anti-perspectival implication arises through pragmatic competition with *come* (Wilkins and Hill, 1995; Sudo, 2018).

196. **Perspectival semantics for *go*:**  $[[\text{go}]]^{C,g} = \lambda x.\lambda e.\text{MOVE}(e) \wedge \text{DEST}(e,x) \wedge x = p_t$

197. **Plain semantics for *go*:**  $[[\text{go}]]^{C,g} = \lambda x.\lambda e.\text{MOVE}(e) \wedge \text{DEST}(e,x)$

The PRSA model shows that the anti-perspectival implicature can arise through pragmatic reasoning alone. To see how this works, we can compare the predictions of the PRSA model when it uses a perspectival semantics for *go* to its predictions when it uses a plain semantics.

#### 7.4.1.1 Dataset

The PRSA model requires a dataset with a set of possible worlds, a set of utterances, and a set of perspectives, along with an interpretation function that maps from combinations of utterances, worlds, and perspectives to truth values. For this example, I assume uniform prior probability distributions over the worlds and perspectives.

#### Perspective set

We will consider a set of worlds with just three entities: Sam, the speaker; Lucy, the listener, and their friend Thelma, a third party who is not involved in the conversation. I will assume that the perspective set contains the speaker's and listener's utterance time perspectives,<sup>8</sup> but not Lucy's. We have seen that there are other possible perspective holders (see Section 2.2.1), but to simplify the demonstration, I will use a minimal perspective set.

---

<sup>8</sup>These are commonly assumed to be automatically entered into the Common Ground (Roberts, 2015).

**Figure 7.11.** Lexical semantics for the candidate utterances

**Lexical semantics of *walk*:**

$[[\text{walk}(x, d)]]^{w,a} = \text{T}$  iff

1. Motion implication:  $[[\exists e.\text{MOVE}(x, e) \wedge \text{DEST}(d, e)]]^{w,a} = \text{T}$

**Lexical semantics of *walk*:**

$[[\text{come}(x, d)]]^{w,a} = \text{T}$  iff

1. Motion implication:  $[[\exists e.\text{MOVE}(x, e) \wedge \text{DEST}(d, e)]]^{w,a} = \text{T}$

2. Anchoring implication:  $\exists y.[[\text{LOC}(y, d)]]^{w,a} = \text{T}$  and  $y$  is a prominent individual holding  $a$ .

**Lexical semantics of *go*, perspectival version:**

$[[\text{go}(x, d)]]^{w,a} = \text{T}$  iff

1. Motion implication:  $[[\exists e.\text{MOVE}(x, e) \wedge \text{DEST}(d, e)]]^{w,a} = \text{T}$

2. Anchoring implication:  $\exists y.[[\text{LOC}(y, d)]]^{w,a} = \text{T}$  and  $y$  is a prominent individual holding  $a$ .

**Lexical semantics of *go*, non-perspectival version:**

$[[\text{go}(x, d)]]^{w,a} = \text{T}$  iff

1. Motion implication:  $[[\exists e.\text{MOVE}(x, e) \wedge \text{DEST}(d, e)]]^{w,a} = \text{T}$

where  $w$  is a world,  $a$  is a perspective,  $d$  is a destination, and  $x$  is an entity.

### Utterance set

The utterance set will consist of three sentence frames: *X is going to Northampton*, *X is coming to Northampton*, and *X is walking to Northampton*, where X represents any of the individuals, for a total of 9 utterances.

Figure 7.11 shows the lexical semantics for each candidate utterance. The lexical semantics of *walk* merely specify that its subject is in motion to the destination. For *come*, the lexical semantics require (1) that its subject is in motion towards the destination and (2) that the perspective holder is located at the destination of motion.<sup>9</sup> There are two versions of *go*: a plain version, whose semantics is identical to that of *walk*, and a perspectival version, which requires that the perspective holder is not at the destination of motion.

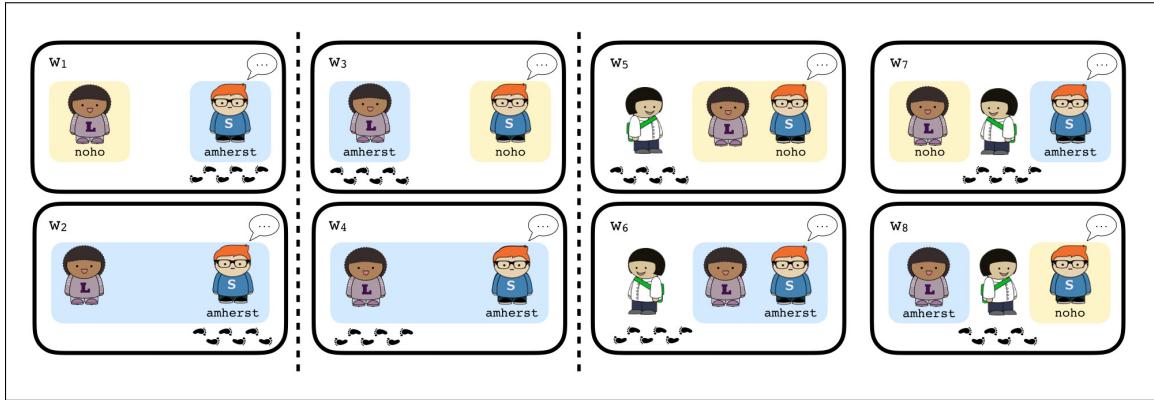
### World set

Each world in the set of possible worlds contain two locations, Northampton and Amherst, and three individuals: Thelma, Sam, and Lucy. We will consider the eight worlds in which

---

<sup>9</sup>The subject and perspective holder cannot be the same individual, since it is logically inconsistent to be both in motion towards a place and already located in that place.

**Figure 7.12.** World set (destination indicated by yellow box).



exactly one person is moving. The set of possible worlds is shown in Figure 7.12. Since we are not considering Thelma’s perspective, her location is irrelevant, and I omit it from the representations of the worlds.

#### 7.4.1.2 Model predictions

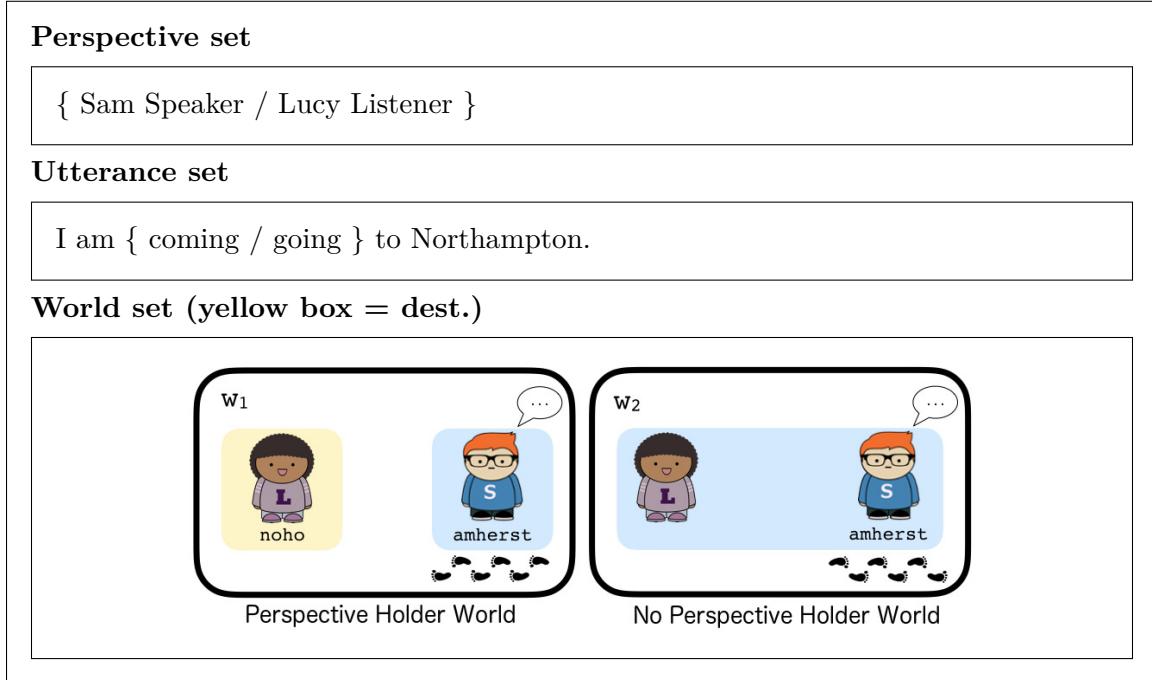
In order to understand the predictions of the PRSA model, let’s consider a simple case: a scenario where the speaker is the mover. There are two worlds consistent with this scenario: World 1 (where Sam is the mover and Lucy is in Northampton) and World 2 (where Sam is the mover and Lucy is in Amherst). I will refer to these respectively as the Perspective Holder world and the No Perspective Holder world, based on whether or not Lucy is at the destination of motion.

There are three sentences that can describe the speaker’s motion event: *I am coming to Northampton*, *I am going to Northampton*, and *I am walking to Northampton*. How does the Pragmatic Listener interpret each of these utterances?

##### **I am coming to Northampton**

When the Pragmatic Listener hears *I am coming to Northampton*, she reasons over world-perspective pairs to decide which is most likely. This utterance is only literally true of two worlds, the Listener world and the None world. Thus, there are really four world-perspective pairs at play: Perspective Holder + Lucy, Perspective Holder + Sam, No Perspective Holder + Lucy, and No Perspective Holder + Sam. However, the speaker cannot be the perspective

**Figure 7.13.** Speaker = mover dataset



holder for *come* when they are also the subject of *come*, which eliminates two of these pairs: Lucy must be the perspective holder.

This still leaves two possibilities: Perspective Holder + Lucy, and No Perspective Holder + Lucy. The lexical semantics of *come*, however, are consistent with only one of these choices: in the None world, Lucy is not located at the destination of motion. Therefore *come* cannot be used to describe this world when Lucy is the perspective holder.<sup>10</sup>

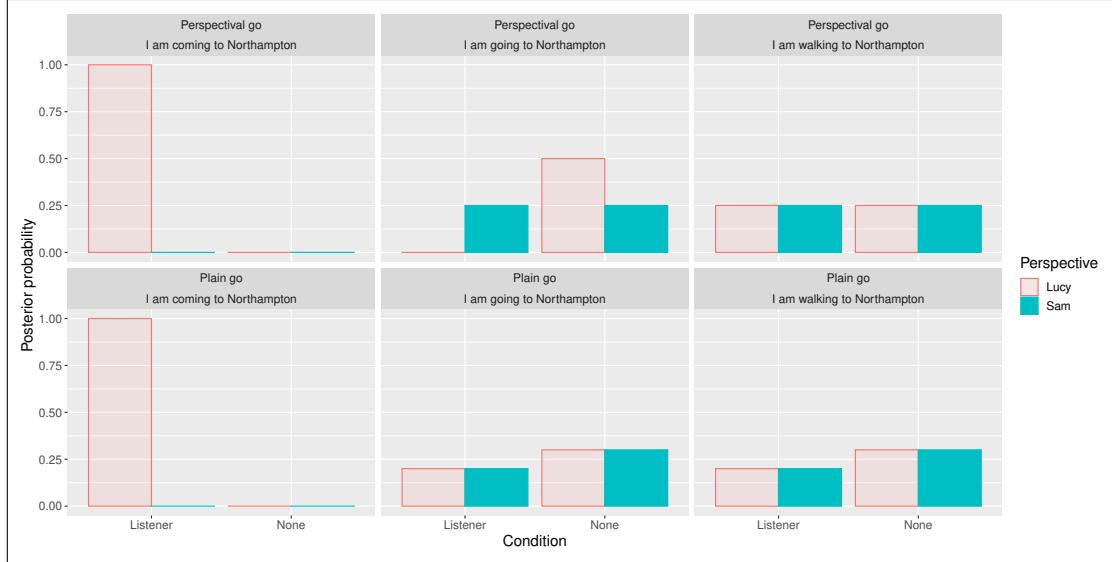
When the Pragmatic Listener hears *I am going to Northampton*, the lexical semantics of *go* come into play. There are again four world-perspective pairs at play: Perspective Holder + Lucy, Perspective Holder + Sam, No Perspective Holder + Lucy, and No Perspective Holder + Sam.

If the semantics for *go* are perspectival, the perspective holder cannot be at the destination of motion. Sam meets this criterion in all of the worlds under consideration, but Lucy meets it only in one: the No Perspective Holder world. Thus, in the perspectival *go* model, there

---

<sup>10</sup>In fact, *come* can never be used to describe the None world, since there is no perspective holder in our perspective set located at the destination of motion.

**Figure 7.14.** Posterior probabilities of world-perspective pairs when mover = speaker and perspective cost = 0



are only three world-perspective pairs allowed by the lexical semantics of *go*: Perspective Holder + Sam, No Perspective Holder + Lucy, and No Perspective Holder + Sam. The marginal posterior probabilities over worlds (calculated by summing over paired perspectives for each world) are highest for No Perspective Holder, since either perspective can be used with *go*.

If the lexical semantics for *go* are non-perspectival, all four perspective pairs are consistent with the lexical semantics of *I am going to Northampton*. However, because *I am coming to Northampton* is a much better alternative if the world is Perspective Holder and the perspective holder is Lucy, the Pragmatic Listener will still assign very little probability to the Perspective Holder + Lucy pair when they hear *I am going to Northampton*.

The model predictions for the perspectival and non-perspectival *go* variants are shown in Figure 7.14. As can be seen, the model predictions look very similar: the anti-perspectival implication of *go* can arise through lexical stipulation or through pragmatic competition.

There are two small differences between the variants. First, in the non-perspectival variant, since we assigned the same lexical semantics to *go* and *walk*, we predict no difference in the use of *I am going to Northampton* and *I am walking to Northampton*. Second, a less

intuitive difference is that when the speaker's perspective is used, the likelihood of the worlds differs for the plain *go* semantics, but not for the perspectival *go* version. This is a result of the joint inference over perspectives and worlds: if the world being described is the Listener world, in the non-perspectival version, using the speaker's perspective and *I am going to Northampton* is possible, but not preferable, because it is better to adopt the listener's perspective and use *come*. In the perspectival version, using the speaker's perspective with *I am going to Northampton* is not a very good choice, since adopting the listener's perspective with *I am coming to Northampton* or *I am going to Northampton* makes the world unambiguous: the speaker must be describing the Listener world and the No Perspective Holder respectively.

Despite these small differences in the Pragmatic Listener's predictions, the main predictions of the PRSA model do not depend on whether a perspectival or a plain semantics for *go* is adopted. As previous theoretical work had posited, and as the PRSA model simulations show, the anti-perspectival interpretation of *go* can arise through pragmatic competition.

#### 7.4.2 Property 2: Context and truth-conditional constraint of perspective

In perspective identification, the discourse context, the truth-conditional content of the utterance, and the available perspective holders are all mutually informative. The PRSA model provides one way of modeling how the listener reasons about all three of these sources of information. A key principle of the PRSA model is that listeners reason simultaneously about speaker's intended meaning and their choice of perspective; this joint inference process also incorporates the listener's knowledge of the discourse context via the prior probability distribution over worlds.

We can contrast this with a simpler heuristic approach. Using Speaker Default alone, the listener assumes that the perspective is that of the speaker, and attempts to interpret the utterance with this perspective. If this fails, the process repeats.

To illustrate why treating utterance interpretation and perspective inference as a joint reasoning process can be advantageous, I present two motivating examples. In the first example, knowledge of the perspective holders' locations aids the listener's inference about

**Figure 7.15.** Dataset for Example (198)

Utterance set				
{ Thelma is coming. / Thelma is walking. / Thelma is going. }				
World set				
	Thelma's destination	Sam's location	Lucy's location	Prior
World 1	Noho	Amherst	Noho	$p(0.23)$
World 2	Amherst	Amherst	Noho	$p(0.23)$
World 3	Easthampton	Amherst	Noho	$p(0.23)$
World 4	Greenfield	Amherst	Noho	$p(0.23)$
World 5	Greenfield	Greenfield	Greenfield	$p(0.02)$
World 6	Easthampton	Easthampton	Easthampton	$p(0.02)$

an implicit argument of the utterance. In the second, world knowledge and the truth conditional content of the sentence guide the listener's inference about the perspective holder.

#### 7.4.2.1 Example 1: Knowledge about location aids destination inference

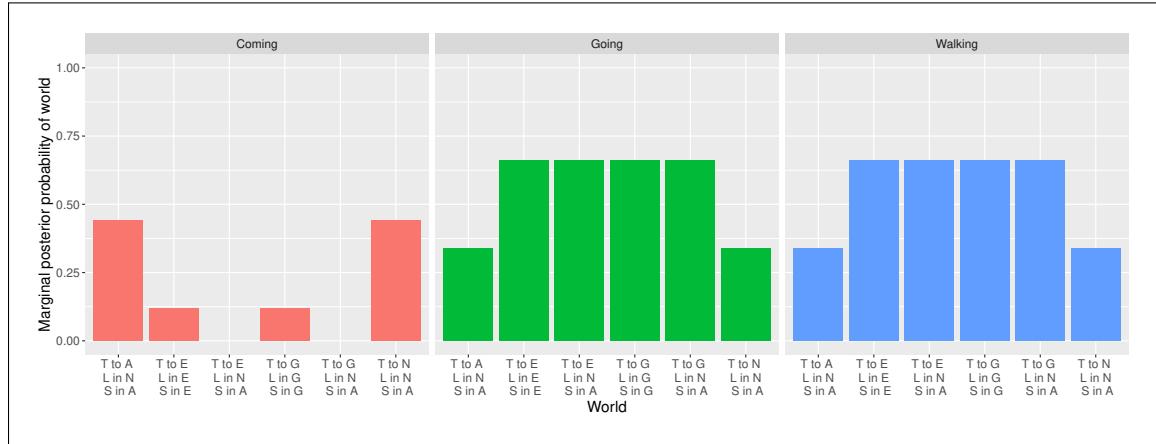
The destination of motion is not always explicit in sentences using *come*. When the destination of *come* is implicit, knowledge about the locations of the potential perspective holders can aid the listener's inference about the destination of motion. For instance, in (198), if Lucy knows her own location and that of Sam, she can infer that Thelma's intended destination is either Amherst (if Sam is using his own perspective) or Northampton (if Sam is using his own).

198. *Lucy Listener is in Northampton. Sam Speaker is in Amherst.*

Thelma is coming.

The PRSA model captures this inference easily. Consider a world set with four destinations: Northampton, Easthampton, Amherst, and Greenfield. Thelma is about equally likely to be moving to any location, but Lucy and Sam are very likely to be Northampton and Amherst respectively (Figure 7.15). We will assume that the utterance set consists of the same three utterances as in Section 7.4.1.1, but with the destination omitted (Figure 7.15).

**Figure 7.16.** Pragmatic Listener posterior marginal probabilities over worlds for (198), plain semantics for *go*



In this scenario, when the Pragmatic Listener hears *Thelma is coming*, she will calculate that the worlds in which Thelma is moving to Greenfield or Easthampton are very unlikely. This is because it is very unlikely for either perspective holder to be located in Greenfield or Easthampton. As Figure 7.16 shows, when *Thelma is coming* is used, the Pragmatic Listener calculates higher marginal posterior probabilities for worlds in which Thelma's destination of motion is Northampton or Amherst.

As this example demonstrates, the Pragmatic Listener in the PRSA model is able to use knowledge of the availability of perspective holders alongside world knowledge to guide their interpretation of the speaker's utterance. Treating perspective identification and utterance interpretation as a joint inference problem improves the Pragmatic Listener's ability to understand the speaker.

#### 7.4.2.2 Example 2: Knowledge about location and destination aid perspective identification

Treating perspective identification and utterance interpretation as a joint inference task also leads the Pragmatic Listener to make better predictions about the speaker's adopted perspective. As we have seen, both world knowledge and the truth-conditional content of a sentence can constrain the set of available perspective holders.

**Figure 7.17.** World set for Example (199)

	Thelma's destination	Sam's location	Lucy's location
World 1	Noho	Amherst	Amherst
World 2	Noho	Noho	Amherst
World 3	Noho	Easthampton	Amherst

**Figure 7.18.** Utterance set for Example (199)

{ Thelma is coming to Northampton. / Thelma is walking to Northampton. / Thelma is going to Northampton. }

When the destination of motion is explicit, it can be used in combination with world knowledge to infer the perspective holder more accurately. Consider the scenario in (199), in which the listener knows that she is not in Northampton.

199. *Lucy Listener is in Amherst.*

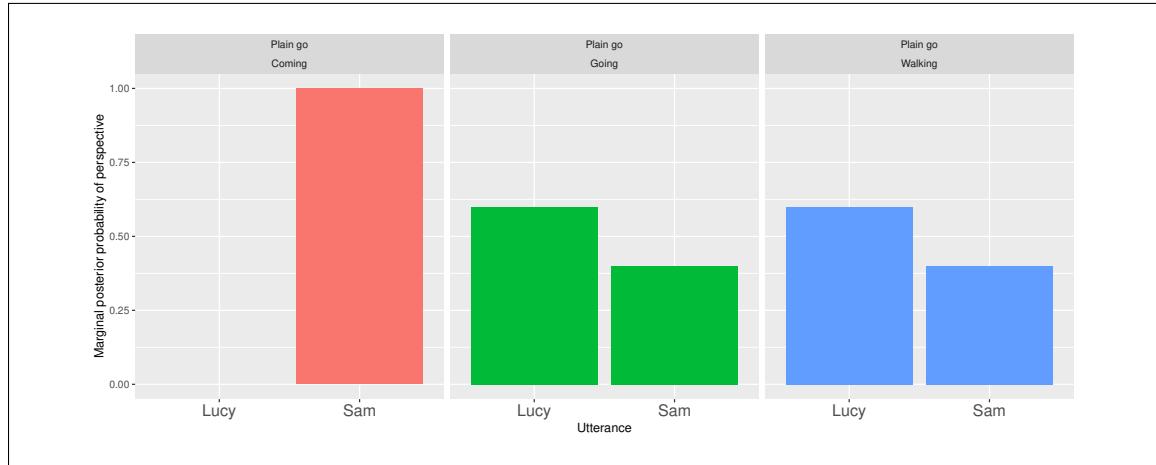
Thelma is coming to Northampton.

When Lucy hears *Thelma is coming to Northampton*, she can use her knowledge of her own location and her knowledge of the truth conditions of *come* to eliminate herself from the set of possible perspective holders. She is likely to infer that the speaker is using his own perspective.

For this example, we will consider the small set of worlds shown in Figure 7.17. In all of the possible worlds, Lucy's location is Amherst, representing the fact that Lucy is aware of her own location. I have also omitted the worlds that are not consistent with the truth conditions of any of the utterances in the utterance set shown in Figure 7.18, such as worlds in which Thelma's destination is not Northampton.

The Pragmatic Listener determines that the speaker must be using their own perspective if they say *Thelma is coming to Northampton*. As shown in Figure 7.19, the Pragmatic Listener calculates that the marginal posterior probability of the speaker perspective is highest.

**Figure 7.19.** Pragmatic Listener posterior marginal probabilities over perspectives for Figure 7.17 world set, plain semantics for *go*



This is because the prior probability distribution over worlds reflects the belief that the listener is not located in Northampton, and therefore, is not a possible perspectival anchor for *come* when Thelma is headed to Northampton.

In contrast to a heuristic like the Speaker Default, the PRSA model is able to consider the discourse context, the truth-conditional content of the utterance, and the set of possible perspective holders simultaneously, allowing the listener to consider the perspectival anchoring while interpreting the utterance meaning, and to use the utterance meaning when considering the perspectival anchoring.

#### 7.4.3 Property 3: Bias towards the speaker's perspective

The Speaker Default heuristic is based in the idea that speakers are biased towards their own perspectives (as discussed in Chapter 6), and therefore, most likely to use their own perspective. Currently, the PRSA model does not capture a preference for the speaker's perspective. In this section, I propose a modification of the PRSA model that implements a defeasible bias favoring the speaker's perspective via a perspective cost function.

This perspective cost function, which is introduced at the level of the Pragmatic Speaker, is a means of implementing a preference for selecting the speaker's perspective that can be defeated if the truth-conditional content of the utterance or the discourse context favor

**Figure 7.20.** Perspectival Rational Speech Acts model (revised)

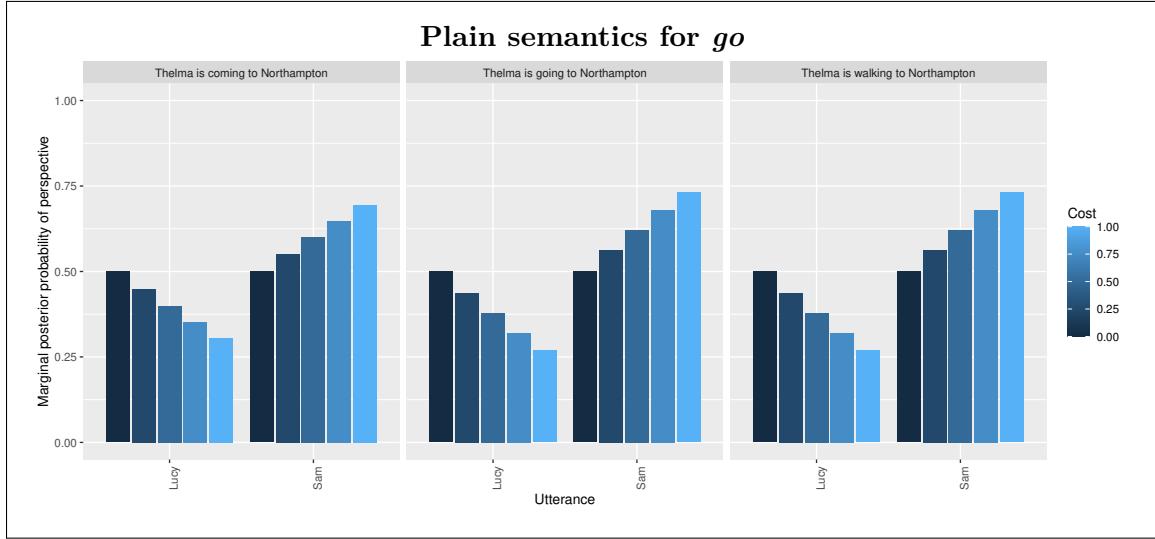
<b>Literal Listener</b>	
$p(w u,a)$	$\propto p(u,a w) p(w)$ by Bayes' Rule
	$\propto p(u,a,w)$ by definition of joint probability
	$\propto p(u a,w)p(a,w)$ by definition of joint probability
	$\propto p(u a,w)p(a)p(w)$ by assumption of independence
	$\propto [[u]]^{a,w} p(w)p(a)$
<b>Pragmatic Speaker</b>	
$p(u,a w)$	$\propto p(w u,a) p(u,a)$ by Bayes' Rule
	$\propto \text{LitList}(w,u,a) p(u,a)$ by definition of literal speaker
	$\propto \text{LitList}(w,u,a) p(u a) p(a)$ by definition of joint probability
	$\propto \text{LitList}(w,u,a) p(u a) p(a) - \text{Cost}_u(u) - \text{Cost}_p(a)$ adding cost functions
<b>Pragmatic Listener</b>	
$p(w,a u)$	$\propto p(u w,a) p(w,a)$ by Bayes' Rule
	$\propto p(w,a,u)$ by Chain Rule
	$\propto p(u,a   w) p(w)$ by Chain Rule
	$\propto \text{PragSpeak}(u,a,w) p(w)$ by definition of literal speaker

another perspective strongly enough. This can be interpreted as an egocentricity bias in the speaker's production process (or rather, the Pragmatic Listener's mental model of the speaker's production process). The revised PRSA model is presented in Figure 7.20.

The perspective cost function penalizes non-speaker perspectives. The strength of this egocentric bias is controlled by a perspective cost parameter. Increasing this parameter makes the speaker more likely to use their own perspective, while setting it to 0 removes the egocentric bias entirely.

A speaker bias could also be implemented by manipulating the prior probability of the speaker's perspective. These two options produce initially similar results in simulations with uniform world and utterance priors. However, over time, they will lead to different effects. The cost function approach assumes that there is a cost to accessing non-speaker perspectives that is fixed over a discourse, no matter how prominent a particular non-speaker perspective becomes. By contrast, the prior approach is well-suited to capturing a discourse-initial bias towards the speaker's perspective that can be mitigated by frequent access to

**Figure 7.21.** Effect of perspective cost on posterior perspective probabilities, subject = Thelma (cost increasing from 0 (left) to 1 (right))



non-speaker perspectives, resulting in increasing prominence for non-speaker perspectives as the discourse proceeds. This can be interpreted as a cognitive penalty for perspective shift (Ferguson et al., 2017), as in Harris (2012)'s Maintain Perspective principle.

I do not think there is conclusive evidence in favor of one approach over the other, but I adopt the cost function approach, leaving exploration of discourse prominence effects for future work.<sup>11</sup>

#### 7.4.3.1 Effect of perspective cost on model predictions

Manipulating the perspective cost parameter setting has a linear effect on the probabilities of speaker versus listener perspectives. Figure 7.21 shows the marginal posterior probabilities over perspectives calculated by the Pragmatic Listener for the Section 7.4.1.1 dataset when the utterance is *Thelma is coming to Northampton*. When the perspective cost is set to 0, the probabilities of the speaker and listener perspectives are equal, but as the perspective cost parameter is increased, the speaker perspective becomes increasingly likely.

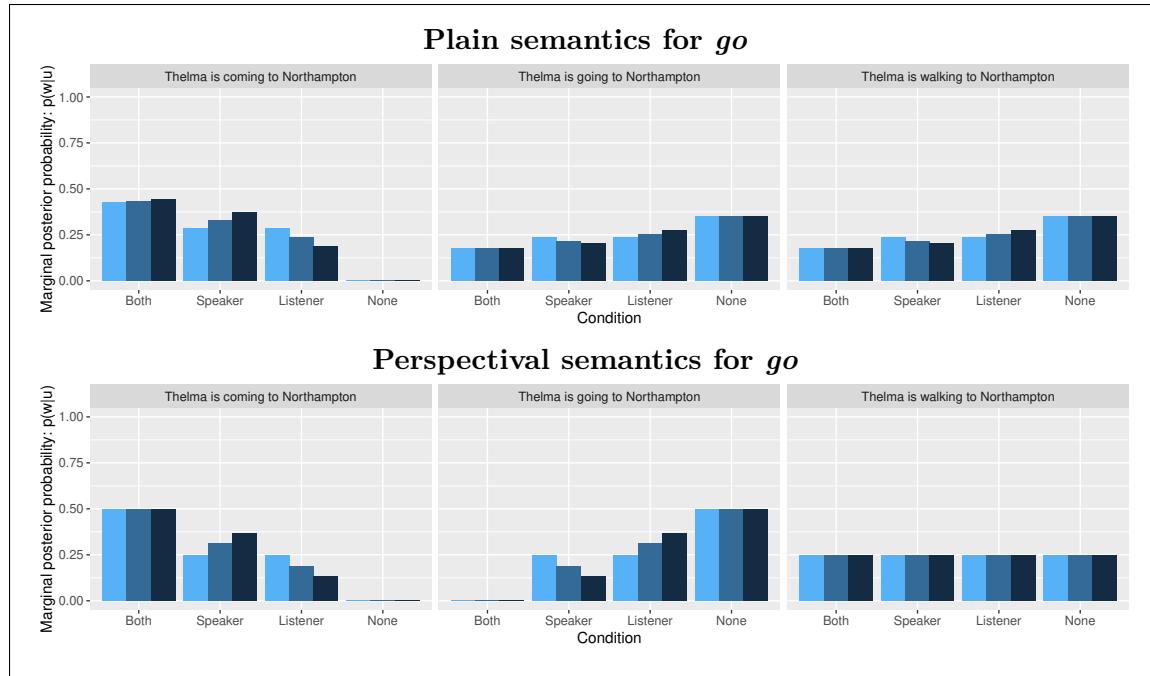
The perspective cost also affects the marginal posterior probabilities of the worlds (Figure 7.22). When the perspective cost is 0 and the utterance is *Thelma is coming to Northampton*,

---

<sup>11</sup>See Section 8.5.4 for some further discussion of using non-uniform priors over perspectives.

the marginal posterior probabilities of the world where only the speaker is in Northampton (Speaker world) and the world where only the listener is in Northampton (Listener world) are equal. As the perspective cost increases, the Speaker world becomes increasingly more likely than the Listener world.

**Figure 7.22.** Pragmatic Listener marginal posterior probabilities of utterances, mover = Thelma (cost parameters = 0, 0.5, 1 from left to right)



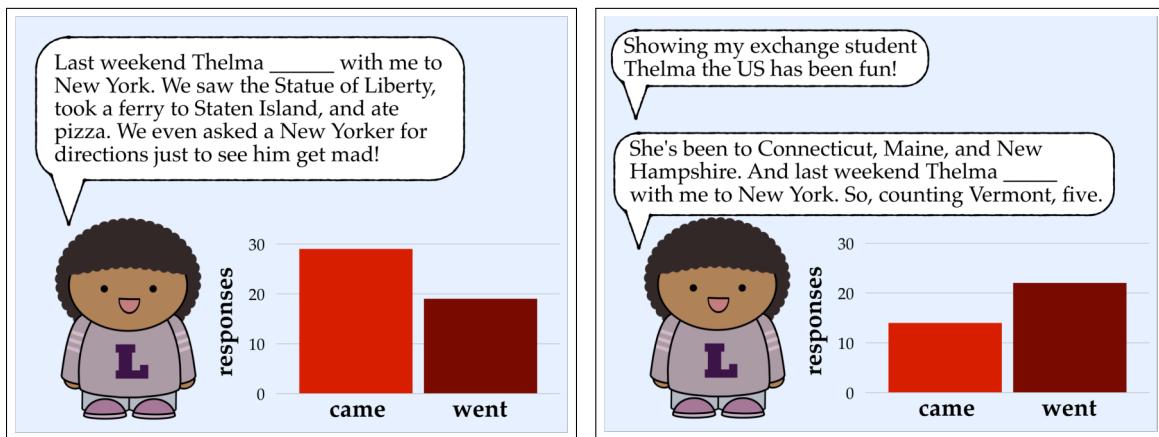
#### 7.4.4 Property 4: Gradience

One strength of the PRSA model is that it generates quantitative predictions about perspective prominence. The PRSA model can predict not just which perspective is most likely to have been used by the speaker, but also the relative likelihood of different perspectives. This makes it easy to map its predictions to gradient experimental data.

This is important for studying perspective inference, because the strength of the comprehender's preference for one perspective over another is variable in many contexts. For instance, I ran an informal poll to gather American English speakers' judgments on whether *came* or *went* is more natural in the sentence in (200).

200. Last weekend Thelma \_\_\_\_ with me to New York.

**Figure 7.23.** Poll responses for (201) (left) and (202) (right)



Their judgments varied depending on the context in which the target sentence was presented. In one poll, it was framed with a short discourse giving details about the New York trip (201). In another, the short discourse focused on the states that Thelma had visited (202).

201. Lucy: **Last weekend Thelma \_\_\_\_ with me to New York.** We saw the Statue of Liberty, took a ferry to Staten Island, and ate pizza. We even asked a New Yorker for directions just to see him get mad!

202. Lucy: Showing my exchange student Thelma the US has been fun!

Sam: How many states has she been to so far?

Lucy: She's been to Connecticut, Maine, and New Hampshire. **And last weekend Thelma \_\_\_\_ with me to New York.** So, counting Vermont, five.

In both cases, there were many votes in favor of each verb, but intuitions shifted based on the discourse context in which the sentence appeared. When presented as in (201), most participants preferred *came*; when presented as in (202), most participants preferred *went* (Figure 7.23). Although these results come from an informal poll, they illustrate the fact that judgments about the perspectival anchoring of an utterance are variable both within a given context and between discourse contexts.

Although the PRSA model would need to be enriched with a better theory of how to represent the discourse contexts of (201) and (202) in order to predict the judgment data shown

in Figure 7.23, this example illustrates why it is beneficial that the output of the PRSA model is probabilistic. While it would be possible to implement a perspective inference system in other frameworks, the RSA framework is particularly well-suited for modeling gradient phenomena like perspective, because it naturally outputs a distribution over possible anchorings. This makes it easy to test the predictions of the PRSA system against evidence from comprehension experiments.

#### 7.4.5 An unexpected prediction of the PRSA model

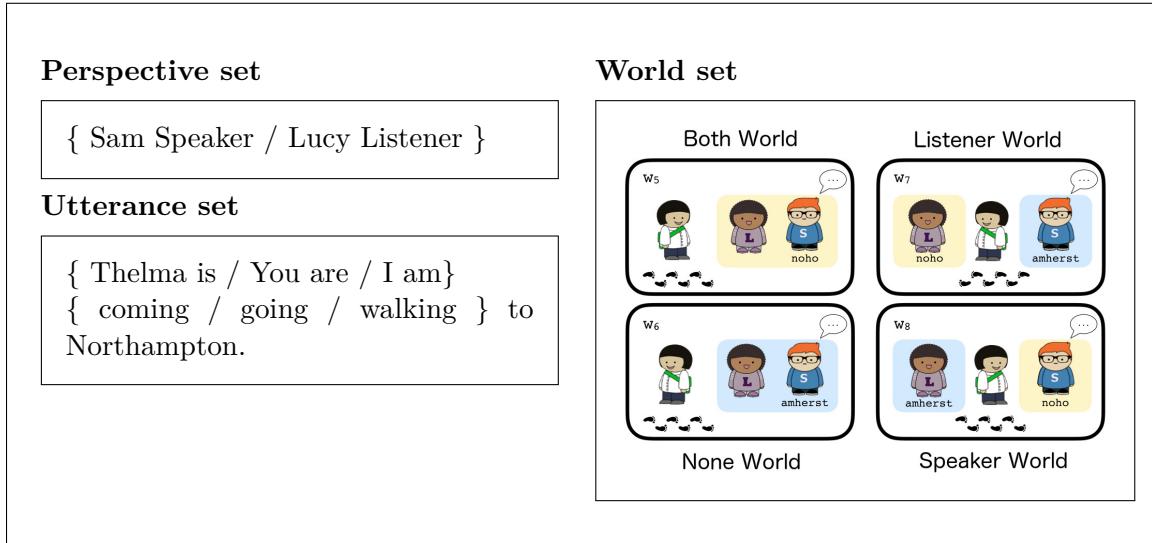
Thus far, the PRSA examples that I have discussed illustrate known properties of perspective in contexts where the judgments are well-established. However, the PRSA model also makes a prediction about the interpretation of perspectival expressions that has not been discussed in previous work. Because the Pragmatic Listener considers multiple perspectives simultaneously, given an utterance using *come*, the PRSA model predicts that the **marginal posterior probability** will be highest for worlds where multiple possible perspective holders are present at the destination of motion.

The Pragmatic Listener calculates the joint probability of a world-perspective pair given an utterance. However, in order to understand the listener’s beliefs about the speaker’s intended meaning, we are actually interested in the probability of a world given an utterance. We can calculate this term by marginalizing over perspective: for each world, we sum over the posterior probability of all of its world-perspective pairs.

What we find is that when a listener is processing a sentence like *Thelma is coming to Northampton* with no prior beliefs about the speaker’s and listener’s locations, the listener will think that it is most likely that both of them are located at the destination of motion.

This prediction arises from the fact that the PRSA listener considers all possible perspectives simultaneously. If there are multiple possible perspectives, a possible world in which all of the perspective holders are located at the destination of motion will be more likely, because that world gets some boost in probability from the possibility that each person located at the destination of motion might be the perspective holder.

**Figure 7.24.** Dataset for marginal posterior probability prediction illustration



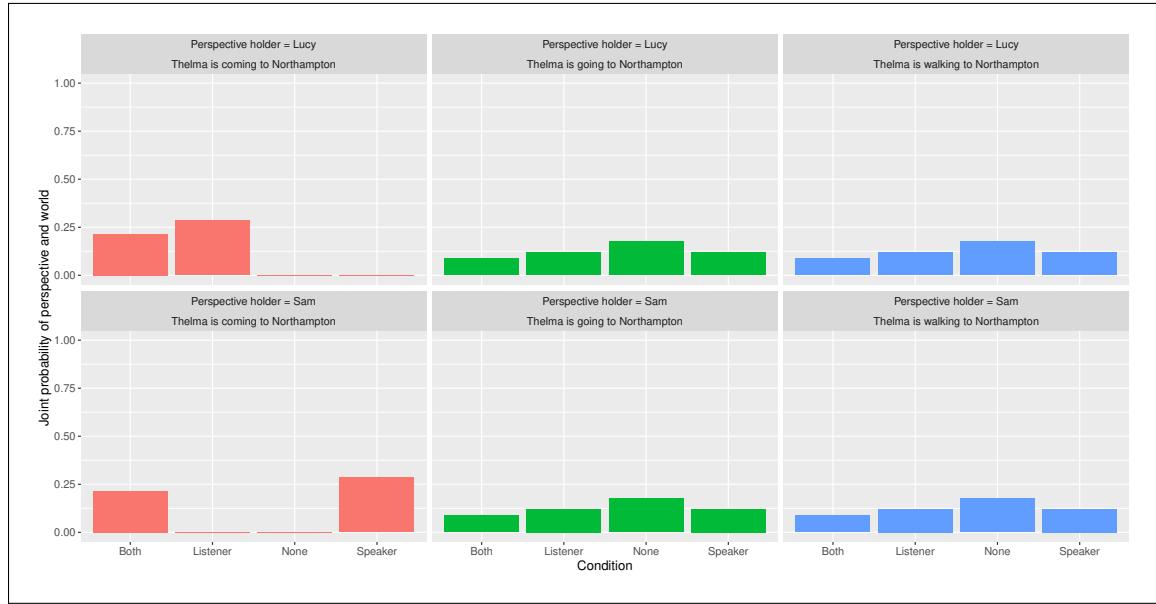
To see how this prediction is derived, we will look again at the model predictions for the Section 7.4.1.1 simulation. The set of utterances, perspective holders, and worlds are repeated in Figure 7.24.

In particular, let us consider the utterance *Thelma is coming to Northampton*. This is an interesting scenario because both perspectives are at play: when Thelma is the subject of the motion verb, neither of the perspectives are ruled out by the lexical semantics of the utterance. Figure 7.25 shows the posterior probabilities for world-perspective pairs predicted by the Pragmatic Listener.

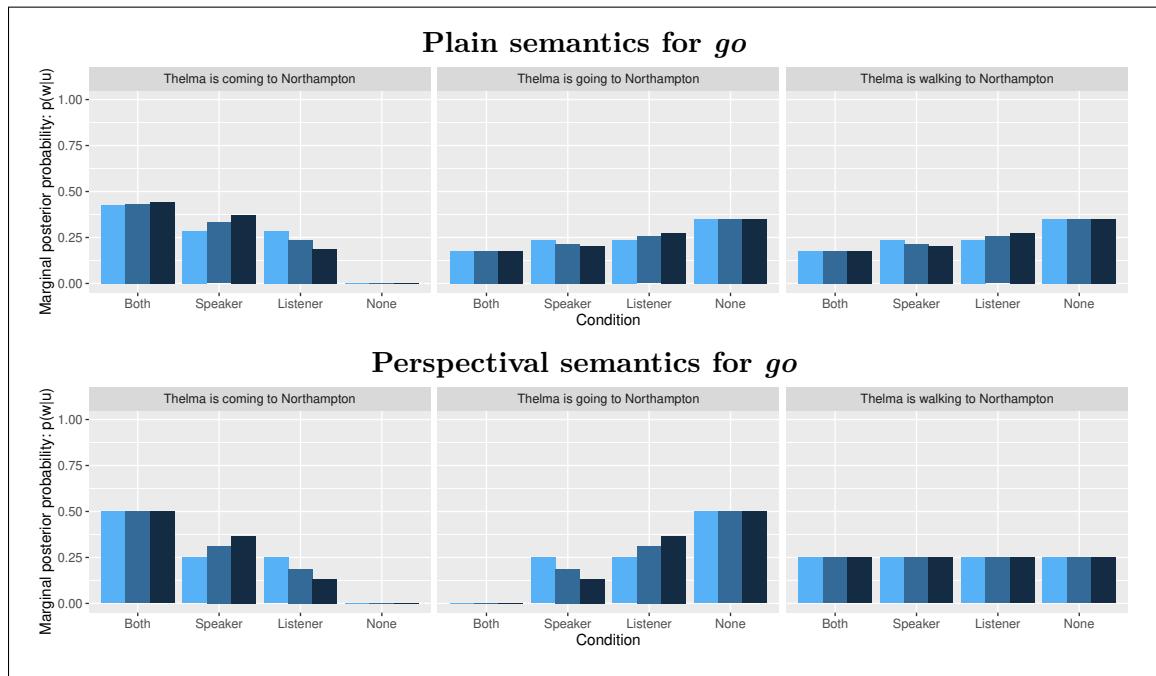
We see that if Lucy is the perspective holder, there are two likely worlds: the Listener world and the Both world. If Sam is the perspective holder, there are also two likely worlds: the Speaker world and the Both world. Because the Both world is a likely world under both possible perspective holders, its marginal posterior probability is highest.

Figure 7.26 shows the marginal posterior probability of each world (calculated by summing over perspectives). As we can see, the marginal posterior probability of the Both world is highest for either version of the semantics of *go*. Moreover, although the strength of this effect varies with the perspective cost setting, the effect is predicted to exist for any parameter cost setting, so long as the listener's perspective is not completely eliminated.

**Figure 7.25.** Posterior probabilities of world-perspective pairs, plain semantics, mover = Thelma and perspective cost = 0



**Figure 7.26.** Pragmatic Listener marginal posterior probabilities of utterances, mover = Thelma (cost parameters = 0, 0.5, 1 from left to right)



This prediction about the marginal posterior probability does not arise from a Speaker Default heuristic approach. According to the Speaker Default, all worlds in which the speaker is located at the destination of motion (the Speaker world and the Both world)

should be equally likely: since the listener's perspective is not considered unless the speaker's is inaccessible, it does not matter where the listener is located.

## 7.5 How do listeners interpret perspectival expressions?

The PRSA model makes a novel prediction concerning the marginal posterior probabilities of worlds when there are multiple available perspectives: given the sentence *Thelma is coming to Northampton*, the most probable world is one where both of the potential perspective holders are located at the destination of motion. This prediction is a consequence of the fact that in the PRSA model, the listener considers multiple perspectives simultaneously.

By contrast, using the Speaker Default heuristic on its own does not lead to this prediction, since the Speaker Default heuristic uses a serial model of perspective selection. Listeners first try to interpret the sentence relative to the speaker's perspective, and then reanalyze if necessary. Thus, given a sentence like *Thelma is coming to Northampton*, the Speaker Default heuristic predicts equal probability over all possible worlds in which the speaker is at the destination of motion. Because the listener's perspective is not considered unless the speaker's perspective is unavailable, it doesn't matter whether the listener is also at the destination of motion.

What do actual listeners do? In this section, I present results from two comprehension studies probing whether listeners reason simultaneously over multiple perspectives, as in the PRSA model, or rely on the simpler Speaker Default heuristic.

### 7.5.1 Experiment 5a method

Experiment 5a investigates the hypothesis that listeners consider multiple perspectives simultaneously using a sentence / image compatibility judgment task. According to the predictions of the PRSA model, when participants interpret *Thelma is coming to Northampton*, they are predicted to prefer worlds where both the speaker and listener are located at the destination of motion. By contrast, if listeners avoid using the more costly reasoning system by leaning on the simpler Speaker Default heuristic, they should assign equal probability to all worlds where the speaker is at the destination of motion.

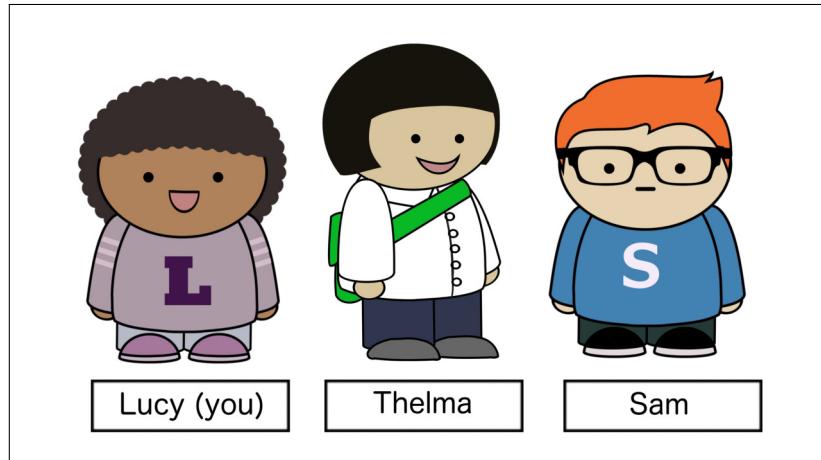
### 7.5.1.1 Participants

Monolingual American English-speaking participants ( $n=80$ ) were recruited through Prolific. Participants who achieved less than 90% accuracy on a spatial control task (described below) were excluded from the experiment.<sup>12</sup> This rejection criterion, as well as the experimental procedures and planned analyses described below, were preregistered through AsPredicted.

### 7.5.1.2 Materials

Participants were introduced to three characters: Thelma, Sam, and Lucy (Figure 7.27), and were asked to imagine themselves as Lucy.

**Figure 7.27.** PRSA experiment characters



Participants were instructed that Sam “sometimes gets confused and says things that don’t make sense,” and that their goal was to decide whether or not what Sam says makes sense according to the picture of the scene.

### Main items

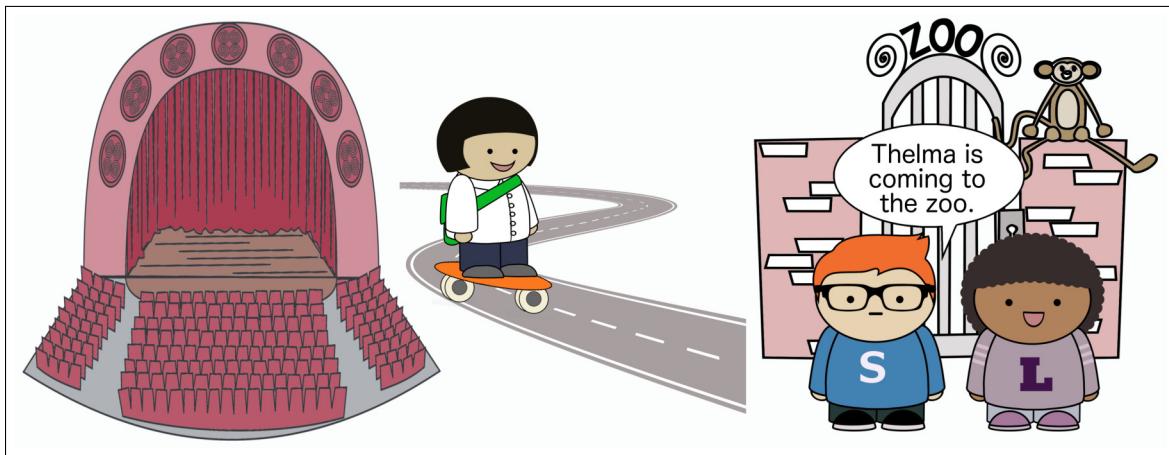
There were 8 conditions, formed by crossing scene type with motion verb type.

There were 4 versions of the scene: one with both the speaker and listener at the destination of motion; one with the speaker at the destination and the listener at the other location;

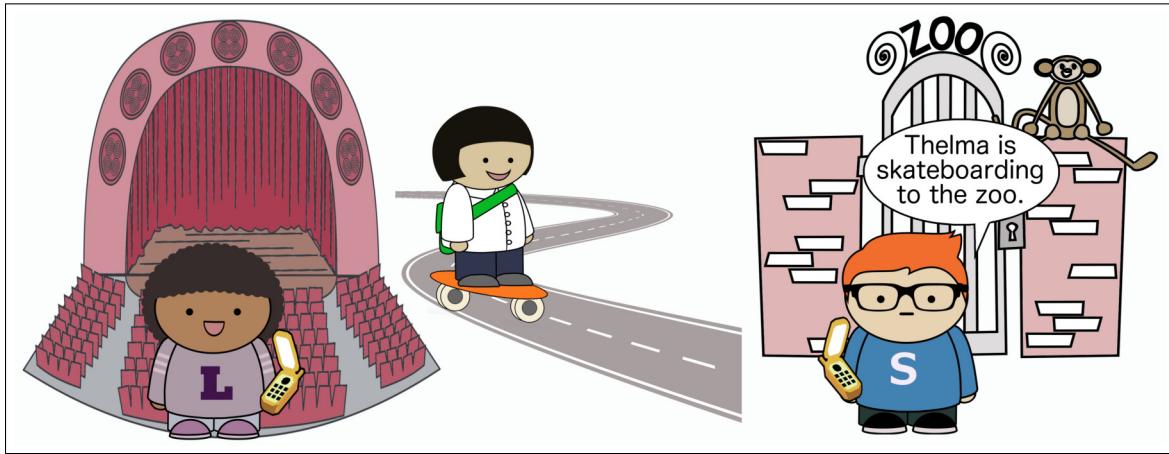
---

<sup>12</sup>128 participants were initially recruited. Of those, 48 participants were excluded due to low accuracy on the spatial task.

**Figure 7.28.** Comprehension stimulus: Perspectival Both condition



**Figure 7.29.** Comprehension stimulus: Plain Speaker condition



one with the listener at the destination and the speaker at the other location; and one where neither the listener nor the speaker were at the destination (both were shown at the other location). Thus, there were four scene types: Both, Speaker, Listener, and None. An example of a Both scene is shown in Figure 7.28.<sup>13</sup>

For scene types where the speaker and listener are not located in the same place, they are depicted talking to each other on the phone (Figure 7.29).

Each scene was matched with two different versions of the sentence: one using *come*, the perspectival condition; and one using a manner-of-motion verb such as *walk* or *drive*.

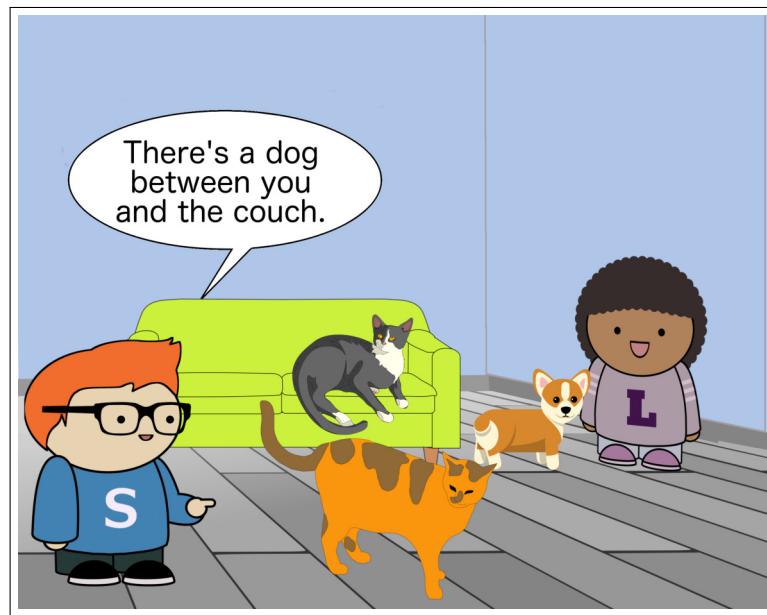
---

<sup>13</sup>The full set of stimuli can be found in Appendix M.

### Spatial control task

A potential concern is that participants might not be able to adopt the spatial perspective of the listener character. In order to make sure that participants were successfully adopting Lucy's perspective, participants also responded to spatial items with contrasts between the participant's visual perspective and Lucy's perspective. An example spatial control item is shown in Figure 7.30.

**Figure 7.30.** Comprehension spatial stimulus



Participants selected true or false for these statements. For the item in Figure 7.30, for example, the animal between the participant and the couch is a cat, but the animal between Lucy and the couch is a dog. Thus, this description-scene pair should be accepted.

#### 7.5.1.3 Procedure

Participants first saw an image of Sam's head with a speech bubble containing the target sentence. Then they saw a scene depicting the conversation, and were given 10 seconds to indicate whether or not the picture and sentence matched. If they did not respond within this time window, they were told that they were too slow, and the experiment moved to the next item.

Both reaction times (how quickly, in milliseconds, participants responded) and percent acceptance (how often participants indicated that the scene and sentence were compatible) were measured.

Each participant saw 6 items in each of the 8 conditions, distributed into 8 Latin Square lists, and combined with 30 fillers (10 true and 20 false) and 20 spatial control items, for a total of 98 items. They were also given 4 training items.

Stimuli were displayed and responses collected using the Ibex Farm platform for web-based experiments (Drummond, 2019). Each experimental session began with an informed consent form and concluded with a demographic survey and a debriefing survey, which allowed participants to report any issues with the survey and contained two bot-check questions.

### **7.5.2 Experiment 5a analysis**

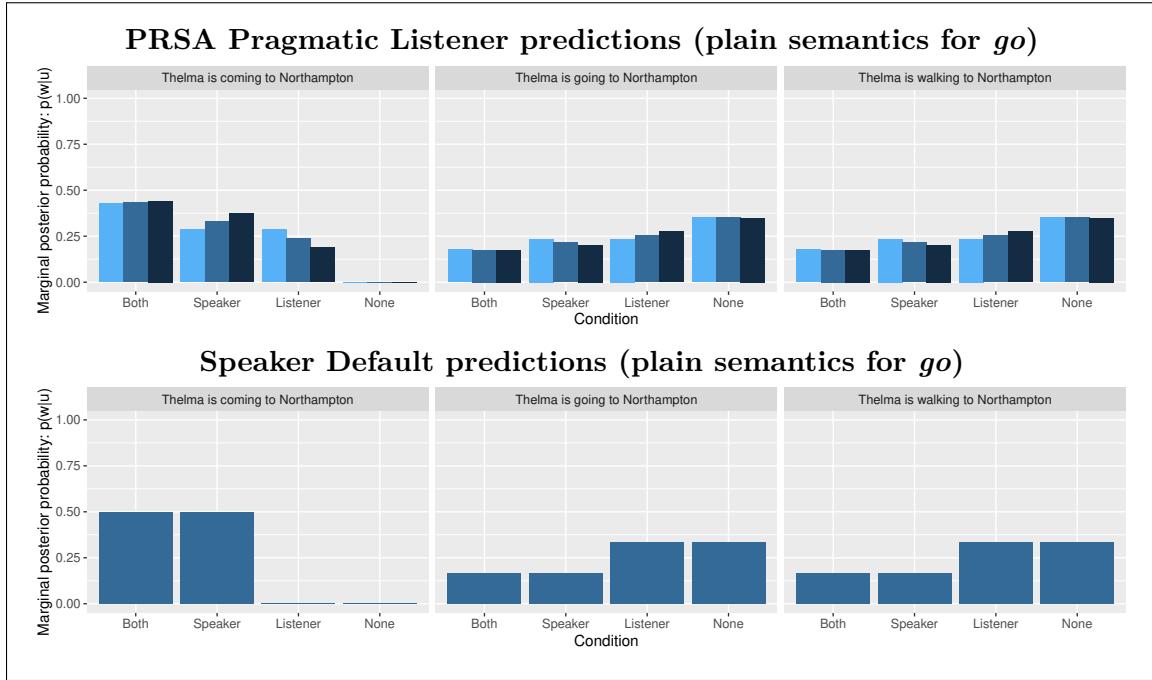
The reaction time results were analyzed using a mixed effects regression model. The maximal random effects structure was used for all models: random intercepts and slopes were included for all fixed-effects predictors, for participants, and for items. All models were fitted using the lme4 package in R (Bates et al., 2015).

Treatment coding was used, treating the Plain Speaker condition as the baseline. This resulted in the following fixed-effects contrasts: Perspectival, 1 for items in the Perspectival condition and 0 for items in the Plain condition; Both, 1 for the Both condition and 0 otherwise; Listener, 1 for the Listener condition and 0 otherwise; and None, 1 for the None condition and 0 otherwise. Treating the Speaker condition as the baseline allows us to interpret the Perspective:Both interaction term as measuring the comparison of interest, since it takes into account the differences between reaction times for scene types in the Plain versus the Perspectival condition.

### **7.5.3 Experiment 5a predictions**

The linking hypothesis is that reaction times are a measure of predictability: if participants are surprised by the scene, then they will be slower to accept the sentence/scene pair. The images are intended as visual representations of possible worlds. If the scene conflicts with

**Figure 7.31.** PRSA (top) and Speaker Default predictions (bottom)



a participant's expectations about the world given the sentence (i.e., the marginal posterior probability over worlds), then the participant should be slower to respond, and more likely to reject the sentence.

The PRSA model predicts highest marginal posterior probability for the world where both the speaker and the listener are at the destination of motion given the sentence *Thelma is coming to Northampton*. According to this hypothesis, reaction times in the Perspectival condition should be fastest for the Both scenes.

By contrast, using the Speaker Default heuristic alone would predict equal marginal posterior probability for all worlds where the speaker is at the destination of motion. Reaction times are expected to be equally fast for the Speaker and Both scenes in the Perspectival condition.

The reaction time measures in the Plain condition are used to control for the possibility of participant bias towards particular scene types. For instance, participants may prefer scenes where both the speaker and listener are depicted in the same location. To measure just the effect of scene type in the Perspectival condition, we look at the differences in reaction

**Figure 7.32.** PRSA and Speaker Default predictions for Experiment 5a

**PRSA model:**

- The Both scenes will be accepted faster and more often than the Speaker scenes in the Perspectival condition
- The Listener scenes will be accepted faster and more often than the None scenes in the Perspectival condition

**Speaker Default heuristic alone:**

- The Both and Speaker scenes will be accepted equally quickly and equally often in the Perspectival condition
- The Listener scenes will be accepted faster and more often than the None scenes in the Perspectival condition

times between the Perspectival and Plain conditions, since both the PRSA and Speaker Default heuristic predict equal posterior probability for the worlds when the sentence is not perspectival.

Although acceptance rate is expected to correlate with reaction time, in that participants should be more likely to reject scenes that do not match their expectations, participants may accept sentence / scene pairs even when the scene is unexpected, so long as it has a felicitous interpretation. In other words, the acceptance rates reflect whether or not the participant believes the scene and sentence are at all compatible, while the reaction times reflect how surprised the participant was by the scene after seeing the sentence.

Both accounts predict that the Perspectival Listener condition should have slower reaction times and lower acceptance rates than the Perspectival Speaker condition. In the RSA model, the extent of this difference depends on the value of the non-speaker perspective cost parameter. If listeners are guided only by the Speaker Default heuristic, processing the Perspectival Listener condition involves first attempting to process according to the speaker's perspective, and then revising, which should result in delayed reaction times.

#### 7.5.4 Experiment 5a results

The reaction time results of Experiment 5a supported the predictions of the PRSA model: reaction times were significantly faster in the Perspectival Both condition compared to the Perspectival Speaker condition. However, the acceptability results were unexpected:

participants accepted the None scene in Perspectival condition even though it violated the perspectival licensing conditions of *come*. The acceptability results therefore could not be used to test the comparison of interest.

#### 7.5.4.1 Reaction time results

Reaction times in the Perspectival Both condition were faster than in the Plain Both condition. In all other conditions, reaction times were slower in the Perspectival condition compared to the Plain condition (Table 7.1).

**Table 7.1.** Experiment 5a results

Scene	Sentence	Mean RT, answer = True	(95% Cosineau CI)	Acceptance rate
Both	Plain	2384	(+/-88)	0.98
	Perspectival	2366	(+/-87)	0.98
Speaker	Plain	2329	(+/-92)	0.98
	Perspectival	2470	(+/-91)	0.98
Listener	Plain	2262	(+/-85)	0.97
	Perspectival	2626	(+/-105)	0.88
None	Plain	2131	(+/-79)	0.98
	Perspectival	2538	(+/-103)	0.83

The finding that reaction times in the Both condition were faster than in the Speaker condition is supported by the mixed-effects model shown in Table 7.2. In a mixed-effects model of reaction time with the Plain Speaker condition as the baseline, the interaction between Both and Perspectival is significant.

**Table 7.2.** Experiment 5a RT mixed effects regression analysis, fixed effects (N=3630)

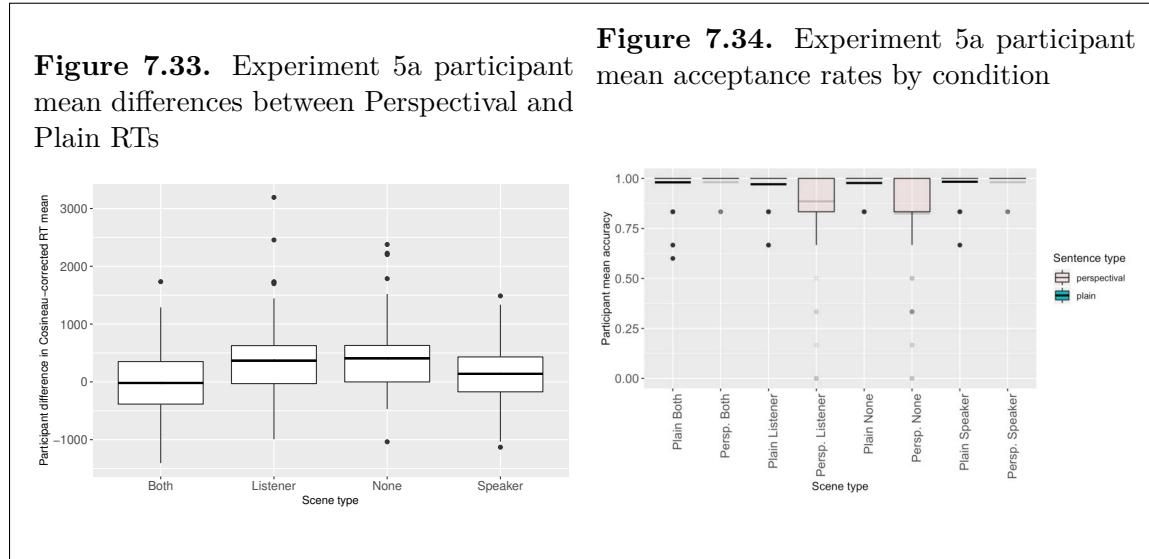
Fixed effects	$\hat{\beta}$	<i>z</i>	<i>p</i>
Perspectival	0.06(+/- 0.02)	3.16	< 0.01
Both	0.03(+/- 0.02)	1.43	0.152
Listener	-0.03(+/- 0.02)	-1.28	0.201
None	-0.08(+/- 0.02)	-4.14	< 0.0001
Perspectival:Both	-0.06(+/- 0.03)	-2.22	< 0.05
Perspectival:Listener	0.03(+/- 0.03)	1.05	0.295
Perspectival:None	0.05(+/- 0.03)	1.65	0.101

The significant effect of the interaction between Perspectival and Both supports the PRSA model's prediction that the Both scene is more likely than the Speaker scene given the

Perspectival sentence. This supports the hypothesis that listeners consider multiple perspectives simultaneously.

#### 7.5.4.2 Acceptance rate results

The acceptance rate results were at ceiling in most conditions, including in the critical conditions (Perspectival Speaker and Perspectival Both). This means that the acceptance data cannot be interpreted as evidence for or against the PRSA model's predictions.



In addition, the acceptance rates for the Perspectival None condition are troubling. Items in this condition were accepted 83% of the time, which is troublingly high, given that *come* is not licensed in this scenario.

#### 7.5.5 Discussion

The results of Experiment 5a provide qualified support for the PRSA model. As it predicts, participants were faster to accept scene-sentence pairs in the Perspectival Both condition than in the Perspectival Speaker condition. This supports the idea that listeners reason simultaneously over multiple perspectives. However, the predicted difference between the Perspectival Both and Speaker condition was not observed in the other experimental measure (acceptance rate). In both conditions of interest, participants' acceptance rates were at ceiling (as they were in all four Plain conditions).

In general, the acceptance rates indicate some experimental design issues. Acceptance rates for the Perspectival None condition were troublingly high, and in the debriefing questions, some participants noted that they had particular trouble deciding whether to reject or accept some of the *come* items.<sup>14</sup> This suggests that when the scene matched the truth conditions of the sentence, but not the perspectival anchoring of *come*, participants were unsure whether to accept or reject.

In addition, the design of Experiment 5a set up a bias towards positive responses. 7 of the 8 conditions were predicted to be accepted, which may have trained participants to accept most of the time. This could have increased the rate of acceptance cross the board, resulting in at-ceiling acceptance rates for the two conditions of interest, as well as an unexpectedly high acceptance rate for items in the None condition.

### 7.5.6 Experiment 5b method

Experiment 5b is a replication of 5a modified to address the response bias and pragmatic violation response concerns in Experiment 5a. Additional fillers were introduced to encourage participants to reject items, as well as additional training items. Participants were also given more explicit instructions about when to reject items.

#### 7.5.6.1 Participants

Monolingual American English-speaking participants ( $n=64$ ) were recruited through Prolific. Participants who achieved less than 90% accuracy on the spatial control task were excluded from the experiment.<sup>15</sup>

#### 7.5.6.2 Materials

Experiment 5b sought to ameliorate the response bias in Experiment 5a by introducing more items that participants would reject. 15 fillers were added, which were designed to be

---

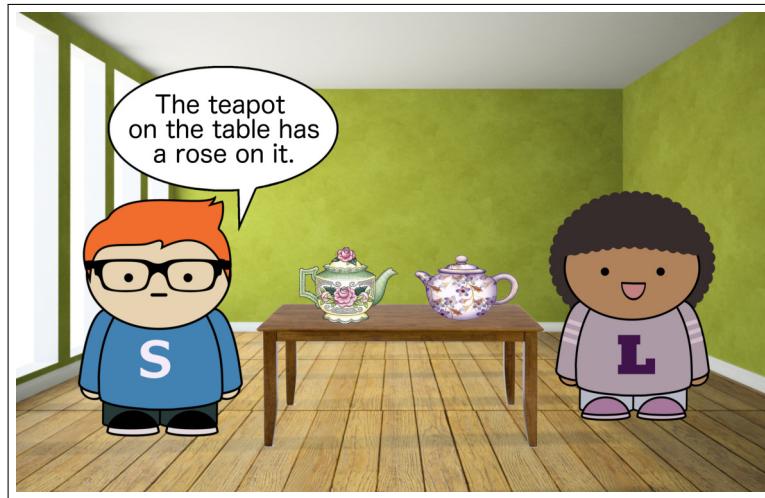
<sup>14</sup>For instance, one participant wrote:

I had some trouble deciding what to do with the “coming to” sentences when Thelma was going there but neither Lucy or Sam were present.

<sup>15</sup>95 participants were recruited, and 31 participants were excluded for low accuracy on the spatial task, for a total of 64 participants.

pragmatically bad: 5 definiteness violations, 3 presupposition violations, 3 scalar implicature cases, and 4 under-specific number cases. Figure 7.35 shows an example definiteness violation filler. The total number of items was therefore increased to 113.<sup>16</sup>

**Figure 7.35.** Pragmatically odd filler example: definiteness violation



#### 7.5.6.3 Procedure

The instructions to participants were modified in order to encourage them to reject pragmatically odd, but truth-conditionally valid descriptions. Participants were instructed that Sam “sometimes says things in a weird way or says things that don’t make sense,” and they were told that their task was to indicate whether or not what Sam says seemed “normal” according to the picture. This was meant to prompt them to reject uses of *come* without a valid perspectival anchoring.

Participants were also given more training with pragmatically odd items, and more feedback about why those items should be rejected. They were shown 6 training items: 2 pragmatic violations, 1 false description, and 3 valid descriptions, including a normal description of an odd scene, to help participants differentiate between unusual scenarios and unusual descriptions.

All other experimental methods remained the same.

---

<sup>16</sup>The full set of stimuli used can be found in Appendix M.

### 7.5.7 Experiment 5b analysis

A mixed effects regression model was fit to the reaction time data as described in Section 7.5.2.

In addition, a mixed effects logistic regression model was fit to the acceptance data. The maximal random effects structure was used: random intercepts and slopes were included for all fixed-effects predictors, for participants, and for items. The model was fitted using the lme4 package in R (Bates et al., 2015).

The same treatment coding was used in the acceptance data model as in the reaction time model, with the Plain Speaker condition treated as the baseline so that the Perspective:Both interaction term can be interpreted to measure the critical comparison of interest between the Perspectival Both and Perspectival Speaker conditions. This resulted in the following fixed-effects contrasts: Perspectival, 1 for items in the Perspectival condition and 0 for items in the Plain condition; Both, 1 for the Both condition and 0 otherwise; Listener, 1 for the Listener condition and 0 otherwise; and None, 1 for the None condition and 0 otherwise.

### 7.5.8 Experiment 5b results

The results of Experiment 5b were somewhat similar to Experiment 5a. The manipulation of the fillers increased the rate of rejections, providing easier-to-interpret rates of acceptance. However, although reaction times were again fastest in the Perspectival Both condition, the interaction of interest was not significant.

**Table 7.3.** Experiment 5b results

Scene	Sentence	Mean RT, answer = True	(95% Cosineau CI)	Acceptance rate
Both	Plain	3205	(+/-124)	0.95
	Perspectival	2896	(+/-121)	0.96
Speaker	Plain	3132	(+/-111)	0.95
	Perspectival	3052	(+/-115)	0.97
Listener	Plain	3082	(+/-119)	0.95
	Perspectival	3652	(+/-155)	0.61
None	Plain	2802	(+/-102)	0.96
	Perspectival	3182	(+/-139)	0.32

### 7.5.8.1 Reaction time results

The reaction time results largely support the previous findings: as in Experiment 5a, reaction times were fastest in the Perspectival Both condition (Table 7.3).

**Table 7.4.** Experiment 5b RT mixed effects regression analysis, fixed effects (N=2581)

Fixed effects	$\hat{\beta}$	<i>z</i>	<i>p</i>
Perspectival	-0.03(+/-0.03)	-1.1	0.28
Both	0.005(+/-0.02)	0.2	0.85
Listener	-0.037(+/-0.03)	-1.4	0.15
None	-0.12(+/-0.03)	-5.1	< 0.0001
Perspectival:Both	-0.06(+/-0.03)	-1.9	0.059
Perspectival:Listener	0.19(+/-0.04)	5.2	< 0.0001
Perspectival:None	0.13(+/-0.05)	2.64	0.01

However, the difference between the Both and Speaker conditions was not as large in this experiment; in the mixed effect regression analysis, the interaction between the Perspectival and Both conditions was below a *p*=0.05 threshold for statistical significance (Table 7.4). This may be a result of the smaller sample size used in this experiment compared with Experiment 5a (64 instead of 80 participants).

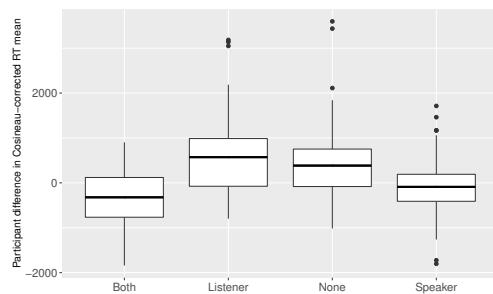
### 7.5.8.2 Acceptance rate results

The addition of extra false fillers was effective at decreasing the acceptance rate for Perspectival None items from 83% to 32%. However, acceptance rates for the Speaker and Both conditions remained at ceiling. A mixed effects logistic regression model (Table 7.5) found no reliable difference between the rates of acceptance in the Perspectival Both and Perspectival Speaker conditions.

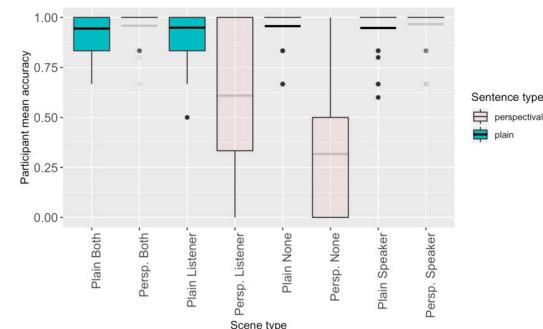
**Table 7.5.** Experiment 5b acceptance mixed effects regression analysis, fixed effects (N=3103)

Fixed effects	$\hat{\beta}$	<i>z</i>	<i>p</i>
Perspectival	0.71(+/-0.4)	1.7	0.09
Both	1.1(+/-0.6)	1.7	0.08
Listener	-0.02(+/-0.3)	-0.06	0.96
None	0.64(+/-0.4)	1.5	0.1
Perspectival:Both	0.09(+/-0.7)	0.13	0.9
Perspectival:Listener	-3.5(+/-0.5)	-7.4	< 0.0001
Perspectival:None	-6.0(+/-0.6)	-10.7	< 0.0001

**Figure 7.36.** Experiment 5b participant mean differences between Perspectival and Plain RTs



**Figure 7.37.** Experiment 5b participant mean acceptance rates by condition



Interestingly, the rate of acceptance of Perspectival Listener items also decreased substantially, from 88% to 61%. Although both the Speaker Default heuristic and the PRSA model predict that listeners should expect the speaker to be located at the destination of motion (possibly in addition to the listener), both models also predict that listener-oriented readings should be available when the speaker's perspective is ruled out by context. This is the case in the Perspectival Listener condition. So although both models predict slower reaction times in this condition (but for different reasons), the low rate of acceptability in this condition is nonetheless surprising.

### 7.5.9 Discussion

The reaction time results from Experiment 5b trended in the direction predicted by the PRSA model: participants were faster to respond to scenes with both the speaker and listener at the destination of motion when they saw a sentence with *come*. However, unlike in Experiment 5a, the interaction of Perspective and Both was not significant. This may be due to a lack of power caused by the smaller number of participants in the replication study.

The addition of the pragmatically bad fillers in Experiment 5b effectively brought down the acceptance rates compared with Experiment 5a: the acceptance rate for the Perspectival None condition dropped by about 50 percentage points. Despite the successful resolution of the response bias observed in Experiment 5a, the difference in the rates of acceptance in

the Perspectival Speaker and Perspectival Both conditions predicted by the PRSA model was not observed. The acceptance rates for both conditions remained at ceiling.

Moreover, the acceptance rate data from Experiment 5b suggest that the speaker perspective is strongly preferred to the listener perspective, since in the Perspectival Listener condition, the acceptance rate was 61%. The Listener condition also had one of the largest differences between Plain and Perspectival reaction times in both experiments, and the interaction between Perspectival and Listener was significant in the Experiment 5b mixed effect model.

The predicted difference between the Perspectival Listener and Perspectival Speaker conditions in the PRSA model varies according the perspective cost setting. The fact that the empirically observed acceptance rate for the Perspectival Listener was so low suggests that the perspective cost parameter should be set fairly high; listeners appear to predict fairly strong egocentricity effects on the part of the speaker.

Unfortunately, strong egocentricity effects make it more difficult to tell whether listeners are really reasoning over perspectives, as in the PRSA model, or relying on a Speaker Default heuristic. This is because as the perspective cost setting increases, the size of the preference for the Both scene predicted by the PRSA model decreases. Since the acceptance results suggest that the speaker preference is quite strong, we would expect the difference between the Perspectival Both and Perspectival Speaker conditions to be fairly small.

This makes it difficult to interpret the results of the comprehension experiments conclusively. On the one hand, the reaction time data provide some support for a preference for scenes where multiple perspectives are located, indicating that listeners may be employing perspectival reasoning in this context. On the other hand, the acceptance rate data did not suggest a preference for the Speaker Both condition. The comprehension data are therefore also fairly consistent with the hypothesis that listeners are using the simpler Speaker Default heuristic approach.

## 7.6 Summary

In Chapter 6, I argued that simple heuristics like the Speaker Default and Maintain Perspective cannot, on their own, account for the interpretation of perspectival expressions. Harris (2012) proposes that the perspective identification system has two parts: an inference system that reasons over the truth conditions of the utterance, the discourse context, and any other information that the listener has; and a simple, less costly set of heuristics. In this chapter, I have proposed the PRSA model as a model of the more powerful perspectival reasoning system. While the PRSA system can only condition on the discourse context in a limited way, it is nonetheless the first working implementation of a perspectival reasoning system that can be used to generate quantitative predictions in a variety of contexts.

Harris (2012) posits that this kind of inference system is too costly for listeners to constantly consult, and suggests that listeners, when possible, rely on the simpler set of heuristics to guide their interpretation. One of the key challenges for his two-stage perspective identification system is how to determine when listeners should switch between the two subsystems. I have argued in Chapter 6 that it is difficult to define the conditions for this switch in a top-down manner, since actual falsity is not frequent enough to serve as the signal.

We might instead try to derive the switch conditions in a bottom-up manner, by looking at different contexts and trying to infer whether listeners are using the heuristic system or the inference system. The comprehension experiments that I have presented in this chapter serve as a first foray into this search space. I have explored how listeners behave in one kind of environment: a novel discourse context where listeners have no particular prior beliefs and where the perspectives of the speaker and of the listener are available.

The results provide tentative evidence that in this context, listeners are reasoning over multiple perspectives simultaneously rather than merely relying on the Speaker Default heuristic. In both comprehension experiments, participants accepted scenes where both speaker and listener were at the destination of motion more quickly following a sentence using *come* than scenes where only the speaker was at the destination. However, this difference was not statistically significant in Experiment 5b, and the acceptance rates for both conditions were at ceiling.

Despite these qualifications, the experimental evidence favors the view that listeners reason simultaneously over multiple perspectives in this context. Future work could manipulate various features of the experimental context to explore whether there are other contexts in which listeners behave heuristically rather than rationally: different results might be found by manipulating the prior beliefs of the listener, the emptiness of the discourse context, or the listener's expectations about the speaker.

The comprehension results also served to quantify the strength of the egocentricity effects that listeners expect. In Experiment 5b, when participants were trained to reject pragmatically bad sentences, they frequently rejected listener-oriented uses of *come*. This lends support to Harris (2012)'s claim that a bias towards the speaker's perspective is an important factor in how perspectival expressions are interpreted. Although the strength of the speaker bias complicated the comparison between the Speaker Default heuristic and PRSA model, it is also useful evidence in its own right, as the data can be fed back into the PRSA model to adjust the perspective cost parameter in order to generate more accurate predictions for future work.

## CHAPTER 8

### A RATIONAL APPROACH TO PERSPECTIVE SELECTION

Part of the process of producing a perspectival expression is the choice of the perspective to use. As I argued in Chapter 6, the producer’s choice of a perspective to use is less constrained than the comprehender’s perspective identification task. This makes it a more difficult process to model. In this chapter, I present a model of perspective selection formulated in the RSA framework. I propose that the speaker’s production process mirrors the listener’s comprehension process: the speaker reasons jointly over perspectives and utterances given some meaning that they wish to convey, just as the listener imagines the speaker to do in the PRSA framework.

The structure of this chapter is as follows. In Section 8.1, I discuss how the RSA framework can be used to model the speaker’s utterance selection process. Although the RSA framework is most commonly applied to comprehension, I argue that it also adequately captures some of the desired properties of cooperative production. In Section 8.2, I propose a joint model of utterance and perspective selection in the RSA framework, mirroring the PRSA model of comprehension proposed in Chapter 7, and explore some of the model’s predictions.

In Section 8.4, I probe the behavior of actual speakers in a production task and compare the results with predictions of the PRSA model. As far as I am aware, there are no other proposed models of perspectival production to compare the PRSA production model against. However, I contrast the predictions of the PRSA model with a simpler heuristic-based system inspired by Harris (2012)’s proposal for perspective identification. I find that the data from the production study supports a view of speakers as strongly guided by a preference for using their own perspectives; the observed behavior does not fit the predictions of the PRSA model.

In Section 8.5, I explore several modifications to the PRSA model of production in hopes of providing a better fit to the experimental data. However, I conclude that none of theoretically-motivated variants of the model capture the strongly egocentric behavior observed in the production study. I conclude with a discussion of the challenges of modeling perspective selection.

## 8.1 The Rational Speech Acts framework as a model of production

Although most contemporary work in the RSA framework applies Bayesian reasoning to comprehension, the earliest work applied it to language production as well (Frank and Goodman, 2012). The RSA model of production mirrors the RSA model of comprehension. Just as listeners reason about the speaker’s production process in the RSA model of comprehension, in the RSA model of production, speakers reason about the listener’s comprehension process in order to maximize the utility of their utterances.

In the vanilla RSA model of production, the speaker observes a possible world and reasons about which utterance best describes the world using a mental model of how the listener interprets utterances. The best utterance is the one that will lead the listener to assign the highest probability to the observed world.

An RSA model of production can be developed by taking the Pragmatic Speaker as a model of the speaker (rather than as the listener’s mental model of the speaker, as in the RSA comprehension model). The Pragmatic Speaker can then be used to generate predictions about speaker behavior. Alternatively, a second-level Pragmatic Speaker can be built on top of the Pragmatic Listener of the RSA comprehension model. Figure 8.1 shows the mathematical model for an RSA model of production that includes both a Pragmatic Speaker and a Second-level Pragmatic Speaker.<sup>1</sup>

---

<sup>1</sup>As in the mathematical models presented in Chapter 7, I omit softmax terms for readability.

**Figure 8.1.** Vanilla RSA model of production (initial)

<b>Literal Listener</b>	
$p(w u)$	$\propto p(u w) p(w)$ $\propto [[m]]^w p(w)$
	by Bayes' Rule
<b>Pragmatic Speaker</b>	
$p(u w)$	$\propto p(w u) p(u)$ $\propto \text{LitList}(w,u) p(u)$ $\propto \text{LitList}(w,u)$
	by Bayes' Rule by definition of Literal Listener
<b>Pragmatic Listener</b>	
$p(w u)$	$\propto p(u w) p(w)$ $\propto \text{PragSpeak}(u,w) p(w)$
	by Bayes' Rule by definition of Pragmatic Speaker
<b>Second-level Pragmatic Speaker</b>	
$p(u w)$	$\propto p(w u) p(u)$ $\propto \text{PragList}(w,u) p(u)$
	by Bayes' Rule by definition of Pragmatic Listener

### 8.1.1 Motivating a rational approach to production

Adopting the RSA framework for production requires accepting the idea that speakers reason about how their utterances will be interpreted. While the RSA model of comprehension is essentially a way of incorporating alternatives into the meanings of utterances, which is necessary in order to derive scalar implicatures, what is the motivation for adopting a rational model of production?

To understand the motivation for a rational approach to production, let us start with some venerable guiding principles for conversational behavior: Grice's Maxims (Grice, 1975).

#### 8.1.1.1 Foundations of cooperative production

Grice (1975) proposes that there are four maxims for cooperative conversation, shown in (203). A cooperative speaker is one whose contribution is appropriately informative, clear, truthful, and relevant.

##### 203. Grice's Conversational Maxims (Grice, 1975)

- **Quantity:** be as informative as you can; as informative as is necessary; and no more.

- **Quality:** be truthful.
- **Relation:** be relevant.
- **Manner:** be clear, brief, and orderly. Avoid obscurity and ambiguity.

We can take Grice's Maxims as a set of desired traits for a model of utterance selection. A cooperative model of production is one that produces utterances satisfying all four maxims. Using this evaluation metric, let us compare the RSA model of production with some simpler alternatives.

### 8.1.1.2 Minimal models of production

The RSA model of production assumes that speakers reason about listeners as part of their sentence selection process. To see why reasoning is necessary in order to produce cooperative utterances, let us consider the simplest model of production: a random sampling algorithm operating over a set of utterances. In the Random Speaker model (Figure 8.2), the speaker's production process is not conditioned on an observed world, the conversational background, or the listener.

**Figure 8.2.** Alternative models of production

<b>Random speaker</b> $p(u w) = p(u)$	<b>Random true speaker</b> $p(u w) = [[u]]^w$
--	--

The Random Speaker will not result in good descriptions of the observed world. Because the utterance selection is not conditioned on the observed world, the speaker will violate both Quality and Relation. They are also unlikely to be informative or clear, since the Random Speaker does not reason about the discourse context. Thus, this model of production violates all four maxims at once.

A slightly less naive model is a sampling model that takes into account the truth of the utterance as a description of the observed world. I will call this model the Random True

model. This model will satisfy Quality, since no false utterances will be produced.<sup>2</sup> It satisfies Relation to some degree, since all utterances will describe the observed world, though it will not satisfy a more sophisticated definition of Relation, such as whether utterances are related to the Question Under Discussion.

However, the Random True model fails to satisfy the other two maxims. It cannot satisfy Quantity because it has no way of distinguishing among utterances that are informative, since all true utterances with equal prior probability are equally likely to be produced. It also cannot satisfy Manner, for the same reason: it has no basis of distinguishing among utterances that are true, so a more specific (clearer) utterance will not be preferred over a less specific one.

In order to satisfy the maxims of Quantity and Manner, the speaker must have a way of selecting among true utterances. To satisfy Quantity, the speaker must have a way of identifying informative utterances. To satisfy Manner, the speaker must have a way of identifying clear utterances. A clear utterance is one that is easy to interpret; this is, essentially, a maxim that is listener-oriented.

### 8.1.2 The RSA model as a cooperative model of sentence production

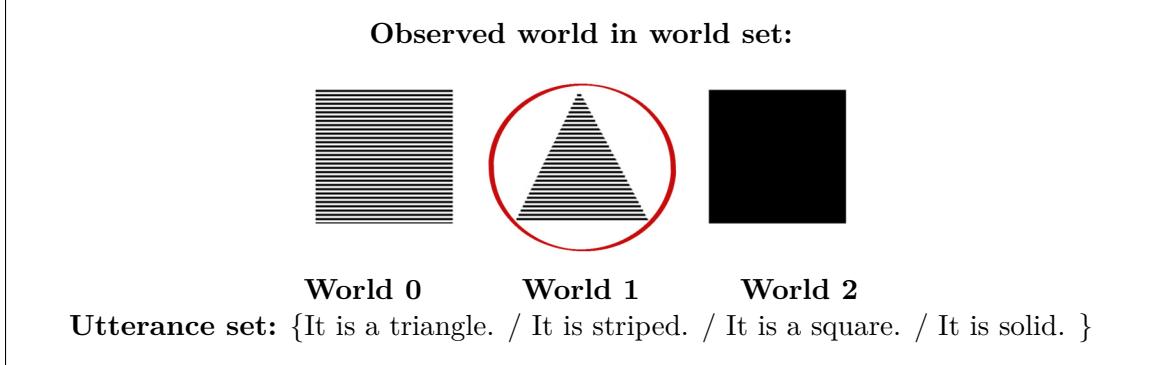
The RSA framework explicitly models the listener's interpretative process. It therefore provides one way of satisfying the maxim of Manner: in order to avoid ambiguity, the listener's interpretative process must be taken into consideration, since the ambiguity or clarity of utterances can only be evaluated according to how they will be interpreted.

Because the RSA model involves recursive reasoning about the behavior of other conversation participants, it is able to handle preferences for clearer alternatives. To see the advantages of the RSA model, let us consider a simple case of selecting a referring expression. In the context illustrated in Figure 8.3, both *striped* and *triangle* are true utterances. However, *triangle* is a clearer utterance, because it picks out a unique shape.

---

<sup>2</sup>Except when all utterances are false; this can be avoided by adding a null utterance to the utterance set.

**Figure 8.3.** RSA referring expression example



**Figure 8.4.** Vanilla RSA Literal Listener calculation for Figure 8.3 dataset

<b>Lit. Listener:</b> $p(w u_0)$ $p(w_0 \text{triangle}) = \frac{0}{0.33} = 0$ $p(w_1 \text{triangle}) = \frac{0.33}{0.33} = 1$ $p(w_2 \text{triangle}) = \frac{0}{0.33} = 0$	<b>Lit. Listener:</b> $p(w u_2)$ $p(w_0 \text{square}) = \frac{0.33}{0.66} = 0.5$ $p(w_1 \text{square}) = \frac{0}{0.66} = 0$ $p(w_2 \text{square}) = \frac{0.33}{0.66} = 0.5$
<b>Lit. Listener:</b> $p(w u_1)$ $p(w_0 \text{striped}) = \frac{0.33}{0.66} = 0.5$ $p(w_1 \text{striped}) = \frac{0.33}{0.66} = 0.5$ $p(w_2 \text{striped}) = \frac{0}{0.66} = 0$	<b>Lit. Listener:</b> $p(w u_3)$ $p(w_0 \text{solid}) = \frac{0}{0.33} = 0$ $p(w_1 \text{solid}) = \frac{0}{0.33} = 0$ $p(w_2 \text{solid}) = \frac{0.33}{0.33} = 1$

The Random True model will produce *triangle* and *striped* at equal rates. Because the RSA model incorporates reasoning about the listener's interpretive behavior, the RSA Pragmatic Speaker will select the utterance that refers less ambiguously. The RSA calculation is shown in Figure 8.5; the Pragmatic Speaker prefers the utterance *triangle* because it is more informative. The RSA model of production therefore satisfies the maxim of Manner, as well as Quality and Relation.

### 8.1.2.1 The limitations of the RSA production model

Above I have argued that the RSA model of production captures three of Grice's Maxims. What about Quantity? This depends somewhat on the definition of informativity. The output of the Pragmatic Speaker is informative in the sense that utterances that describe the observed world are preferred. However, true informativity rests on the state of the discourse context. An informative utterance is one that distinguishes among possible worlds that are

**Figure 8.5.** Vanilla RSA Pragmatic Speaker calculation for Figure 8.3 dataset

Prag. Speaker: $p(u w_0)$	Prag. Speaker: $p(u w_1)$	Prag. Speaker: $p(u w_2)$
$p(\text{triangle} w_0) = \frac{0}{1} = 0$	$p(\text{triangle} w_1) = \frac{1}{1.5} = 0.66$	$p(\text{triangle} w_2) = \frac{0}{2.5} = 0$
$p(\text{striped} w_0) = \frac{0.5}{1} = 0.5$	$p(\text{striped} w_1) = \frac{0.5}{1.5} = 0.33$	$p(\text{striped} w_2) = \frac{0}{2.5} = 0$
$p(\text{square} w_0) = \frac{0.5}{1} = 0.5$	$p(\text{square} w_1) = \frac{0.5}{1.5} = 0$	$p(\text{square} w_2) = \frac{0.5}{2.5} = 0.33$
$p(\text{solid} w_0) = \frac{0}{1} = 0$	$p(\text{solid} w_1) = \frac{0}{1.5} = 0$	$p(\text{solid} w_2) = \frac{1}{1.5} = 0.66$

being considered; in other words, one that makes an update to the probability distribution over worlds.

The RSA framework does incorporate the discourse context in a limited way, through the probability distribution over possible worlds. In the RSA model, the speaker takes into account both the truth of a message according to the observed world, and the likelihood of observing that possible world (through the speaker's model of the listener). The speaker's utterance selection therefore takes into account how the listener is going to interpret the utterance in light of the current discourse context.

However, in order to incorporate a pressure towards informative updates to the discourse context, we would need to understand how discourse updates proceed in the RSA framework. So far we have only considered a single conversational move: either the speaker's production of a single utterance or the listener's interpretation of a single utterance. Each of these steps produces a posterior distribution, but I have not provided a mechanism for updating the discourse context according to these posterior distributions. In order to produce truly informative utterances, the Pragmatic Speaker would need to reason not just about the posterior distributions over worlds produced by the Pragmatic Listener, but also about the difference between the prior and posterior distributions. The RSA framework therefore satisfies three of the Grice's Maxims fully, and one partially.

### 8.1.2.2 Modifications to the RSA model of production

The RSA model of production presented in Figure 8.1 captures several desired properties of cooperative utterance selection. However, this behavior comes at the cost of a fairly

significant assumption: that speakers behave maximally rationally. The empirical evidence in support of this assumption is mixed.<sup>3</sup> In practice, the assumption of maximal rationality is often relaxed.

First, as discussed in Chapter 7, an utterance cost function is often introduced into the Pragmatic Speaker. This cost function penalizes utterances with high production cost. Utterance length is often taken as metric of production cost, but this cost function could also be used to implement other factors that affect the ease of production.<sup>4</sup>

Second, a utility parameter  $\beta$  is often added to the Pragmatic Speaker. This parameter controls whether the Pragmatic Speaker always selects the maximally likely utterance. When the utility parameter  $\beta$  is set to 1, the Pragmatic Speaker produces utterances with a frequency proportional to their likelihood; when it is set to 0, the Pragmatic Speaker selects utterances at random (Degen et al., 2019). I use a utility parameter of 1 and omit it from the definition of the model shown in Figure 8.6.

**Figure 8.6.** Vanilla RSA model of production (final)

<b>Literal Listener</b>	
$p(w u) \propto p(u w) p(w)$	by Bayes' Rule
$\propto [[m]]^w p(w)$	
<b>Pragmatic Speaker</b>	
$p(u w) \propto p(w u) p(u)$	by Bayes' Rule
$\propto \text{LitList}(w,u) p(u)$	by definition of literal listener
$\propto \text{LitList}(w,u) - \text{Cost}_u(u)$	cost function added
<b>Pragmatic Listener</b>	
$p(w u) \propto p(u w) p(w)$	by Bayes' Rule
$\propto \text{PragSpeak}(u,w) p(w)$	by definition of pragmatic speaker
<b>Second-level Pragmatic Speaker</b>	
$p(u w) \propto p(w u) p(u)$	by Bayes' Rule
$\propto \text{PragList}(w,u) p(u)$	by definition of pragmatic listener

<sup>3</sup>As exemplified by some of the findings discussed in Chapter 6, such as Heller et al. (2012)'s finding that speakers failed to consider the Common Ground about 5% of the time.

<sup>4</sup>The use of an utterance cost function also serves to align the RSA production model more closely with the maxim of Quantity, since over-informative expressions will be dispreferred if they are more costly.

### 8.1.3 Summary

The RSA production model is not a fully satisfactory model of production, but it does capture some of the important properties that have been posited to guide production. In the remainder of this chapter, I present a joint model of perspective and utterance selection formulated in the RSA framework, and explore how well it fits actual perspectival production behavior. I show that the production behavior that is predicted by the PRSA model diverges from the behavior of actual speakers in important ways. I explore several modifications of the basic PRSA model in an attempt to better fit the experimental data, and conclude that the observed asymmetry between speaker and listener behavior is fundamentally challenging for RSA approaches to production.

## 8.2 A RSA model of utterance and perspective selection

In the vanilla RSA model of production presented above, the goal of the speaker is to select the utterance that maximizes their chance of communicating successfully with the listener. When producing utterances with perspectival expressions, however, the speaker has two decisions to make: a choice between utterances, and a choice between perspectives. In this section, I describe a PRSA model of perspectival production. The core assumption is that speakers reason jointly over pairs of utterances and perspectives.

### 8.2.1 The PRSA production model

The PRSA model of production assumes that the speaker jointly reasons about which perspective to adopt and which utterance to use. Just as the PRSA model of comprehension involves recursive reasoning, the PRSA model of production involves recursive reasoning about the listener's comprehension process. Given an observed world, the PRSA speaker seeks to jointly maximize the choice of perspective and utterance based on a mental model of how the listener will interpret the utterance.

We can consider two models of production: one in which the speaker assumes that the listener has direct access to the selected perspective (the Pragmatic Speaker) and one in which the speaker assumes the listener is jointly reasoning over meanings and perspectives

**Figure 8.7.** Perspectival Rational Speech Acts model of production

<b>Literal Listener</b>	
$p(w u,a)$	$\propto p(u,a w) p(w)$
	$\propto p(u,a,w)$
	$\propto p(u a,w)p(a,w)$
	$\propto p(u a,w)p(a)p(w)$
	$\propto [[u]]^{a,w} p(w)p(a)$
<b>Pragmatic Speaker</b>	
$p(u,a w)$	$\propto p(w u,a) p(u,a)$
	$\propto \text{LitList}(w,u,a) p(u,a)$
	$\propto \text{LitList}(w,u,a) p(u a) p(a)$
	$\propto \text{LitList}(w,u,a) p(u a) p(a) -$
	$\text{Cost}_u(u) - \text{Cost}_p(a)$
<b>Pragmatic Listener</b>	
$p(w,a u)$	$\propto p(u w,a) p(w,a)$
	$\propto p(w,a,u)$
	$\propto p(u,a   w) p(w)$
	$\propto \text{PragSpeak}(u,a,w) p(w)$
<b>Second-level Pragmatic Speaker</b>	
$p(u,a w)$	$\propto p(w u,a) p(u,a)$
	$\propto p(w,a,u)$
	$\propto p(w,a   u) p(u)$
	$\propto \text{PragList}(u,a,w) p(u)$

(the Second-level Pragmatic Speaker). Figure 8.7 shows the PRSA model of production with both versions of the Pragmatic Speaker.<sup>5</sup>

The Pragmatic Speaker reasons jointly over perspectives and utterances using a limited mental model of the listener's comprehension process. It assumes that their listener has direct access to the perspective that they have selected: in the Pragmatic Speaker's mind, the listener's task is simply to figure out which world is most likely given an utterance and a perspective. This involves interpreting the utterance according the given perspective and a possible world; the prior probability of the world is also taken into account.

---

<sup>5</sup> As in the mathematical models presented in Chapter 7, I omit softmax terms for readability.

The Second-level Pragmatic Speaker, on the other hand, entertains a more complex mental model of the listener. The Second-level Pragmatic Speaker assumes that their listener is jointly reasoning over worlds and perspectives, just as the Pragmatic Listener in the PRSA comprehension model does. In other words, the Second-level Pragmatic Speaker is aware that their listener does not have direct access to the perspective that they select, while the Pragmatic Speaker makes a false assumption about the listener's comprehension process (but does a simpler calculation because of this simplifying assumption).

#### **8.2.1.1 Implementing the perspective cost function**

For the Second-level Pragmatic Speaker, there is a choice to make about the perspective cost function. The model shown in Figure 8.7 implements another recursive layer on top of the PRSA model of comprehension presented in Chapter 7. This means that the speaker represented by the Second-level Pragmatic Speaker imagines that the listener is aware of the perspective cost. We might also imagine that the speaker is subject to a perspective cost, but does not believe that the listener is aware of it. In this case, we would move the perspective cost from the first Pragmatic Speaker to the Second-level Pragmatic Speaker.

#### **8.2.2 Summary**

The PRSA model of production that I have proposed is a mirror image of the PRSA model of comprehension outlined in Chapter 7. This symmetry is a useful property in certain ways; if the production and comprehension processes really mirror each other, then speakers should refine their mental models of their listeners through their own experience as listeners (and vice versa). From a modeling standpoint, it is also convenient, because it means that the parameters for the production model can be estimated from comprehension data, and vice versa. However, this is an assumption that requires testing.

### **8.3 Capturing perspective shift with the PRSA production model**

In Chapter 6, I discussed the challenges of designing a rule-based system for perspective selection. I argued that the process of perspective selection is less constrained than that of perspective identification, because the truth conditions of the utterance do not provide a

constraint on the choice of perspective. A key challenge for models of perspective selection is to derive a speaker bias without ruling out the possibility of perspective shift altogether.

In this section, I argue that the joint reasoning process of the PRSA production model meets this challenge. The PRSA model of production captures three key properties of perspective selection: (1) competition between *come* and *go*, (2) an egocentric bias towards the speaker’s perspective, and (3) variability in perspective shift. In the three sections, I discuss these properties.

### 8.3.1 Informativity-driven competition

One of the advantages of the RSA framework is that it naturally captures competition between alternatives that is driven by differences in informativity. Because the Pragmatic Speaker seeks to maximize the chance that the listener will infer their intended meaning, when two utterances are both true, but one is more informative, the Pragmatic Speaker will select the more informative one.

This arises from the Pragmatic Speaker’s consideration of the listener’s interpretation process. An utterance that is consistent with multiple worlds is less likely to guide the listener to the correct world than an utterance that is consistent only with the intended world.

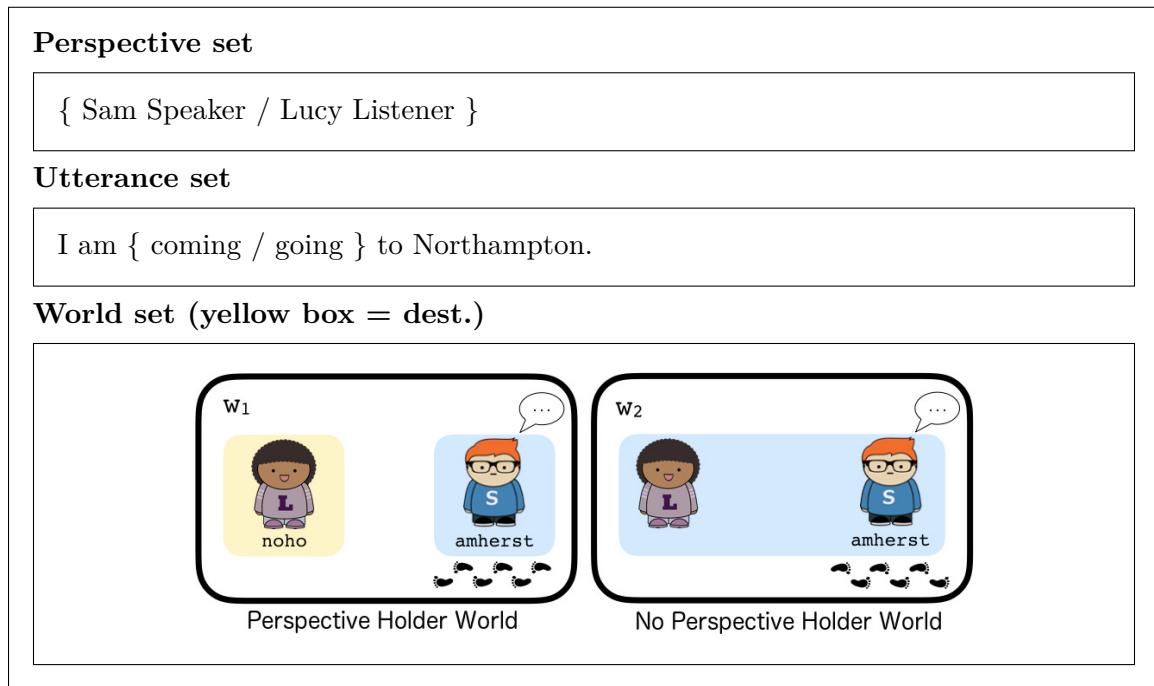
To see how this works, consider the small example dataset in Figure 8.8. There are only two individuals, the speaker and the listener, and both of their perspectives are available. For simplicity, we will consider just two utterances, *I am coming to Northampton* and *I am going to Northampton*, and adopt the non-perspectival semantics for *go*.

When the Pragmatic Speaker observes the Listener world, in which the listener is at the destination, they will produce *I am coming to Northampton*, even though according to the plain semantics of *go*, *I am going to Northampton* is also a true description of this world. Figure 8.9 shows the PRSA Pragmatic Speaker’s predictions for the Listener world: if we marginalize over perspectives, the most likely utterance is *I am coming to Northampton*.<sup>6</sup>

---

<sup>6</sup>The listener perspective paired with *I am coming to Northampton* is also the overall most likely utterance to describe the Listener world, since *I am coming to Northampton* has no probability when paired with the speaker’s perspective.

**Figure 8.8.** Dataset for informativity-driven competition illustration



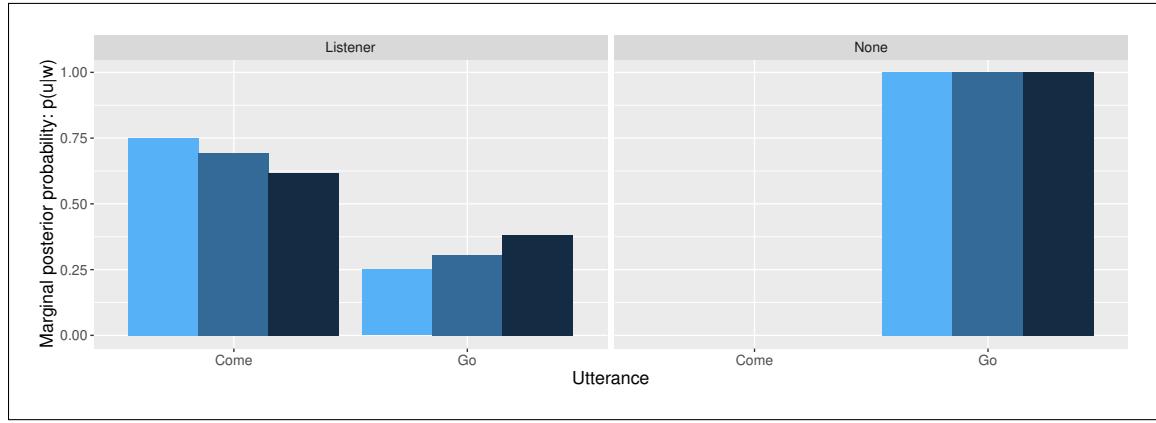
This is because *I am going to Northampton* is also true in the None world, the world where the listener is not at the destination of motion. Intuitively, this is because when the Pragmatic Speaker simulates the Literal Listener's evaluation of each candidate utterance, it finds that the listener is more likely to pick out the correct world given *I am coming to Northampton*, since it discriminates between the two possible worlds, unlike *I am going to Northampton*, which is a true description of both.

This particular example is specific to the plain semantics for *go*, since if a perspectival semantics is used, *go* and *come* will be equally informative. However, the ability to capture competitions between utterances that differ in informativity is a general property of the RSA framework: similar examples can be devised for either analysis of *go*.

### 8.3.2 A defeasible speaker bias

In Chapter 6, I discussed simple heuristic approaches for perspective selection, like the Speaker Default heuristic. I argued that the Speaker Default model is not a good model of production because it predicts speaker-anchoring of perspectival expressions in all contexts.

**Figure 8.9.** PRSA Pragmatic Speaker marginal posterior probabilities given Figure 8.8 dataset (plain semantics for *go*, cost parameters = 0, 0.5, 1 from left to right)



204. **Speaker Default model of production:** when using a perspectival expression, use your own perspective. If this results in falsehood, select a different perspective.

Using only the Speaker Default heuristic prevents perspective shift from occurring because the speaker, unlike the listener, has choice over the perspectival expression. Since the speaker can always select an utterance that is consistent with their own perspective, falsity will never arise, and, consequently, there is no cue for perspective shift.

Of course, we could consider other heuristics rather than switching to a reasoning system. However, if we omit a Speaker Default from the set of heuristics, then we will not capture the egocentricity effects for which we have seen a variety of empirical support (Chapter 6). Heuristic approaches to perspectival selection face a dilemma: either employ a Speaker Default, and eliminate the possibility of perspective shift, or omit it, and eliminate the bias towards the speaker's perspective.

By contrast, the RSA model of production allows a speaker bias to be incorporated in a defeasible way, via the perspective cost mechanism. When all else is equal and the perspective cost is set higher than 0, the speaker's perspective will be used. However, if there is a reason to prefer a non-speaker oriented utterance over a speaker-oriented utterance, it is possible to overcome the cost of a non-speaker perspective.

To understand why, let us again consider the dataset presented in Figure 8.8. If the speaker seeks to describe the Perspective Holder world, the Speaker Default model will predict the utterance *I am going to Northampton*. This follows from the fact that the Speaker Default model always prefers the speaker’s perspective: if the speaker’s perspective is used, *I am coming to Northampton* is false, since the perspectival anchoring requirement of *come* is not satisfied.

The PRSA model of production, on the other hand, predicts that the speaker will use *I am coming to Northampton* to describe the Perspective Holder world. As discussed in the previous section, this is because *coming* is a more informative alternative than *going*, which is true in both the Perspective Holder world and the No Perspective Holder world. Thus, the fact that the PRSA model of production incorporates reasoning over the listener’s interpretation process gives it a way of overcoming the speaker bias when motivated by a sufficient reduction in ambiguity. By contrast, if speakers employ the Speaker Default heuristic only, they will never adopt any perspective other than their own.

### 8.3.3 Variability in perspective selection

As discussed above, one of the advantages of the PRSA model of production that I am proposing is that its speaker bias is defeasible. As I have argued, a speaker default heuristic on its own would lead speakers to always use their own perspectives; this is inconsistent with the many cases of perspective shift that I have already presented. In addition, judgments about the perspectival anchoring of perspectival expressions are often gradient: for instance, in the comprehension experiments in Chapter 7, listener-anchored interpretations of *come* were judged less acceptable than speaker-anchored interpretations, but more acceptable than the absence of any valid perspective holder.

Gradient judgments are challenging to model in a rule-based system, since either a rule is in play or it is not. However, they are easy to capture in a probabilistic model like the PRSA model that I have proposed: the probabilities of producing an utterance can be directly mapped onto the rate of its production in an experimental context.

The advantage of being able to model gradient data, of course, is not limited to probabilistic approaches within the RSA framework; the gradience does not emerge from the central assumption about rationality of the RSA framework, but rather, from the fact that the prominence of the perspective holders in the discourse context is treated probabilistically. It would be possible to put forward to rule-based model that could capture gradient phenomena, by having probabilistic constraints rather than rules.<sup>7</sup>

### 8.3.4 Probing the predictions of the PRSA production model

I have argued that the PRSA model of production captures some key properties of perspective selection. Another benefit of RSA models, as I argued in Chapter 7, is that they are easy to implement computationally in order to generate predictions. In that chapter, I discussed a novel prediction made by the PRSA model of comprehension, which arose because of the model's assumption that listeners entertain multiple perspectives simultaneously.

Because the PRSA production model is built on top of the PRSA comprehension model, its predictions largely mirror those discussed in Chapter 7. The PRSA model of production gives rise to a parallel prediction because it assumes that speakers reason over multiple perspectives simultaneously when selecting utterances.

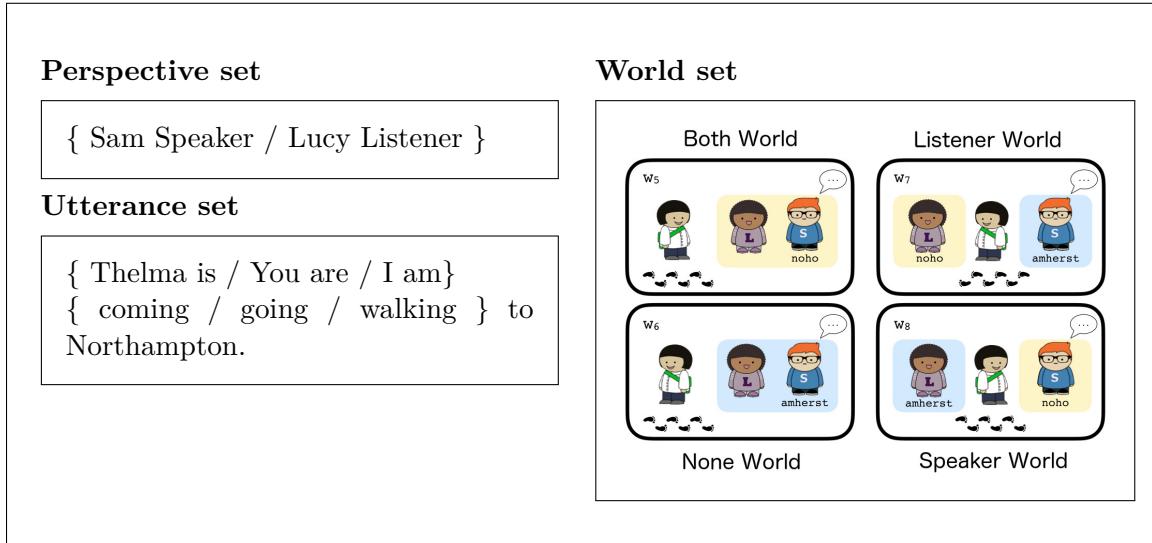
This prediction is relevant in contexts where multiple possible perspectives are at play. Let us return to the dataset discussed in Chapter 7, presented again in Figure 8.10. There are two perspective holders, the speaker and the listener, and a set of nine utterances, created by crossing three possible subjects with three motion verbs. If we assume that the person in motion is Thelma, and that her destination is Northampton, we have a set of four worlds to consider: the Both world, where both the speaker and listener are at the destination of motion; the Speaker world, the Listener world, and the None world.

The PRSA comprehension model predicted that when listeners interpret *Thelma is coming to Northampton*, they should calculate that the most probable world is one where both perspective holders are located at the destination of motion (the Both world). This is

---

<sup>7</sup>It might be possible to pursue this approach to salvage the Speaker Default heuristic approach.

**Figure 8.10.** Dataset for marginal posterior probability prediction illustration



because the Both world receives some boost in probability from the possibility that the listener is the perspective holder and is located there, and some boost in probability from the possibility that the speaker is the perspective holder. The results of Experiments 5a and 5b provided tentative support for this hypothesis.

The PRSA production model makes a prediction that mirrors this key prediction of the PRSA comprehension model. When describing worlds where Thelma is the mover, there are three potentially truthful utterances for the Pragmatic Speaker to consider: *Thelma is coming to Northampton*, *Thelma is going to Northampton*, and *Thelma is walking to Northampton*. Because both the first- and second-level Pragmatic Speakers consider multiple perspectives simultaneously, they are more likely to use *Thelma is coming to Northampton* to describe the Both world, in which both possible perspective holders are located at the destination of motion.

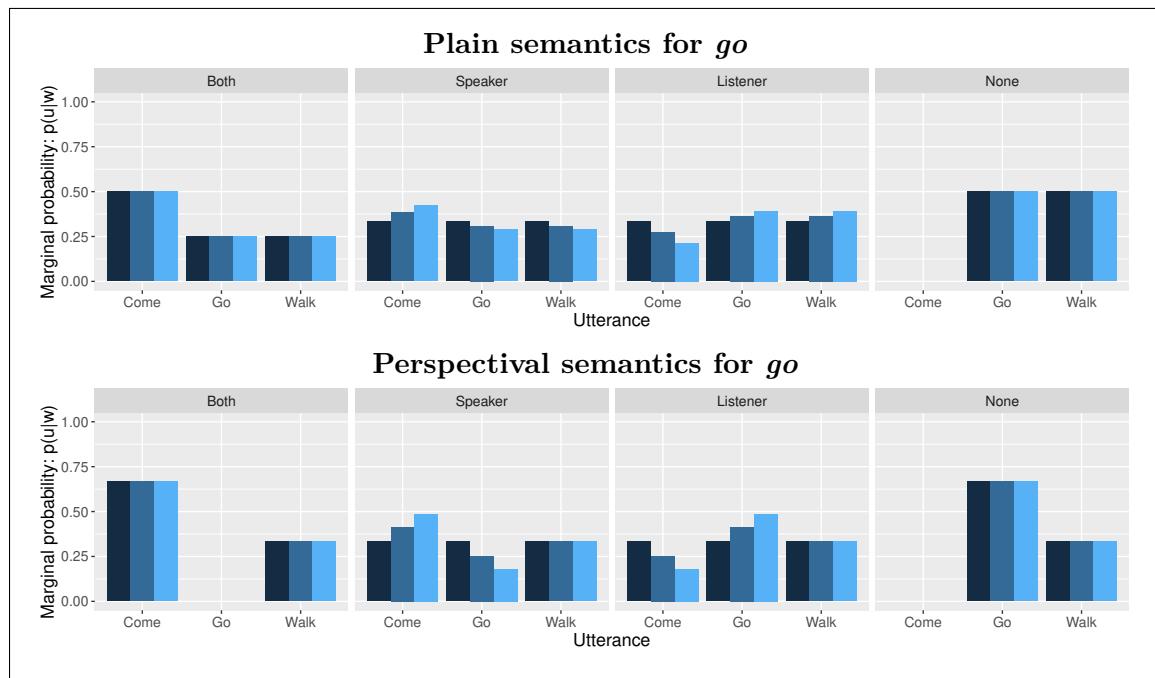
#### 8.3.4.1 Pragmatic Speaker predictions

The Pragmatic Speaker reasons about the best utterance-perspective pair using a simplified mental model of the listener's comprehension process; it assumes that the Literal Listener has direct access to the perspective adopted by the speaker. Despite this simpler reasoning model, the Pragmatic Speaker still gives rise to the prediction of interest.

Because the utterance *Thelma is coming to Northampton* is consistent with either the listener's or speaker's perspective, the Pragmatic Speaker is most likely to use this utterance to describe World 5, the world where both the speaker and listener are at the destination of motion. Figure 8.11 shows the Pragmatic Speaker's marginal posterior probabilities for utterances when the world sampled is one in which Thelma is the mover.

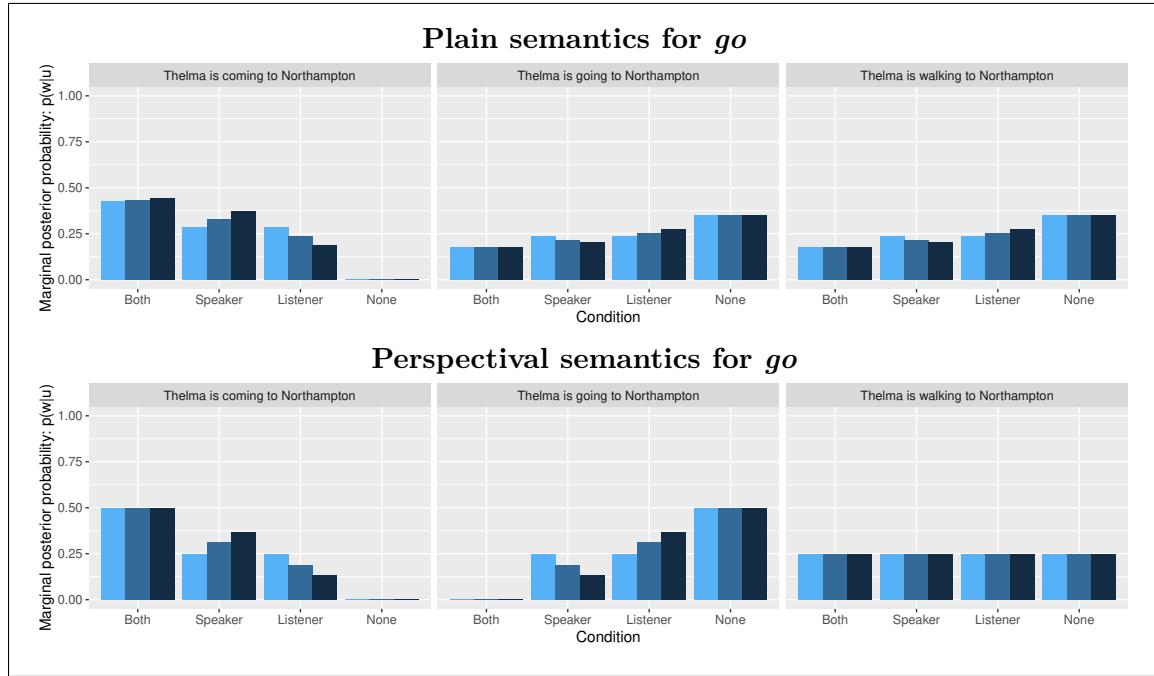
As Figure 8.11 shows, the predictions of the PRSA production model in this scenario mirror those of the PRSA comprehension model: the marginal posterior probabilities of sentences given worlds in Figure 8.11 mirror the PRSA comprehension model's Pragmatic Speaker's marginal posterior probabilities of worlds given sentences from Chapter 7 (Figure 8.12).

**Figure 8.11.** Pragmatic Speaker marginal posterior probabilities of utterances, mover = Thelma (cost parameters = 0, 0.5, 1 from left to right)



The Pragmatic Speaker is most likely to produce *come* for the world in which both speaker and listener are at the destination (the Both world). They are also likely to use *come* for the worlds where just one of the perspective holders is at the destination of motion (the Speaker and Listener worlds), though *go* and *walk* are also somewhat likely to be used. For these worlds, the setting of the perspective cost parameter affects the relative probabilities

**Figure 8.12.** Pragmatic Listener marginal posterior probabilities of worlds, mover = Thelma (cost parameters = 0, 0.5, 1 from left to right)



of *go* and *come*: when perspective cost is high, *come* is most probable for the Speaker world and *go* for the Listener world.

When the None world is observed, *come* is never produced, since its truth conditions are violated. The perspectival *go* version of the model predicts that *go* is most likely; the plain version predicts *walk* and *go* are equally likely, since their semantics is the same.<sup>8</sup>

Thus, the production model Pragmatic Speaker makes a prediction about the marginal probabilities of utterances that mirrors the comprehension model Pragmatic Listener's prediction about the marginal probabilities of worlds. Because the Pragmatic Listener jointly reasons over perspectives and utterances, they are most likely to produce *come* when observing the world in which multiple perspective holders are at the destination of motion.

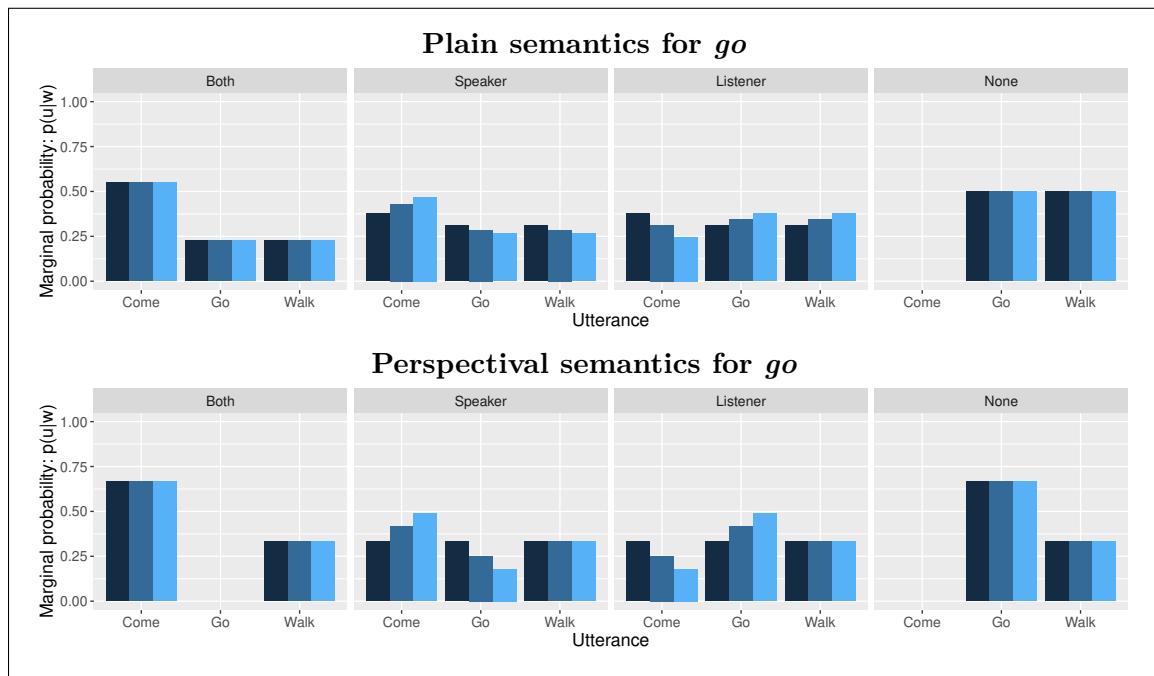
<sup>8</sup>To me this indicates a weakness in the RSA way of deriving alternatives without lexical scales (Bergen et al., 2016): there seems to be a sense in which *go* and *come* are in competition that is not true of *come* and manner-of-motion verbs. Perhaps this can be shown to arise in a RSA system by a process of conventionalization.

### 8.3.4.2 Second-level Pragmatic Speaker predictions

The Second-level Pragmatic Speaker employs a more sophisticated recursive reasoning process. Their mental model of the listener assumes that the listener does not have direct access to the selected perspective, but is reasoning jointly about the perspective and the speaker's intended meaning. Thus, the Second-level Pragmatic Speaker expects the listener to behave like the Pragmatic Listener of the PRSA comprehension model.

Despite the more complex mental model of the listener, the predictions of the Second-level Pragmatic Speaker are similar to those of the plain Pragmatic Speaker (Figure 8.13). The utterance *Thelma is coming to Northampton* is most likely to be used to describe the world in which both the listener and speaker are at the destination of motion.<sup>9</sup>

**Figure 8.13.** Second-level Pragmatic Speaker marginal posterior probabilities of utterances, mover = Thelma (cost parameters = 0, 0.5, 1 from left to right)



The marginal posterior probability prediction is quite robust. It does not matter whether the Pragmatic Speaker uses a particularly good model of the listener: the Pragmatic Speaker

<sup>9</sup>The predictions shown in Figure 8.13 are generated from a model where the perspective cost is included in the Second-level Pragmatic Speaker and not the base Pragmatic Speaker. However, implementing the perspective cost in the earlier stage Pragmatic Speaker or even at both levels does not change the results very much. A comparison of the predictions can be found in Appendix A.1.

who reasons about a Literal Listener makes predictions that are very similar to those of the Second-level Pragmatic Speaker, who reasons about a Pragmatic Listener.

#### 8.3.4.3 Summary

The PRSA model of production posits that the speaker’s utterance and perspective selection process involves reasoning about the listener’s interpretative process. This leads to symmetry between the predictions of the PRSA comprehension and production models. Just as the PRSA comprehension model predicts highest marginal posterior probability for worlds in which multiple perspective holders are at the destination of motion when the listener hears *Thelma is coming*, the PRSA production model predicts that *Thelma is coming* is most likely to be used to describe a world in which multiple perspective holders are at the destination of motion.

### 8.4 How do speakers select perspectival expressions?

One of the foundations of the RSA framework is its recursive nature: speakers and listeners reason about each other’s behavior. This leads to symmetry between the predictions of the PRSA production and comprehension models.<sup>10</sup> Because of this symmetry, we expect the results of a production experiment to mirror those of the comprehension experiments.

In the PRSA model, the Pragmatic Speaker seeks the best utterance-perspective pair to describe the world that they have observed. The posterior probabilities of utterances from the Pragmatic Speaker can be interpreted as predictions about speaker production: if the Pragmatic Speaker assigns 50% probability to *Thelma is walking* given a particular world, a speaker should produce this utterance 50% of the time. According to this linking hypothesis, the PRSA Pragmatic Speaker predicts that speakers should use *Thelma is coming* more in worlds where both the speaker and listener are at the destination of motion.

This prediction springs from the fact that the PRSA model assumes that speakers consider multiple perspectives in order to pick the one that best furthers their communicative goals. By contrast, a heuristically-guided speaker would not reason over perspectives. The

---

<sup>10</sup>This symmetry is bounded by the degree of rationality we assume for speakers and listeners.

Speaker Default heuristic, for instance, would lead a speaker to use *come* only when it is consistent with their own perspective. This predicts equal rates of production of *Thelma is coming to Northampton* for all worlds where the speaker is in Northampton, since the heuristically-guided speaker never considers the listener's perspective unless the speaker's is unavailable.<sup>11</sup>

A Pragmatic Speaker who can reason about the optimal choice of perspective is appealing for its ability to capture a range of perspective shift behavior. However, as Harris (2012) argues, a perspective inference system might be too costly to run constantly. So far we have not seen evidence that speakers reason about perspectives, beyond the observation that simple heuristic approaches to production fail because speakers do indeed shift perspectives sometimes. Between the full PRSA model, which proposes both that speakers reason over multiple perspectives at once and also that they reason by simulating the comprehension process, and the simplest heuristic approaches that I have dismissed, there may be many forms of compromise.

In order to explore this possible middle ground, it is necessary to gather data on the extent to which speakers behave rationally when selecting perspectives and perspectival expressions. In this section, I present a production experiment that probes the rationality of perspective selection in the context of perspectival motion verbs. I contrast the predictions of the PRSA production model with the behavior expected if speakers rely on a simpler Speaker Default heuristic. Despite the fact that the simple Speaker Default heuristic cannot explain why speakers ever shift perspective, I find that it is a relatively accurate predictor of speaker behavior in the environment tested.

---

<sup>11</sup>Which, I have argued in Section 8.3.2, is never; however, even if a heuristically-guided speaker used a more sophisticated Speaker Default rule that could shift to the listener's perspective in the scenario considered in Section 8.3.2, the heuristic model would not make the same prediction about the rate of *come* use.

### **8.4.1 Experiment 6 method**

Experiment 6 is a production experiment designed to test the hypothesis of the PRSA production model about the marginal posterior probability of utterances against the behavior expected if speakers are guided by the Speaker Default heuristic alone.

#### **8.4.1.1 Participants**

Monolingual American English-speaking participants ( $n=40$ ) were recruited through Prolific. One participant was excluded from the study because they reported mixing up the speaker and listener characters. This rejection criterion was not preregistered because the circumstance was unforeseen. However, the experimental procedures, data coding methods, and planned analyses described below were preregistered through the Open Science Foundation.

#### **8.4.1.2 Materials**

Experiment 6 has four conditions, corresponding to the four scene types used in Experiments 5a and 5b. There was no manipulation of motion verb type in this experiment, since the verb is elicited from the participant.

As in the comprehension experiments presented in Chapter 7, the scenes consist of two locations and three characters: the speaker, Sam; the listener; Lucy; and Thelma, who is in motion towards the righthand location.

There are four conditions for each scene: the Both condition, which shows both the speaker and listener at the righthand location (the destination of motion); the Speaker condition, which shows the speaker on the right and the listener on the left; the Listener condition, which shows the listener on the right and the speaker on the left; and the None condition, which shows the speaker and listener at the lefthand location and no one at the destination of motion. An example stimulus in the Listener condition is shown in Figure 8.14.

The illustrations were slightly modified between the comprehension and production experiments. In the production experiment, the speech bubble above the speaker contained the prompt *Thelma is... ,* rather than an entire sentence, and participants were asked to com-

**Figure 8.14.** Experiment 6 example stimulus



plete the sentence. In addition, in order to discourage participants from focusing on the manner of Thelma's motion, she was shown walking in all of the images.

The same filler items and spatial control items were used as in the comprehension experiments, modified in the same way.

#### 8.4.1.3 Procedure

In the production study, participants were asked to imagine themselves as Sam instead of as Thelma. Unlike the comprehension studies, in which the sentence was shown before the scene, in the production study, participants were shown the scene with a partially completed sentence inside the speaker's speech bubble (Figure 8.14).

Participants were asked to complete the speaker's sentence as naturally as possible according to the context depicted and using the prompt shown in the speech bubble. In the main conditions, the sentence prompt that the participants were given was "Thelma is ...." In the filler and spatial control items, the prompt varied based on the scene.

Because the motion verb was not given to the participants, there were only 4 conditions in this study, corresponding to the 4 versions of each scene (Both, Speaker, Listener, None).

Performance on the spatial control task was not used as an exclusion criterion in this study, due to the difficulty of prompting participants to describe the critical contrast between the speaker's visual perspective and their own point-of-view.

Stimuli were displayed and responses collected using the Ibex Farm platform for web-based experiments (Drummond, 2019). Each experimental session began with an informed consent form and concluded with a demographic survey and a debriefing survey, which allowed participants to report any issues with the survey and contained two bot-check questions.

#### 8.4.2 Experiment 6 analysis

##### 8.4.2.1 Data coding

Participant responses were coded for 11 categories, as shown in Figure 8.15. There were two annotators: the author and an annotator who was blind to the purpose of the experiment. The inter-annotator agreement scores by category are shown in Table 8.1.

**Figure 8.15.** Experiment 6 data coding categories

- Go: 1 if there is a non-light verb use of *go*
- Come: 1 if there is a non-light verb use of *come*
- Motion: 1 if there is a manner-of-motion verb
- Other: 1 if Go, Come, and Motion are all 0
- Bring: 1 if the verb *bring* is used
- Take: 1 if the verb *take* is used
- Here: 1 if *here* is used
- Mood: 1 if a statement; 2 if a question
- Tense: verb tense / aspect where -1 = past, 0 = present, 1 = future, 2 = modal / stative / other
- Subject: 1 if the subject of the verb is different than the subject of the prompt
- Destination: indicates the destination of the motion event if there is one

The primary dependent variables of interest are the rates of *come*, *go*, and manner-of-motion use. However, the categorization of tense, mood, subject, and destination are important for excluding responses that do not describe the scene in progress.

**Table 8.1.** Experiment 6 inner-annotator agreement scores by category

Category	Cohen's $\kappa$
Go	1.0
Come	1.0
Motion	1.0
Other	1.0
Bring	1.0
Take	1.0
Here	1.0
Mood	1.0
Tense	1.0
Subj.	1.0
Dest.	0.89

Using these annotations, any responses that were questions, did not describe an ongoing event, or had a different subject or destination than intended were excluded. For instance, the response shown in (205) was excluded because it was completed as a question.

205. *Thelma is ... coming to the pet store? Nice.*

Motion descriptions with destinations other than the depicted destination were also excluded; these were assumed to result from misinterpretations of the scene.

#### 8.4.2.2 Regression analysis

The rate of responses using *come* was analyzed using two mixed effects logistic regression models. The maximal random effects structure was used in both models: random intercepts and slopes were included for all fixed-effects predictors, for participants, and for items. The models were fit to the coded data using the lme4 package in R (Bates et al., 2015).

The first model was designed to compare the Speaker and Both conditions. Treatment coding was used, treating the Speaker condition as the baseline. This resulted in the following fixed-effects contrasts: Both, 1 for the Both condition and 0 otherwise; Listener, 1 for the Listener condition and 0 otherwise; and None, 1 for the None condition and 0 otherwise.

The second model was designed to compare the Listener and None conditions. Speaker and Listener were coded as fixed-effects predictors as follows: Speaker: 1 for presence of the

speaker at the destination of motion in the scene and 0 otherwise; Listener: 1 for presence of the listener at the destination of motion in the scene and 0 otherwise.<sup>12</sup>

#### 8.4.3 Experiment 6 predictions

The linking hypothesis is that speakers should produce utterances in proportion to the marginal posterior probabilities of utterances predicted by the PRSA Pragmatic Speaker. The PRSA model predicts that when speakers observe the world where both speaker and listener are in Northampton, they should produce *Thelma is coming to Northampton* 50-66% of the time (depending on the lexical semantics of *go* and the perspective cost setting), *Thelma is going to Northampton* 0-25% of the time, and *Thelma is walking to Northampton* 25-33% of the time (Figure 8.2).

**Table 8.2.** PRSA Pragmatic Speaker predicted proportion of utterances by scene

Scene	go semantics	come	go	walk
Both	Plain	50%	25%	25%
	Perspectival	66%	0%	33%
Speaker	Plain	33%-42%	29%-33%	29%-33%
	Perspectival	33%-49%	18%-33%	33%
Listener	Plain	21%-33%	33%-39%	33%-39%
	Perspectival	18%-33%	33%-49%	33%
None	Plain	0%	50%	50%
	Perspectival	0%	66%	33%

The critical prediction is that the rate of *come* responses should be highest in the Both condition, since even when the perspective cost is set very high, the rate of *come* responses in the Speaker condition can never exceed that of the Both condition. The PRSA model also predicts that *come* should be used to some extent in the listener-only context, unless the perspective cost is so high that the listener's perspective is effectively unavailable.

By contrast, if speakers rely exclusively on the Speaker Default heuristic, they should produce *come* equally often in the Both and Speaker conditions. The Speaker Default model also predicts that speakers will never produce *come* in the Listener condition. In this con-

---

<sup>12</sup>This model was not preregistered; it was motivated by the low rate of *come* responses in the Listener condition.

dition, the perspective of the speaker is available, but it cannot license *come*, so speakers should instead produce *go* or a non-perspectival alternative.

One interpretative difficulty is that if the perspective cost is set so high that the listener's perspective is never used, it becomes very difficult to tell whether speakers are reasoning about perspective selection or relying only on the Speaker Default heuristic. If the speaker's perspective is extremely unlikely, the PRSA model will no longer predict observably higher rates of *come* for the Both world compared to the Speaker world. More generally, if the distribution over perspectives is highly skewed towards the speaker or the speaker cost is high enough, the listener's perspective will be sampled so infrequently that the perspective set will appear to contain only the speaker's perspective. The PRSA model will then predict behavior identical to that of speakers using only the Speaker Default heuristic.

Because of this, there are three possible outcomes for a production study: (1) if *come* is used more in the Both condition, it supports the PRSA model; (2) if *come* is used equally in the Both and Speaker conditions and never in the Listener condition, it supports either the Speaker Default heuristic or an extremely high perspective cost; and (3) if the *come* is used equally in the Both and Speaker condition and sometimes in the Listener condition, neither model is an exact fit.

#### 8.4.4 Experiment 6 results

The proportion of *come*, *go*, and manner-of-motion verbs used in motion verb descriptions are shown in Table 8.3; the raw responses are plotted in Figure 8.16.

Participant completions using *come* occurred most frequently in the Speaker condition. A high rate of manner-of-motion completions was found across conditions (42%-71%).

**Table 8.3.** Experiment 6 proportion of all responses by condition and type

Condition	<i>come</i>	<i>go</i>	other motion	non-motion
Both	25%	9%	65%	30%
Speaker	31%	12%	56%	25%
Listener	4%	25%	71%	26%
None	0.8%	57%	42%	21%

Within the Both condition, the most frequent responses were manner-of-motion descriptions, followed by *come*. The make-up of responses in the Speaker condition was similar: a majority of manner-of-motion responses, followed by *come* responses, and a low rate of *go* completions.

**Table 8.4.** Experiment 6 proportion of motion responses by condition and type

Condition	<i>come</i>	<i>go</i>	other motion	N motion responses
Both	27%	10%	63%	335
Speaker	30%	13%	58%	359
Listener	4%	27%	69%	353
None	0.8%	59%	40%	381

The rate of manner-of-motion completions was highest in the Listener condition (71%). The proportion of Listener completions using *come* was very low (4%).

**Table 8.5.** Experiment 6 proportion of all deictic responses by condition and type

Condition	<i>come</i>	<i>go</i>	N deictic responses
Both	72%	28%	126
Speaker	70%	30%	153
Listener	14%	86%	110
None	1%	99%	228

The None condition was the only condition in which manner-of-motion completions were not the most common category. In the None condition, *go* was strongly preferred. The proportion of None condition responses involving *come* was very small, as expected.

Although the proportion of *come* responses was larger in the Speaker condition, in the mixed effects model with the Speaker condition treated as the baseline, the difference between the rate of *come* responses in the Speaker and Both conditions was not significant (Table 8.6).

**Table 8.6.** Experiment 6 *come* response mixed effects logistic regression analysis with Speaker treated as baseline

Fixed effects (N=1872)	$\hat{\beta}$	<i>z</i>	<i>p</i>
Both	-0.26(+/-0.18)	-1.4	0.15
Listener	-4.18(+/-1.1)	-3.8	<b>0.0002</b>
None	-4.24(+/-0.6)	-7.0	< 0.0001

**Table 8.7.** Experiment 6 *come* response mixed effects logistic regression analysis with Listener and Speaker as fixed-effects

Fixed effects (N=1872)	$\hat{\beta}$	<i>z</i>	<i>p</i>
Listener	1.4(+/-0.7)	2.1	<b>0.04</b>
Speaker	3.9(+/-0.6)	6.0	< <b>0.0001</b>
Listener:Speaker	-1.8(+/-0.7)	-2.6	<b>0.008</b>

Despite the low rate of *come* responses in the Listener condition, the mixed effects model where the presence of the Speaker and Listener at the destination were treated as fixed-effects revealed a significant difference in the rate of *come* responses in the None and Listener conditions (Figure 8.7). This suggests that there is a small but reliable difference between these conditions.

#### 8.4.5 Discussion

##### 8.4.5.1 Speaker preference

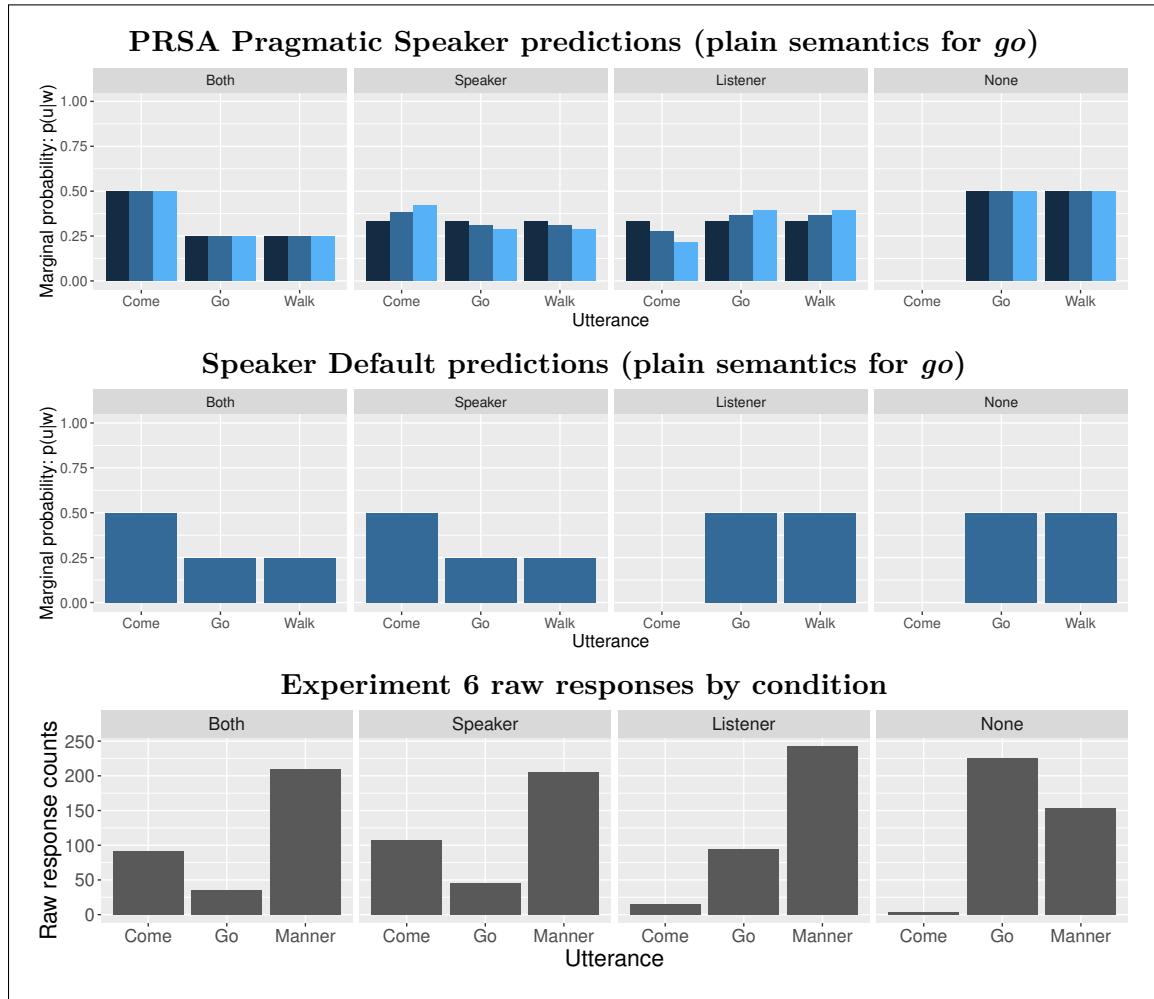
The main finding of interest is that speakers produced more responses using *come* in the Speaker condition than the Both condition. This is contrary to the predictions of the PRSA model, regardless of perspective cost settings (Figure 8.16). Although the Speaker Default heuristic predicts equal rates of *come* in both conditions, rather than higher rates in the Speaker condition, the difference between the two conditions was not significant in the mixed effects model. Therefore, these results are consistent with the behavior expected by speakers using a Speaker Default heuristic alone.

One confounding possibility is that a PRSA model with a very high perspective cost setting would also predict this behavior. In Experiment 5b, we saw evidence in support of a high speaker cost. Therefore, these results are not conclusive evidence against the PRSA model of production, but they do not favor it.

##### 8.4.5.2 Perspectival avoidance

The rate of *come* use in the Listener condition was very low, which is again consistent either with the predictions of a PRSA model with a very high perspective cost or the use of a Speaker Default heuristic. Under the Speaker Default heuristic, speakers are not expected to ever use *come* in this condition, since the speaker's perspective is available but inconsistent

**Figure 8.16.** PRSA predictions (top), Speaker Default predictions (middle), and experimental results (bottom)



with *come*. Although the observed rate of *come* responses in the Listener condition was significantly different from that in the None condition, the rate was quite low.

The Listener results also reveal one unexpected trend: manner-of-motion verb completions were more frequent in the Listener condition than in other conditions. In fact, the overall rate of deictic motion verb completions in this condition was smaller, suggesting that speakers may avoid perspectival expressions altogether in this condition.

Although this effect was not predicted, this data is consistent with the view that speakers are reasoning over alternatives: perhaps using *come* gives rise to the assumption that the speaker is at the destination and using *go* gives rise to the assumption that neither speaker

or listener is at the destination. In that case, using a manner-of-motion verb is a way of avoiding either failure mode. Despite the intuitive appeal of this line of reasoning, the PRSA model does not currently make this prediction.

#### 8.4.5.3 Lexical semantics of *go*

One additional finding of interest was that 9% of the responses in the Both condition used *go*. This could be interpreted as evidence in favor of a non-perspectival semantics for *go*, since if the truth conditions of *go* require that the perspective holder is not at the destination of motion, we would not expect any *go* responses in this condition. However, some portion of these responses may be due to incorrect coding of future auxiliary *go* as a motion verb.

#### 8.4.5.4 Summary

These results do not lend support to a PRSA model of production, although they do not eliminate it completely. The results in the Listener condition provide evidence in support of a strong bias towards the speaker's perspective. This bias could be encoded as a very high perspective cost in the PRSA model, which might account for the observed lack of a difference in rate of *come* responses between the Speaker and Both conditions, but it is also consistent with a speaker who is guided by a Speaker Default heuristic alone.

This leads to a dilemma about the Speaker Default heuristic model. On the one hand, I have argued that the production process cannot rely on a Speaker Default heuristic alone, since this predicts that perspective shift should never occur. On the other hand, the Speaker Default heuristic provides a good fit for the experimentally observed behavior of speakers.

One way to reconcile these two findings is to adopt a two-stage system of production, as Harris (2012) proposes for perspective comprehension. Instances of perspective shift could be explained by appealing to a perspectival reasoning system, of which the PRSA production model is one possible instantiation. The Speaker Default heuristic would then guide production most of the time; the experimental results could be interpreted as evidence that speakers rely heavily on the faster heuristic system in the context tested.

In this sense, the production results do not rule out a role for the PRSA system, since it is one possible model of the more costly reasoning system. However, the asymmetry observed between production and comprehension results is nonetheless troubling for a RSA approach to perspective identification and selection. One of the foundational assumptions of the RSA framework is that conversation participants reason about each others' behavior, an assumption that should give rise to symmetry between the production and comprehension results. Even if the production results do not lead us to discard a rational approach entirely, the existence of a mismatch in rationality between speakers and listeners challenges this central tenet of the RSA framework

Of course, it is possible that the observed asymmetry springs from experimental design issues and does not reflect real world behavior. Perhaps participants in the production experiment are not really simulating conversational behavior. In the comprehension studies, the spatial control task was used to measure whether participants were capable of taking the spatial perspective of the listener. Although the same spatial items were included in this study, many participants did not describe the contrast of interest, so their performance could not be used as an exclusion criterion. Future work replicating the production study would be helpful in determining whether the findings were affected by the experimental design.

Barring some issue in the experimental design, however, if we accept the experimental results, we are faced with a puzzle. On the one hand, the comprehension results are well-described by the PRSA model, and not by the Speaker Default heuristic. On the other hand, the production results are very different from the predictions of the PRSA model, and provide better support for a heuristically guided speaker.

## 8.5 Extensions to the PRSA model

A number of extensions or modifications to the PRSA model are possible. In this section, I discuss the effect of manipulating the priors on worlds and perspectives, adding lexical uncertainty into the model, and implementing the perspective cost in different ways. I show that although it is possible to enrich the PRSA model in various ways, none of the

extensions that I have explored resolve the inconsistencies between the production behavior predicted by the PRSA model and the experimentally observed production behavior.

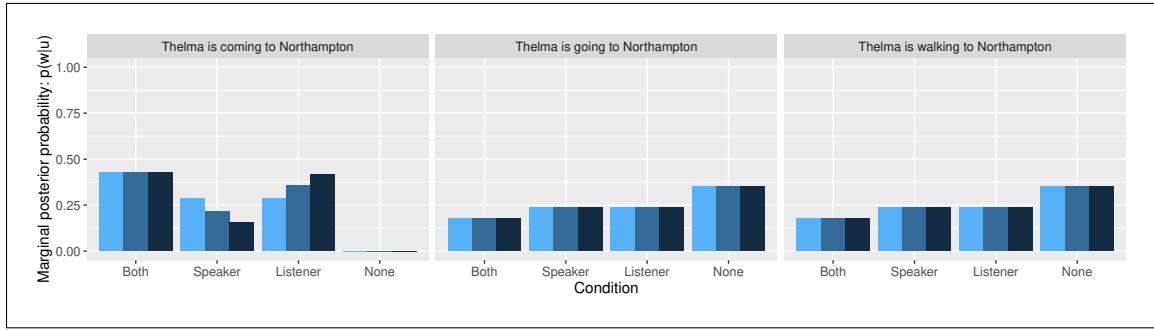
### 8.5.1 Separate perspective costs for speaker and listener

In the PRSA model proposed in Section 7.1.1, egocentricity effects were implemented in the Pragmatic Speaker as a perspective cost function. Because the Pragmatic Speaker is both our model of speaker behavior and the listener's model of speaker behavior, our model assumes that the listener is aware of the speaker's egocentric bias.

**Figure 8.17.** Egocentric Perspectival Rational Speech Acts model

<b>Literal Listener</b>	
$p(w u,a)$	$\propto p(u,a w) p(w)$ by Bayes' Rule
	$\propto p(u,a,w)$ by definition of joint probability
	$\propto p(u a,w)p(a,w)$ by definition of joint probability
	$\propto p(u a,w)p(a)p(w)$ by assumption of independence
	$\propto [[u]]^{a,w} p(w)p(a)$
<b>Listener's Pragmatic Speaker</b>	
$p(u,a w)$	$\propto p(w u,a) p(u,a)$ by Bayes' Rule
	$\propto \text{LitList}(w,u,a) p(u,a)$ by definition of Literal Listener
	$\propto \text{LitList}(w,u,a) p(u a) p(a)$ by definition of joint probability
	$\propto \text{softmax}(\text{LitList}(w,u,a) p(u a) p(a) - \text{Cost}_u(u))$ adding cost functions
<b>Pragmatic (Actual) Speaker</b>	
$p(u,a w)$	$\propto p(w u,a) p(u,a)$ by Bayes' Rule
	$\propto \text{LitList}(w,u,a) p(u,a)$ by definition of Literal Listener
	$\propto \text{LitList}(w,u,a) p(u a) p(a)$ by definition of joint probability
	$\propto \text{softmax}(\text{LitList}(w,u,a) p(u a) p(a) - \text{Cost}_u(u))$ adding cost functions
	$\propto \text{softmax}(\text{LitList}(w,u,a) p(u a) p(a) - \text{Cost}_u(u) - \text{Cost}_s(a))$ egocentricity bias
<b>Pragmatic (Actual) Listener</b>	
$p(w,a u)$	$\propto p(u w,a) p(w,a)$ by Bayes' Rule
	$\propto p(w,a,u)$ by Chain Rule
	$\propto p(u,a   w) p(w)$ by Chain Rule
	$\propto \text{PragSpeak}(u,a,w) p(w)$ by definition of Pragmatic Speaker
	$\propto \text{PragSpeak}(u,a,w) p(w) - \text{Cost}_l(a)$ egocentricity bias

**Figure 8.18.** Egocentric Pragmatic Listener marginal posterior probabilities over worlds, subject = Thelma and plain *go* semantics (cost parameters = 0, 0.5, 1 from left to right)



Perhaps egocentricity is a cognitive bias that is actually separate from the rational reasoning process of conversation participants. In this case, we should expect both speakers and listeners to be affected by the bias towards their own perspective. This means that the listener must have a model of the speaker that is separate from the actual speaker's behavior.

In this Egocentric Perspectival Rational Speech Acts model (Figure 8.17), the Pragmatic Listener's model of the Pragmatic Speaker is separate from the actual Pragmatic Speaker and does not include a perspective cost. Both the actual Pragmatic Speaker and Pragmatic Listener are biased towards their own perspectives and include perspective cost functions. The speaker's perspective cost function ( $\text{Cost}_s$ ) penalizes non-speaker perspectives, while the listener's perspective cost function ( $\text{Cost}_l$ ) penalizes non-listener perspectives. The weights of these perspective cost functions can vary independently.

Although this model is appealing because of its cognitive plausibility, its predictions do not fit the experimental data. The Pragmatic Listener still predicts highest marginal posterior probability for the Both scene, but it now predicts higher probability for the Listener scene than the Speaker scene (Figure 8.18).

The Pragmatic Speaker's predictions are unchanged, since we have not altered the Pragmatic Speaker. Separating out the speaker and listener models does not provide a solution to the asymmetry between the comprehension and production results.

### 8.5.2 Lexical uncertainty model: semantics of *go*

Throughout this discussion, I have contrasted the predictions of a model with perspectival semantics for *go* and one where *go* simply means move. Neither of these is entirely satisfactory: the plain *go* model is better able to capture the more frequent use of *walk* by speakers in the Listener condition, but fails to differentiate *walk* from *go*.

One possibility is that there is lexical uncertainty about the semantics of *go*. Bergen et al. (2012) proposed a RSA model for scalar implicatures that incorporates uncertainty over lexical semantics. In a lexical uncertainty model, the lexical semantics for utterances that is fed into the interpretation function is no longer fixed. Instead, we have a probability distribution over the lexical semantics of each utterance.

In the perspectival case, we incorporate uncertainty only over the lexical semantics for *go*.<sup>13</sup> Instead of using either the perspectival or non-perspectival semantics for *go*, we can sample one of these semantics for *go* probabilistically.

Figure 8.19 shows the predictions of the lexical uncertainty model when the perspectival and plain semantics are equally likely. Like the plain *go* model, the lexical uncertainty model correctly predicts that *walk* is used more frequently in the Listener condition than the Speaker condition. However, like the perspectival *go* model, it does not predict equal probabilities for *go* and *walk*, since their semantics are not identical.

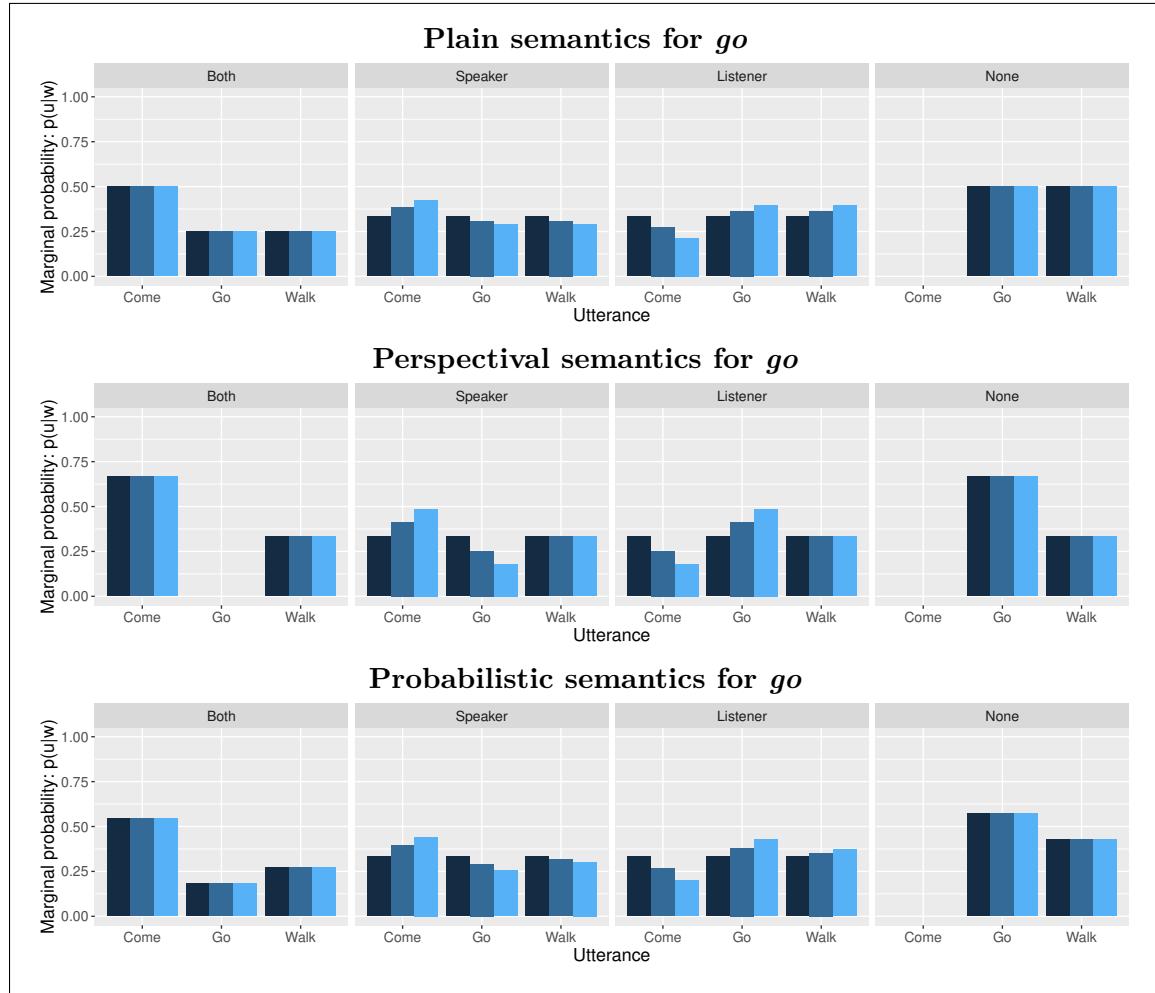
The predictions of the lexical uncertainty model are a better fit for the experimental results than either the perspectival or plain *go* models. However, proposing lexical uncertainty over the semantics of *go* should be motivated by empirical evidence of variation in the semantics of *go*: a model like this seems most plausible in a situation where there is established interspeaker variation in the lexicon. I do not have evidence of this within the population of American English speakers.<sup>14</sup>

---

<sup>13</sup>Though we could imagine taking a lexical uncertainty approach for the semantics of *come* as well, if, for instance, there was evidence of interspeaker variation in whether listener-anchoring of *come* was possible.

<sup>14</sup>Although there is well-documented cross-linguistic variation (Wilkins and Hill, 1995).

**Figure 8.19.** Lexical Uncertainty Pragmatic Speaker marginal posterior probabilities for utterances, subject = Thelma (cost parameters = 0, 0.5, 1 from left to right)



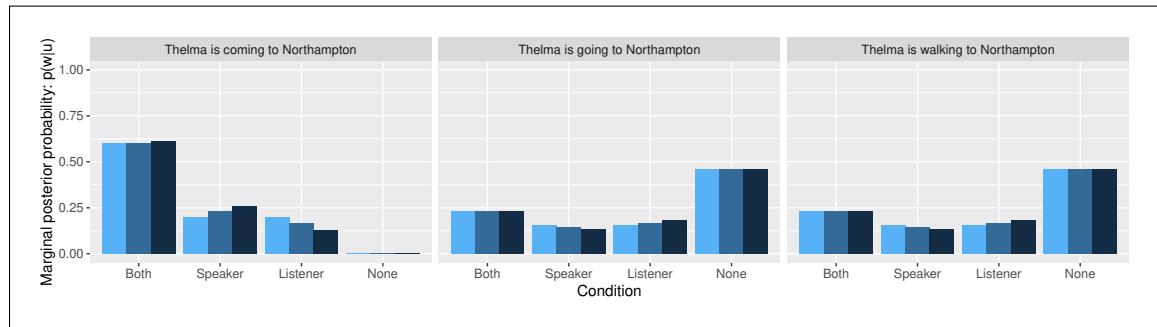
### 8.5.3 Effect of world priors: bias towards shared speaker/listener location

In the model simulations reported thus far, I have used a uniform prior distribution over possible worlds. One concern in the experimental work, however, was that participants might prefer some scenes over others, a bias expressible as a non-uniform prior over worlds.

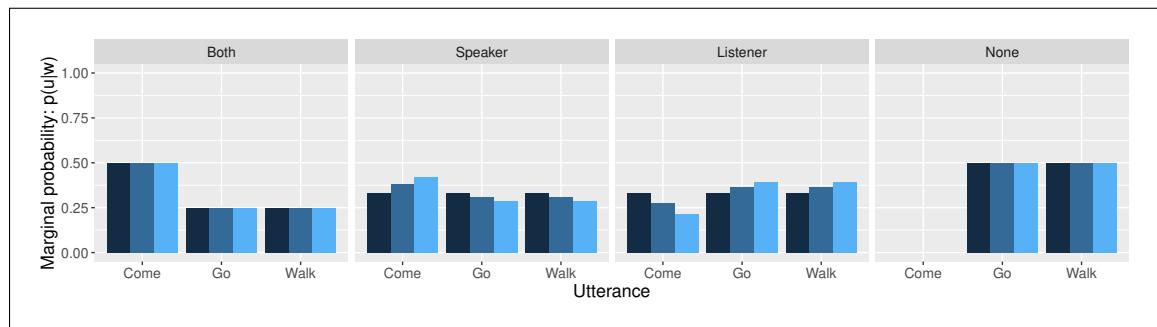
One plausible scene bias is to prefer scenes where both the speaker and listener are in the same place, since a prototypical conversation may be a face-to-face one. This bias interferes with the main comparison of interest, since a general preference for the scenes in the Both condition could provide false evidence in support of the PRSA predictions over the Speaker Default heuristic's predictions.

Figure 8.20 shows the model predictions for the Pragmatic Listener when the Both and None worlds are more likely than the Speaker and Listener worlds. When the prior distribution over worlds favors worlds in which the speaker and listener are in the same place, the marginal posterior probability of the Both scene is even higher than in a PRSA model with uniform priors over worlds.

**Figure 8.20.** Pragmatic Listener marginal posterior probabilities over worlds when Both and None scene priors are higher, subject = Thelma and plain *go* semantics (cost parameters = 0, 0.5, 1 from left to right)



**Figure 8.21.** Pragmatic Speaker marginal posterior probabilities for utterances when Both and None scene priors are higher, subject = Thelma and plain *go* semantics (cost parameters = 0, 0.5, 1 from left to right)



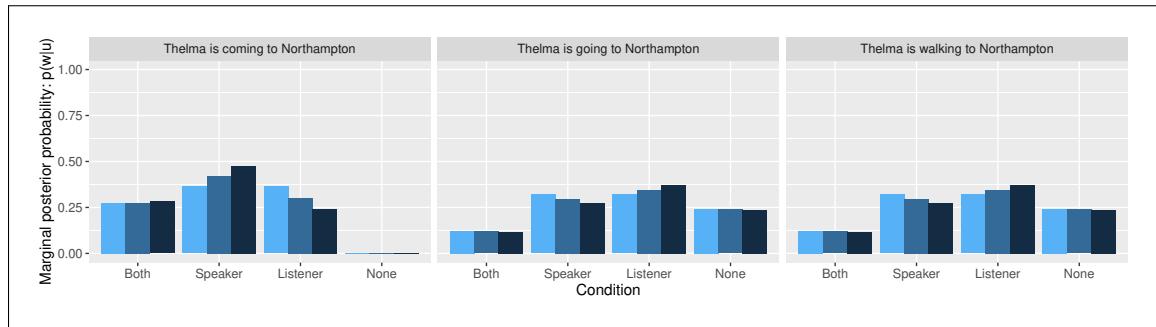
Manipulating the priors over worlds does not affect the marginal posterior probabilities for utterances involving Thelma calculated by the Pragmatic Speaker (Figure 8.21).<sup>15</sup>

It is also possible that participants have the opposite bias: in a scenario where both the speaker and listener are in the same place, it might be hard to imagine that the speaker has information that the listener does not, given their shared visual perspective on the scene.

<sup>15</sup>It is difficult to give an intuition for this. See Appendix A.2 for details.

Figure 8.22 shows the model predictions for the Pragmatic Listener when the Speaker and Listener worlds are more likely than the Both and None worlds.

**Figure 8.22.** Pragmatic Listener marginal posterior probabilities over worlds when Speaker and Listener scene priors are higher, subject = Thelma and plain *go* semantics (cost parameters = 0, 0.5, 1 from left to right)



When the prior distribution over worlds favors worlds in which the speaker and listener are in the same place, the marginal posterior probability of the Both scene given *Thelma is coming to Northampton* is lower than that of the Speaker scene; it is even lower than that of the Listener scene, for certain perspective cost settings.

The two scene biases lead to opposite predictions about manner-of-motion descriptions (the Plain condition). If participants expect the speaker and listener to be located together, we would expect fastest reaction times for *walk* for the None scene, followed by the Both scene, and then the Listener. The slowest reactions times are predicted in the Speaker scene (or, with a perspectival semantics for *go*, equally slow in the Speaker and Listener conditions). On the other hand, if participants expect them to be apart, reaction times should be fastest in the Listener condition, followed by the Speaker condition, and then the None condition. The reaction times should be slowest in the Both condition (or, equally slow in the Both and None conditions).

The actual comprehension results do not support either scene bias. Reaction times for utterances in the Plain condition in Experiments 5a and 5b were fastest in the None condition, followed by the Listener and then Speaker condition. The slowest reaction times in both experiments are in the Both condition. This is not predicted by either of PRSA models

with non-uniform priors over worlds, since in both modified models, the Both and None worlds should pattern similarly (in the Plain condition).

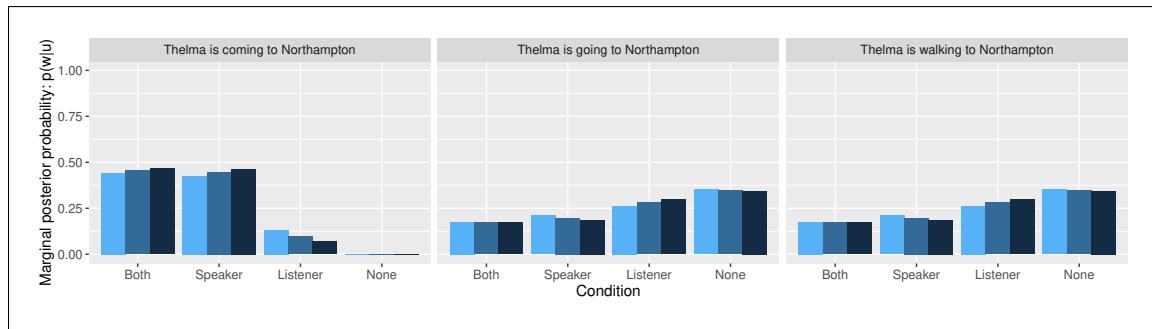
Manipulating the priors over the possible worlds can affect the predictions of the PRSA about the marginal posterior probabilities over worlds; it can even reverse the key prediction that the Both world should be more probable than the Speaker world given *Thelma is coming to Northampton*. However, the reaction time data in the comprehension experiments does not lend support for the kind of prior distribution over worlds necessary for this result.

#### 8.5.4 Effect of perspective priors: Dirichlet distribution favoring speaker

The prior probability distribution over perspectives can also be manipulated. One possibility is to implement an egocentricity bias through a Dirichlet prior over perspectives, rather than through the perspective cost function.

For the Pragmatic Listener, a Dirichlet prior over perspectives favoring the speaker will lead to similar effects on the marginal posterior probabilities over worlds as a heavy perspective cost parameter setting (Figure 8.23).

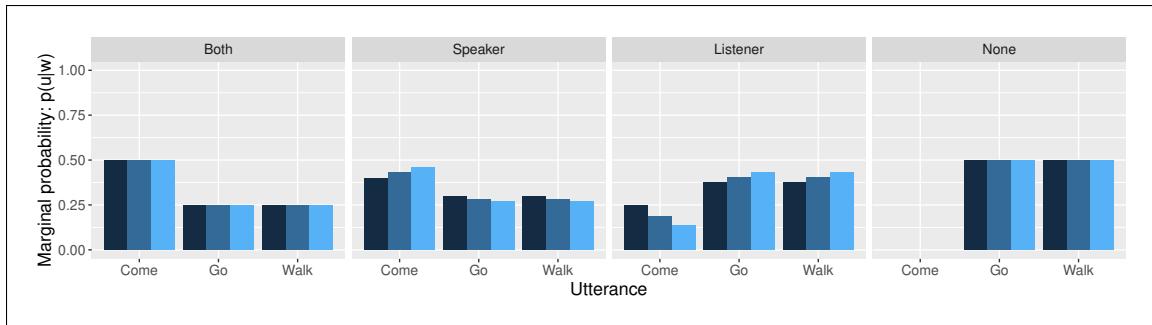
**Figure 8.23.** Pragmatic Listener marginal posterior probabilities over worlds with a Dirichlet prior over perspectives favoring the speaker, subject = Thelma and plain *go* semantics (cost parameters = 0, 0.5, 1 from left to right)



Similarly, for the Pragmatic Speaker, a prior distribution that favors the speaker will act similarly to a perspective cost function: it increases the likelihood of producing *come* in the Speaker condition, and decreases it in the Listener condition (Figure 8.24).

Thus, a speaker-biased prior over perspectives impacts the predictions of the PRSA model similarly to the perspective cost function.

**Figure 8.24.** Pragmatic Speaker marginal posterior probabilities over utterances with a Dirichlet prior over perspectives favoring the speaker, subject = Thelma and plain *go* semantics (cost parameters = 0, 0.5, 1 from left to right)



## 8.6 Summary

The production of perspectival expressions involves two decisions: the selection of an utterance and the selection of a perspective. In this chapter, I have proposed a model of the perspective production process in the RSA framework which posits that speakers reason jointly over perspectives and utterances in order to maximize their communicative success. The PRSA model proposes that speakers reason based on a mental model of how the listener will interpret their utterance. Adopting the Bayesian RSA framework provides a model of utterance selection that is flexible (as demonstrated by the variety of extensions proposed in Section 8.5), capable of generating testable predictions, and capable of capturing the egocentricity as a defeasible speaker bias, in contrast to heuristic-based approaches to perspective selection, which struggle to predict perspective shift in any circumstance.

However, I have also presented results from a production study that are challenging for the PRSA model to explain. The results provided evidence of a very strong bias towards the speaker's own perspective, which makes it difficult to test whether speakers are reasoning over perspectives, or using a simpler Speaker Default heuristic.

In tandem with Chapter 7, the findings of Chapter 8 suggest an asymmetry between the comprehension and production of perspectival expressions that is troubling for the PRSA approach. The experimental results reveal an asymmetry between speaker and listener behavior: when listeners hear a perspectival utterance, they assign highest probability to

the world in which both speaker and listener are at the destination of motion, but speakers do not produce *come* utterances to describe this world as frequently as the listeners expect.

This asymmetry is troubling for a RSA model, since speakers and listeners are actively reasoning about each other's behavior. Over time, listeners and speakers should be able to fine-tune their models of each other's processes, especially given that an individual generally has experience as both a speaker and a listener. It would even be surprising to find a mismatch between the speaker and listener estimates of the perspective cost, if, as the RSA framework posits, conversation participants are constantly refining their mental models of each other's behavior.

Although it is plausible for speakers and listeners to employ distinct systems of perspective identification and selection, it is unexpected in a RSA approach, which posits fundamental symmetry between the reasoning processes of the conversation participants. While the individual PRSA models proposed might be salvaged as separate models of perspective identification and perspective selection (particularly as components within a two-stage systems), the experimental findings are not consistent with the broader vision of the RSA framework for a symmetrical, recursive model of production and comprehension.

In addition to the asymmetry issue, there are a number of unanswered questions about perspective production and comprehension. Harris (2012)'s two-stage model remains an attractive proposal, since it provides a way of balancing the attested cases of perspective shift that seem to require perspectival reasoning with the empirically observed strong speaker bias, which can emerge from reliance on a simpler heuristic approach. However, a key unresolved question is how the switch between the two systems is triggered. As I have argued, it is not enough to switch between systems when the truth conditions of an utterance are inconsistent with the heuristically-indicated perspective. Some other cue is needed.

Another large set of questions relate to the impact of the discourse factors discussed in Chapter 6 on perspective identification and selection. Capturing the impact of these factors would necessitate enriching the PRSA model of discourse. Although the existing PRSA model is conditioned in a limited way on the discourse context (in the form of the priors

on perspectives and worlds), it cannot explain perspectival reasoning and selection over long stretches of discourse, since it lacks a process for integrating the contribution of an utterance back into the discourse context. As I argued at the beginning of this chapter, this prevents RSA models of production from capturing Grice's Maxim of Quantity. In order to incorporate discourse factors, the PRSA model would need to be integrated into a larger theory of conversational update that describes how the speaker and listener process a single discourse move and update the conversational scoreboard.

## CHAPTER 9

### CONCLUSION

This dissertation has explored a number of questions related to perspective. Holding different perspectives is a central part of the human condition. The fact that our knowledge is always incomplete means that our points-of-view will always differ, even in matters of fact, let alone in other aspects of our experience, such as perceptions and preferences. It is therefore unsurprising that many aspects of language involve perspective.

Throughout this dissertation, I have explored some of the ways that natural language encodes perspective. I have focused on perspectival expressions: expressions whose meaning depends in part on the point-of-view of an individual. Perspectival expressions are interesting because they allow multiple possible perspective holders. I take this optionality to be a key property of perspectival expressions, and have focused this dissertation on exploring two of the avenues of inquiry it opens up.

One of these avenues focuses on the question of how the perspective holder is encoded in the semantics of perspectival expressions. In Part I, I developed a theoretical landscape of perspectival expressions consisting of four families: the lexical family, the indexical family, the logophoric family, and the anaphoric family. The analysis underpinning each family rests, with few additional assumptions, on commonly adopted mechanisms for context sensitivity. Although I have not focused on testing whether all four theoretical families of perspectival expressions are attested, I presented two case studies: one on American English *come*, in which I argued for an anaphoric approach, and one on non-utterance time *tomorrow*, in which I argued for either a logophoric or anaphoric approach.

Another avenue of inquiry focuses on the challenge that the optionality over perspective holders poses for conversation participants. Since there are often multiple perspective hold-

ers available to anchor a perspectival expression, the speaker must have some kind of process for selecting a perspective to use. Similarly, the optionality over perspective holders poses an interpretive challenge from the point-of-view of the listener. As I argued in Chapter 7, the speaker’s intended meaning and their adopted perspective are often mutually constraining (and therefore, mutually informative). The listener must have some process for identifying the speaker’s adopted perspective as part of the comprehension process. In Chapters 7 and 8, I developed computational models of these processes in the Rational Speech Acts framework, a powerful framework for modeling pragmatic competition in conversation.

In the following section, I briefly summarize the main findings of the dissertation, before turning to a variety of open questions related to the landscape of perspectival expressions and how they are processed.

## 9.1 Main findings

### 9.1.0.1 The theoretical landscape of perspectival expressions

One of the main contributions of this dissertation is a theoretical typology of perspective. I have made a distinction between two kinds of perspectival phenomena: perspectival expressions, which reference the perspectives of discourse-given individuals, and perspectival environments, which control the introduction and prominence of perspective holders. My focus throughout the dissertation has been on the first kind of phenomenon, though I briefly discussed perspectival environments in Chapter 3 and factors that affect the prominence of perspectives in Chapter 6.

Perspectival expressions refer relative to a perspective. In Chapter 3, I proposed four ways that the perspective holder might be encoded in the semantics of perspectival expressions: lexically, indexically, logophorically, and anaphorically. Each of these approaches relies, for the most part, on independently motivated semantic mechanisms. I am not really proposing a new semantic class of expressions, since, as I discussed in Chapter 5, each of these approaches corresponds to an already-established class of context sensitivity, but trying to define a theoretical space of semantic hypotheses for perspectival expressions.

I view each of the four approaches as a theoretical family of hypotheses. I have attempted to describe a parent analysis for each family that uses a minimal set of assumptions. The child hypotheses in each of these families may be visualized as nearer or farther from this center analysis based on the additional assumptions that they would require in order to capture the attested behavior of perspectival expressions. For instance, I detailed three variations on a logophoric analysis of *tomorrow* that use different methods to derive utterance time readings. Other children in this family could also be imagined.

#### 9.1.0.2 A set of perspectival diagnostics

One of my goals in describing a theoretical typology of perspectival expressions is to facilitate the analysis of novel perspectival expressions. In Chapter 3, I explored the predictions of each of the four proposed families of perspectival expressions in order to build a set of perspectival diagnostics. I refined this set of diagnostics in Chapter 5 after demonstrating its use in two cases studies: in Chapter 3, a short case study of a canonical perspectival expression, American English *come*, and in Chapter 4, a more involved case study of an expression not previously recognized as perspectival, American English *tomorrow*. The final set of diagnostics is presented again in Table 9.1.

**Table 9.1.** Predictions by account

	Lex.	Idx.	Idx.+Log.	Obl. Log.	Opt. Log.	Ana.
Singleton perspective set	✓	X	X	X	X	X
Restricted perspective set	X	✓	✓	✓	✓	✓
Obligatory <i>de se</i> anchoring	◊	✓	✓	✓	✓	◊
Shift outside finite CP	✓	X	✓	✓	✓	✓
Shift outside XP with subj.	✓	X	X	X	✓	✓
Shift Together in finite CP	X	✓	X	X	X	◊
Shift Together in XP with subj.	X	✓	X	✓	✓	◊
Anchoring across utterances	X	X	X	X	✓	✓
Covariation in quant. contexts	X	X	✓	✓	✓	✓

My hope is that this set of diagnostics can be applied to probe the semantics of novel perspectival expressions, whether in a fieldwork setting, where the semantics of the expression have not previously been studied, or in a situation where unexpected behavior is observed

for an already well-studied context-sensitive expression, as in the case of non-utterance time readings of American English *tomorrow*. Although not all of the diagnostics may apply to every language, my hope is that the set of diagnostics will serve as a guide for narrowing the space of semantic hypotheses for a given perspectival expression.

My work on *tomorrow* highlights how critical it is to probe the behavior of context-sensitive expressions carefully before drawing conclusions: the licensing conditions for different classes of context sensitivity, both perspectival and non-perspectival, overlap very closely. In order to adjudicate between two competing analyses, it is vital to test the handful of environments that do not fall in their intersection.

#### **9.1.0.3 A perspectival semantics for *tomorrow***

In Chapter 4, I developed a perspectival analysis of American English *tomorrow*, motivated by the finding that many speakers accept non-utterance time readings of *tomorrow*. I explored the licensing conditions for such readings in a series of acceptability judgment studies. I showed that many speakers accept non-utterance time readings of *tomorrow* in contexts where shifted readings of indexicals are unattested, such as in quantificational binding contexts and outside of attitude reports. I further argued that the quantificational binding data is inconsistent with a simple time-anaphoric account, and supports a perspectival approach. Having narrowed the space of perspectival hypotheses to a logophoric or anaphoric account, I proposed a perspective-anaphoric semantics for *tomorrow*.

Several key findings about the semantic encoding of perspective emerged from this case study. First, this account highlights that perspective sensitivity may be found even in well-studied context-sensitive expressions, because the licensing conditions of different kinds of context sensitivity overlap to a great extent. Second, the experimental work revealed a surprising degree of interspeaker variability in judgments of non-utterance time readings of *tomorrow*. The source and scope of this variation remains unidentified. Third, the difficulty in adjudicating between the logophoric and perspective-anaphoric families of approaches for *tomorrow* exemplifies the challenges of precisely diagnosing the semantics of a given perspectival expression. Despite the fine-grained experimental exploration of *tomorrow*'s

licensing conditions, there remain a number of viable ways of treating its semantics within the logophoric-anaphoric hypothesis space.

#### **9.1.0.4 Processing and producing perspectives**

In Part II, I turned to questions of how perspectival expressions are used and interpreted. After surveying previous approaches to perspective processing and some of the factors proposed to affect perspective prominence in Chapter 6, I turned to the question of how listeners interpret perspectival expressions.

In Chapter 7, I proposed a model of perspective identification based in the Rational Speech Acts framework. I posited that listeners use Bayesian inference to simulate the speaker's production process as they jointly reason about the speaker's adopted perspective and their intended meaning. I presented experimental evidence that tentatively supports a key assumption of this model: that listeners reason simultaneously over multiple perspectives.

In Chapter 8, I proposed a symmetrical model of perspective selection in which speakers use Bayesian inference to reason jointly about the best utterance and perspective to use based on a mental model of the listener's comprehension process. However, the evidence from a production study of perspectival motion verbs did not support the predictions of this model. This presents a challenge to the symmetry between speaker and listener behavior predicted by the Rational Speech Acts framework.

A main contribution of this part of the dissertation is the development of implemented computational models of perspectival reasoning. These models can be used to simulate perspectival reasoning on conversation and generate empirically testable quantitative predictions about speaker and listener behavior. They are also flexible models that can be easily adapted to new contexts and to incorporate different kinds of biases. In Chapter 8, I also discussed several variants on the PRSA model, although the results did not resolve the puzzling asymmetry between speaker and listener behavior. The PRSA models that I have proposed can be used on their own, or as part of a multi-stage model, such as the two-stage joint heuristic and reasoning model Harris (2012) posits for perspectival processing.

## 9.2 Open questions

I have explored two sets of questions related to perspective in this dissertation. However, there are many open questions that remain. Some of these are longstanding issues of interest; others are sparked by the findings of this dissertation.

### 9.2.1 The actual landscape of perspectival expressions

In Part I, I laid out a theoretical landscape of perspectival expressions: four ways that the perspective holder can be encoded using independently motivated semantic mechanisms. However, I have not gone very far in illustrating the actual landscape of perspectival expressions. Are all four of these theoretical classes of perspectival expressions actually found in natural language?

Answering this question will require extensive cross-linguistic work. Throughout Part I, I have tried to highlight the importance of applying fine-grained diagnostics. Because the licensing conditions of different classes of perspective sensitivity (and context sensitivity) overlap so closely, particular care must be taken to test their behavior in the critical environments that distinguish among them.

In order to build a typology of attested perspectival expressions, it may be necessary to revisit some expressions that have already been classified. As my case study on *tomorrow* revealed, applying perspectival diagnostics to well-studied expressions can yield surprising results. Although the case studies that I explored pointed towards the logophoric and anaphoric families of analyses, other classes of perspectival expressions may exist as well, and may even be masquerading in the guise of another class of context sensitivity.

### 9.2.2 Uniformity of perspective across languages

The case studies in this dissertation have focused on expressions in American English. Looking at expressions in other languages may produce very different findings. We have already seen evidence from Charnavel (2018) that the semantic encoding of perspectival motion verbs varies across languages. In order to flesh out the landscape of attested perspectival expressions, we must draw upon data from many other languages.

The question of cross-linguistic uniformity of perspective is also important to the questions that I explored in Part II. Just as there is cross-linguistic variation in the semantic encoding of perspectival expressions, we might also expect cross-linguistic variation in the factors that affect perspective prominence. In particular, we saw in Chapter 2 that some languages restrict the set of perspective holders for perspectival motion verbs more severely than English. This will affect the perspective identification and selection processes, at least in the PRSA approach that I have proposed, since a reduced set of perspective holders impacts the model's predictions.

Since we have seen evidence that egocentricity and perspective maintenance are cognitive pressures, we might not expect them to vary cross-linguistically. However, the impact of discourse factors like topicality, subjecthood, and empathy on perspective prominence may well vary cross-linguistically, since these are linguistic pressures. I hope that future work will focus more on these factors. The fact that they are likely sites of cross-linguistic variation makes work on them more pressing, since if that is the case, findings from one language will not transfer well to another language.

### 9.2.3 Uniformity of perspective within a language

Another question we might ask about the actual landscape of perspectival expressions is how consistent the encoding of perspective is within a given language. Do individual languages use multiple methods of encoding the perspective holder, or is the strategy for encoding the perspective holder a point of cross-linguistic variation only?

In this dissertation, I have focused on perspectival expressions in American English. In both of the case studies in Part I, I found evidence in support of a perspective-anaphoric analysis. One possibility is that this is the only strategy that English uses to encode the perspective holder. Or it may be that all methods of perspective encoding are available, and different expressions in English make use of different methods.

It is also unknown whether the set of perspective holders is consistent within a language. Do perspectival *tomorrow* and *come* allow all the same perspective holders? If so, we should expect shifted readings of *tomorrow* to be available in every context where non-

speaker anchorings of *come* are possible. Two perspective-anaphoric expressions in identical environments should refer relative to the same prominent perspective, since their referents are determined by the Common Ground. However, it is also possible that the expression-specific component of meaning might impose restrictions on the set of candidate perspective holders; for instance, if the speaker is the subject, their perspective cannot anchor *come*, but it may still be an available anchor for *tomorrow*.

The predictions are even less clear for expressions in different perspectival classes. A perspectival indexical and a perspective-anaphoric expression derive their referent in different ways, so they do not necessarily share the same perspective set. However, it is also possible that the set of perspective holders is determined by the language and is consistent across all perspectival expressions.

#### **9.2.4 Interspeaker variability in perspective identification**

One of the most puzzling findings in Part I was the attested interspeaker variability in ratings of non-utterance time *tomorrow* uses. This variability does not seem to be tied to any of the basic demographic variables measured. However, it is possible that there is some as yet unidentified sociolinguistic correlate.

This issue could be pursued in a number of different directions. First, there is the question of whether the variability is due to the existence of two distinct grammars (or lexical entries for *tomorrow*), or to variability in the prominence of perspectives. Is *tomorrow* perspectival for all speakers, but some speakers have more difficulty accessing non-speaker perspectives?

Second, gathering experimental evidence of how speakers interpret different kinds of perspectival expressions could reveal whether this interspeaker variability is a common property of perspectival phenomena, or limited to American English *tomorrow*. This is related to the first question: if the interspeaker variability is due to two lexical entries for *tomorrow*, then we would not expect to observe it with other perspectival expressions; if it is due to processing differences, we might expect to observe it for all perspectival expressions.

### **9.2.5 Incorporating discourse factors**

The production and comprehension models that I have proposed incorporate information about the discourse context in a very limited way: through the prior probability distributions over worlds and perspectives. However, in Chapter 6, we saw that many different kinds of discourse factors have been proposed to affect the prominence of perspective holders. One important area for future work is to incorporate the discourse context into the production and comprehension models more fully.

My approach to the question of perspective identification in Part II may have seemed counter-intuitive, given that I focused mostly on perspective-anaphoric expressions in Part I. Why not approach the perspective identification problem using one of the many systems that have been proposed for pronoun resolution? Instead, I proposed a more general system, partly motivated by the observation that such a system is necessary even for perspectival expressions that are not anaphoric.

One interesting direction for future work is whether the reasoning-based system that I have proposed can be combined with an anaphoric resolution system that makes use of discourse factors like topicality and subjecthood.<sup>1</sup> This might be one way of combining richer contextual information, which clearly plays a role in perspective identification, with the desirable features of the PRSA system, such as probabilistic handling of alternatives and joint reasoning over perspectives and meanings.

### **9.2.6 Incorporating social factors**

Another future direction for the PRSA model is to incorporate social factors. In the PRSA model, the perspective cost function implements a bias towards the speaker's perspective. However, listeners might adjust their expectations about the speaker's likelihood to shift perspectives based on various social cues. For instance, when someone is speaking to their supervisor, they may be more likely to accommodate their perspective as a mark of respect. Incorporating situation-dependent social information, like the relationship between the speaker and listener, might improve the listener's predictions about the speaker's in-

---

<sup>1</sup>For instance, Kehler and Rohde (2013).

tended meaning. One way of doing this would be to assume that the perspective is speaker-dependent; listeners might adjust it throughout a conversation to better fit the behavior of an individual speaker.

Social factors might also impact the speaker's behavior. Production often involves a trade-off between ease of production and communicative success. Speakers might behave more rationally when prompted by evidence that they are not communicating well with their listener. A speaker's beliefs about the listener's ability to shift perspectives might also influence their perspective selection process. For instance, parents might be more likely to adopt the listener's perspective when speaking to children, since children may struggle to shift perspective. Or an instructor might accommodate the class's perspective in order to communicate more effectively, like when instructors say *Today we will learn X* even though they already understand X. Whether or not social factors like these play a role in the speaker's production process is an interesting question for future empirical work.

### **9.2.7 Perspective production: an open challenge**

Although I found some support for the predictions of the PRSA model of comprehension, the experimental evidence did not support the predictions of the PRSA model of production. Speakers rarely adopted the listener's perspective, contrary to the expectations of the listeners in the comprehension experiments. I argued that this was problematic for the RSA framework, since it fundamentally predicts symmetry between speakers and listeners.

Modeling perspectival production remains a challenge. While the results of the production study show that speakers are more strongly egocentric than the PRSA model predicts, they should not be interpreted as evidence that speakers rely only on a Speaker Default. As I argued in Chapter 6, relying on simple heuristics like a Speaker Default alone will never predict perspective shift, yet, perspective shift happens.

### **9.2.8 Why shift perspective shift?**

The question of why speakers shift perspective remains one of the most puzzling and least studied questions about perspective. The emphasis in so much of the literature (both within linguistics and in adjacent fields) is on the difficulty of adopting non-speaker perspectives.

Relatively little is known about its benefits. Since we have evidence both that speakers adopt other perspectives and that this is cognitively costly, then (assuming that speaker choices are influenced by cognitive pressures) there must be some benefit to shifting perspective.

### 9.3 Last thoughts

Natural language incorporates many different kinds of context sensitivity. Sensitivity to perspective is one kind of context sensitivity that is displayed by a diverse set of linguistic elements. In this dissertation, I have explored one category of perspective sensitivity: expressions whose meaning depends on the point-of-view of a perspective holder selected by the speaker. I have focused on two sets of questions about perspectival expressions. In Part I, I explored how the perspective holder is encoded in the semantics of perspectival expressions. I divided the theoretical landscape of perspectival expressions into four semantic families, and proposed a set of diagnostics for probing their semantics. In Part II. I turned to questions of perspective processing. I proposed twin models of perspective production and comprehension to address how speakers select a perspective to use and how listeners identify the speaker's adopted perspective.

Throughout this dissertation, I have also tried to demonstrate the value of using experimental and computational methods in studying the semantics and pragmatics of perspective. Although these methodologies are valuable in many domains, the fact that there are multiple available perspectives in most contexts makes them particularly important for measuring and modeling gradient acceptability of perspectival expressions.

There are many interesting questions about perspective that remain unanswered. I hope that this dissertation will aid further study of perspectival expressions at a number of levels. The set of diagnostics proposed in Part I is meant both as a first step towards a typology of perspectival expressions, and as a guide for analyzing novel perspectival expressions. The computational models proposed in Part II can be adapted to novel datasets and modified in various ways, or combined with other strategies in order to build a multi-stage model of processing. The data from the comprehension and production studies can also be used to refine and test alternative models of perspective processing.

Point-of-view is a thread woven through much of natural language. The perspective sensitivity of a linguistic phenomenon can be obvious and central to the meaning of the expression, as in the case of perspectival motion verbs, or very subtle, as in the way that good writers delicately flavor a discourse through the choice of aspect and argument structure. I have focused on expressions whose semantics directly refer to a perspective, but in order to develop a complete picture of the role of perspective in natural language, these overtly perspectival expressions must be integrated into a unified framework alongside perspective shift environments and subtler perspective-enhancing discourse effects.

## APPENDIX A

### DEMOGRAPHIC ANALYSIS OF NON-UTTERANCE TIME *TOMORROW* READINGS

One interesting finding from Chapter 4 was the interspeaker variation in acceptance of non-utterance time readings of *tomorrow*. In this appendix, I discuss demographic trends and individual differences among the participants. None of the population-level dimensions of variance that were measured show a strong correlation with acceptance of non-utterance time readings, though there is some effect of geographic location and age. The present sample sizes are too small to infer much, but future work could explore these factors further.

#### A.1 Demographic measures

Participants in Experiments 1-4 completed a short demographic survey at the conclusion of the experiment in which they reported their age, state of residence, and gender.

##### A.1.1 Acceptability ratings by geographic location

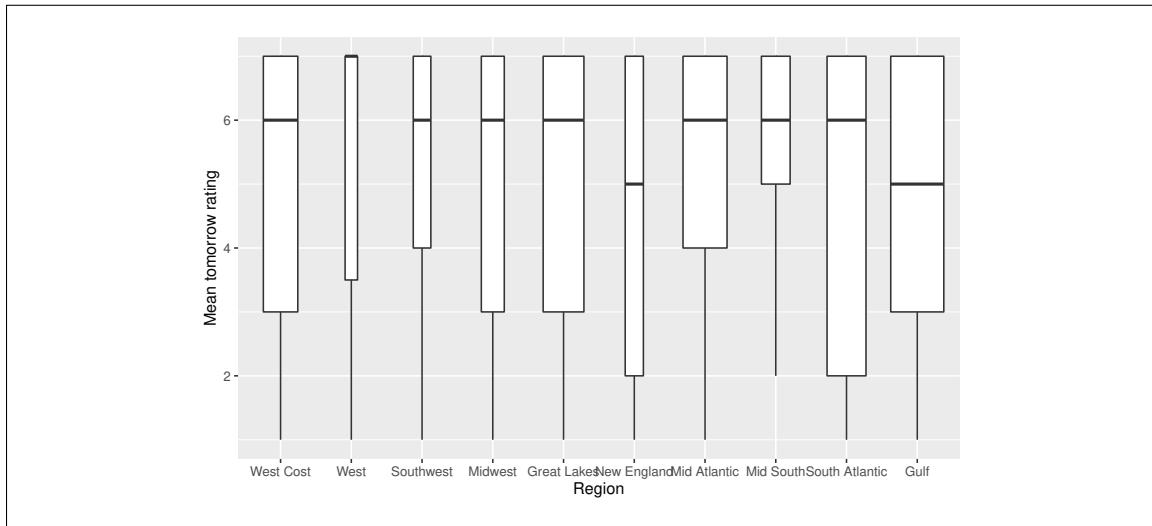
Participants were asked about their state of residence. There is some variance in *tomorrow* ratings by geographic location. Figure A.1 shows the ratings for *tomorrow* in all experiments by geographic region.<sup>1</sup>

This looks fairly uniform, with perhaps lower ratings in the Gulf states.

---

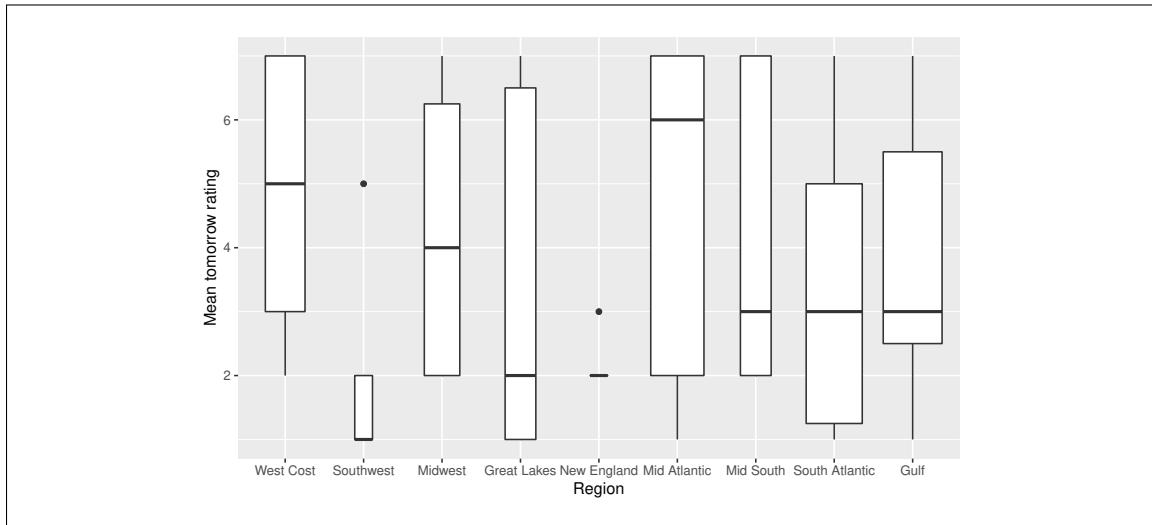
<sup>1</sup>States were binned into regions as follows:  
West Coast: Alaska, California, Hawaii, Oregon, Washington.  
West: Idaho, Montana, Wyoming.  
Southwest: Arizona, Colorado, Nevada, New Mexico, Utah.  
Midwest: Iowa, Kansas, Nebraska, North Dakota, South Dakota.  
Great Lakes: Illinois, Indiana, Michigan, Minnesota, Ohio, Wisconsin.  
New England: Maine, Massachusetts, New Hampshire, Rhode Island, Vermont.  
Mid Atlantic: Connecticut, Delaware, New Jersey, New York, Pennsylvania.  
Mid South: Arkansas, Kentucky, Missouri, Oklahoma, Tennessee, West Virginia.  
South Atlantic: District of Columbia, Maryland, North Carolina, South Carolina, Virginia.  
Gulf: Alabama, Florida, Georgia, Louisiana, Mississippi, Texas.

**Figure A.1.** Ratings for *tomorrow* items in all experiments by geographic region

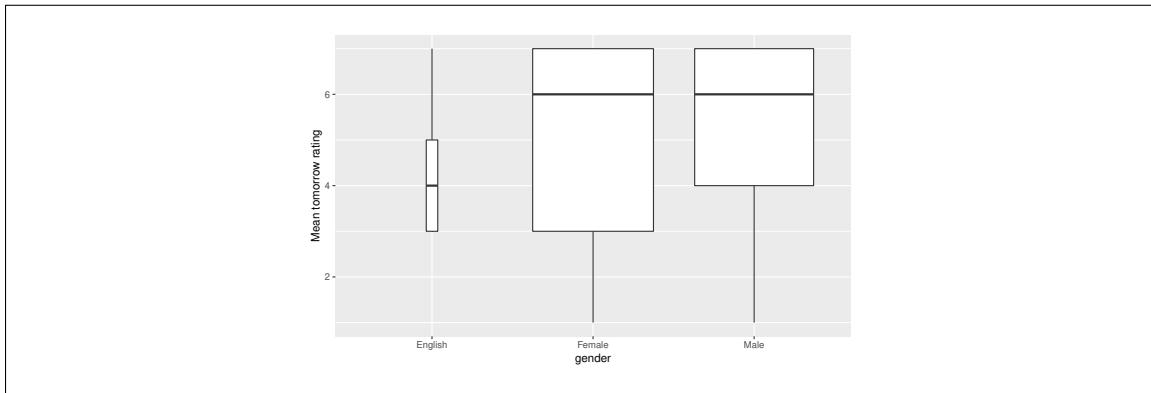


However, the ratings for *tomorrow* items in Experiment 4 look somewhat influenced by region; Figure A.2 shows the *tomorrow* ratings in Experiment 4 by region. The *tomorrow* ratings in this experiment were much lower in the southern states compared to the mid-Atlantic states. However, the small sample sizes make it difficult to explore this further.

**Figure A.2.** Ratings for *tomorrow* items in Experiment 4 by geographic region



**Figure A.3.** Ratings for *tomorrow* items in all experiments by gender



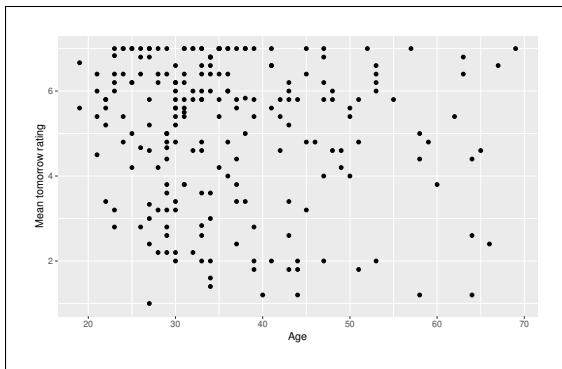
### A.1.2 Acceptability ratings by gender

Participants were asked about their gender. Although the form allowed free responses, all participants except one identified as male or female. As Figure A.3 shows, there does not seem to be any significant variance in *tomorrow* ratings by gender.

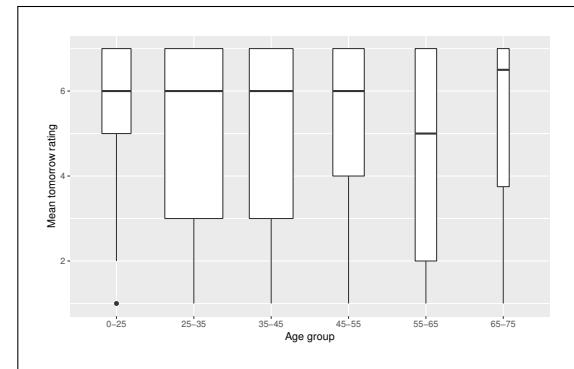
### A.1.3 Acceptability ratings by age

Participants were also asked to give their age. There does not seem to be a strong correlation between age and ratings for *tomorrow* items. Figures A.4 and Figure A.5 show the mean responses of participants plotted by age.

**Figure A.4.** Ratings for *tomorrow* items in all experiments by age



**Figure A.5.** Ratings for *tomorrow* items in all experiments by age group



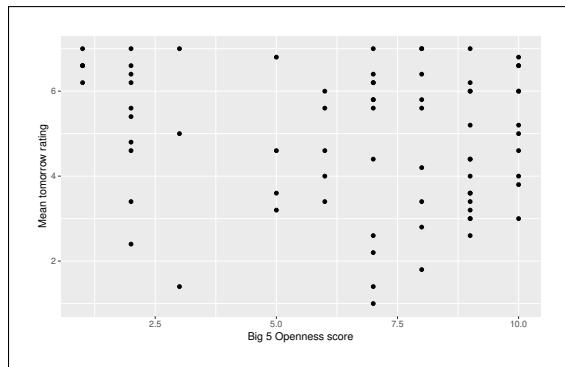
## A.2 Individual differences measures

In addition to the demographic questions asked in all experiments, participants in Experiment 2 took part in three tasks measuring individual differences.

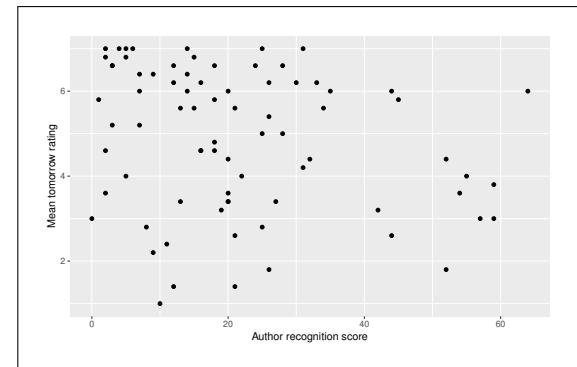
### A.2.1 Acceptability ratings by Big Five Openness characteristics

In Experiment 2, participants were evaluated with respect to the Openness construct of the Big Five personality questionnaire (John and Srivastava, 1999). The Openness construct measures characteristics like imagination, curiosity, and willingness to go against conventions.

**Figure A.6.** Experiment 2 ratings for *tomorrow* items by Big Five Openness scores



**Figure A.7.** Experiment 2 ratings for *tomorrow* items by Author Recognition scores



Participants were asked to evaluate how well various statements described them. Their responses were gathered using a five-point Likert scale where 1 indicated that they definitely agreed with the statement, and 5 indicated that they definitely disagreed. An example statement is shown in (206).

206. I am someone who likes to reflect and play with ideas.

As shown in Figure A.6, there was no strong correlation between Openness scores and the ratings that participants gave to *tomorrow* items ( $\rho=-0.18$ ).

### A.2.2 Acceptability ratings by Author Recognition performance

The literary familiarity of participants in Experiment 2 was evaluated using the Author Recognition Task (Acheson et al., 2008). This task involves a list of 65 names of authors

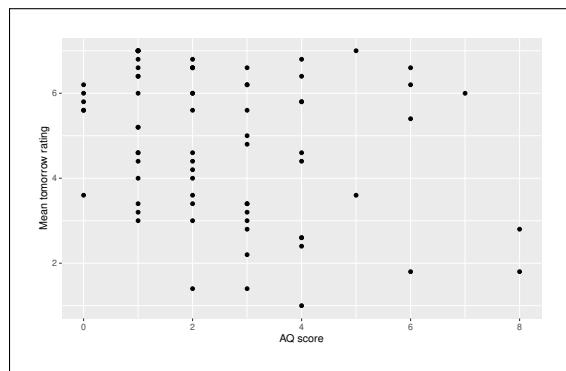
and 65 distractor names. Participants are asked to identify how many of the names they recognize as belonging to authors.

This was measured because participants who have much literary experience might be more familiar with perspective shift environments like Free Indirect Discourse, and might therefore be more likely to accept shifted readings of *tomorrow*. As shown in Figure A.7, however, there was no positive correlation between participants' scores on the Author Recognition Task and the ratings that they gave to *tomorrow* items ( $\rho=-0.23$ ).

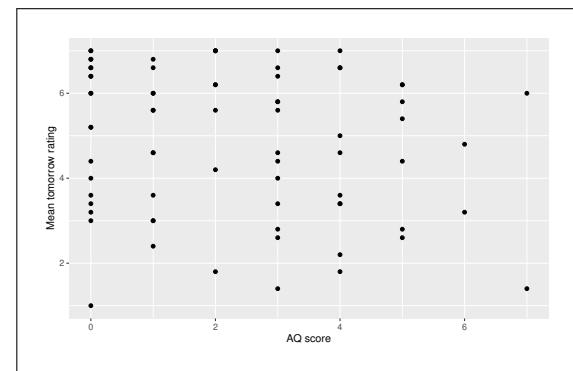
### A.2.3 Acceptability ratings by AQ scores

Participants in Experiment 2 were also evaluated using the Imagination and Communication portions of the Autism Quotient (AQ) questionnaire (Baron-Cohen et al., 2001). These subcategories are composed of twenty self-descriptive statements. Participants were asked to indicate how well each statement fit them using a four-point Likert scale, where 1 indicated that they definitely agreed and 4 indicated that they definitely disagreed. Responses of 1 or 2 were coded as matching the statement. Two example statements are shown below.

**Figure A.8.** Experiment 2 ratings for *tomorrow* items by AQ Imagination scores



**Figure A.9.** Experiment 2 ratings for *tomorrow* items by AQ Communication scores



207. **Imagination:** I find it difficult to imagine what it would be like to be someone else.

208. **Communication:** I know how to tell if someone listening to me is getting bored.

The Imagination and Communication subcategories were chosen because they relate to the ability to imagine other people's points-of-view. Participants who self-identify as strong in this ability might be more likely to accept shifted readings of *tomorrow*. However, as

Figures A.8 and A.9 show, neither portion of the AQ questionnaire correlated positively with ratings for *tomorrow* ( $\rho=-0.16$  for Communication;  $\rho=-0.23$  for Imagination).

## APPENDIX B

### LOGOPHORIC ANALYSES OF *TOMORROW*

In Chapter 4, I presented some native speaker judgments that supported a perspective-anaphoric analysis of *tomorrow* over a logophoric approach. However, as shown by the experimental work in that chapter, there is much interspeaker variation in judgments on non-utterance time readings of *tomorrow*. The judgments that I provided should therefore be treated with a healthy amount of skepticism. Future experimental work may reveal that speakers do not accept instances of non-utterance time *tomorrow* licensed across utterance boundaries. If this is the case, then a logophoric approach should be pursued. In this appendix, I sketch three variants on a logophoric analysis of *tomorrow*.

#### B.1 Mechanisms for the logophoric approach

The logophoric approach assumes that the perspectival variable is governed by a logophoric operator. I draw on Charnavel (2019)'s treatment of exempt anaphors, which proposes that a logophoric operator may be projected in any spellout domain (TP, vP, DP, and any XP that has a subject). The logophoric operator is a syntactic head  $OP_{LOG}$  that selects a silent logophoric pronoun  $pro_{LOG}$  as subject. The proposed configuration is shown in 209.

$$209. [pro_{log_i} [OP_{log} [\alpha \dots \text{exempt anaphor}_i \dots]]]$$

The logophoric operator imposes a restriction on its complement: the complement must be interpreted from the first-person perspective of the subject of the logophoric operator. The semantics of Charnavel (2019)'s logophoric operator are shown in 210.

$$210. [[OP_{log}]]^{C,g} = \lambda\alpha.\lambda x.\alpha \text{ from } x\text{'s first-personal perspective.}$$

I spell out this requirement in the following way. First, I propose that logophoric perspectival expressions contain perspectival variables bound by lambda operators. Thus, logophoric

perspectival expressions are functions from perspectives to their ordinary category. For instance, the perspectival version of a verb that would ordinarily be of type  $< e, t >$ , will be of type  $< u, < e, t >>$ , as shown in (211) using motion verbs as an example.

211. Perspectival versus plain motion verb semantics:

- (a)  $[[\text{move to the bank}]]^{C,g} = \lambda e_v.\text{MOVE}(e) \wedge \text{DEST}(e, b)$
- (b)  $[[\text{come to the bank}]]^{C,g} = \lambda a_u.\lambda e_v.\text{MOVE}(e) \wedge \text{DEST}(e, \text{LOC}(a))$

I give this example for illustrative purposes only; as I argued in Chapter 3, I view perspectival motion verbs in English as anaphoric, not logophoric.<sup>1</sup>

The logophoric semantics of *tomorrow* similarly contains an unsaturated perspectival variable, as shown in (212).

212. Logophoric semantics for *tomorrow*:

$$[[\text{tomorrow}_{log}]]^{C,g} = \lambda a_u.\lambda Q_{<v,t>}.\lambda e_v.Q(e) \wedge \tau(e) \subset \iota t_i.\text{DAY-AFTER}(t, \text{TIME}(a))$$

where  $\text{TIME}(a)$  returns the temporal location index of perspective  $a$

I propose that the logophoric operator take as its arguments a perspectival complement and a subject, and asserts that the perspectival variable of the complement represents a perspective of the subject. This is shown in (213) below.

213.  $[[OP_{log}]]^{C,g} = \lambda \alpha_{<u,<v,t>>}.\lambda x_e.\exists a_u.\text{PERSPECTIVE}(a, x) \wedge \alpha(a)$ , where  $\text{PERSPECTIVE}(a,x)$   
is true if  $a$  is a first-person perspective of  $x$ .<sup>2</sup>

I posit that the complement of the logophoric operator must be a predicate of type  $< u, < v, t >>$ . I am not necessarily committed to restricting the complement to a perspectival event description. However, I do want to restrict the complement to containing only one unsaturated perspectival variable. In my view, what it means for the complement of

<sup>1</sup>This semantics may, however, be viewed as an implementation of Charnavel (2018)'s logophoric proposal for perspectival motion verbs in French.

<sup>2</sup>Note that  $a$ , though existentially bound, is still contextually restricted; thus,  $a$  must be a discourse-given perspective, as in the anaphoric account.

the logophoric operator to be from the first-personal perspective of its subject is that all perspectival variables in the complement refer to the subject's perspective.<sup>3</sup>

Introducing unsaturated perspectival arguments requires some additional rules of composition.<sup>4</sup> I will propose a PERSPECTIVAL FUNCTION APPLICATION rule as defined in (214). Its purpose is to allow unsaturated perspectival arguments to percolate up through a derivation until they reach a logophoric operator.

**214. Perspectival Function Application (PFA):** If  $\alpha$  is a branching node with daughters  $\beta$  and  $\gamma$ , then for any assignment  $C, g$  and any types  $x, y$ :

- (a) If  $\beta$  is a function  $\lambda a_u. \lambda b_x. \delta(b)$  of type  $< u, < x, y >>$  and  $\gamma$  is a function of type  $x$ , then  $\beta(\gamma)$  is the function  $\lambda a'_u. \delta(\gamma)[a \leftarrow a']$  of type  $< u, y >$ .
- (b) If  $\beta$  is a function  $\lambda b_x. \delta(b)$  of type  $< x, y >$  and  $\gamma$  is a function  $\lambda a_u. \epsilon$  of type  $< u, x >$ , then  $\beta(\gamma)$  is the function  $\lambda a_u. \delta(\epsilon[a \leftarrow a'])$  of type  $< u, y >$ .
- (c) If  $\beta$  is a function  $\lambda a_u. \lambda b_x. \delta(b)$  of type  $< u, < x, y >>$  and  $\gamma$  is a function  $\lambda a'_u. \epsilon$  of type  $< u, x >$ , then  $\beta(\gamma)$  is the function  $\lambda a_u. \delta(\epsilon[a' \leftarrow a])$  of type  $< u, y >$ .
- (d) If  $\beta$  is a function  $\lambda a_u. \lambda b_x. \delta(b)$  of type  $< u, < x, y >>$  and  $\gamma$  is a function  $\lambda a'_u. \epsilon$  of type  $< u, x >$ , then  $\beta(\gamma)$  is the function  $\lambda a_u. \lambda a'_u. \delta(\epsilon)$  of type  $< u, < u, y >>$ .

There are three cases. First, there may be a perspectival function that must be composed with a non-perspectival argument. This is handled by the first subrule, which allows a perspectival function to compose with an argument without saturating its perspectival variable (Figure B.1).

<sup>3</sup>This is also a way of deriving a locality constraint on perspectival variables: all unsaturated perspectival variables will end up bound by the nearest logophoric operator that scopes over them.

<sup>4</sup>This is not necessary for the perspective-anaphoric analysis, even though it uses perspective variables, because they are always free variables. As shown in (214), the presence of a free perspectival variable in an expression does not affect its type.

214. (a)  $[[\text{John sings}]]^{C,g} = \lambda e_v. \text{sing}(e) \wedge \text{AGENT}(e, j)$   
 (b)  $[[\text{John sings tomorrow}]]^{C,g} = \lambda e_v. \text{sing}(e) \wedge \text{AGENT}(e, j) \wedge \tau(e) \subset \lambda t_i. \text{DAY-AFTER}(t, \text{TIME}(a))$

**Figure B.1.** Perspectival function applied to a non-perspectival argument by PFA

$$\begin{array}{c} \lambda a. \beta \wedge \gamma(a) \text{ by PFA subrule 1} \\ \swarrow \qquad \searrow \\ \lambda a. \lambda \alpha. \alpha \wedge \gamma(a) \quad \beta \end{array}$$

Second, there may be non-perspectival functions that need to be composed with perspectival arguments. This is handled by the second subrule, which composes the two arguments while passing on the unsaturated perspectival argument to the result (Figure B.2).

**Figure B.2.** Non-perspectival function applied to a perspectival argument by PFA

$$\begin{array}{c} \lambda a. \beta(a) \wedge \gamma \text{ by PFA subrule 2} \\ \swarrow \qquad \searrow \\ \lambda \alpha. \alpha \wedge \gamma \quad \lambda a. \beta(a) \end{array}$$

Lastly, there may be both a perspectival function and a perspectival argument that need to be composed. This is handled by the third and fourth subrules. Either the perspectival variables are identified as part of function composition (Figure B.3), or they are both passed on to the result of the composition (Figure B.4).

**Figure B.3.** Perspectival function applied to a perspectival argument by PFA, perspective identification

$$\begin{array}{c} \lambda a. \beta(a) \wedge \gamma(a) \text{ by PFA subrule 3} \\ \swarrow \qquad \searrow \\ \lambda a. \lambda \alpha. \alpha \wedge \gamma(a) \quad \lambda a'. \beta(a') \end{array}$$

**Figure B.4.** Perspectival function applied to a perspectival argument by PFA, no perspective identification

$$\begin{array}{c} \lambda a. \lambda a'. \beta(a) \wedge \gamma(a') \text{ by PFA subrule 4} \\ \swarrow \qquad \searrow \\ \lambda a. \lambda \alpha. \alpha \wedge \gamma(a) \quad \lambda a'. \beta(a') \end{array}$$

The outcome in all cases is that the unsaturated perspectival variable is passed on to the resulting function.

Although there are several subrules for Perspectival Function Application, the goal is to propose the minimal set of composition roles needed to allow unsaturated perspectival arguments to be passed upwards through a derivation, until they can be resolved by a

logophoric operator. The body of the expressions compose in an ordinary manner. If there is an unsaturated perspectival argument, it is carried through to the result. If there are two unsaturated perspectival arguments, they can either be identified with each other or not.

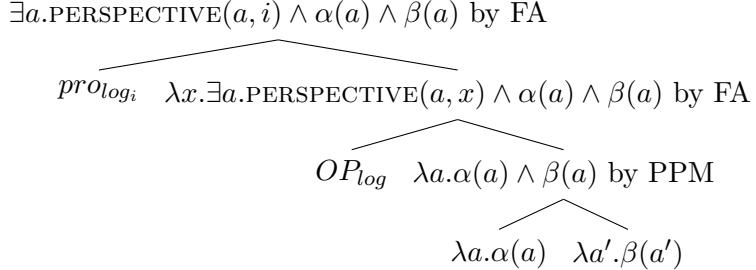
In addition to Perspectival Function Application, it may also be necessary to have a perspectival variant of Predicate Modification. I propose the following PERSPECTIVAL PREDICATE MODIFICATION rule, which allows predicates to combine regardless of whether they carry the same number of unsaturated perspectival arguments.

**215. Perspectival Predicate Modification (PPM):** If  $\alpha$  is a branching node with daughters  $\beta$  and  $\gamma$ , then for any assignment  $C, g$  and any types  $x, y$ :

- (a) If  $\beta$  is a function  $\lambda z_x.\delta(z)$  of type  $\langle x, y \rangle$  and  $\gamma$  is a function  $\lambda z'_x.\epsilon(z')$  of type  $\langle x, y \rangle$ , then  $\beta \oplus_p \gamma$  is the function  $\lambda z_x.\delta(z) \wedge \epsilon(z)$
- (b) If  $\beta$  is a function  $\lambda a_u.\delta(a)$  of type  $\langle u, x \rangle$  and  $\gamma$  is a function  $\lambda a'_u.\epsilon(a')$  of type  $\langle u, x \rangle$ , then  $\beta \oplus_p \gamma$  is the function  $\lambda a''_u.\delta(a'') \oplus_p \epsilon(a'')$
- (c) If  $\beta$  is a function  $\lambda a_u.\delta(a)$  of type  $\langle u, x \rangle$  and  $\gamma$  is a function  $\epsilon$  of type  $x$ , then  $\beta \oplus_p \gamma$  is the function  $\lambda a_u.\delta(a) \oplus_p \epsilon$
- (d) If  $\beta$  is a function  $\delta$  of type  $x$  and  $\gamma$  is a function  $\lambda a_u.\epsilon$  of type  $\langle u, x \rangle$ , then  $\beta \oplus_p \gamma$  is the function  $\lambda a_u.\delta \oplus_p \epsilon(a)$

This rule is defined recursively: if neither predicate has an unsaturated perspectival argument, the predicates combine via normal predicate modification (subrule 1). If both predicates have a perspectival argument, the perspectival arguments are identified and the

**Figure B.5.** Derivation with two perspectival expressions combining via PPM



body of the predicates are combined via predicate modification (subrule 2). If only one of the predicates has a perspectival argument, the body of the two expressions are composed via predicate modification and the perspectival argument is percolated to the result (subrules 3 and 4).

These perspectival rules of composition allow perspectival variables to percolate up a derivation until they reach a logophoric operator. Although the fourth subrules of PFA allows perspectival expressions to combine without identifying their unsaturated perspective variables with each other, the selectional restriction on the logophoric operator will generally prevent the number of unsaturated perspectival arguments from expanding, since it selects arguments with a single unsaturated perspectival argument.

## B.2 Three variations on a logophoric analysis of *tomorrow*

The logophoric family of analyses treats perspectival expressions as containing bound perspectival variables. Shifted readings of *tomorrow* are predicted to arise when *tomorrow* is in the scope of a logophoric operator whose subject has a non-utterance time perspective. Given the set of mechanisms for encoding perspective introduced above, which are motivated independently by Charnavel (2019)'s work on exempt anaphors, there are three possible analyses logophoric analyses for *tomorrow*.

One option is to assume that there is a speaker-oriented top-level logophoric projection (Speas and Tenny, 2003). If there is no lower logophoric projection to saturate the perspectival argument of *tomorrow*, a top-level logophoric phrase is projected with the speaker as the subject. This will derive indexical-like behavior in the absence of lower logophoric operators, since ordinarily, the temporal perspective of the speaker is identical to utterance time. This approach predicts obligatory Shift Together effects for multiple instances of *come* within the same binding domain and that non-utterance time readings of *tomorrow* will only arise in logophoric environments.

Another way of deriving utterance time readings in the logophoric approach is to posit a process for salvaging unsaturated perspectival arguments by converting them into free variables. In this approach, *tomorrow* carries an unsaturated perspectival variable, but

there is a last resort mechanism for resolving its value anaphorically (similar to Jacobson (1999)'s treatment of pronouns). This predicts Shift Together effects for multiple instances of *tomorrow* within the same logophoric environment, but allows shifted readings to arise outside of the scope of logophoric operators.

A third option for deriving utterance time readings within the logophoric approach is to posit lexical ambiguity. Perhaps there are two versions of *tomorrow*: an indexical one and a logophoric one. When *tomorrow* is in a logophoric environment, its logophoric semantics are licensed; logophoric *tomorrow* cannot be used outside of the scope of a logophoric operator because its perspectival variable is obligatorily bound. When *tomorrow* occurs outside of a logophoric environment, the indexical variant is used, resulting in a utterance time reading. This predicts that shifted readings will only arise in logophoric environments, but it does not predict obligatory Shift Together effects, since the indexical version of *tomorrow* is available inside of logophoric environments as well.

In the next few sections, I demonstrate how each of these analyses would work.

### **B.2.1 Version 1: Shift Together is obligatory and shifted readings only happen in logophoric environments**

In the logophoric approach, the perspectival variable in the semantics of *tomorrow* is obligatorily bound by a logophoric operator. Shifted readings arise when *tomorrow* is in the scope of a logophoric operator projected in some spellout domain. In this approach, the perspectival variable of *tomorrow* is bound by a lambda abstractor, as shown in (216).

#### **216. Logophoric semantics for *tomorrow*:**

$$[[\text{tomorrow}_{log}]]^{C,g} = \lambda a_u. \lambda Q_{<v,t>}. \lambda e_v. Q(e) \wedge \tau(e) \subset \iota t_i. \text{DAY-AFTER}(t, \text{TIME}(a))$$

There are multiple options for how utterance time readings arise in a logophoric treatment of *tomorrow*. One possibility is that the perspectival variable is always resolved by a logophoric operator. In cases where there is no lower logophoric projection, it is resolved by a top-level

speaker-oriented logophoric projection (Speas and Tenny, 2003).<sup>5</sup> This will derive utterance time interpretations for instances of *tomorrow* that do not occur in a lower logophoric projection, since the speaker is located at utterance time.

Shifted readings arise only if *tomorrow* is in the scope of a lower, non-speaker oriented logophoric operator. This makes a perspective other than that of the speaker at utterance time available to bind the perspectival variable of *tomorrow*. Example (217) shows how a derivation for shifted *tomorrow* would proceed under this account.

First, the logophoric operator takes *John sings tomorrow* as its complement. This complement meets the logophoric operator's selectional restriction, since it contains a single unsaturated perspectival variable. This variable is then existentially bound and restricted to perspectives of the subject of the logophoric operator (217g).

Next, the logophoric operator combines with its subject. In this example, its subject is a logophoric pronoun that is co-indexed with the attitude holder, Mary.<sup>6</sup> The perspectival variable is now restricted to ranging over perspectives that Mary holds.

217.  $[\log P \text{ prolog}_{\text{speaker}} OP_{\log} [\text{Mary}_i \text{ said } [\log P \text{ prolog}_i OP_{\log} \text{ that John sings tomorrow.}]]]$
- (a) Assignment function:  $g(i) = \text{Mary}$
  - (b)  $[[\text{tomorrow}_{\log}]]^{C,g} = \lambda a_u. \lambda Q_{<v,t>}. \lambda e_v. Q(e) \wedge \tau(e) \subset \iota t_i. \text{DAY-AFTER}(t, \text{TIME}(a))$
  - (c)  $[[\text{prolog}_i]]^{C,g} = g(i) = m$  by Assignment Function
  - (d)  $[[OP_{\log}]]^{C,g} = \lambda \alpha_{<u,<v,t>>}. \lambda x_e. \exists a_u. \text{PERSPECTIVE}(a, x) \wedge \alpha(a)$
  - (e)  $[[\text{John sings}]]^{C,g} = \lambda e. \text{SING}(e) \wedge \text{AGENT}(e, j)$
  - (f)  $[[\text{John sings tomorrow}]]^{C,g} = \lambda a. \lambda e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, \text{TIME}(a)) \wedge \text{SINGS}(e) \wedge \text{AGENT}(e, j)$  By Perspectival Function Application

---

<sup>5</sup>This strategy cannot be available in all languages with logophoric operators, however, since some logophors cannot receive speaker-oriented interpretations in matrix clauses.

<sup>6</sup>I assume that the logophoric pronoun evaluates to the attitude holder through co-indexing, but this could also be achieved by direct binding of the logophoric pronoun. I am not committed to a particular view of this binding process.

- (g)  $[[OP_{log} \text{ John sings tomorrow}]]^{C,g} = \lambda x'.\lambda e.\exists a.\text{PERSPECTIVE}(a,x) \wedge \tau(e) \subset \iota t.\text{DAY-AFTER}(t, \text{TIME}(a)) \wedge \text{SING}(e) \wedge \text{AGENT}(e,j)$  By Function Application
- (h)  $[[\text{pro}_{log_i} OP_{log} \text{ John sings tomorrow}]]^{C,g} = \lambda e.\exists a.\text{PERSPECTIVE}(a,m) \wedge \tau(e) \subset \iota t.\text{DAY-AFTER}(t, \text{TIME}(a)) \wedge \text{SING}(e) \wedge \text{AGENT}(e,j)$  By Function Application
- (i)  $[[\text{said } \text{pro}_{log_i} OP_{log} \text{ John sings tomorrow}]]^{C,g} = \lambda e'.\text{SAY}(e) \wedge \text{THEME}(e', (\lambda e.\exists a.\text{PERSPECTIVE}(a,m) \wedge \tau(e) \subset \iota t.\text{DAY-AFTER}(t, \text{TIME}(a)) \wedge \text{SING}(e) \wedge \text{AGENT}(e,j)))$  By Function Application
- (j)  $[[v \text{ said } \text{pro}_{log_i} OP_{log} \text{ John sings tomorrow}]]^{C,g} = \lambda x.\lambda e'.\text{SAY}(e) \wedge \text{AGENT}(e',x) \wedge \text{THEME}(e', (\lambda e.\exists a.\text{PERSPECTIVE}(a,m) \wedge \tau(e) \subset \iota t.\text{DAY-AFTER}(t, \text{TIME}(a)) \wedge \text{SING}(e) \wedge \text{AGENT}(e,j)))$  By Function Application
- (k)  $[[ \text{Mary said } \text{pro}_{log_i} OP_{log} \text{ John sings tomorrow}]]^{C,g} = \lambda e'.\text{say}(e') \wedge \text{AGENT}(e',m) \wedge \text{THEME}(e', (\lambda e.\exists a.\text{PERSPECTIVE}(a,m) \wedge \tau(e) \subset \iota t.\text{DAY-AFTER}(t, \text{TIME}(a)) \wedge \text{SING}(e) \wedge \text{AGENT}(e,j)))$  By Function Application  
 $= \exists e'.\text{say}(e') \wedge \text{AGENT}(e',m) \wedge \text{THEME}(e', (\lambda e.\exists a.\text{PERSPECTIVE}(a,m) \wedge \tau(e) \subset \iota t.\text{DAY-AFTER}(t, \text{TIME}(a)) \wedge \text{SING}(e) \wedge \text{AGENT}(e,j)))$  By Existential Closure

For a non-shifted reading, there must be a top-level logophoric projection to resolve the unsaturated perspectival variable. (218) shows how an utterance time interpretation of *tomorrow* would arise. In this case, the unsaturated perspectival variable of *tomorrow* is passed upwards through the derivation via Perspectival Function Application. There is a top-level logophoric projection anchored to the speaker. The top-level logophoric operator ultimately binds the perspectival variable and restricts it to perspectives of the speaker.

218.  $[\log P \text{ pro}_{log_i} OP_{log} [\text{Mary}_i \text{ said that John sings tomorrow.}]]$

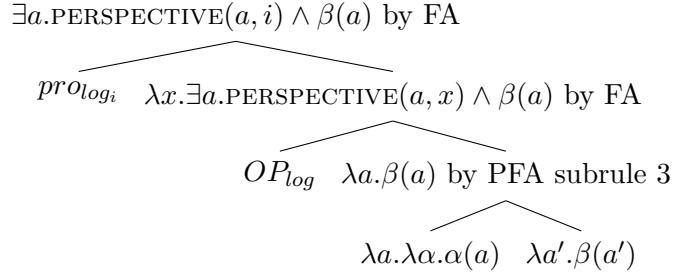
- (a) Assignment function:  $g(i) = \text{speaker}$
- (b)  $[[\text{tomorrow}_{log}]]^{C,g} = \lambda a_u.\lambda Q_{<v,t>}.\lambda e_v.Q(e) \wedge \tau(e) \subset \iota t_i.\text{DAY-AFTER}(t, \text{TIME}(a))$
- (c)  $[[\text{pro}_{log_i}]]^{C,g} = g(i) = \text{speaker}$  By Assignment Function
- (d)  $[[OP_{log}]]^{C,g} = \lambda \alpha_{<u,<v,t>>}.\lambda x_e.\exists a_u.\text{PERSPECTIVE}(a,x) \wedge \alpha(a)$

- (e)  $[[\text{John sings}]]^{C,g} = \lambda e. \text{SING}(e) \wedge \text{AGENT}(e, j)$
- (f)  $[[\text{John sings tomorrow}]]^{C,g} = \lambda a. \lambda e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, \text{TIME}(a)) \wedge \text{SING}(e) \wedge \text{AGENT}(e, j)$   
By Function Application
- (g)  $[[\text{said John sings tomorrow}]]^{C,g} =$   
 $\lambda a. \lambda e'. \text{SAY}(e) \wedge \text{THEME}(e', \lambda e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, \text{TIME}(a)) \wedge \text{SING}(e) \wedge \text{AGENT}(e, j))$   
By Perspectival Function Application
- (h)  $[[v \text{ said John sings tomorrow}]]^{C,g} =$   
 $\lambda a. \lambda x. \lambda e'. \text{SAY}(e) \wedge \text{AGENT}(e, x) \wedge \text{THEME}(e', \lambda e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, \text{TIME}(a)) \wedge \text{SING}(e) \wedge \text{AGENT}(e, j))$   
By Perspectival Function Application
- (i)  $[[\text{Mary said John sings tomorrow}]]^{C,g} =$   
 $\lambda a. \lambda e'. \text{say}(e') \wedge \text{AGENT}(e', m) \wedge \text{THEME}(e', \lambda e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, \text{TIME}(a)) \wedge \text{SING}(e) \wedge \text{AGENT}(e, j))$   
By Perspectival Function Application
- (j)  $[[OP_{log} \text{ Mary said John sings tomorrow}]]^{C,g} =$   
 $\lambda x. \lambda e'. \exists a. \text{PERSPECTIVE}(a, x) \wedge \text{say}(e') \wedge \text{AGENT}(e', m) \wedge \text{THEME}(e', \lambda e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, \text{TIME}(a)) \wedge \text{SING}(e) \wedge \text{AGENT}(e, j))$   
By Function Application
- (k)  $[[\text{pro}_{log_i} OP_{log} \text{ Mary said John sings tomorrow}]]^{C,g} =$   
 $\lambda e'. \exists a. \text{PERSPECTIVE}(a, \text{speaker}) \wedge \text{say}(e') \wedge \text{AGENT}(e', m) \wedge \text{THEME}(e', \lambda e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, \text{TIME}(a)) \wedge \text{SING}(e) \wedge \text{AGENT}(e, j))$   
By Function Application

What about Shift Together effects? I assume that the logophoric operator carries a selectional restriction: its complement must have only one unsaturated perspectival variable. This is one way of deriving Charnavel (2019)'s locality constraint semantically: if all perspectival variables must be identified with each other prior to projecting a logophoric operator, then the logophoric operator will bind all perspectival variables in its scope. If another logophoric operator is projected above the first, the perspectival variables will be unaffected by the higher logophoric operator, since they have already been bound by the most local operator.

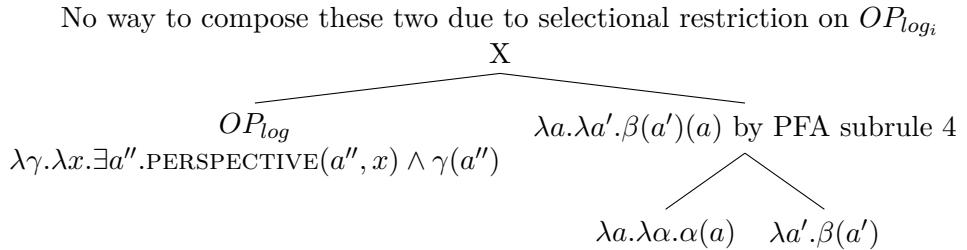
This selectional restriction derives mandatory Shift Together effects. When there are multiple perspectival expressions, a logophoric operator cannot be projected unless their perspectival variables have been identified with each other. Figure B.6 shows how a derivation involving two perspectival expressions can proceed: the functions combine via the third subrule of Perspectival Function Application so that their unsaturated perspectival arguments are identified with each other. This ensures that the complement satisfies the logophoric operator's selectional restriction.

**Figure B.6.** Derivation with two perspectival expressions combining via PFA



By contrast, if two perspectival functions are composed via the fourth subrule of Perspectival Function Application, which does not identify the perspectival variables with each other, then the derivation will fail. As shown in Figure B.7, if there are two perspectival variables in the complement, then it does not satisfy the selectional restriction of the logophoric operator and the derivation fails.

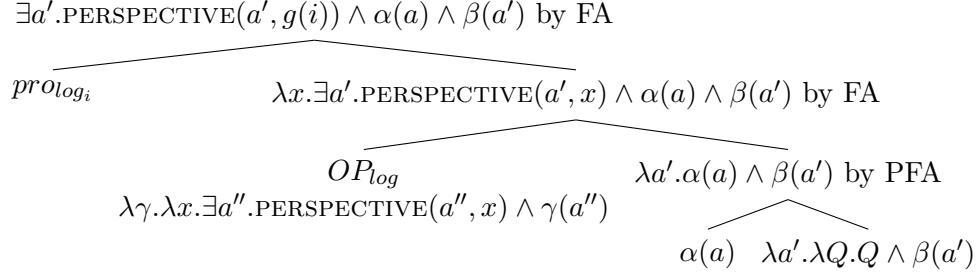
**Figure B.7.** Failed logophoric projection over two non-unified perspectival variables



While this account predicts mandatory Shift Together effects for all logophoric perspectival variables, it does not predict mandatory Shift Together effects between logophoric and anaphoric perspectival expressions in the same domain. For instance, if both *come* and logophoric *tomorrow* appear in the scope of a logophoric operator, there is no guarantee that

the perspectival variable of *come* will be resolved to the subject of the logophoric operator, since this is handled by the anaphora resolution algorithm. This is shown schematically in Figure B.8, though the denotational details of each expression are elided.

**Figure B.8.** Derivation showing lack of Shift Together effects between logophoric and anaphoric perspectival variables



### B.2.2 Version 2: Shift Together is obligatory and shifted readings happen out of logophoric environments

Another possibility is to adopt a logophoric approach, but treat the derivation of utterance time readings differently. Instead of positing a top-level logophoric operator, we could posit a last-resort approach that converts unsaturated perspectival variables into free variables. This would allow them to be resolved anaphorically. I will call this operation Perspective Pronominalization:

219. **Perspective Pronominalization:** convert an unsaturated perspectival variable to a free perspectival variable.

$$\lambda a_u.\beta(a) \rightarrow \beta(a') \text{ by Perspective Pronominalization}$$

This process can only be invoked at the top level of a derivation. Essentially, if a perspectival variable makes it to the top of a derivation without being saturated, then it can receive a value directly from the discourse context.<sup>7</sup>

With these modifications in place, we can now build an account where *tomorrow* behaves logophorically in the scope of a logophoric operator, but anaphorically in all other environments. As in the approach in Section A.2.1, I assume a logophoric semantics for *tomorrow*.

---

<sup>7</sup>Similar to Pauline Jacobson's approach to pronouns (Jacobson, 1999).

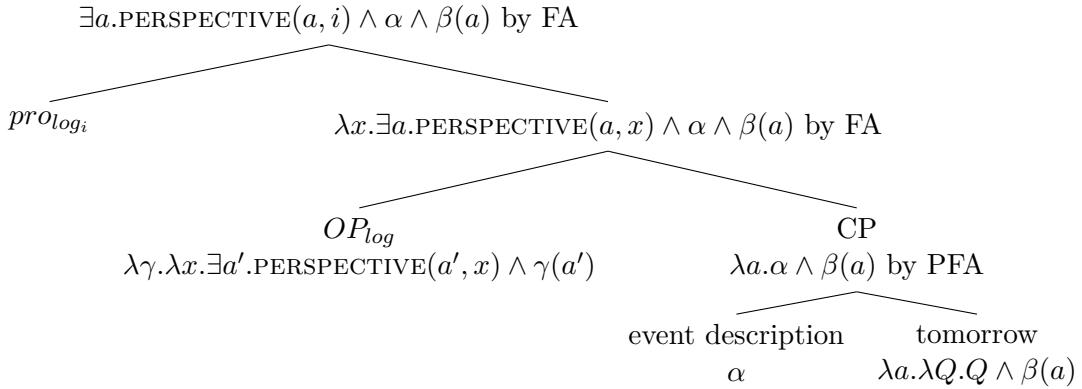
220. Logophoric semantics for *tomorrow*:

$$[[\text{tomorrow}_{log}]]^{C,g} = \lambda a_u. \lambda Q_{<v,t>}. \lambda e_v. Q(e) \wedge \tau(e) \subset \iota t_i. \text{DAY-AFTER}(t, \text{TIME}(a))$$

I also assume, as in Section A.2.1, that the logophoric operator has a selectional restriction prohibiting it from taking complements with more than one unsaturated perspectival variable.

The derivation of non-utterance time readings of *tomorrow* proceeds identically as in Section A.2.1. As shown schematically in Figure B.9, the perspectival variable of *tomorrow* percolates upwards until it is bound by the logophoric operator.

**Figure B.9.** Non-utterance time derivation for Version 3



When *tomorrow* is not in the scope of a logophoric operator, however, its unsaturated perspectival variable is converted into a free variable. This free variable is then resolved via the anaphora resolution process. An example of how utterance time readings of *tomorrow* are derived in this system is shown in (221).

This derivation proceeds similarly to the utterance time reading in (218) in Section A.2.1 until the eighth step (221h). Instead of combining with a top-level logophoric operator, the unsaturated perspectival argument is converted into a free variable by Perspective Pronominalization. Its value is then determined by the assignment function.

221. Mary said that John sings tomorrow.

$$(a) \quad \begin{aligned} \text{Perspective set} = & \{ \quad a_{\text{speaker}} : < \{w_1, \dots\}, t_{\text{speaker}}, l_{\text{speaker}} > \\ & a_{\text{Mary}} : < \{w_1, \dots\}, \text{TIME}(e_{\text{say}}), \text{LOC}(e_{\text{say}}) > \} \end{aligned}$$

- (b) Assignment function:  $g(a) = a_{\text{speaker}}$
- (c)  $[[\text{tomorrow}]]^{C,g} = \lambda a. \lambda Q. \lambda e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, \text{TIME}(a)) \wedge Q(e)$
- (d)  $[[\text{John sings}]]^{C,g} = \lambda e. \text{SING}(e) \wedge \text{AGENT}(e, j)$
- (e)  $[[\text{John sings tomorrow}]]^{C,g} = \lambda a. \lambda e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, \text{TIME}(a)) \wedge \text{SING}(e) \wedge \text{AGENT}(e, j)$   
By Perspectival Function Application
- (f)  $[[\text{said John sings tomorrow}]]^{C,g} =$   
 $\lambda a. \lambda e'. \text{SAY}(e') \wedge \text{THEME}(e', \lambda e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, \text{TIME}(a)) \wedge \text{SINGS}(e) \wedge \text{AGENT}(e, j))$   
by Perspectival Function Application
- (g)  $[[v \text{ said John sings tomorrow}]]^{C,g} =$   
 $\lambda a. \lambda x. \lambda e'. \text{SAY}(e') \wedge \text{AGENT}(e', x) \wedge \text{THEME}(e', \lambda e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, \text{TIME}(a)) \wedge \text{SING}(e) \wedge \text{AGENT}(e, j))$   
by Perspectival Function Application
- (h)  $[[\text{Mary said John sings tomorrow}]]^{C,g}$   
 $= \lambda a. \lambda e'. \text{SAY}(e') \wedge \text{AGENT}(e, m) \wedge \text{THEME}(e', \lambda e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, \text{TIME}(a)) \wedge \text{SING}(e) \wedge \text{AGENT}(e, j))$   
By Perspectival Function Application  
 $= \lambda a. \exists e'. \text{SAY}(e') \wedge \text{AGENT}(e, m) \wedge \text{THEME}(e', \lambda e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, \text{TIME}(a)) \wedge \text{SING}(e) \wedge \text{AGENT}(e, j))$   
By Existential Closure  
 $= \exists e'. \text{SAY}(e') \wedge \text{AGENT}(e, m) \wedge \text{THEME}(e', \lambda e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, \text{TIME}(a)) \wedge \text{SING}(e) \wedge \text{AGENT}(e, j))$   
By Perspective Pronominalization  
 $= \exists e'. \text{SAY}(e') \wedge \text{AGENT}(e, m) \wedge \text{THEME}(e', \lambda e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, \text{TIME}(a_{\text{speaker}})) \wedge \text{SING}(e) \wedge \text{AGENT}(e, j))$   
By Assignment Function  
 $= \exists e'. \text{SAY}(e') \wedge \text{AGENT}(e, m) \wedge \text{THEME}(e', \lambda e. \tau(e) \subset \iota t. \text{DAY-AFTER}(t, t_{\text{speaker}}) \wedge \text{SING}(e) \wedge \text{AGENT}(e, j))$   
By evaluation of the speaker's perspective's temporal field

Unlike in the account proposed in Section A.2.1, in this account, non-utterance time readings can arise outside of logophoric environments as well as inside of them. This is a consequence of allowing bound perspectival variables to be converted into free variables. Just as in the fully anaphoric account, the anaphora resolution algorithm determines how free perspectival variables are resolved. Consequently, the free perspectival variable in (221) can be resolved to either the perspective of the speaker or the attitude holder. Thus, in this approach, it is

possible for *tomorrow* to receive shifted interpretations outside of logophoric environments, as well as inside them.

This has ramifications for this account's predictions about Shift Together effects. In this account, as in the account in Section A.2.1, obligatory Shift Together effects are predicted for multiple instances of *tomorrow* in the scope of the same logophoric environment. However, in this account, Shift Together effects are not predicted to hold for perspectival variables resolved via Perspective Pronominalization, since they are essentially converted into pronouns.

To see how this works, let us consider the derivation with two logophoric perspectival expressions shown in Figure B.10.

**Figure B.10.** No Shift Together effects under Perspective Pronominalization

$$\begin{aligned}
 & \lambda a. \lambda a'. \alpha \wedge \beta(\text{TIME}(a)) \wedge \delta(\text{TIME}(a')) \text{ by PFA} \\
 & = \alpha \wedge \beta(\text{TIME}(a)) \wedge \delta(\text{TIME}(a')) \text{ by Perspective Pronominalization} \\
 & = \alpha \wedge \beta(\text{TIME}(a_{\text{speaker}})) \wedge \delta(\text{TIME}(a_{\text{Mary}})) \text{ by Assignment Function}
 \end{aligned}$$

```

graph TD
    Root["\u03bba.\u03bb a'. \u03b1 \u2227 \u03b2(TIME(a)) \u2227 \u03b4(TIME(a')) by PFA"] --> L["\u03bb a. \u03b1 \u2227 \u03b2(TIME(a)) by PFA"]
    Root --> R["\u03bb a'. \u03bb P. P \u2227 \u03b4(TIME(a'))"]
    L --> Alpha["\u03b1"]
    L --> TomorrowL["tomorrow"]
    Alpha --> AlphaLabel["\u03b1"]
    TomorrowL --> TomorrowLabel["tomorrow"]
    TomorrowLabel --> Bottom["\u03bb a. \u03bb Q. Q \u2227 \u03b2(TIME(a))"]
  
```

The perspectival functions compose via the fourth subrule of Perspectival Function Application, which keeps their unsaturated perspectival variables distinct. This blocks the projection of a logophoric operator.<sup>8</sup> Both unsaturated perspectival variables percolate up to the top of the derivation, at which point they are converted into free variables through Perspective Pronominalization. However, since they have different binders, they may be converted into indices that receive different values from the assignment function. Thus, in

---

<sup>8</sup>Note that this type of Perspectival Function Application can never be invoked in the account proposed in Section A.2.1, since the result violates the selectional restriction of the logophoric operator and there is no other way of resolving perspectival variables.

this account, Shift Together effects are predicted to hold within logophoric environments, but not outside of them.

### B.2.3 Version 3: Shift Together is not obligatory and shifted readings only happen in logophoric environments

A third way to derive utterance time readings in a logophoric account of *tomorrow* is to invoke lexical ambiguity. We could posit that *tomorrow* has both a pure indexical lexical entry and one with a logophoric semantics, as shown in (223).

$$222. [[\text{tomorrow}_{idx}]]^{C,g} = \lambda Q_{<v,t>}. \lambda e_v.Q(e) \wedge \tau(e) \subset \iota t_i.\text{DAY-AFTER}(t, C_t)$$

$$223. [[\text{tomorrow}_{log}]]^{C,g} = \lambda a_u. \lambda Q_{<v,t>}. \lambda e_v.Q(e) \wedge \tau(e) \subset \iota t_i.\text{DAY-AFTER}(t, \text{TIME}(a))$$

This approach derives non-utterance time readings of *tomorrow* identically as in Sections A.2.1 and A.2.2. However, the utterance time readings proceed differently. Rather than positing a top-level logophoric operator, as in Section A.2.1, or a way of converting unsaturated perspectival arguments into free variables, as in Section A.2.2, in this approach, we will assume that utterance time readings arise from an indexical version of *tomorrow*.

When *tomorrow* receives a shifted reading, it is the logophoric variant in the scope of a logophoric operator. If there is no logophoric operator, then the indexical variant of *tomorrow* must be used. An utterance time derivation using an indexical semantics for *tomorrow* is shown in (224).

224. Mary said that John sings tomorrow.

$$(a) [[\text{tomorrow (indexical)}]]^{C,g} = \lambda Q_{<v,t>}. \lambda e_v.Q(e) \wedge \tau(e) \subset \iota t_i.\text{DAY-AFTER}(t, C_t)$$

$$(b) [[\text{John sings}]]^{C,g} = \lambda e.\text{SING}(e) \wedge \text{AGENT}(e, j)$$

$$(c) [[\text{John sings tomorrow}]]^{C,g} = \lambda e. \tau(e) \subset \iota t.\text{DAY-AFTER}(t, C_t) \wedge \text{SING}(e) \wedge \text{AGENT}(e, j)$$

By Function Application

$$(d) [[\text{Mary said John sings tomorrow}]]^{C,g}$$

$$= \lambda e'. \text{say}(e') \wedge \text{AGENT}(e', m) \wedge \text{THEME}(e', \lambda e. \tau(e)) \subset \iota t.\text{DAY-AFTER}(t, C_t) \wedge \text{SING}(e) \wedge \text{AGENT}(e, j))$$

By Function Application

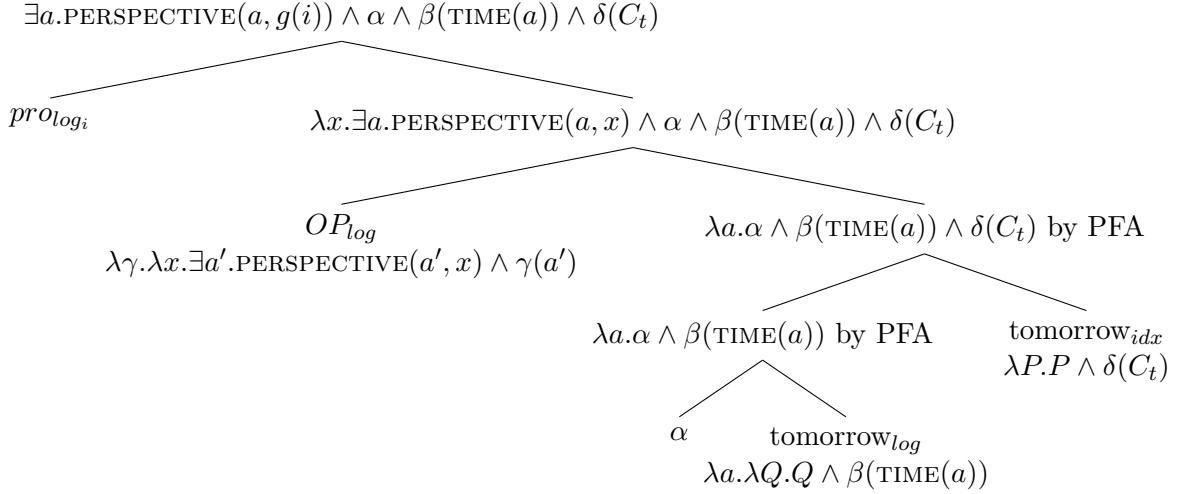
$$= \exists e'.say(e') \wedge \text{AGENT}(e', m) \wedge \text{THEME}(e', \lambda e.\tau(e)) \subset \iota t.\text{DAY-AFTER}(t, C_t) \wedge \text{SING}(e) \wedge \text{AGENT}(e, j))$$

By Existential Closure

In this proposal, there is no way to resolve a perspectival variable that is left unsaturated at the top level of a derivation. If the logophoric version of *tomorrow* is projected and there is no logophoric operator to resolve its perspectival variable, the derivation will crash. Thus, shifted readings can only arise in logophoric environments, as in the approach pursued in Section A.2.1.

Unlike in the approach proposed in Section A.2.1, however, when *tomorrow* is in a logophoric environment, it need not be logophoric; the indexical variant can also be used. If there are two instances of *tomorrow* in the scope of a logophoric operator, one may be indexical while the other is logophoric. This would lead to one receiving a shifted reading while the other receives an utterance time reading. This is shown schematically in Figure (B.11).

**Figure B.11.** Derivation showing lack of Shift Together effects in logophoric environments



This leads to the prediction that Shift Together effects are not obligatory within logophoric environments. Multiple instances of *tomorrow* in the same logophoric domain do not necessarily receive a uniform interpretation, since some of the instances may be logophoric while others are indexical.<sup>9</sup>

---

<sup>9</sup>However, Shift Together effects could still be implemented in this system by a stipulation in the semantics of the logophoric operator. For instance, we could posit that because logophoric environments express the perspective of their subject, when there is a choice of two perspectival expressions, the one consistent with

This approach is one kind of lexical ambiguity approach. We could also ask whether it makes sense to consider other kinds of lexical ambiguity, for instance, a combination of the anaphoric approach with the indexical approach. However, in practice, there would be no way to tell the resulting analysis apart from a solely anaphoric analysis, since all the diagnostics will point to anaphoricity. A less restrictive analysis enables *tomorrow* to ‘escape out of’ the predictions of the more restricted one, in the sense that its licensing conditions are a superset of the licensing conditions of a more restricted approach. This is not the case for the indexical/logophoric combination approach, because both approaches impose some constraints, but it is the case for most other lexical ambiguity possibilities.

#### B.2.4 Summary

I have proposed three variants on a logophoric semantics for *tomorrow*. The three logophoric variants, together with the perspective-anaphoric analysis proposed in Chapter 4, correspond to each of the four logical outcomes of testing for obligatory Shift Together effects within logophoric environments and for shifted readings outside of logophoric environments.

---

the perspective of the logophoric subject is preferred. This would be a competition-based explanation in the style of Reinhart and Reuland (1993)’s approach to pronominalization.

## APPENDIX C

### ***DE SE JUDGMENTS AND NON-UTTERANCE TIME TOMORROW***

In Chapter 4, I omitted discussion of one perspectival diagnostic: whether *tomorrow* is obligatorily *de se*-interpreted. In order to test this, we must set up a scenario where the perspective holder is mistaken about their own temporal location. In (225), Marge is confused about the date. Her self-ascribed temporal location is June 23rd, but her actual temporal location is June 22nd. If *tomorrow* is obligatorily *de se*-interpreted, it should only be able to be interpreted as referring to June 24th, not June 23rd. (225) is infelicitous for this reason: *tomorrow* refers to June 23rd, the day after the actual temporal location of Marge's event time self, not her self-ascribed temporal location.

225. *On June 22nd, which is a Tuesday, Marge is confused about the date. She thinks it is the 23rd. She is expecting an important package on the 23rd, so she waits at home all day for it. The next week, she tells her friend John:*

226. #I had a huge mix-up about dates last week. I stayed home on Tuesday and waited all day for a package. I didn't realize it was scheduled to come tomorrow!

By contrast, in (225), Marge's self-ascribed temporal location is the 22nd, and her actual location is the 23rd. In this context, *tomorrow* should only be able to refer to the 23rd, not the 24th. Since *tomorrow* refers to the 23rd in (225), it is felicitous.

227. *On June 23rd, which is a Tuesday, Marge is confused about the date. She thinks it is the 22nd. She is expecting an important package on the 23rd, and plans to take the next day off to wait for it, but it comes while she is at work on the 23rd. The next week, she tells her friend John:*

228. I had a huge mix-up about dates last week. I went to work on Tuesday because I thought my package was coming tomorrow. I didn't realize it was scheduled to come that day!

These judgments are based on my own intuitions. Similar examples would have to be explored experimentally in order to confirm that *tomorrow* is obligatorily *de se*-interpreted.

## APPENDIX D

### PRSA MODEL SPECIFICATION AND IMPLEMENTATION

In this Appendix, I describe in more detail the implementation and specifications for the PRSA models proposed in Chapters 7 and 8.

#### D.1 Dataset and parameter settings

A PRSA dataset must include the following components: a set of possible worlds along with its prior probability distribution, a set of utterances, and a set of perspectives along with its prior probability distribution. In addition, the interpretation function, which maps utterances to truth values, must be specified, and, in models that include cost functions, the cost function weight must be set.

##### D.1.1 Possible worlds

Each possible world is a tuple of fields with binary values that indicate the truth of a proposition in that world.<sup>1</sup> For computational purposes, only a finite number of propositions can be encoded in each world; propositions that are not encoded can be viewed as having been resolved in the prior discourse, or as ones about whose truth the conversation participants are completely agnostic. Either of these views is safe so long as the truth of all utterances under consideration does not depend on the missing proposition.

##### D.1.2 Utterances

The set of utterances affects the model’s calculations by manipulating the alternatives that are considered. For the case study that I have focused on, perspectival motion verbs, I included a non-perspectival alternative whose semantics overlapped (partially or completely,

---

<sup>1</sup> Alternatively, each world can be modeled as a vector of binary values.

depending on the semantics used for *go*) with the verbs of interest. This is important because omitting a relevant competitor affects not just the numerical posterior probabilities calculated by the model, but also the interactions among competing utterances.

### D.1.3 Perspectives

In the simulations that I presented, I limited the perspectives under consideration to the speaker and the listener. This is a simplification: the perspectives of other individuals may also be at play. There are a number of strategies for broadening the perspective set. One is to include the perspectives of all discourse-given individuals. Another is to start with a perspective set consisting of just the conversation participants, and allow others to be introduced by linguistic operators: for instance, Roberts (2015) proposes that attitude verbs introduce the perspectives of their subjects into the Common Ground.

### D.1.4 Prior probability distributions over worlds

In the simulations reported, I use uniform prior probability distributions except where otherwise noted. The uniform prior over worlds is theoretically motivated: it simulates an empty Common Ground where the speaker and listener have no publicly shared beliefs about any propositions.

In reality, speakers and listeners do not enter conversations as blank slates: they have individual beliefs and assumptions about the beliefs of the other conversation participants. To model these private commitments, the PRSA model could be enriched with separate priors over worlds for each conversation participant in addition to the shared distribution over worlds of the current model. This would also allow us to model cases where conversation participants do not fully trust each others' contributions. However, it is a significant enrichment of the model, and I leave this for future work.

I do experiment with some non-uniform priors over worlds in Section 8.5.3, to test whether the experimental results could be explained by a bias towards a particular scene type.

### D.1.5 Prior probability distributions over perspectives

The choice of a prior probability distribution for perspectives is less obvious. As discussed in Chapter 7, an alternative to the perspective cost function is to use a speaker-biased prior on perspectives. I experiment with this approach in Section 8.5.4.

### D.1.6 Perspective cost function

For perspective cost, I explored settings of  $\{0, 0.25, 0.5, 0.75, 1.0\}$ . A setting of 0 indicates that there is no penalty for adopting non-speaker perspectives. There is no upper bound on the cost setting, so values above 1 could also be explored.

## D.2 Implementation details

The simulations reported were run using an implementation of the PRSA model in the WebPPL probabilistic programming language.

### D.2.1 Sampling

Simulations can be run in different ways. Because exact inference is often intractable for Bayesian models, I approximate the posterior distributions using Markov Chain Monte Carlo sampling. The simulations reported were generated from 100,000 sampling iterations.

### D.2.2 Softmax computation

A technical challenge posed by certain datasets is the existence of incompatible perspective-world pairs for which there is no true utterance. When this is the case, the softmax in the Literal Listener cannot be computed directly, since this would involve division by zero. This is a known challenge of RSA modeling that can be addressed in a number of ways.<sup>2</sup> In the PRSA model that I propose, this problem arises at the level of the Pragmatic Speaker. The solution that I have adopted is to first sample a perspective, then condition the set of utterances on that perspective.

---

<sup>2</sup>See Bergen et al. (2016) for discussion of four approaches to this problem, though the joint inference involved in the PRSA model poses additional challenges.

## D.3 Model limitations

Models developed in the RSA framework have a number of limitations. I summarize them in this section.

### D.3.1 Manual specification of dataset components

One of the limitations of the RSA framework at present is that the components of the dataset must be manually specified. This is both labor-intensive, limiting the applicability of the RSA approach to larger datasets, and a possible source of error.

#### D.3.1.1 Utterance set

One of the drawbacks of the RSA framework is that the predictions of the model are sensitive to the alternatives included in the utterance set. Omitting a relevant competitor can change the model’s calculations in non-obvious ways, because it affects not just the probabilities of the utterances and the worlds they describe, but also the competitive interactions between utterances.

#### D.3.1.2 World set

Omitting relevant possible worlds can also affect the model’s predictions. However, this is less of a concern, since it is more likely that a relevant competitor utterance will be overlooked than that a relevant possible world will be omitted.

The set of possible worlds can be derived from a set of propositions by enumerating all possible combinations of their truth values. In general, it is only necessary to encode the propositions relevant to the set of utterances under consideration. If there is no utterance under consideration whose truth depends on a proposition, then the proposition can safely be omitted from the representation of worlds. To see why, consider a world set in which worlds only encode one proposition. This world set will contain two worlds, which we can refer to as True World and False World. If the world set is enriched to encode a second proposition, we obtain four worlds: True-True, True-False, False-True, and False-False. If we now apply an update based on an utterance that conditions only on the first field, True-True and True-False will have equal posterior probability, as will False-True and False-False.

If we apply the update first and then enrich the world set, the result is the same: True's posterior probability will be evenly distributed between True-True and True-False; False's posterior probability will be evenly distributed between False-False and False-True. Thus, as long as no conversational update is conditional on an omitted proposition, the world set can always be enriched with the proposition at a later time if it becomes relevant.

#### D.3.1.3 Perspective set

Another limitation is that the perspective set considered might not be an accurate reflection of the perspectives at play in a discourse. There is much that is currently unknown about the set of available perspectives. One possibility is that all discourse-given individual contribute a perspective that should be included in the perspective set. Another possibility is that perspectives beyond the speaker and listener are introduced by linguistic operators; for instance, Roberts (2015) proposes that attitude verbs introduce the perspectives of their subjects into the Common Ground. Future work on the availability of perspectives could shed light on this issue.

#### D.3.1.4 Lexical semantics and the interpretation function

The lexical semantics of the utterances and their interpretation function must also be manually specified. Thus, any predictions generated by an RSA model rest on assumptions about the analysis of the semantics of the utterances.

On the one hand, this makes RSA models useful for comparing different semantic analyses; for instance, I show that the anti-perspectival implication of *go* can arise via pragmatic competition, as suggested by previous theorists (Wilkins and Hill, 1995; Sudo, 2018).

On the other hand, it limits the applicability of the RSA model to larger datasets, where it would be prohibitively time-consuming to specify a lexical semantics for every possible competitor. There is some work on hybrid machine learning-RSA approaches in which the semantic representations are learned rather than specified that seeks to address this issue (Monroe et al., 2017; Cohn-Gordon et al., 2018; Monroe, 2018; Nie et al., 2020).

### **D.3.2 Limited incorporation of discourse context**

The PRSA model that I have proposed is capable of taking into account only a very limited amount of information about the discourse context: the prior probability distribution over possible worlds. This models one component of the Common Ground: the shared public beliefs of the conversation participants. However, it does not capture other components of the discourse context, such as topicality, Questions Under Discussion, or even which individuals and entities are discourse-given. In order to handle the full range of factors that may influence the selection and identification of perspectives, the current framework would have to be enriched with a richer representation of the discourse context.

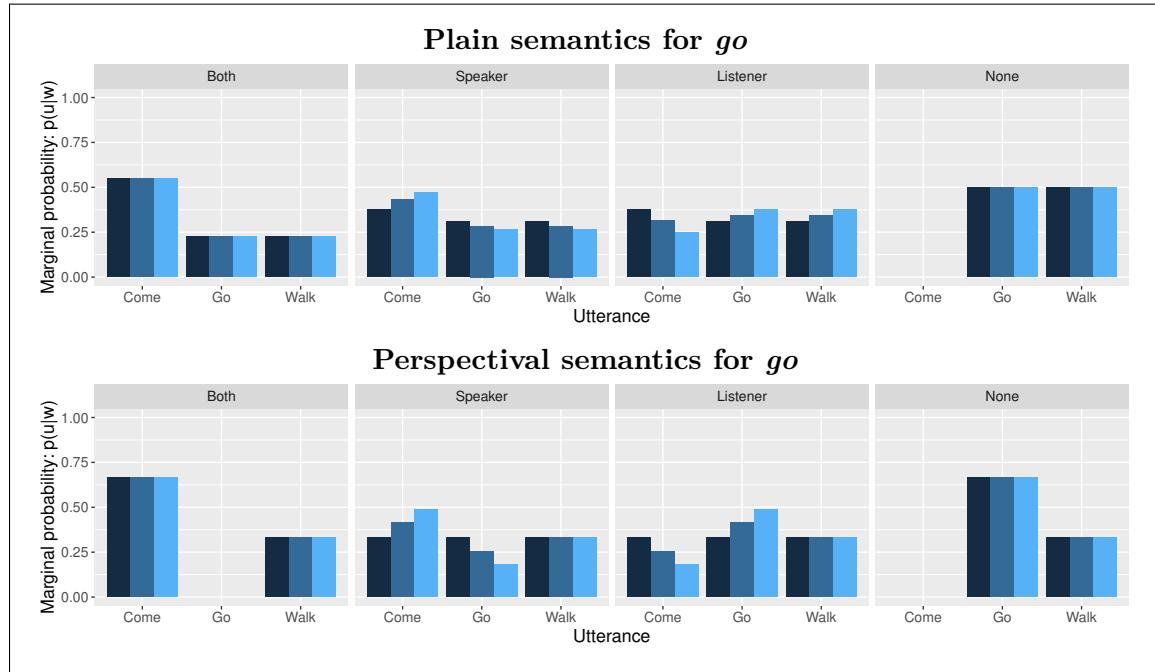
## APPENDIX E

### ADDITIONAL RSA FINDINGS

#### E.1 Comparing implementations of the Pragmatic Speaker

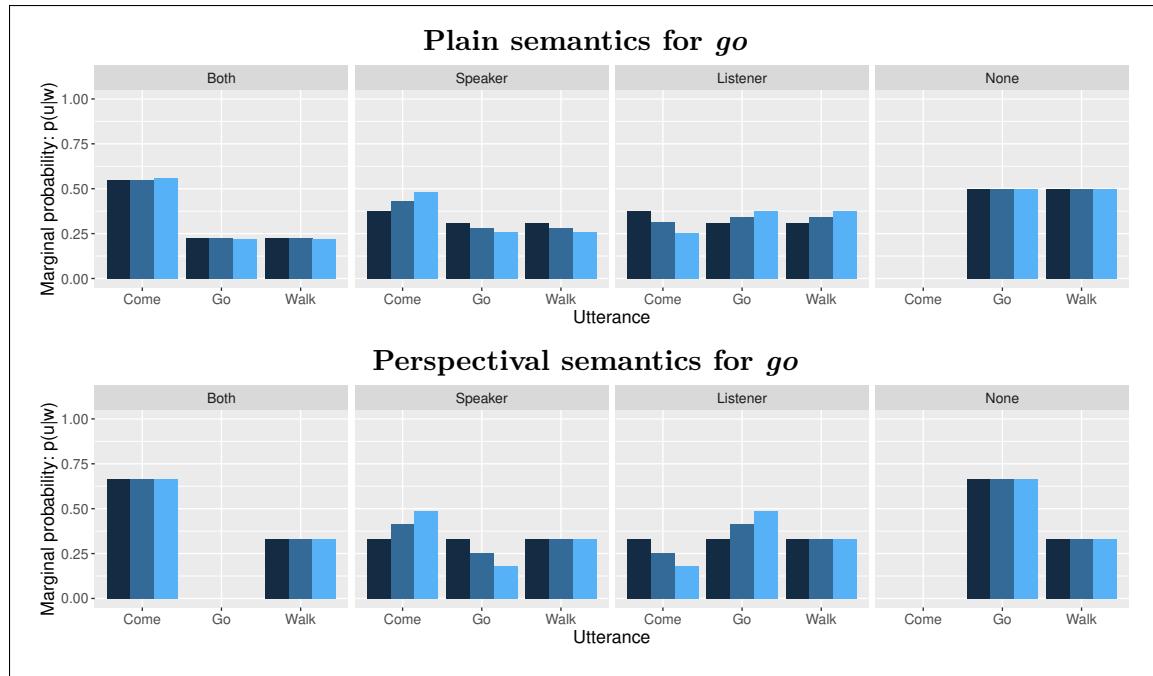
The perspective cost can be implemented at various levels in the PRSA production model. One possibility is to include it in the second-level Pragmatic Speaker but not the first-level Pragmatic Speaker. These predictions are shown in Figure E.1. This means assuming that there is a perspective cost, but that the speaker does not think the listener is aware of it.

**Figure E.1.** Second-level Pragmatic Speaker marginal posterior probabilities of utterances, mover = Thelma and perspective cost at second level (cost parameters = 0, 0.5, 1 from left to right)

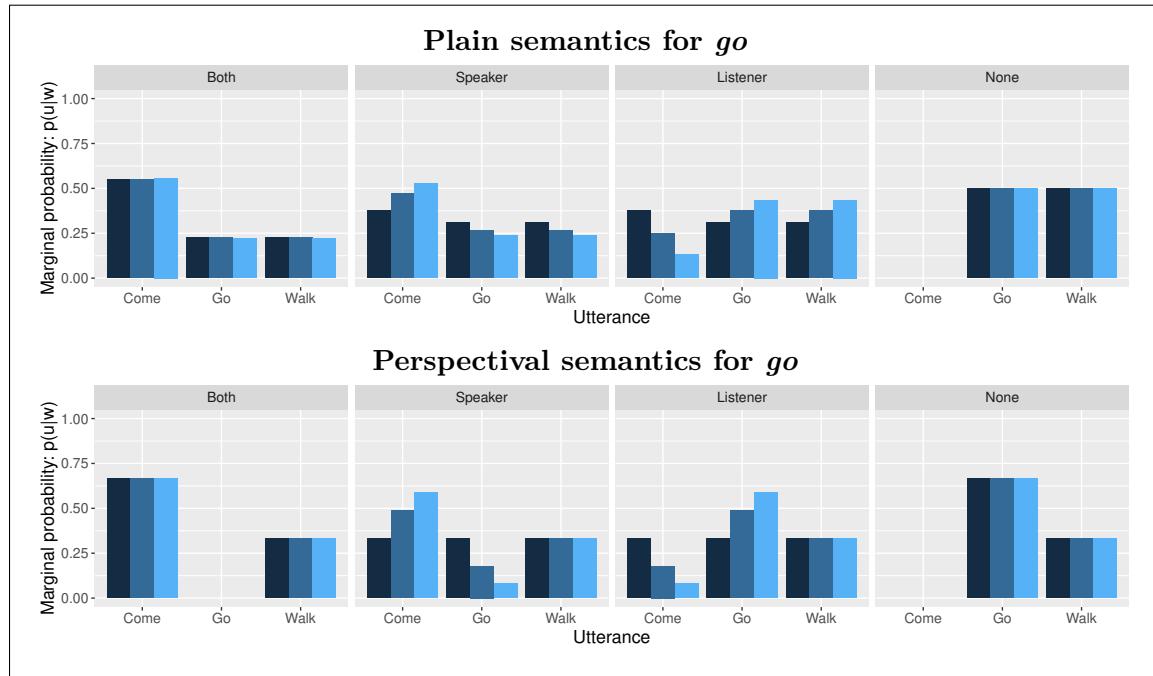


Another possibility is to implement it in the first-level Pragmatic Speaker, which means that the speaker believes that the listener is aware of the perspective cost. The predictions generated by this model are shown in Figure E.2.

**Figure E.2.** Second-level Pragmatic Speaker marginal posterior probabilities of utterances, mover = Thelma and perspective cost at first level (cost parameters = 0, 0.5, 1 from left to right)



**Figure E.3.** Second-level Pragmatic Speaker marginal posterior probabilities of utterances, mover = Thelma and perspective cost at both levels (cost parameters = 0, 0.5, 1 from left to right)



A third possibility is to implement the cost in both places. The predictions of this model are shown in Figure E.3.

As these figures show, the level that perspective cost is implemented in does not have a large impact on the predictions of the model. Introducing it at the first level weakens the effect of the perspective cost a little, while implementing it at both levels strengthens it.

## E.2 Pragmatic Speaker predictions given non-uniform priors over worlds

Manipulating the prior distribution over worlds does not change the predictions of the Pragmatic Speaker for sentences involving Thelma. To demonstrate this, I will walk through the calculations for the Literal Listener and Pragmatic Speaker.

Consider the following prior distribution over worlds:

**Figure E.4.** Example non-uniform priors over worlds

World	Loc.	Lucy	Loc.	Sam	Mover	Prior
$w_1$	A		A		Lucy	0.0625
$w_2$	A		A		Thelma	0.0625
$w_3$	A		A		Sam	0.0625
$w_4$	N		N		Thelma	0.0625
$w_5$	A		N		Lucy	0.125
$w_6$	A		N		Thelma	0.125
$w_7$	N		A		Thelma	0.125
$w_8$	N		A		Sam	0.125

In other words, we have a non-uniform distribution favoring worlds in which the speaker and listener are not in the same location. Figure E.5 shows the Literal Listener calculation for sentences involving Thelma with a uniform distribution over worlds. Figure E.6 shows the Literal Listener calculation using the non-uniform prior distribution over worlds given in Figure E.4.

In the Pragmatic Speaker step, we use the calculations of the the Literal Listener to estimate the probability of a world given an utterance and a perspective. For a single perspective-utterance pair, we consider how likely the observed world is given this pair.

**Figure E.5.** Literal Listener calculation with uniform priors over worlds, plain semantics for *go*

**Utterance = Thelma is coming, perspective = Sam**

$$p(w_4|T \text{ is coming}, p = S) = \frac{p(w_4)}{(p(w_4) + p(w_6))} = 0.5$$

$$p(w_6|T \text{ is coming}, p = S) = \frac{p(w_6)}{(p(w_4) + p(w_6))} = 0.5$$

**Utterance = Thelma is coming, perspective = Lucy**

$$p(w_4|T \text{ is coming}, p = L) = \frac{p(w_4)}{(p(w_4) + p(w_7))} = 0.5$$

$$p(w_7|T \text{ is coming}, p = L) = \frac{p(w_7)}{(p(w_4) + p(w_7))} = 0.5$$

**Utterance = Thelma is walking, perspective = Sam**

$$p(w_2|T \text{ is walking}, p = S) = \frac{p(w_2)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.25$$

$$p(w_4|T \text{ is walking}, p = S) = \frac{p(w_4)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.25$$

$$p(w_6|T \text{ is walking}, p = S) = \frac{p(w_6)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.25$$

$$p(w_7|T \text{ is walking}, p = S) = \frac{p(w_7)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.25$$

**Utterance = Thelma is walking, perspective = Lucy**

$$p(w_2|T \text{ is walking}, p = L) = \frac{p(w_2)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.25$$

$$p(w_4|T \text{ is walking}, p = L) = \frac{p(w_4)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.25$$

$$p(w_6|T \text{ is walking}, p = L) = \frac{p(w_6)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.25$$

$$p(w_7|T \text{ is walking}, p = L) = \frac{p(w_7)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.25$$

**Utterance = Thelma is going, perspective = Sam**

$$p(w_2|T \text{ is walking}, p = S) = \frac{p(w_2)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.25$$

$$p(w_4|T \text{ is walking}, p = S) = \frac{p(w_4)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.25$$

$$p(w_6|T \text{ is walking}, p = S) = \frac{p(w_6)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.25$$

$$p(w_7|T \text{ is walking}, p = S) = \frac{p(w_7)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.25$$

**Utterance = Thelma is going, perspective = Lucy**

$$p(w_2|T \text{ is walking}, p = L) = \frac{p(w_2)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.25$$

$$p(w_4|T \text{ is walking}, p = L) = \frac{p(w_4)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.25$$

$$p(w_6|T \text{ is walking}, p = L) = \frac{p(w_6)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.25$$

$$p(w_7|T \text{ is walking}, p = L) = \frac{p(w_7)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.25$$

In order to understand whether the non-uniform distribution over worlds impacts the results of the Pragmatic Speaker, then, we need only compare between the output of the Literal Listener shown in Figures E.6 and E.5, holding the world constant.

**Figure E.6.** Literal Listener calculation with non-uniform priors over worlds, plain semantics for *go*

**Utterance = Thelma is coming, perspective = Sam**

$$p(w_4|T \text{ is coming}, p = S) = \frac{p(w_4)}{(p(w_4) + p(w_6))} = 0.33$$

$$p(w_6|T \text{ is coming}, p = S) = \frac{p(w_6)}{(p(w_4) + p(w_6))} = 0.66$$

**Utterance = Thelma is coming, perspective = Lucy**

$$p(w_4|T \text{ is coming}, p = L) = \frac{p(w_4)}{(p(w_4) + p(w_7))} = 0.33$$

$$p(w_7|T \text{ is coming}, p = L) = \frac{p(w_7)}{(p(w_4) + p(w_7))} = 0.66$$

**Utterance = Thelma is walking, perspective = Sam**

$$p(w_2|T \text{ is walking}, p = S) = \frac{p(w_2)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.167$$

$$p(w_4|T \text{ is walking}, p = S) = \frac{p(w_4)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.167$$

$$p(w_6|T \text{ is walking}, p = S) = \frac{p(w_6)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.33$$

$$p(w_7|T \text{ is walking}, p = S) = \frac{p(w_7)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.33$$

**Utterance = Thelma is walking, perspective = Lucy**

$$p(w_2|T \text{ is walking}, p = L) = \frac{p(w_2)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.167$$

$$p(w_4|T \text{ is walking}, p = L) = \frac{p(w_4)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.167$$

$$p(w_6|T \text{ is walking}, p = L) = \frac{p(w_6)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.33$$

$$p(w_7|T \text{ is walking}, p = L) = \frac{p(w_7)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.33$$

**Utterance = Thelma is going, perspective = Sam**

$$p(w_2|T \text{ is walking}, p = S) = \frac{p(w_2)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.167$$

$$p(w_4|T \text{ is walking}, p = S) = \frac{p(w_4)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.167$$

$$p(w_6|T \text{ is walking}, p = S) = \frac{p(w_6)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.33$$

$$p(w_7|T \text{ is walking}, p = S) = \frac{p(w_7)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.33$$

**Utterance = Thelma is going, perspective = Lucy**

$$p(w_2|T \text{ is walking}, p = L) = \frac{p(w_2)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.167$$

$$p(w_4|T \text{ is walking}, p = L) = \frac{p(w_4)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.167$$

$$p(w_6|T \text{ is walking}, p = L) = \frac{p(w_6)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.33$$

$$p(w_7|T \text{ is walking}, p = L) = \frac{p(w_7)}{(p(w_2) + p(w_4) + p(w_6) + p(w_7))} = 0.33$$

We find that for any world, the ratio of its conditional likelihood for each utterance according

to the Literal Speaker is the same, regardless of the priors.

**Figure E.7.** Pragmatic Speaker calculation with uniform priors over worlds for sentences involving Thelma

$W_2$		
$p(\text{Thelma is coming}, S w_2) =$	$0 =$	$0$
$p(\text{Thelma is coming}, L w_2) =$	$0 =$	$0$
$p(\text{Thelma is walking}, S w_2) =$	$0.25p(u a)p(a) =$	$p(u a)p(a)$
$p(\text{Thelma is walking}, L w_2) =$	$0.25p(u a)p(a) =$	$p(u a)p(a)$
$p(\text{Thelma is going}, S w_2) =$	$0.25p(u a)p(a) =$	$p(u a)p(a)$
$p(\text{Thelma is going}, L w_2) =$	$0.25p(u a)p(a) =$	$p(u a)p(a)$
$W_4$		
$p(\text{Thelma is coming}, S w_4) =$	$0.5p(u a)p(a) =$	$p(u a)p(a)$
$p(\text{Thelma is coming}, L w_4) =$	$0.5p(u a)p(a) =$	$p(u a)p(a)$
$p(\text{Thelma is walking}, S w_4) =$	$0.25p(u a)p(a) =$	$0.5p(u a)p(a)$
$p(\text{Thelma is walking}, L w_4) =$	$0.25p(u a)p(a) =$	$0.5p(u a)p(a)$
$p(\text{Thelma is going}, S w_4) =$	$0.25p(u a)p(a) =$	$0.5p(u a)p(a)$
$p(\text{Thelma is going}, L w_4) =$	$0.25p(u a)p(a) =$	$0.5p(u a)p(a)$
$W_6$		
$p(\text{Thelma is coming}, S w_6) =$	$0.5p(u a)p(a) =$	$p(u a)p(a)$
$p(\text{Thelma is coming}, L w_6) =$	$0 =$	$0$
$p(\text{Thelma is walking}, S w_6) =$	$0.25p(u a)p(a) =$	$0.5p(u a)p(a)$
$p(\text{Thelma is walking}, L w_6) =$	$0.25p(u a)p(a) =$	$0.5p(u a)p(a)$
$p(\text{Thelma is going}, S w_6) =$	$0.25p(u a)p(a) =$	$0.5p(u a)p(a)$
$p(\text{Thelma is going}, L w_6) =$	$0.25p(u a)p(a) =$	$0.5p(u a)p(a)$
$W_7$		
$p(\text{Thelma is coming}, S w_7) =$	$0 =$	$0$
$p(\text{Thelma is coming}, L w_7) =$	$0.5p(u a)p(a) =$	$p(u a)p(a)$
$p(\text{Thelma is walking}, S w_7) =$	$0.25p(u a)p(a) =$	$0.5p(u a)p(a)$
$p(\text{Thelma is walking}, L w_7) =$	$0.25p(u a)p(a) =$	$0.5p(u a)p(a)$
$p(\text{Thelma is going}, S w_7) =$	$0.25p(u a)p(a) =$	$0.5p(u a)p(a)$
$p(\text{Thelma is going}, L w_7) =$	$0.25p(u a)p(a) =$	$0.5p(u a)p(a)$

Thus, manipulating the priors in this way has no effect on the predictions of the Pragmatic Speaker, since the same likelihood ratio between competitor utterance-perspective pairs for a given world is preserved.

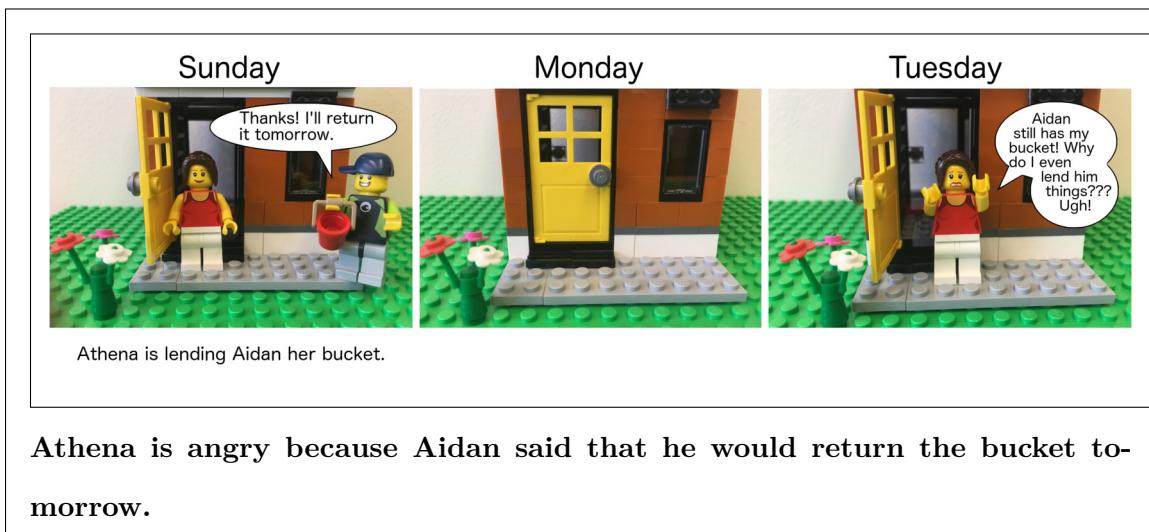
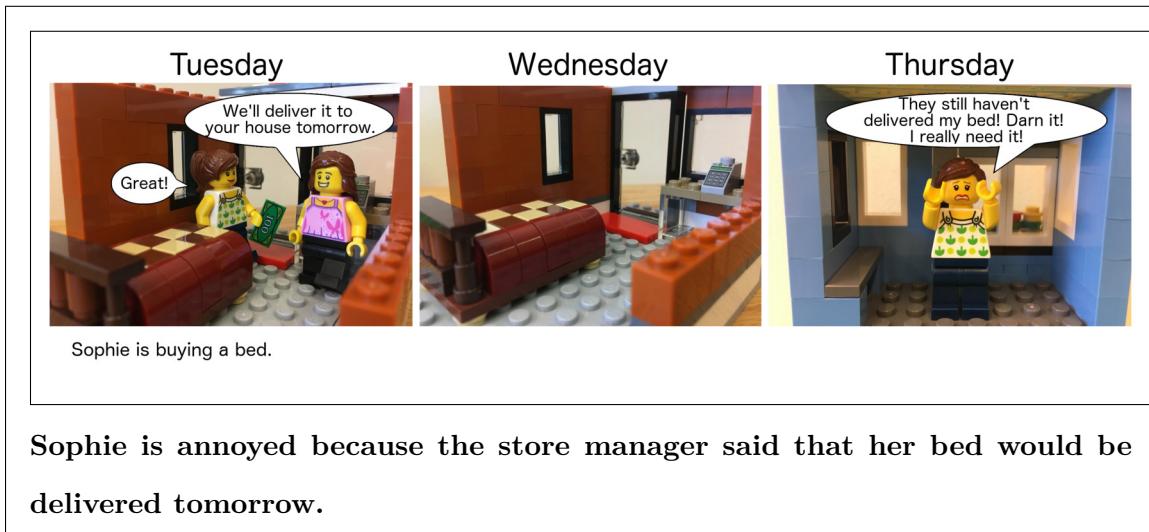
**Figure E.8.** Pragmatic Speaker calculation with non-uniform priors over worlds for sentences involving Thelma

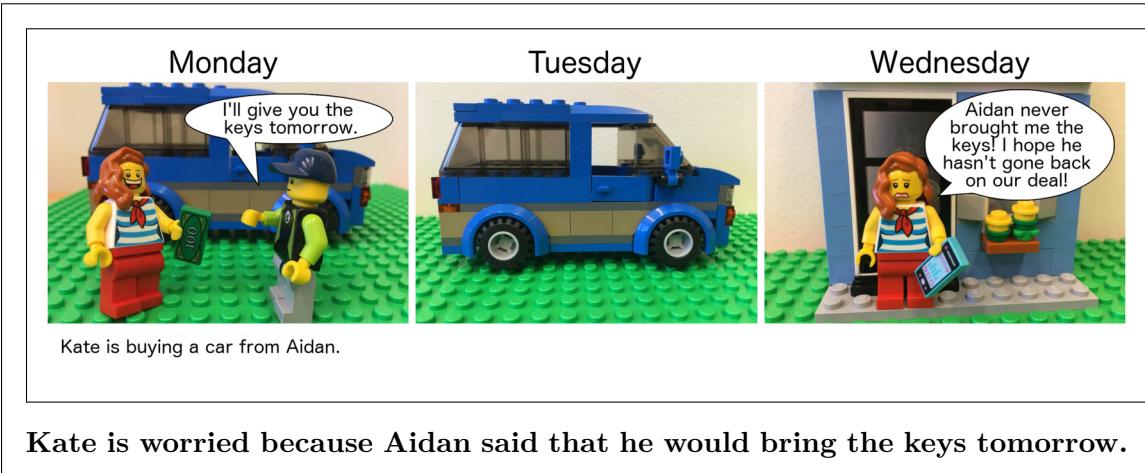
$W_2$		
$p(\text{Thelma is coming}, S w_2) =$	$0 =$	$0$
$p(\text{Thelma is coming}, L w_2) =$	$0 =$	$0$
$p(\text{Thelma is walking}, S w_2) =$	$0.167p(u a)p(a) =$	$p(u a)p(a)$
$p(\text{Thelma is walking}, L w_2) =$	$0.167p(u a)p(a) =$	$p(u a)p(a)$
$p(\text{Thelma is going}, S w_2) =$	$0.167p(u a)p(a) =$	$p(u a)p(a)$
$p(\text{Thelma is going}, L w_2) =$	$0.167p(u a)p(a) =$	$p(u a)p(a)$
$W_4$		
$p(\text{Thelma is coming}, S w_4) =$	$0.33p(u a)p(a) =$	$p(u a)p(a)$
$p(\text{Thelma is coming}, L w_4) =$	$0.33p(u a)p(a) =$	$p(u a)p(a)$
$p(\text{Thelma is walking}, S w_4) =$	$0.167p(u a)p(a) =$	$0.5p(u a)p(a)$
$p(\text{Thelma is walking}, L w_4) =$	$0.167p(u a)p(a) =$	$0.5p(u a)p(a)$
$p(\text{Thelma is going}, S w_4) =$	$0.167p(u a)p(a) =$	$0.5p(u a)p(a)$
$p(\text{Thelma is going}, L w_4) =$	$0.167p(u a)p(a) =$	$0.5p(u a)p(a)$
$W_6$		
$p(\text{Thelma is coming}, S w_6) =$	$0.66p(u a)p(a) =$	$p(u a)p(a)$
$p(\text{Thelma is coming}, L w_6) =$	$0 =$	$0$
$p(\text{Thelma is walking}, S w_6) =$	$0.33p(u a)p(a) =$	$0.5p(u a)p(a)$
$p(\text{Thelma is walking}, L w_6) =$	$0.33p(u a)p(a) =$	$0.5p(u a)p(a)$
$p(\text{Thelma is going}, S w_6) =$	$0.33p(u a)p(a) =$	$0.5p(u a)p(a)$
$p(\text{Thelma is going}, L w_6) =$	$0.33p(u a)p(a) =$	$0.5p(u a)p(a)$
$W_7$		
$p(\text{Thelma is coming}, S w_7) =$	$0 =$	$0$
$p(\text{Thelma is coming}, L w_7) =$	$0.66p(u a)p(a) =$	$p(u a)p(a)$
$p(\text{Thelma is walking}, S w_7) =$	$0.33p(u a)p(a) =$	$0.5p(u a)p(a)$
$p(\text{Thelma is walking}, L w_7) =$	$0.33p(u a)p(a) =$	$0.5p(u a)p(a)$
$p(\text{Thelma is going}, S w_7) =$	$0.33p(u a)p(a) =$	$0.5p(u a)p(a)$
$p(\text{Thelma is going}, L w_7) =$	$0.33p(u a)p(a) =$	$0.5p(u a)p(a)$

## APPENDIX F

### EXPERIMENT 1 ITEMS

#### F.1 Main items

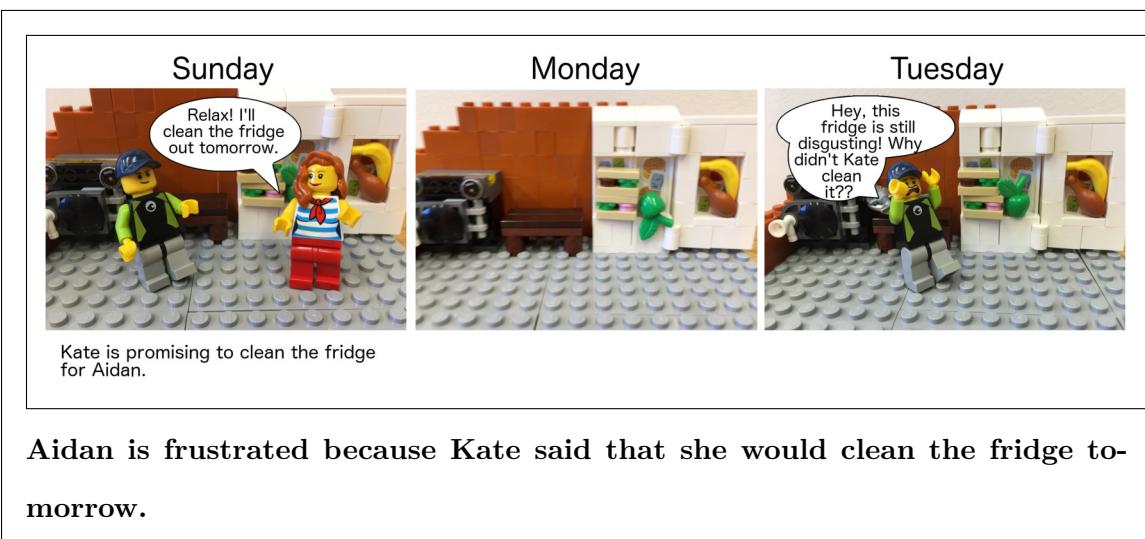


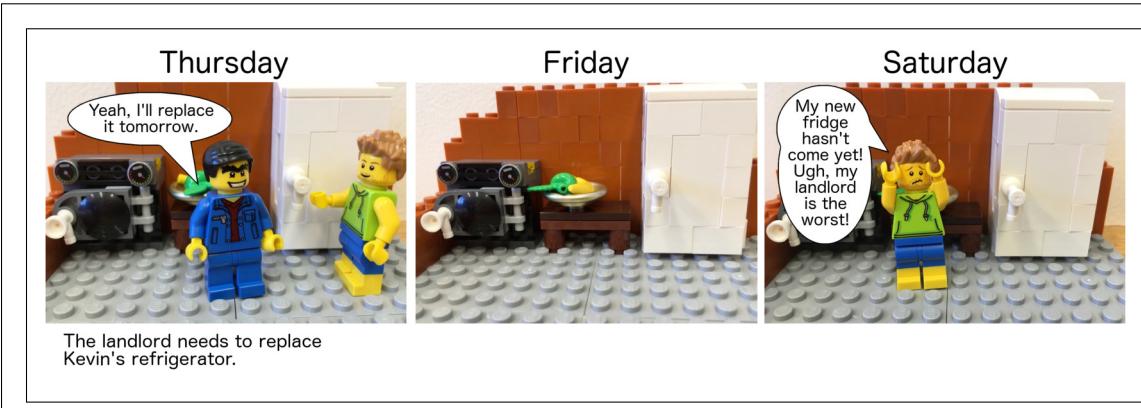


Kate is worried because Aidan said that he would bring the keys tomorrow.

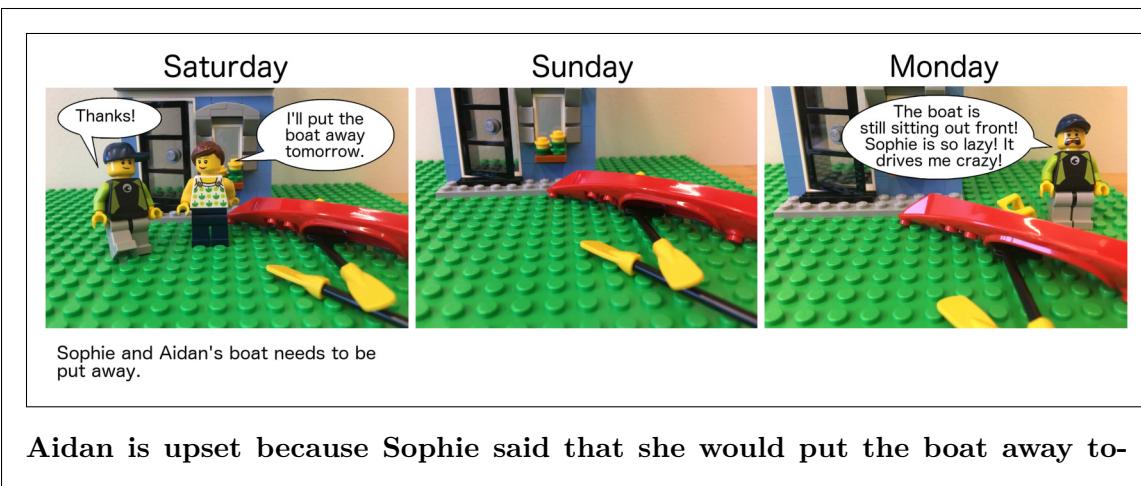


Kevin is annoyed because Sophie said she would bring the pan back tomorrow.

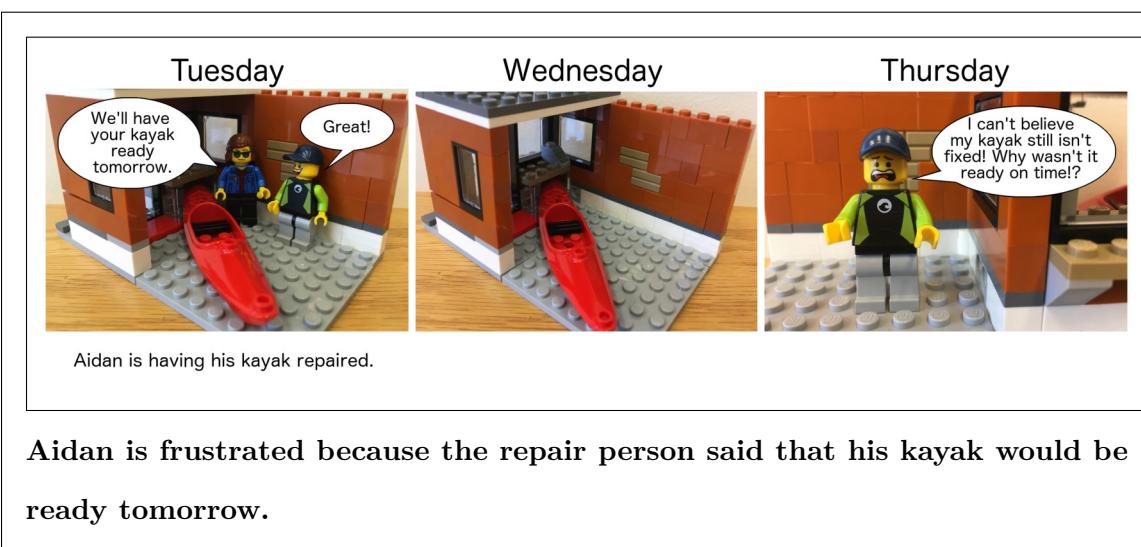




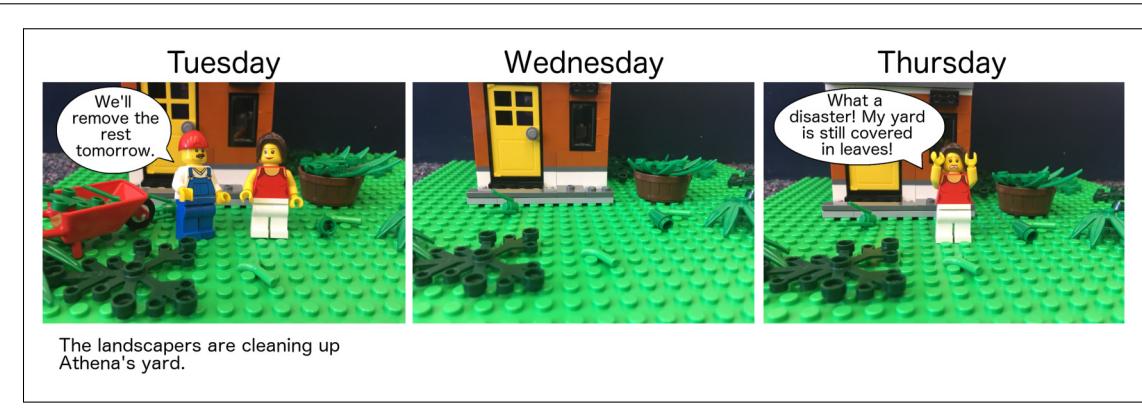
**Kevin is upset because his landlord said that he would replace his fridge tomorrow.**



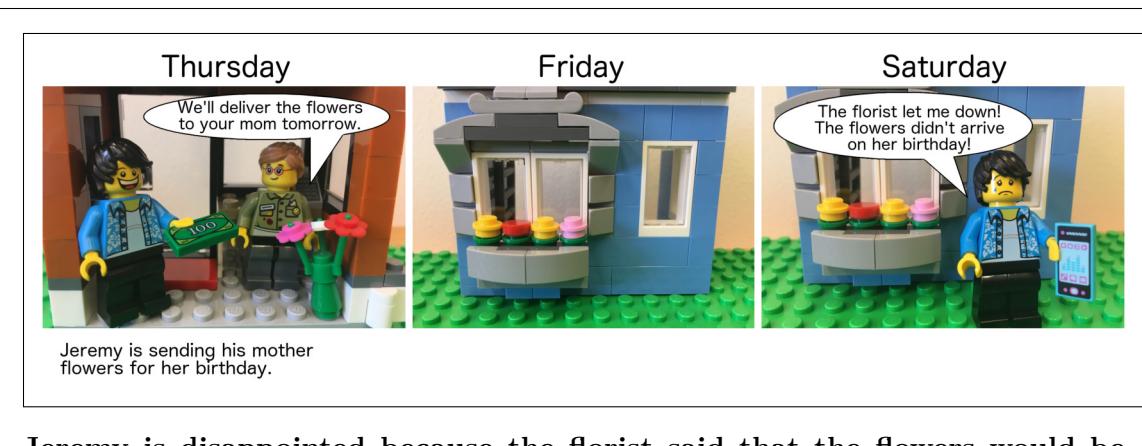
**Aidan is upset because Sophie said that she would put the boat away tomorrow.**



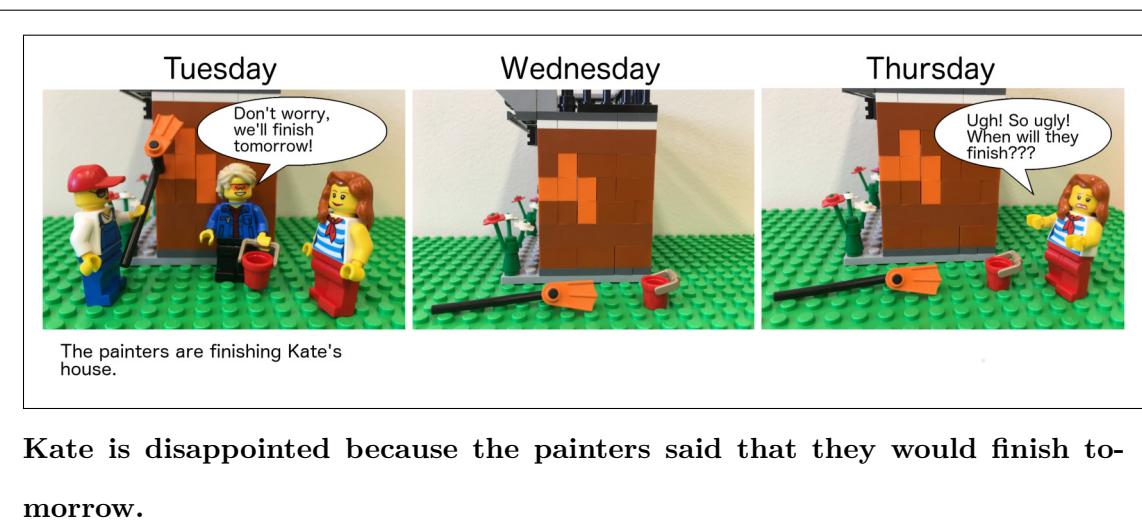
**Aidan is frustrated because the repair person said that his kayak would be ready tomorrow.**



Athena is frustrated because the landscapers said that they would remove the leaves tomorrow.



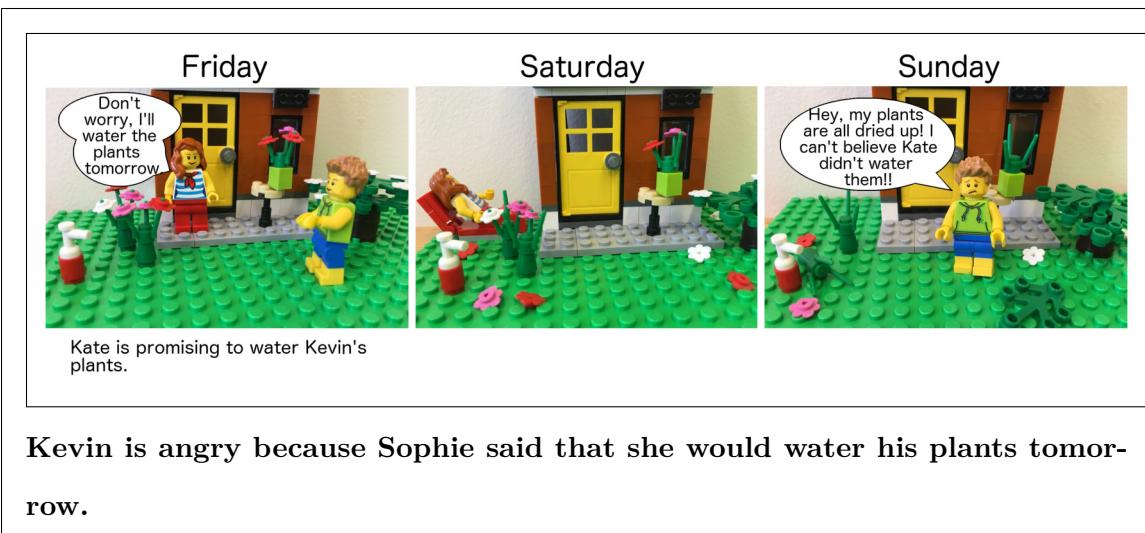
Jeremy is disappointed because the florist said that the flowers would be delivered tomorrow.



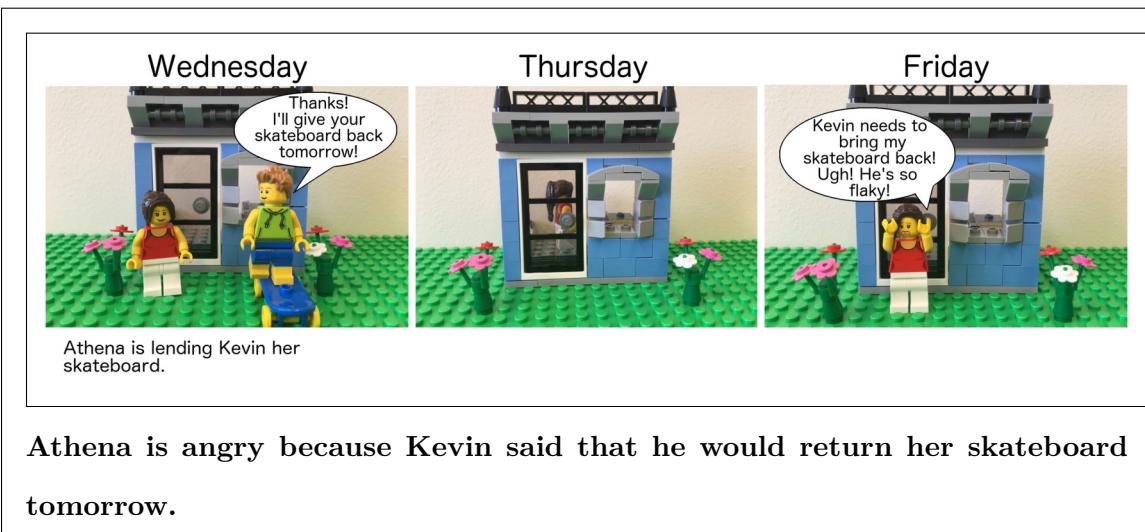
Kate is disappointed because the painters said that they would finish tomorrow.



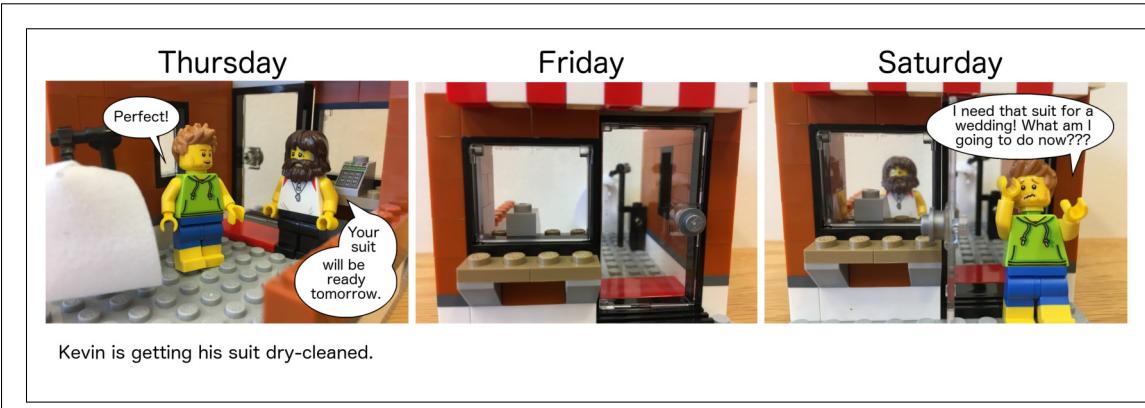
**Kevin is disappointed because the store manager said that they would have iPhones tomorrow.**



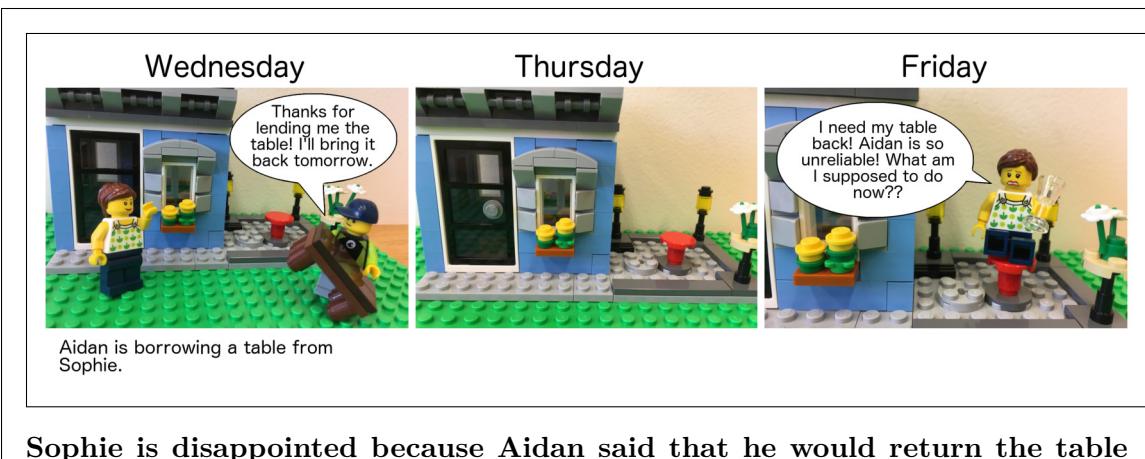
**Kevin is angry because Sophie said that she would water his plants tomorrow.**



**Athena is angry because Kevin said that he would return her skateboard tomorrow.**



**Kevin is angry because the dry-cleaners said that his suit would be ready tomorrow.**



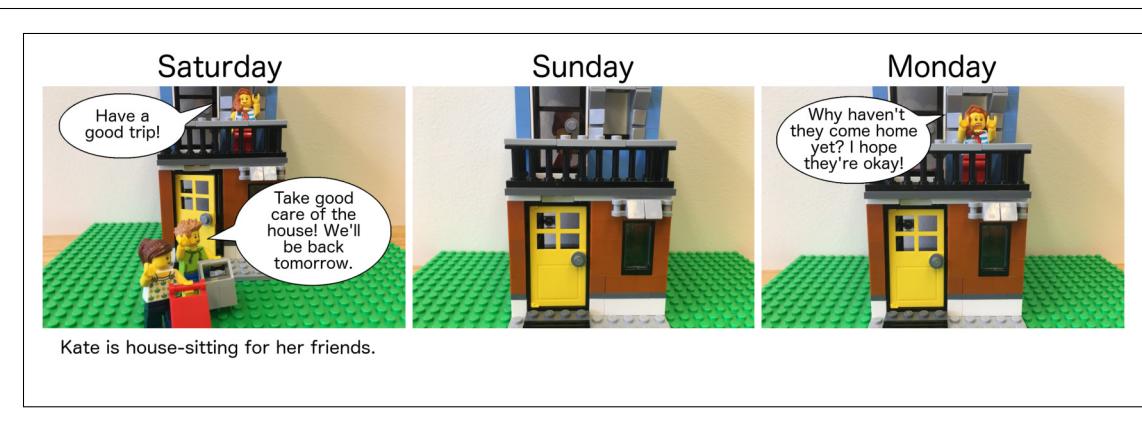
**Sophie is disappointed because Aidan said that he would return the table tomorrow.**



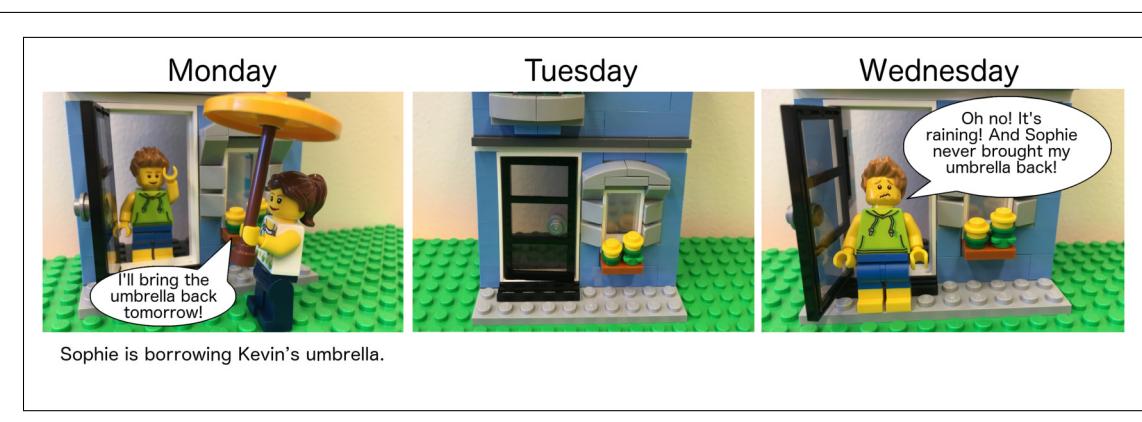
**Kate is angry because Jeremy said that he would buy toilet paper tomorrow.**



Aidan is disappointed because the contractors said that they would remove the tree tomorrow.

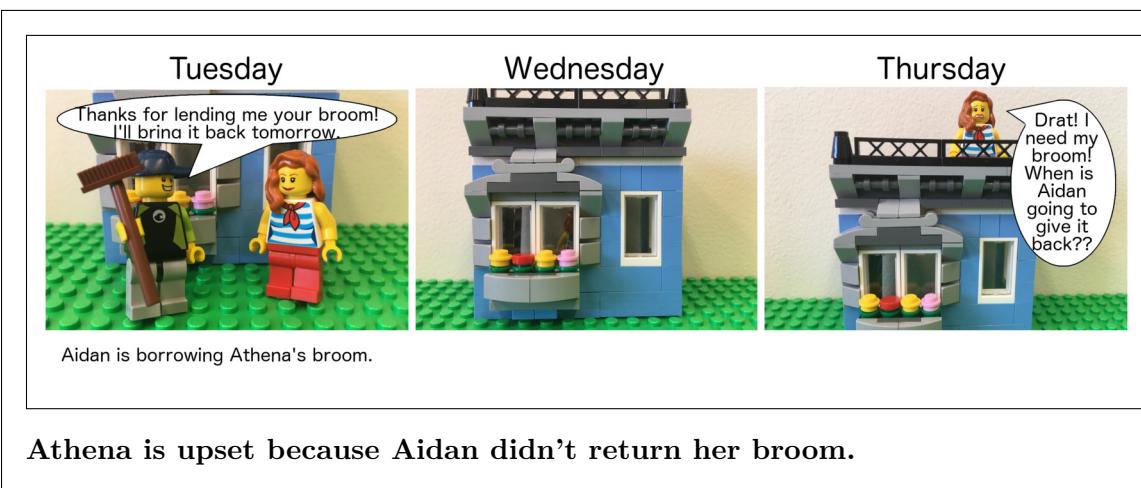
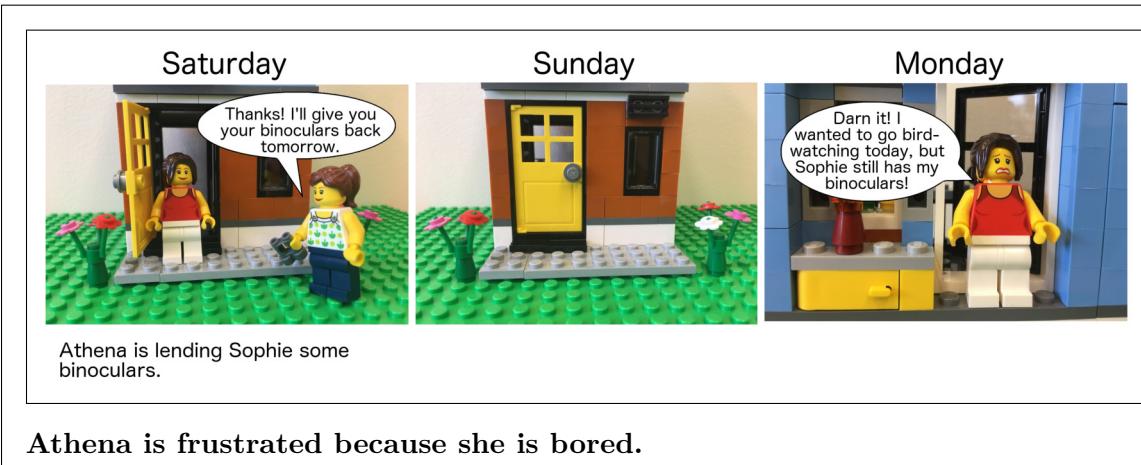
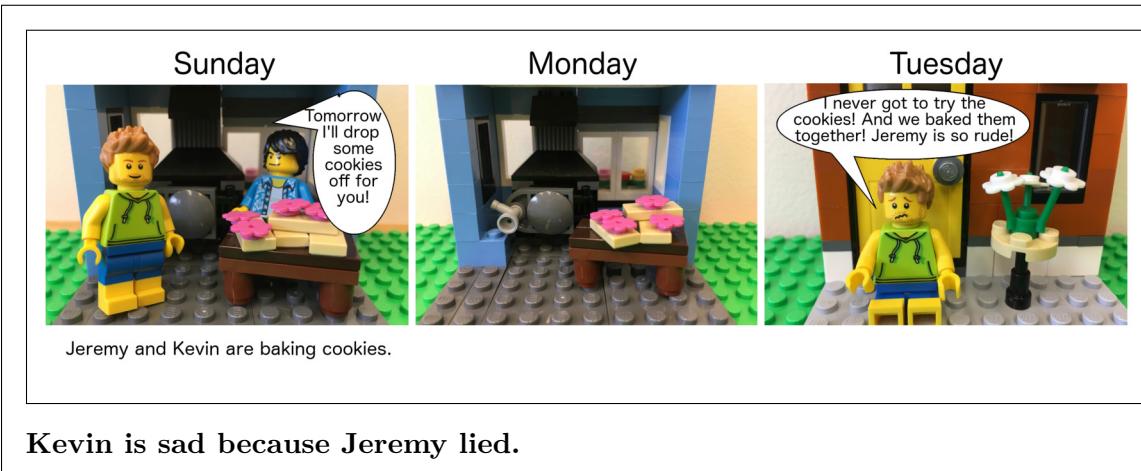


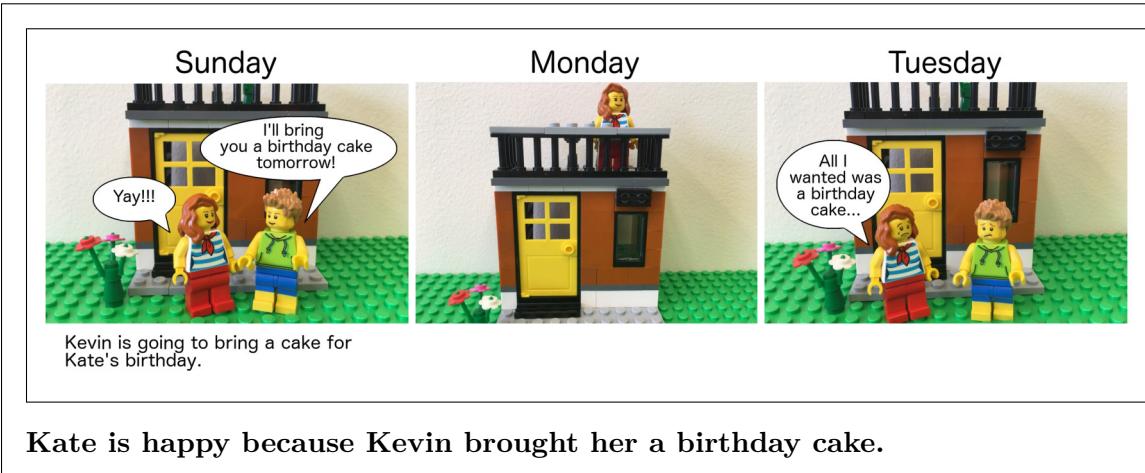
Kate is worried because her friends said that they would be back tomorrow.



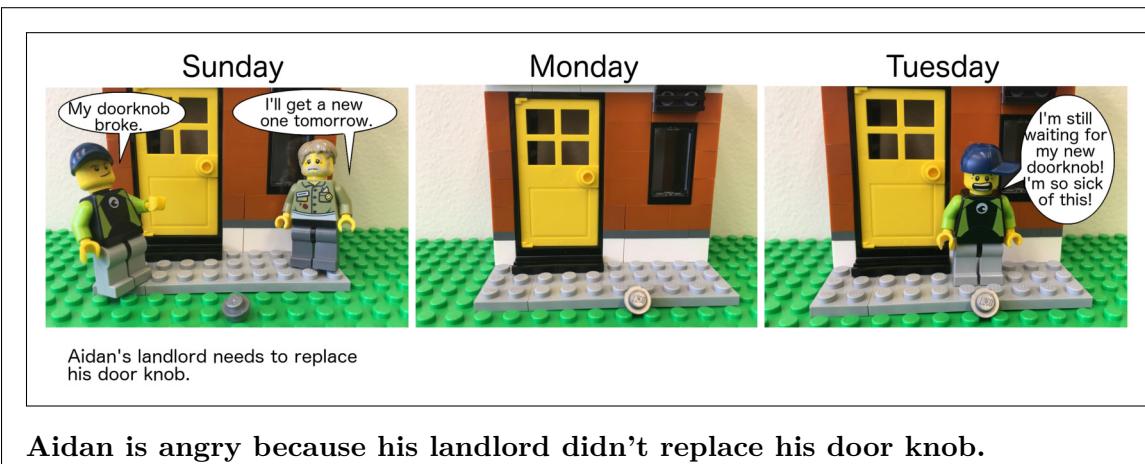
Kevin is upset because Sophie said that she would bring his umbrella back tomorrow.

## F.2 Fillers

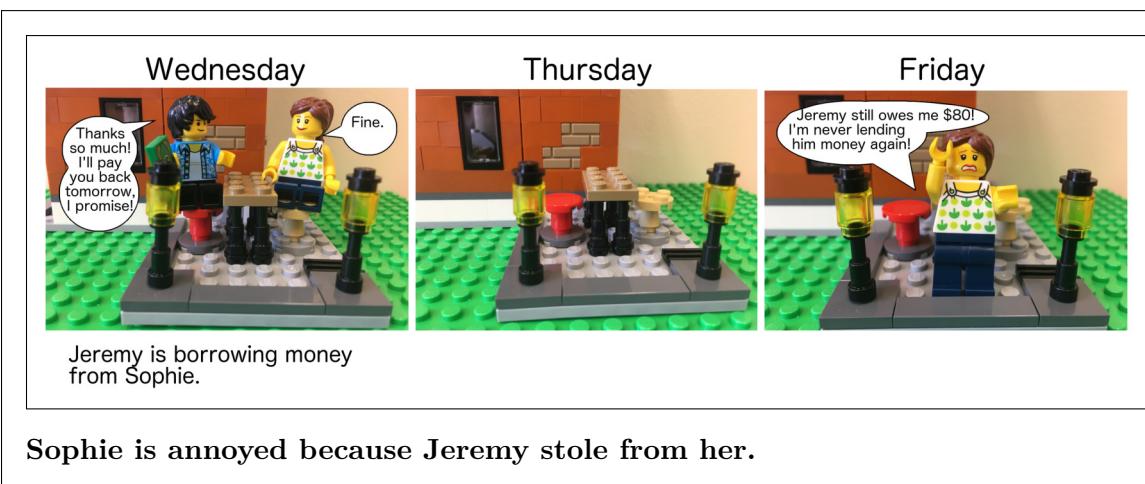




Kate is happy because Kevin brought her a birthday cake.



Aidan is angry because his landlord didn't replace his door knob.



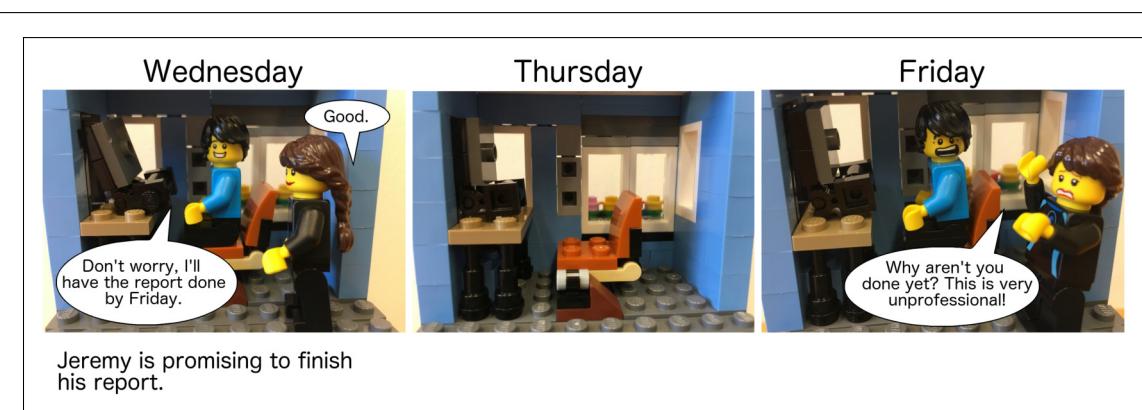
Sophie is annoyed because Jeremy stole from her.



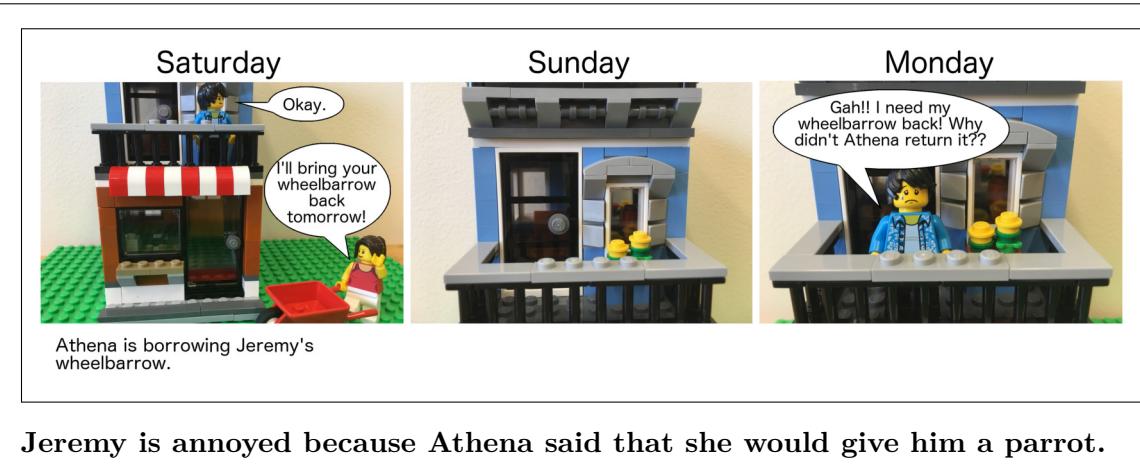
Kate is annoyed because Sophie said that she would bring crackers.



Jeremy is pleased because the contractors finished his railing.



Jeremy's boss is angry that he hasn't finished his report.

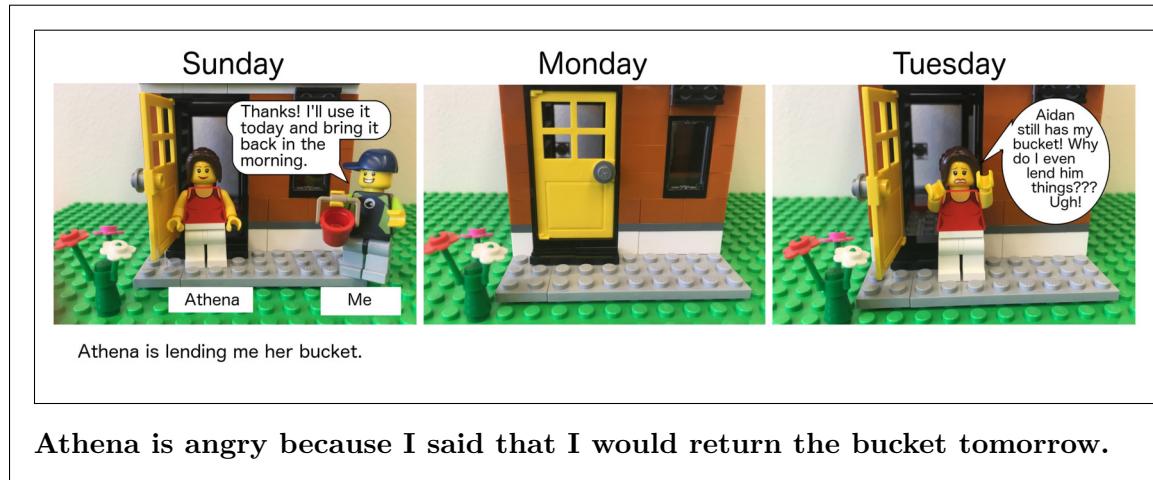
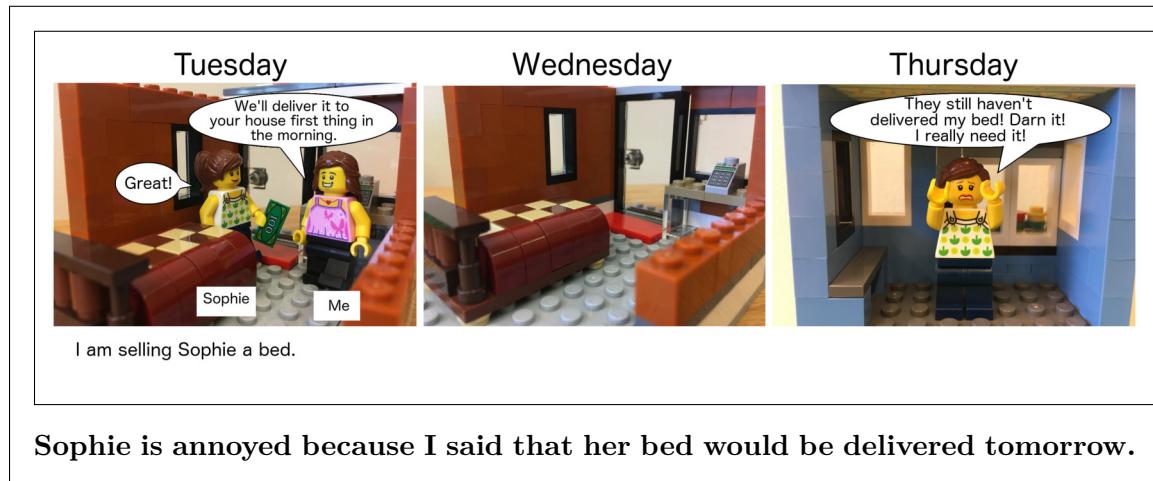


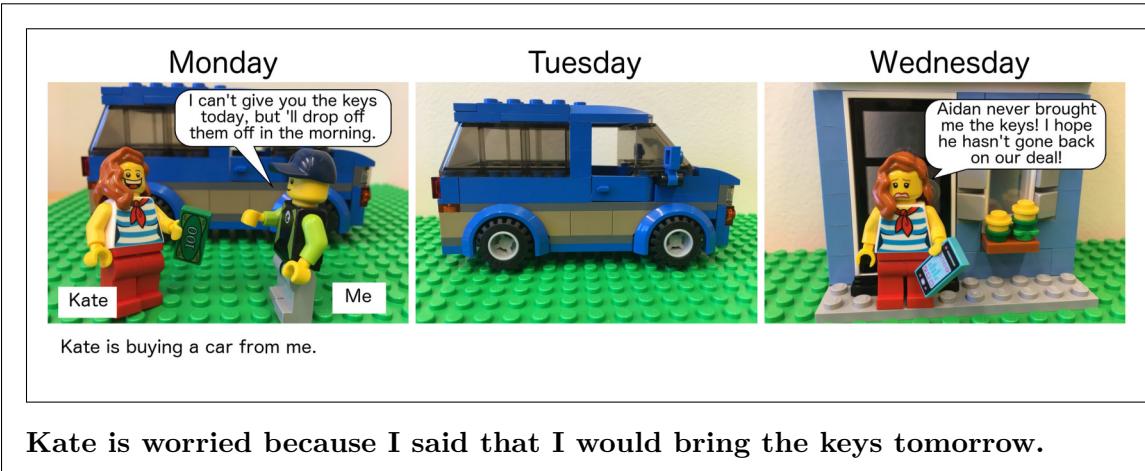
**Jeremy is annoyed because Athena said that she would give him a parrot.**

## APPENDIX G

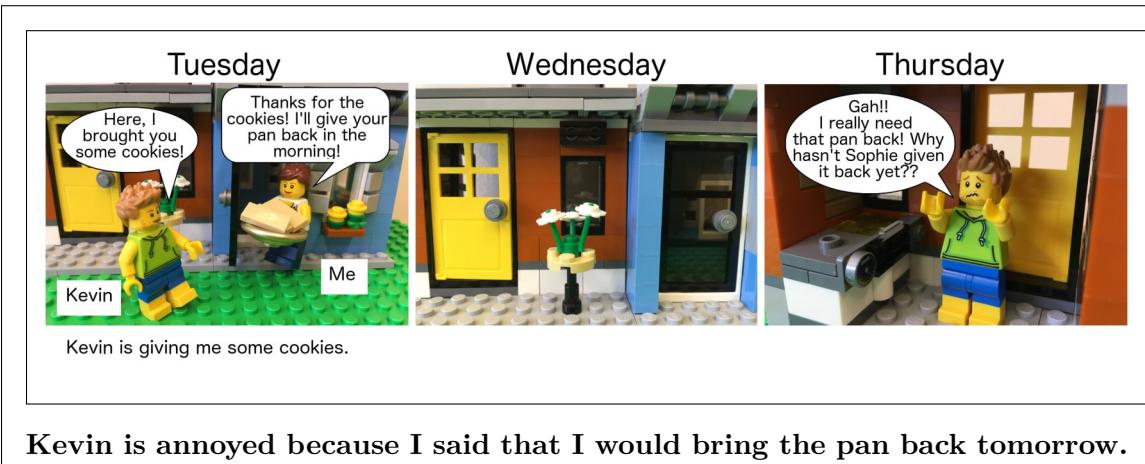
### EXPERIMENT 2 ITEMS

#### G.1 Main items

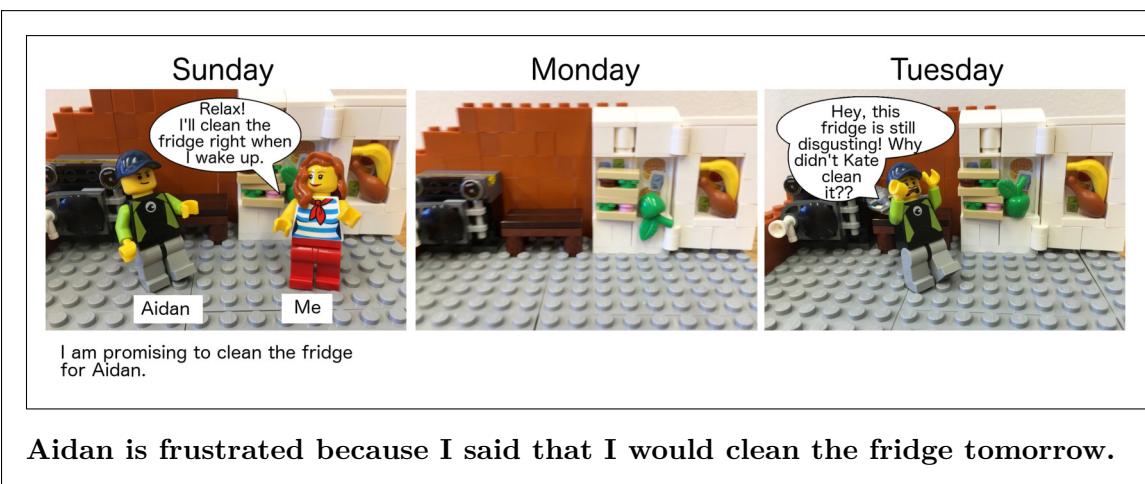




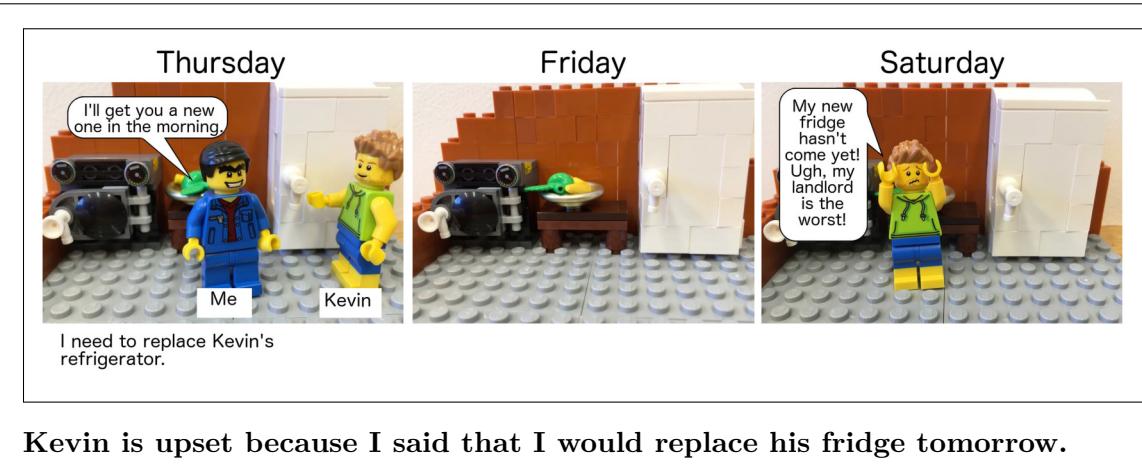
**Kate is worried because I said that I would bring the keys tomorrow.**



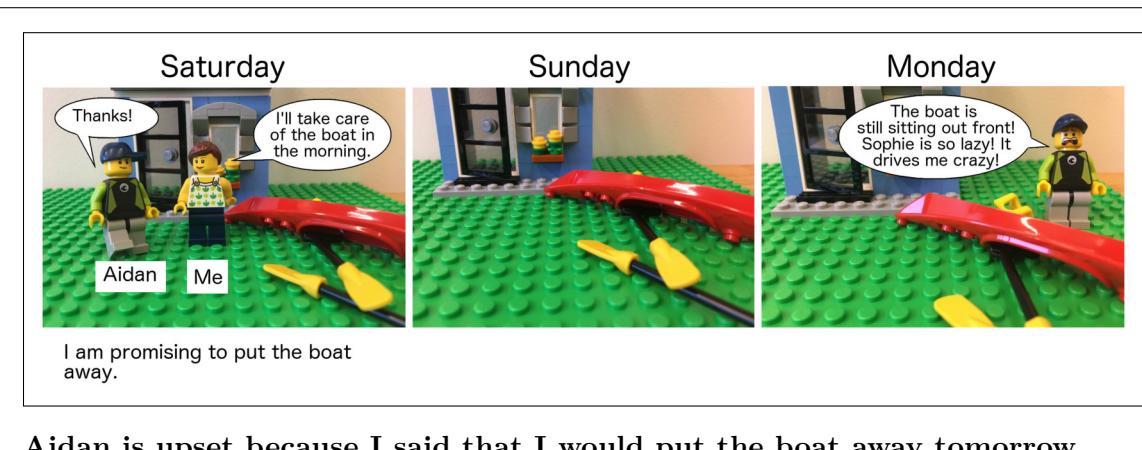
**Kevin is annoyed because I said that I would bring the pan back tomorrow.**



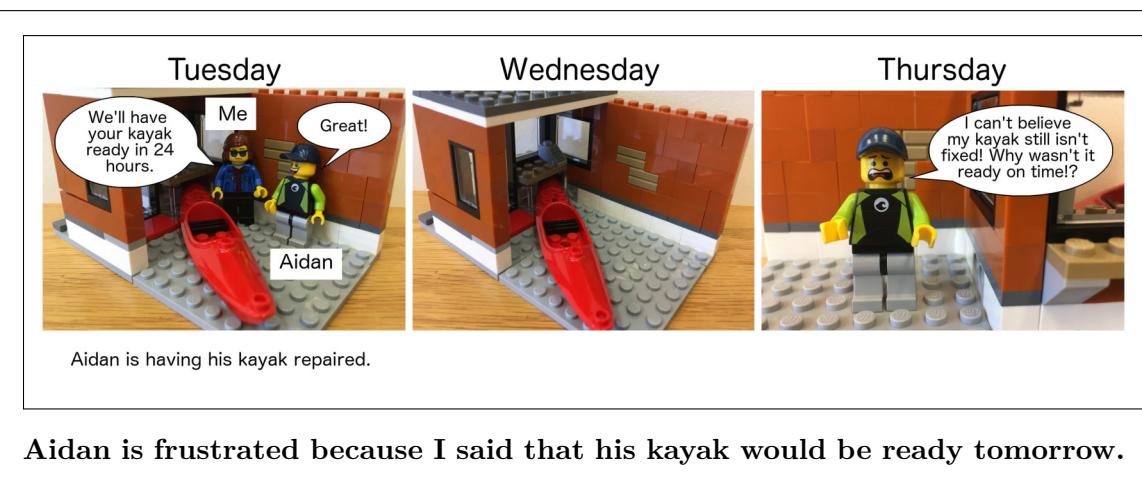
**Aidan is frustrated because I said that I would clean the fridge tomorrow.**



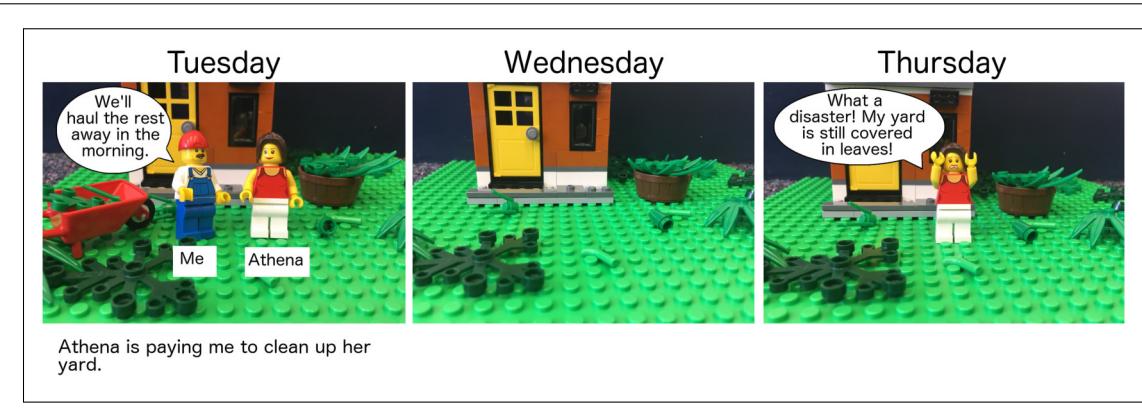
**Kevin is upset because I said that I would replace his fridge tomorrow.**



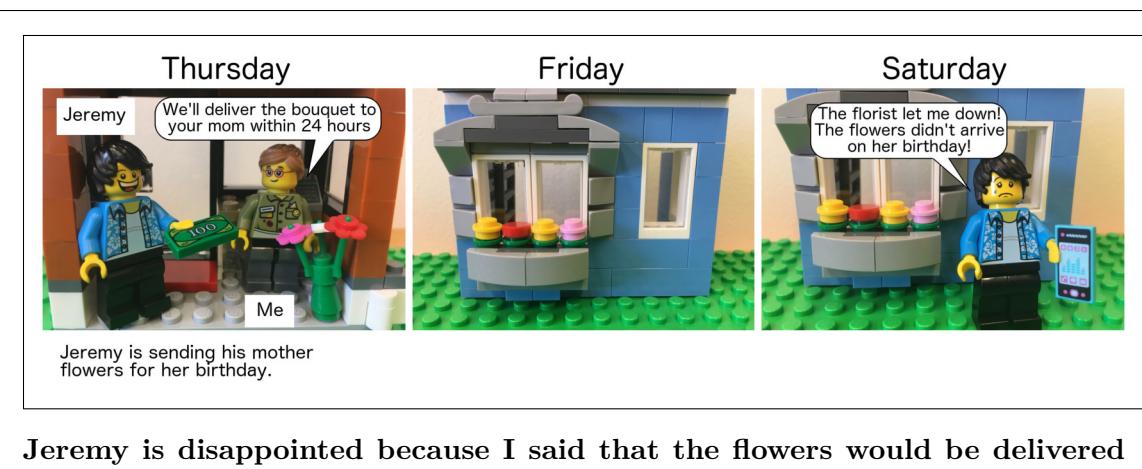
**Aidan is upset because I said that I would put the boat away tomorrow.**



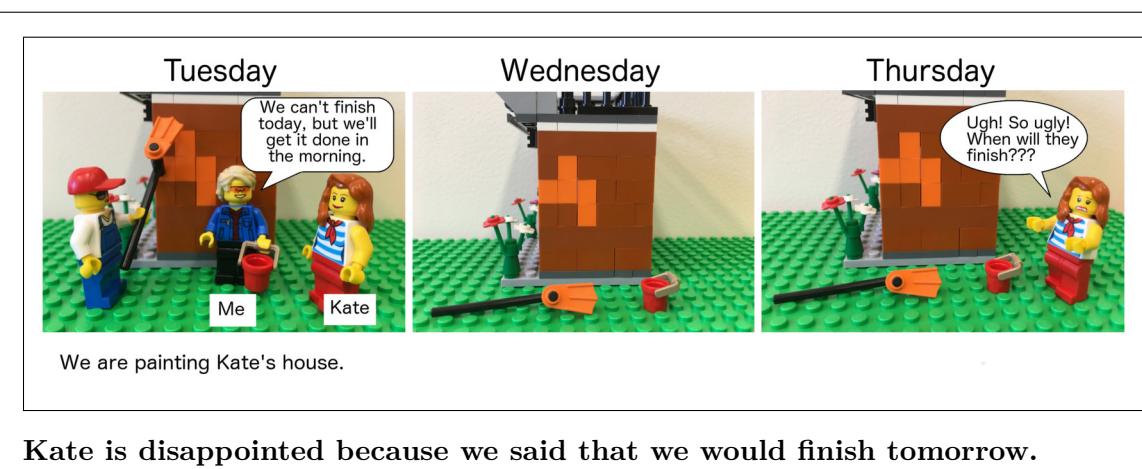
**Aidan is frustrated because I said that his kayak would be ready tomorrow.**



Athena is frustrated because we said that we would remove the leaves tomorrow.



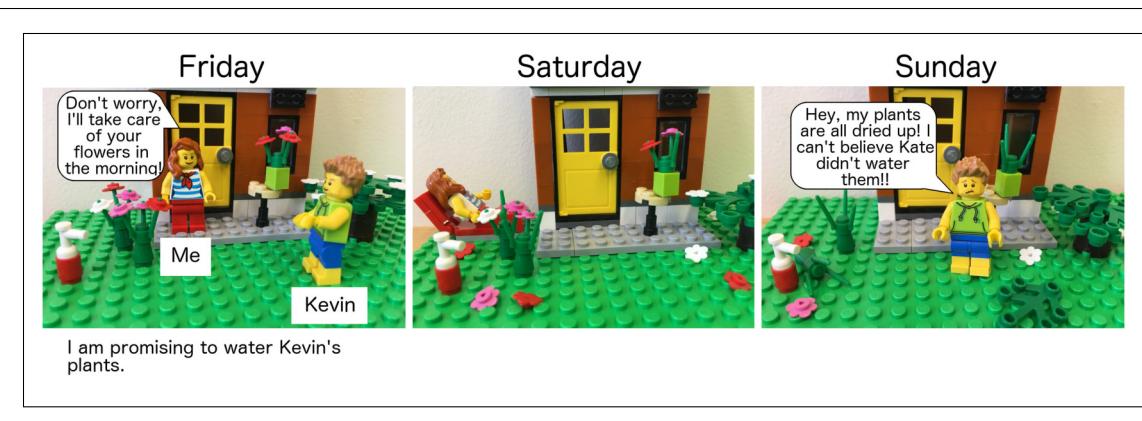
Jeremy is disappointed because I said that the flowers would be delivered tomorrow.



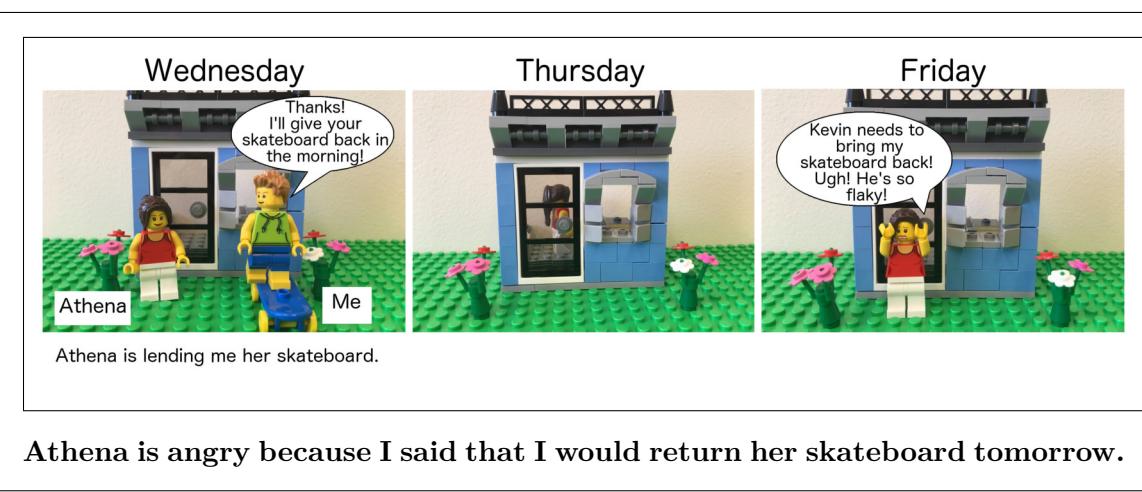
Kate is disappointed because we said that we would finish tomorrow.



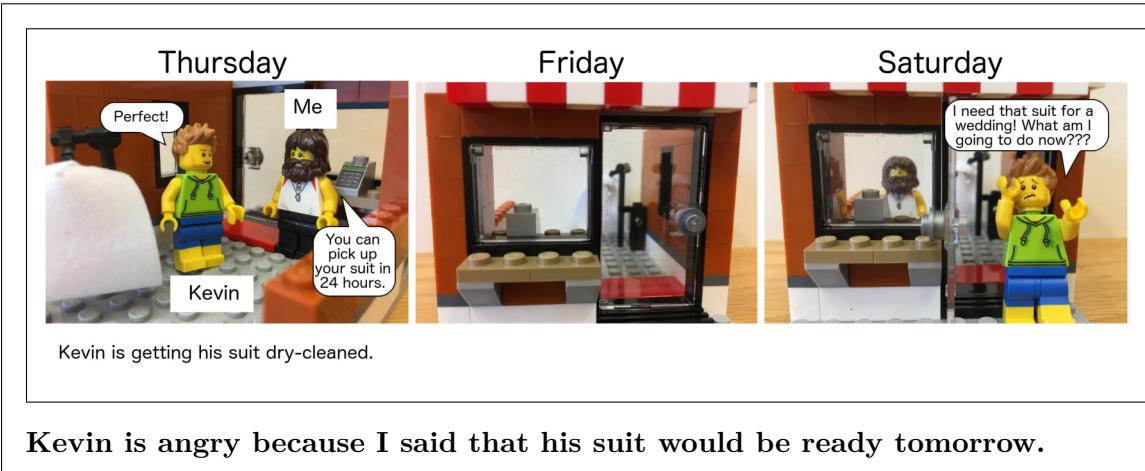
**Kevin is disappointed because I said that the store would have iPhones tomorrow.**



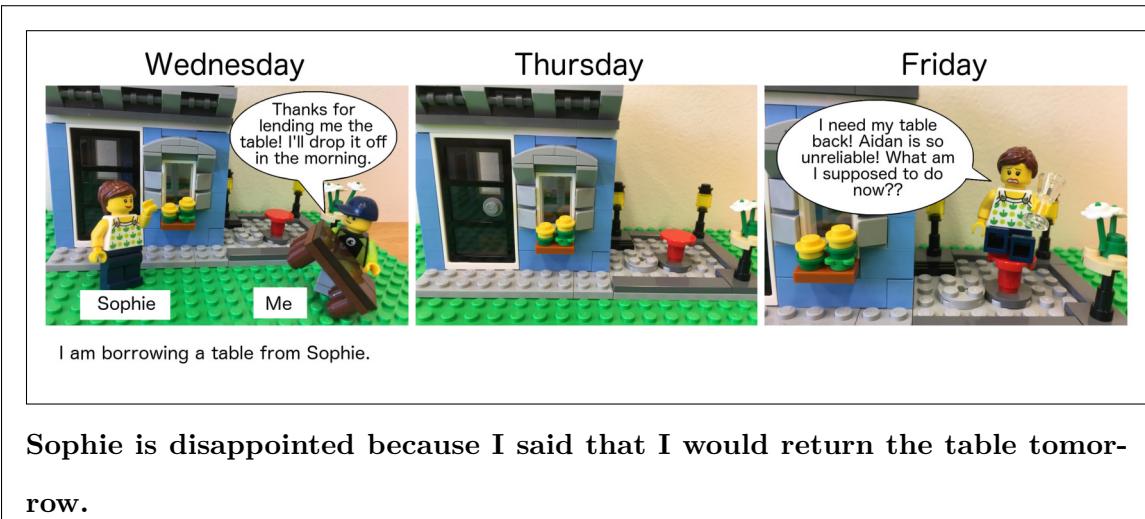
**Kevin is angry because I said that I would water his plants tomorrow.**



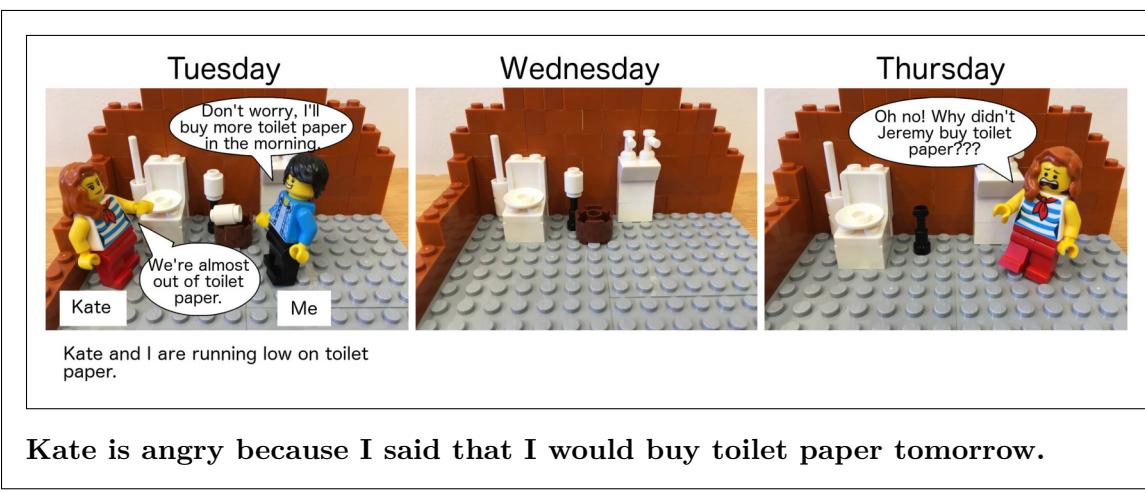
**Athena is angry because I said that I would return her skateboard tomorrow.**

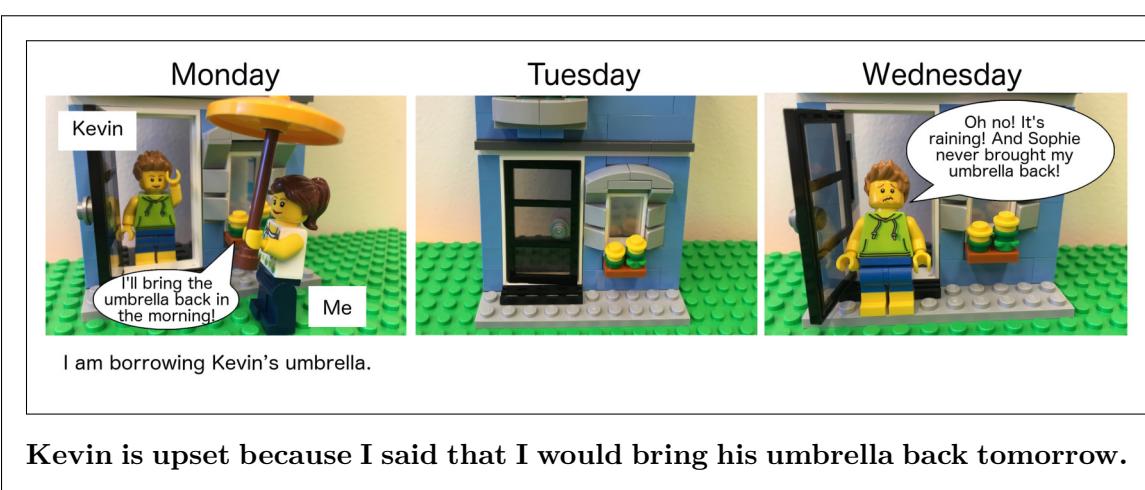
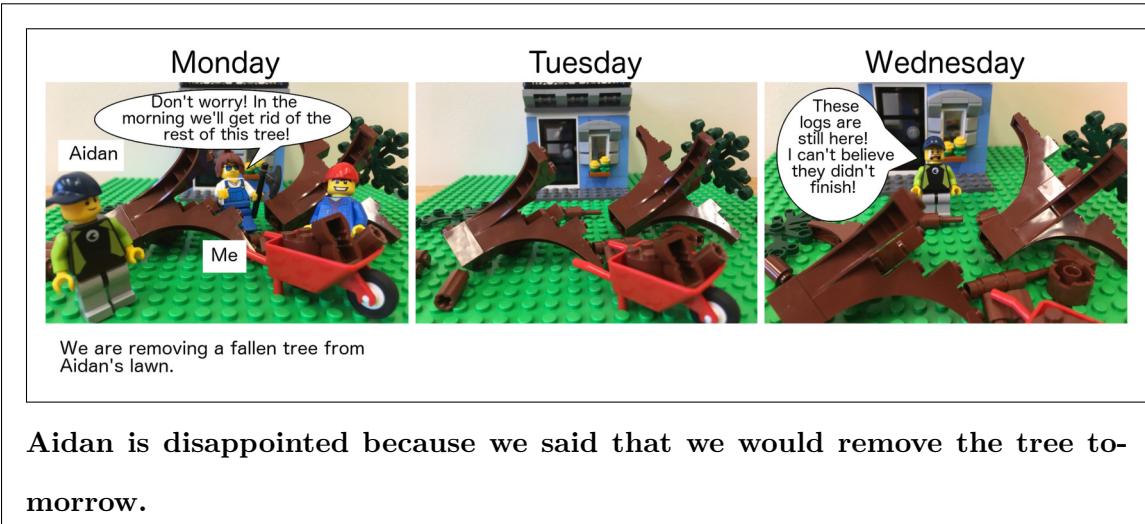


**Kevin is angry because I said that his suit would be ready tomorrow.**



**Sophie is disappointed because I said that I would return the table tomorrow.**





## G.2 Fillers

Sunday	Monday	Tuesday
 Kevin Me		 I never got to try the cookies! And we baked them together! Jeremy is so rude!
Kevin and I are baking cookies.		

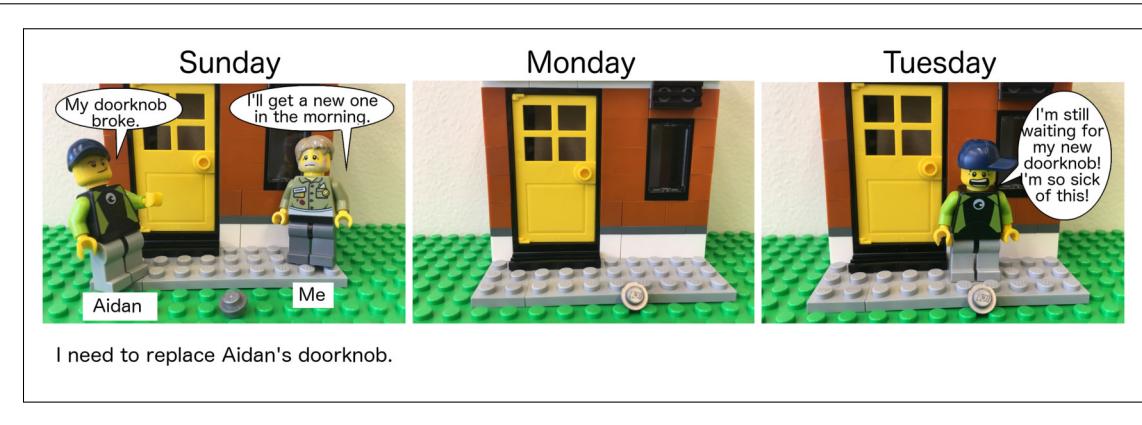
**Kevin is sad because I lied.**

Saturday	Sunday	Monday
 Athena Me		 Darn it! I wanted to go bird-watching today, but Sophie still has my binoculars!
Athena is lending me some binoculars.		

**Athena is frustrated because she is bored.**

Sunday	Monday	Tuesday
 Yay!!! I'll bring you a birthday cake for breakfast! Kate Me		 All I wanted was a birthday cake...
I am promising to bring a cake for Kate's birthday.		

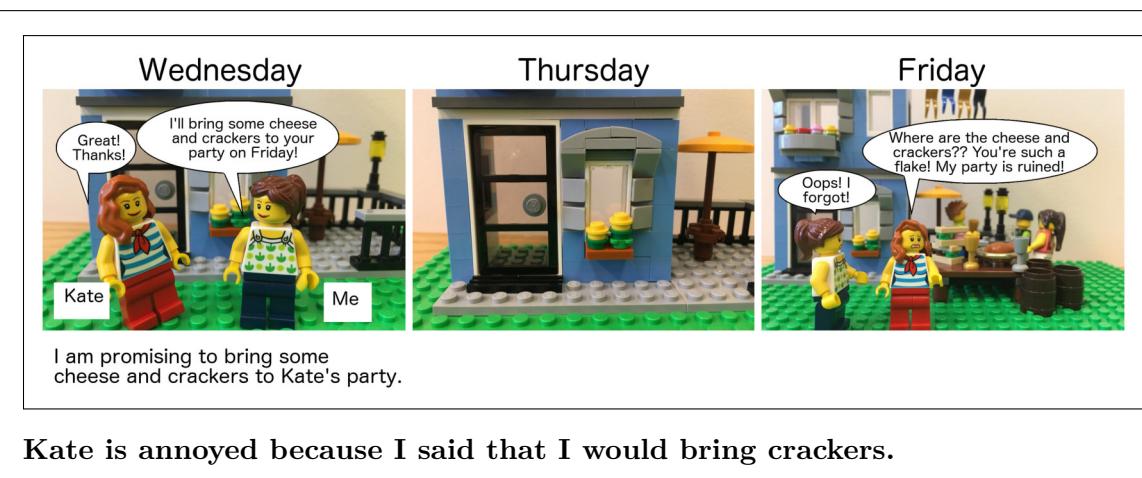
**Kate is happy because I brought her a birthday cake.**



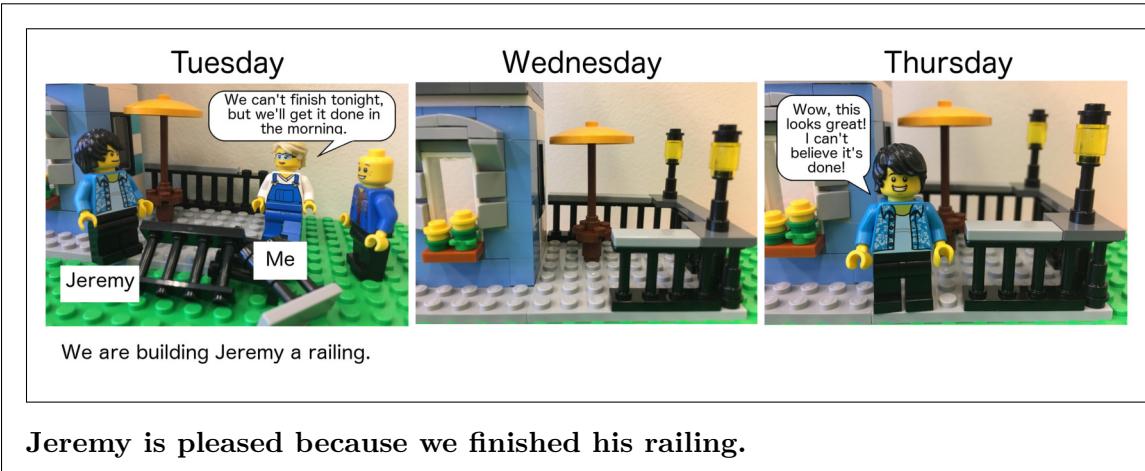
Aidan is angry because I didn't replace his door knob.



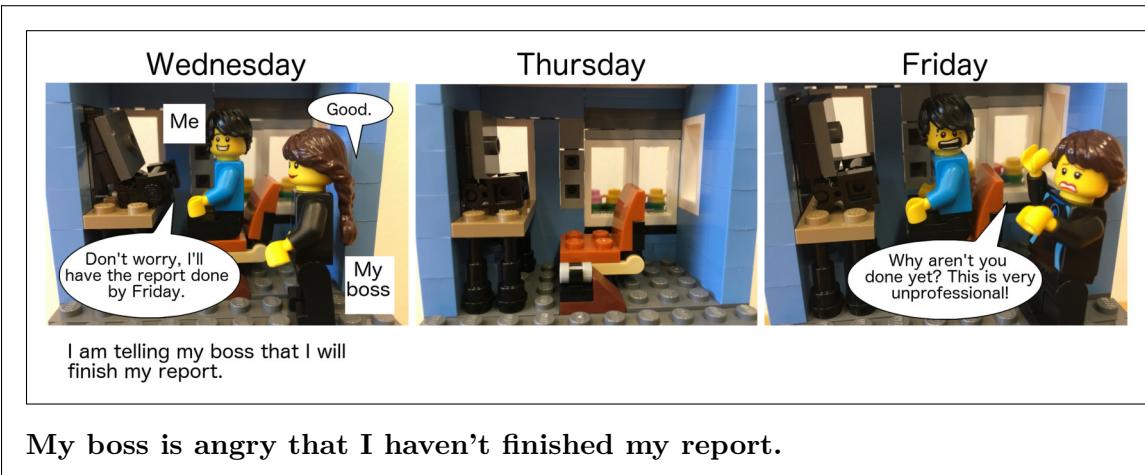
Sophie is annoyed because I stole from her.



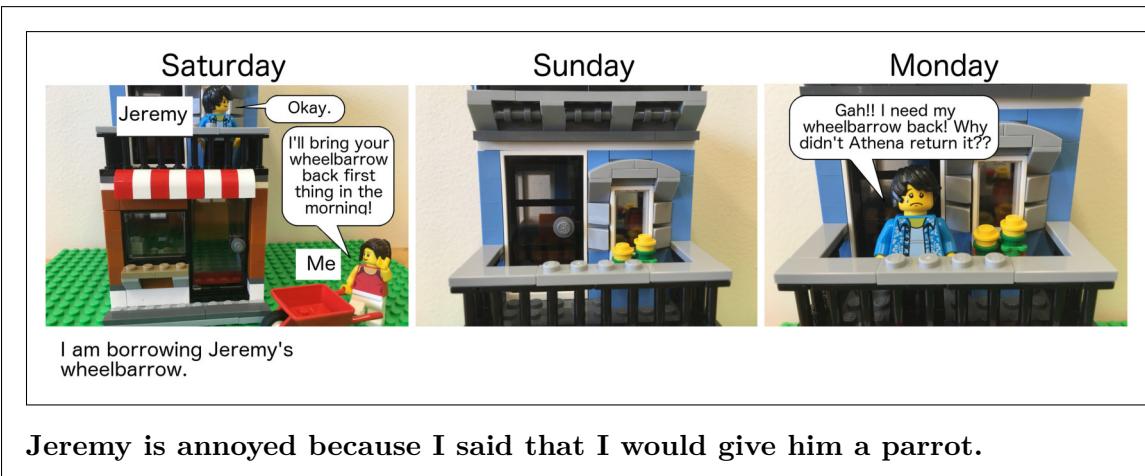
Kate is annoyed because I said that I would bring crackers.



**Jeremy is pleased because we finished his railing.**



**My boss is angry that I haven't finished my report.**

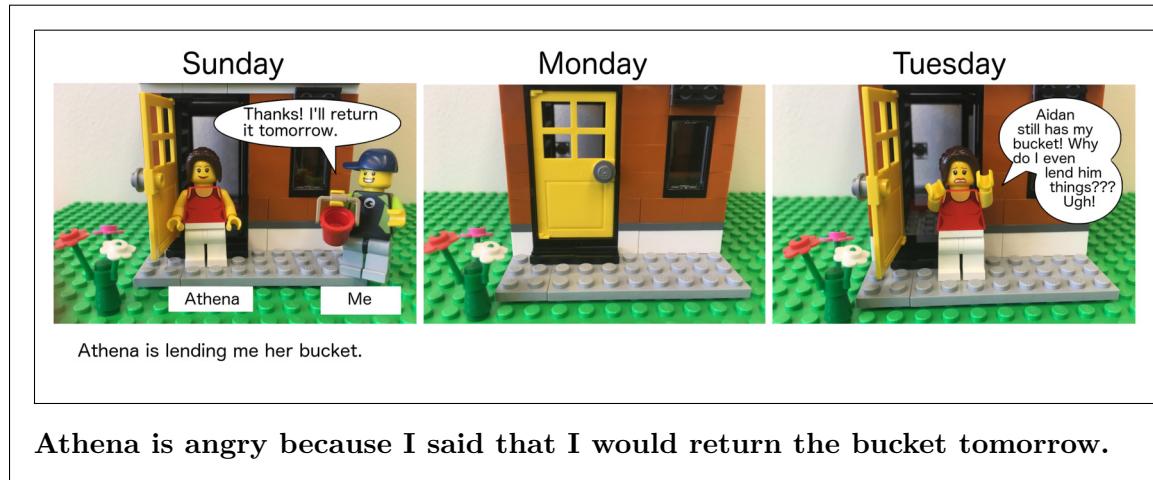
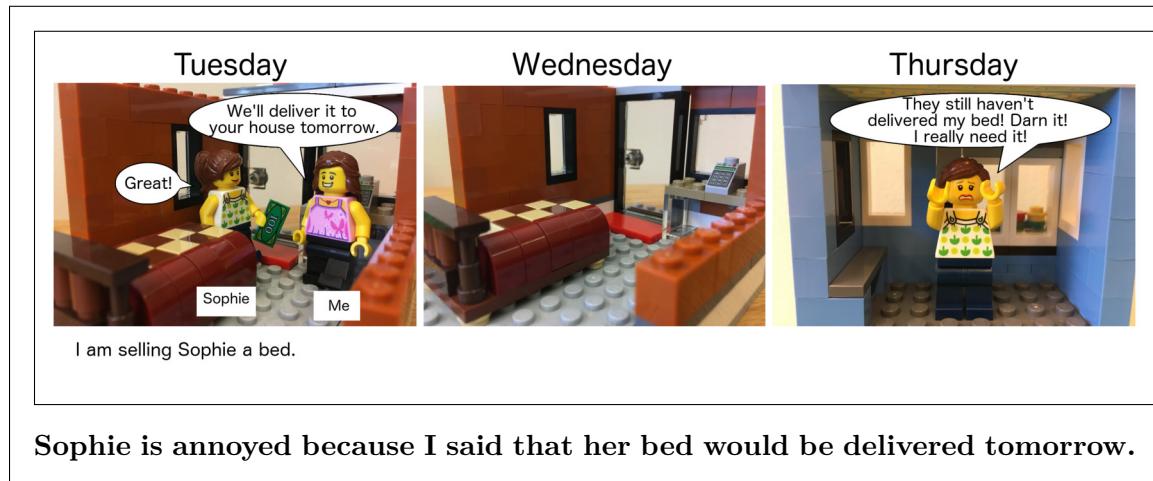


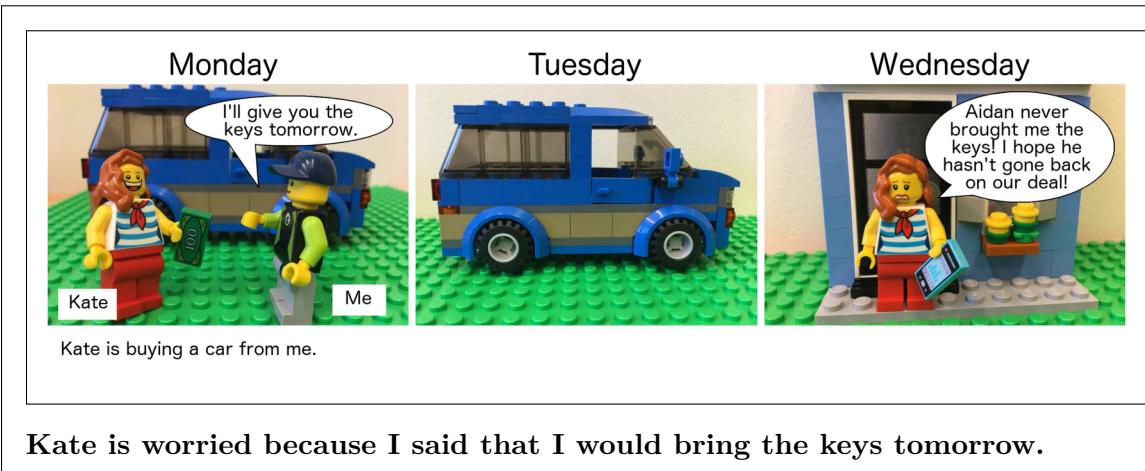
**Jeremy is annoyed because I said that I would give him a parrot.**

## APPENDIX H

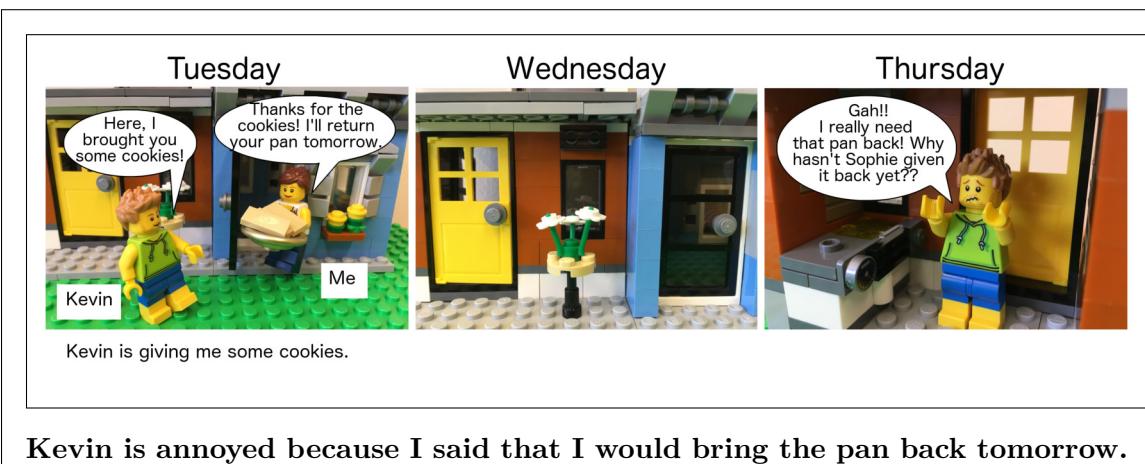
### EXPERIMENT 3A ITEMS

#### H.1 Main items

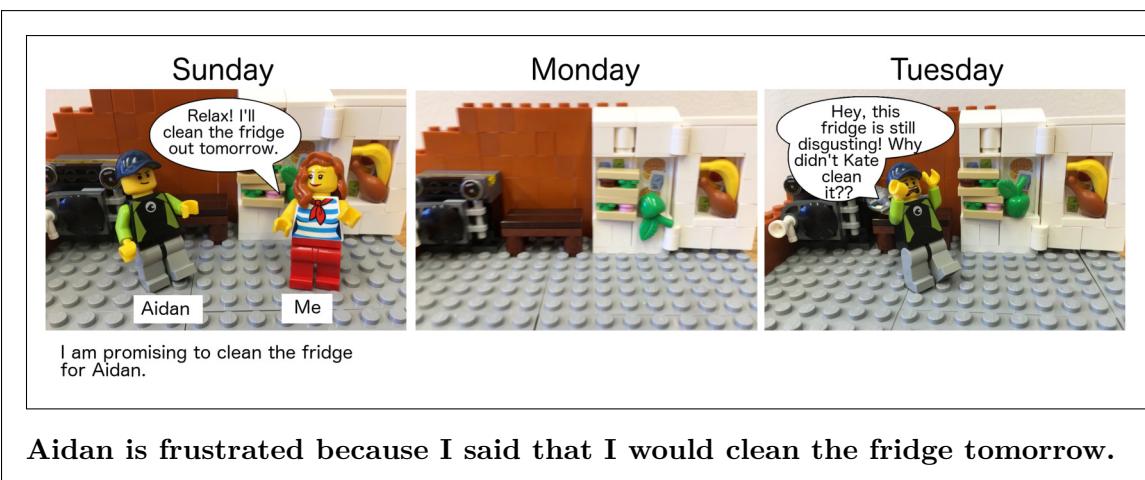




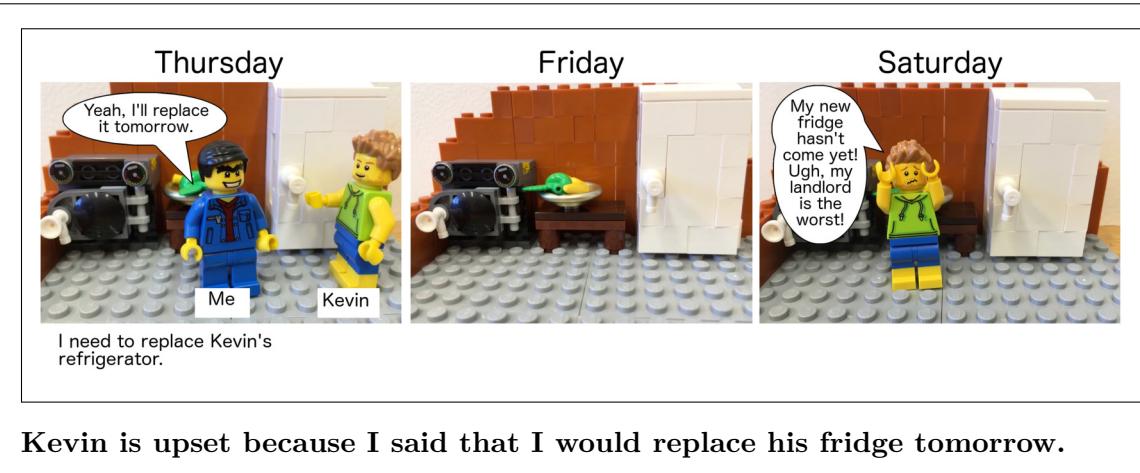
**Kate is worried because I said that I would bring the keys tomorrow.**



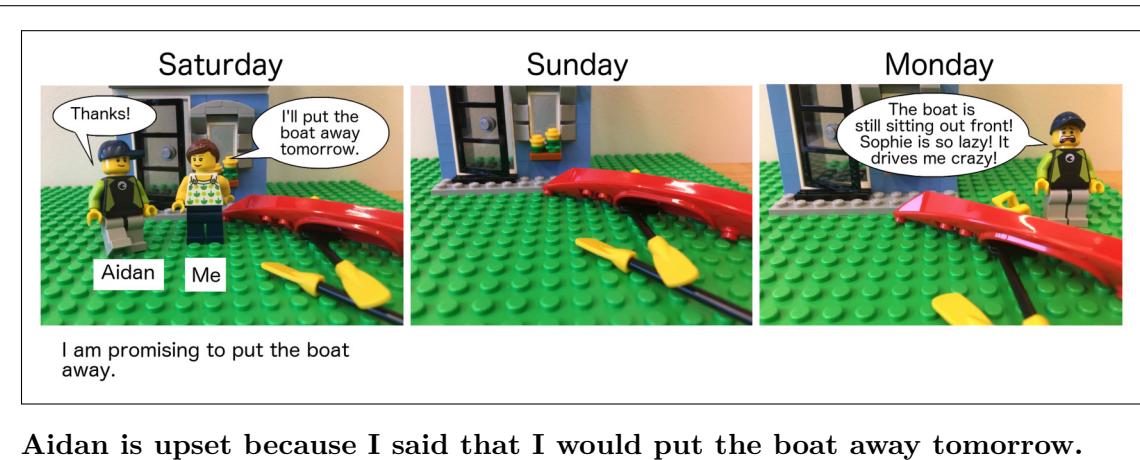
**Kevin is annoyed because I said that I would bring the pan back tomorrow.**



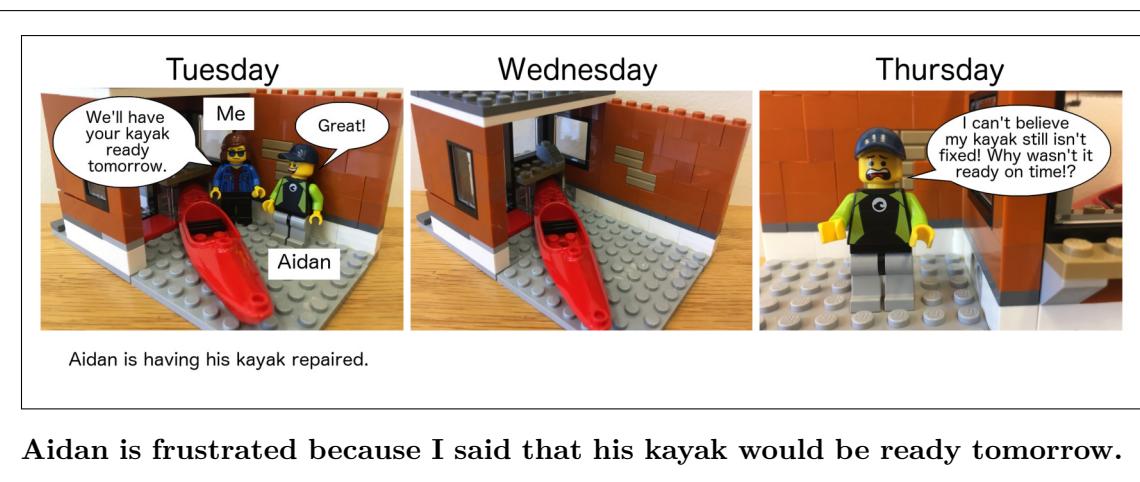
**Aidan is frustrated because I said that I would clean the fridge tomorrow.**



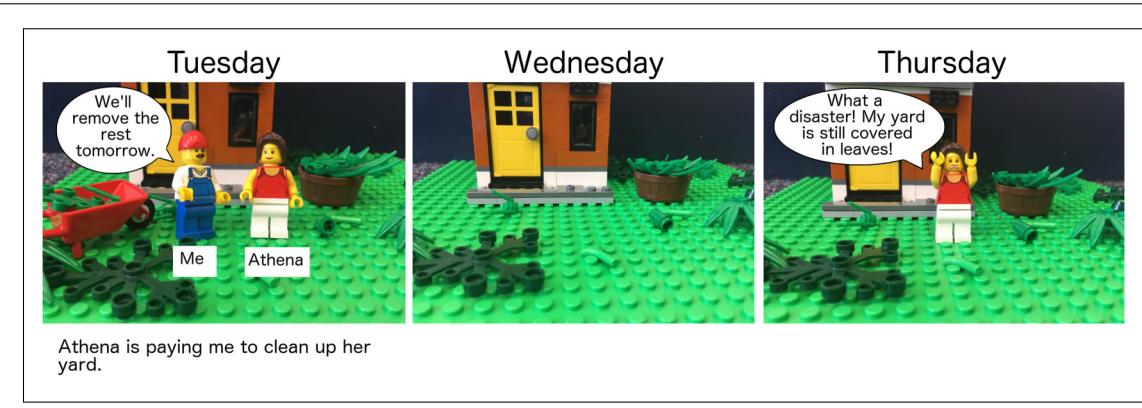
**Kevin is upset because I said that I would replace his fridge tomorrow.**



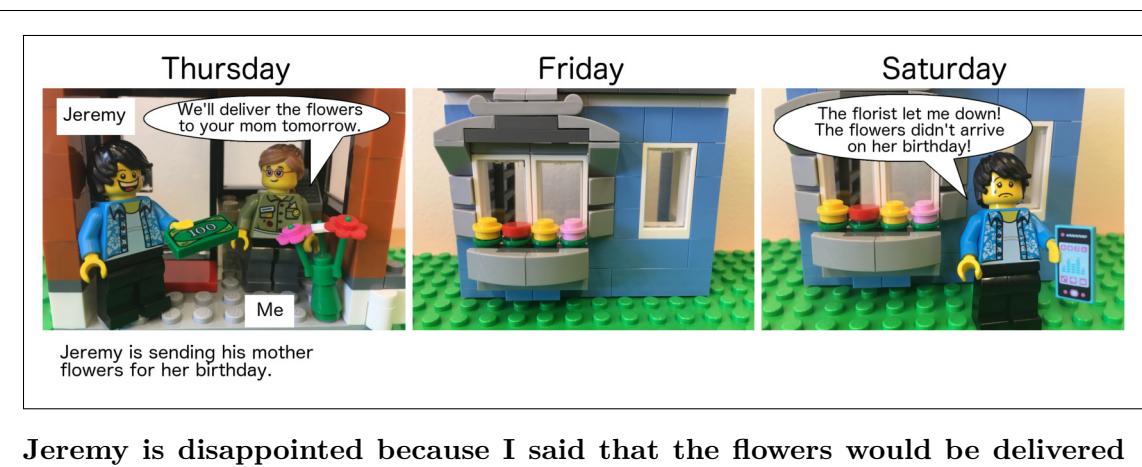
**Aidan is upset because I said that I would put the boat away tomorrow.**



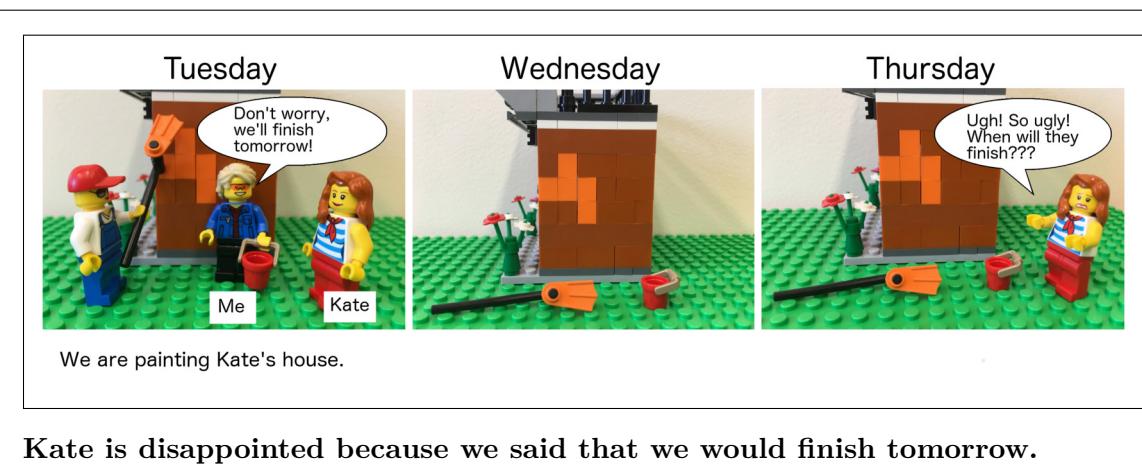
**Aidan is frustrated because I said that his kayak would be ready tomorrow.**



Athena is frustrated because we said that we would remove the leaves tomorrow.



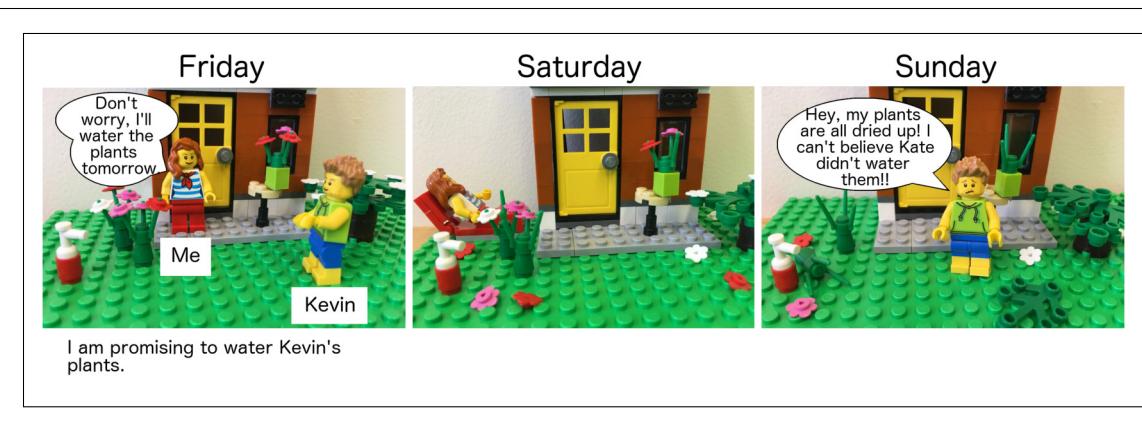
Jeremy is disappointed because I said that the flowers would be delivered tomorrow.



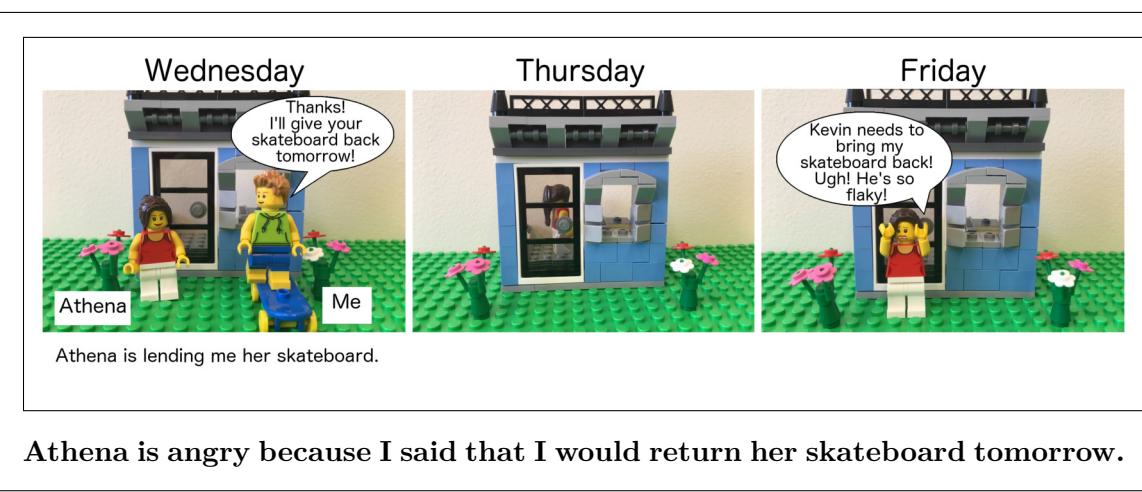
Kate is disappointed because we said that we would finish tomorrow.



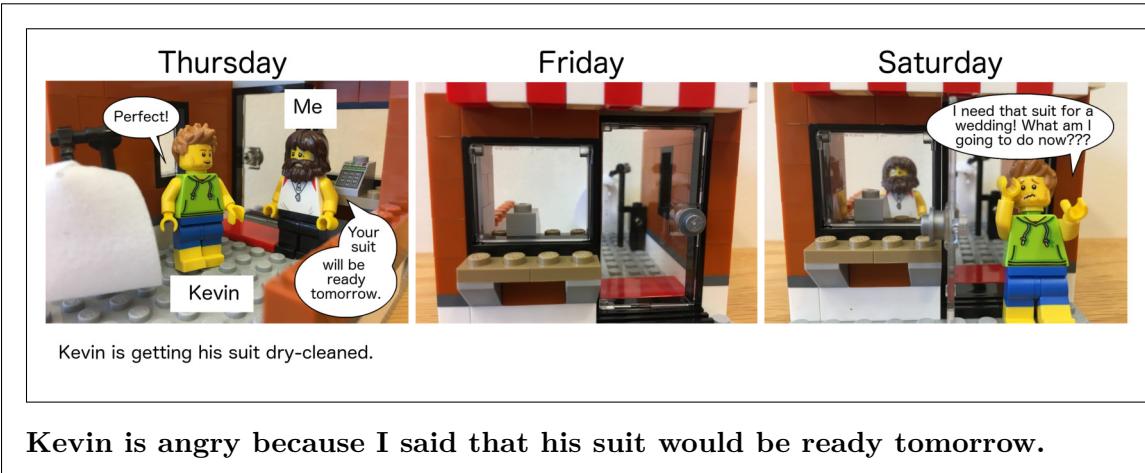
**Kevin is disappointed because I said that the store would have iPhones tomorrow.**



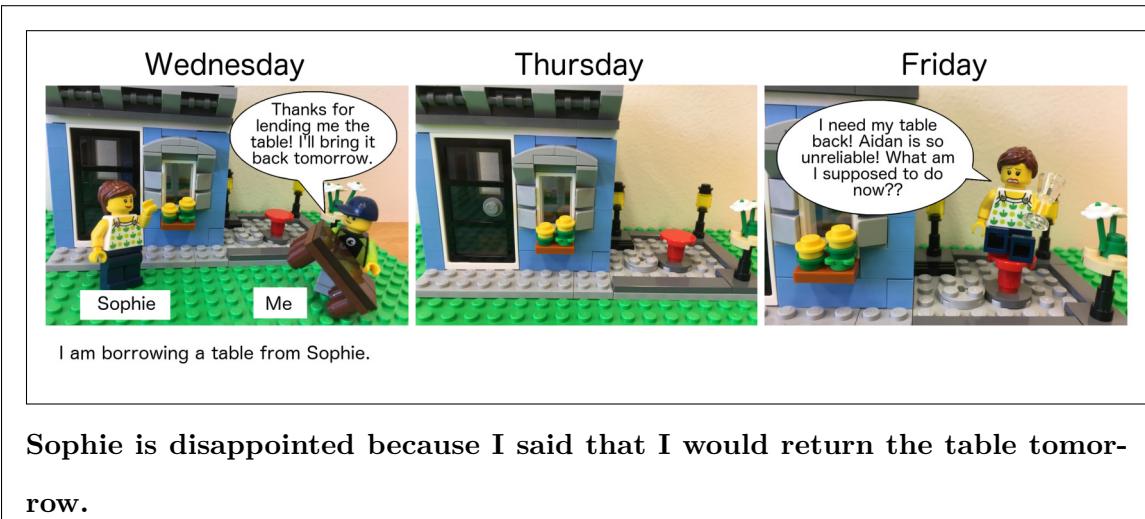
**Kevin is angry because I said that I would water his plants tomorrow.**



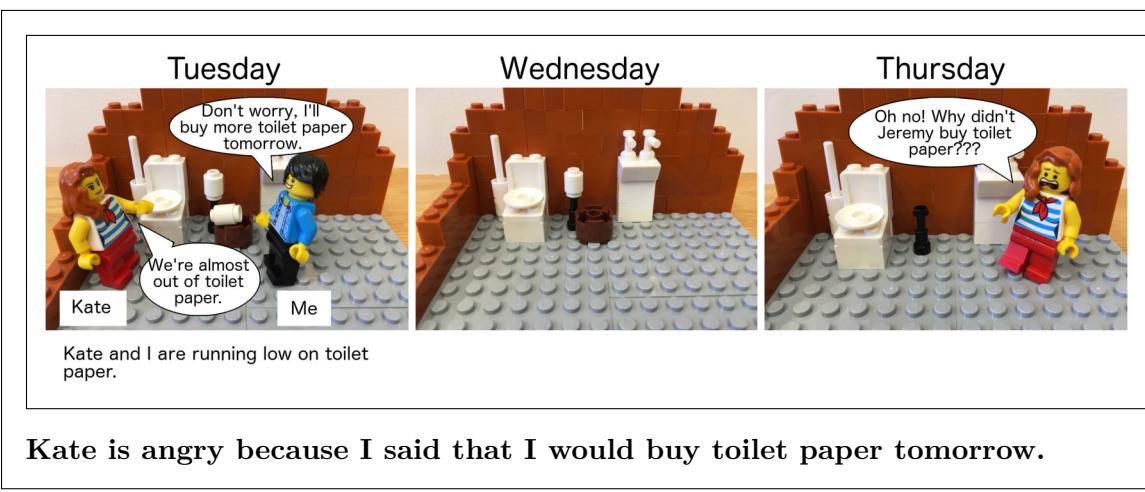
**Athena is angry because I said that I would return her skateboard tomorrow.**



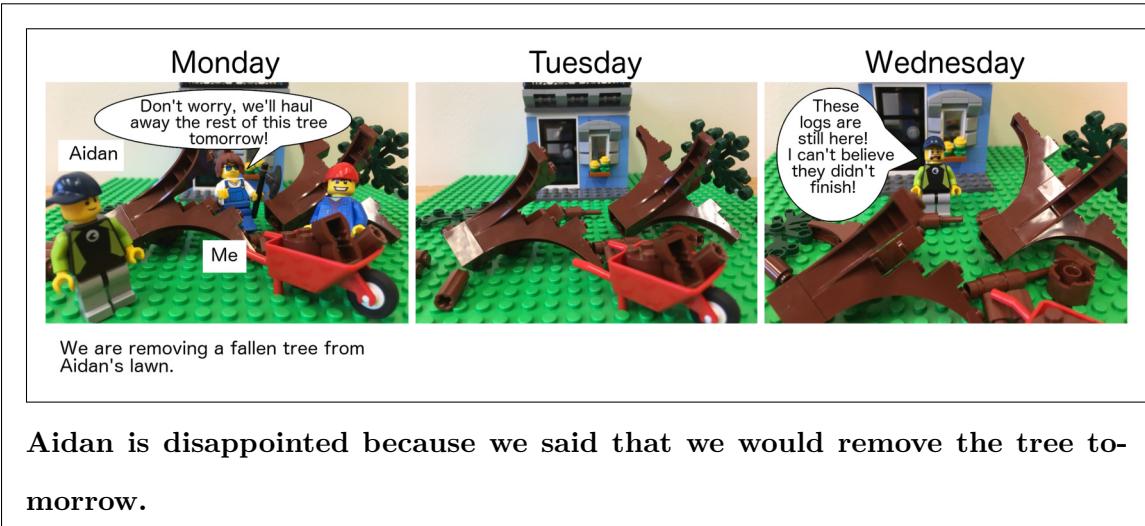
**Kevin is angry because I said that his suit would be ready tomorrow.**



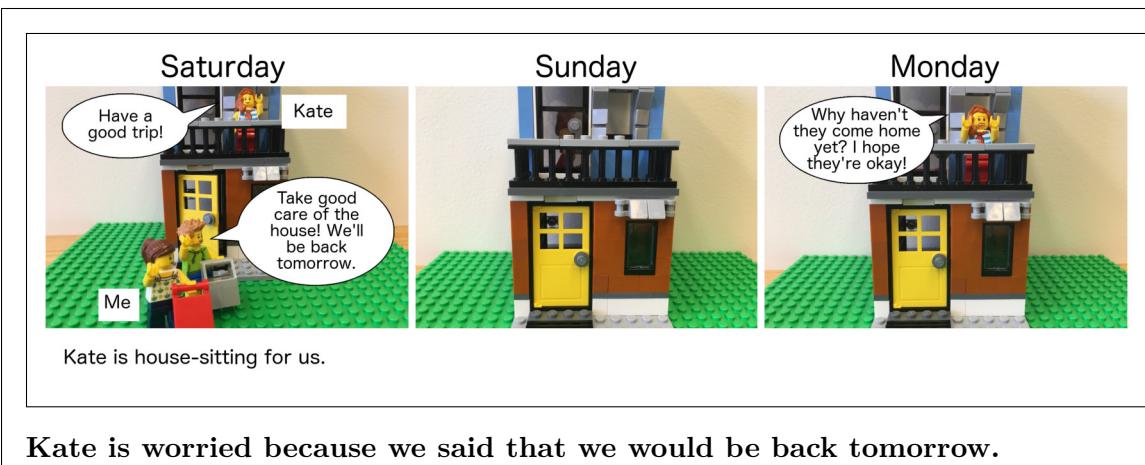
**Sophie is disappointed because I said that I would return the table tomorrow.**



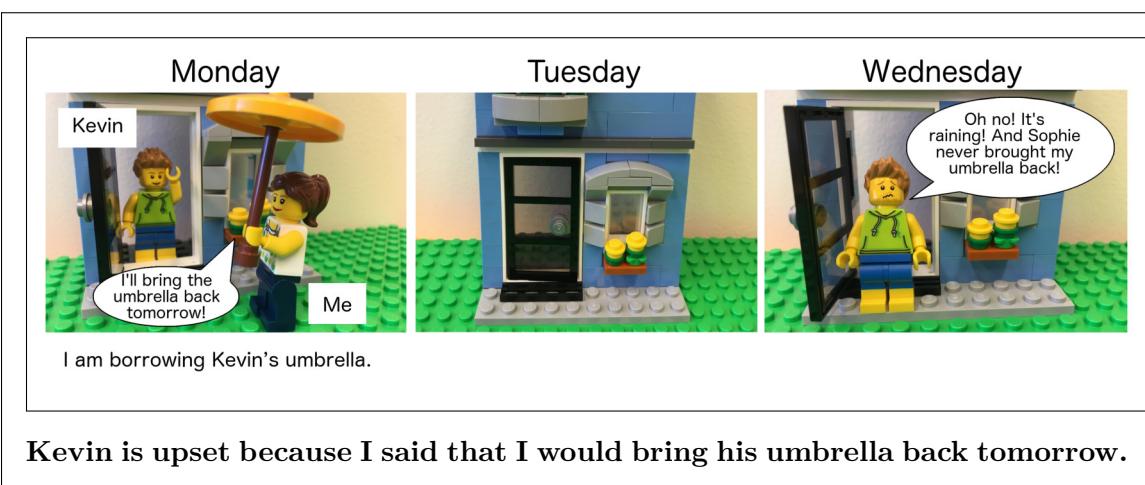
**Kate is angry because I said that I would buy toilet paper tomorrow.**



Aidan is disappointed because we said that we would remove the tree tomorrow.

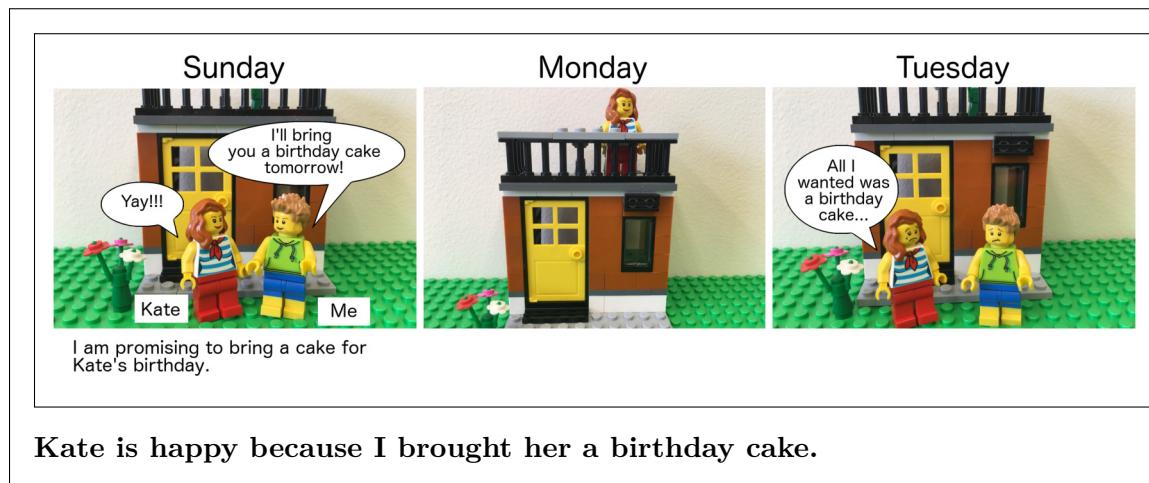
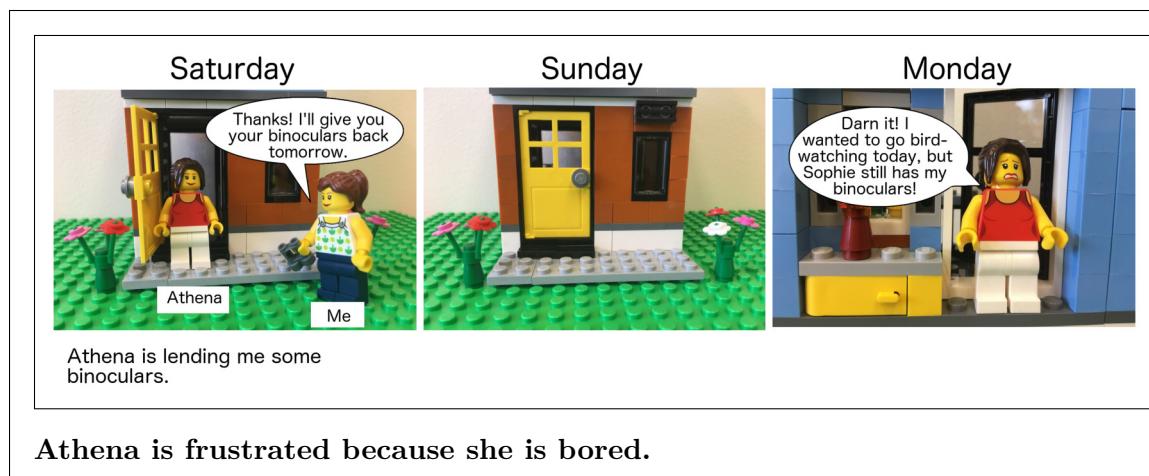
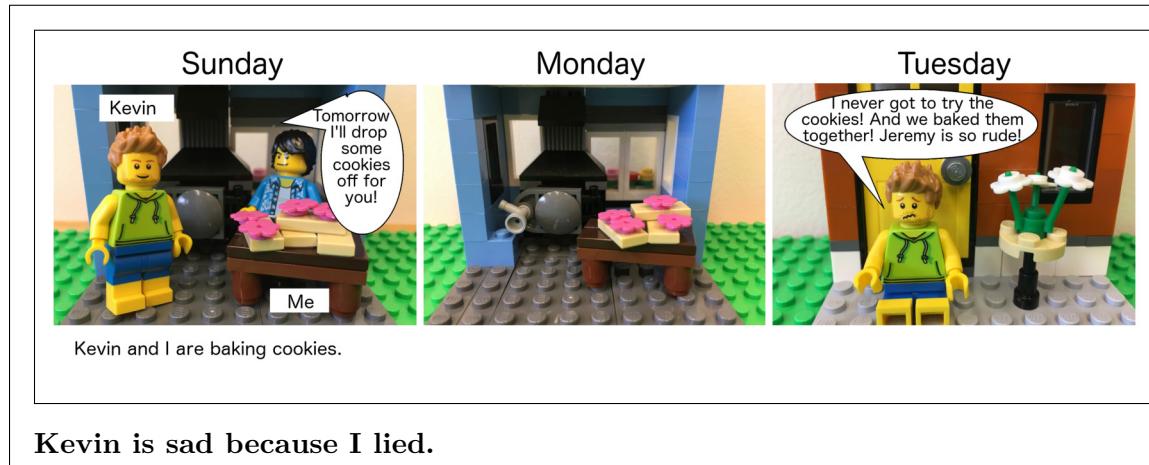


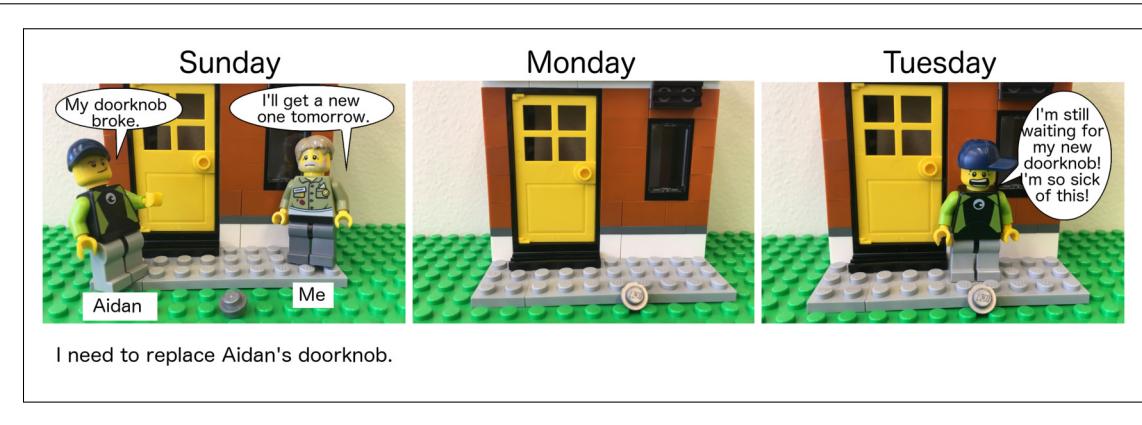
Kate is worried because we said that we would be back tomorrow.



Kevin is upset because I said that I would bring his umbrella back tomorrow.

## H.2 Fillers





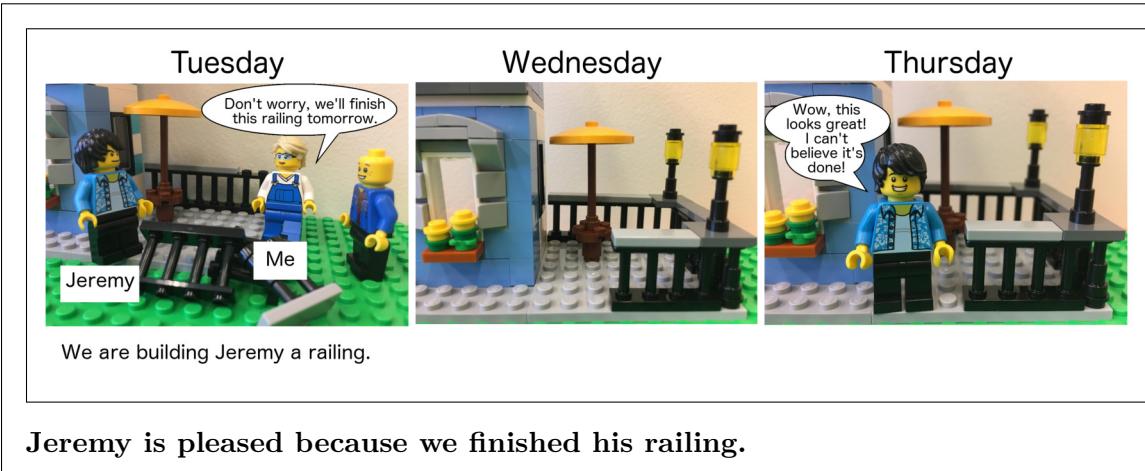
Aidan is angry because I didn't replace his door knob.



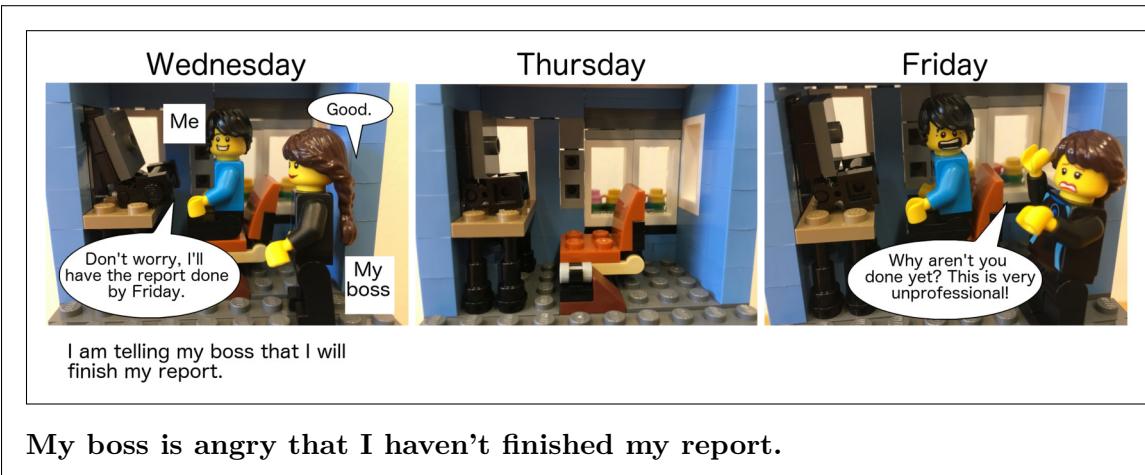
Sophie is annoyed because I stole from her.



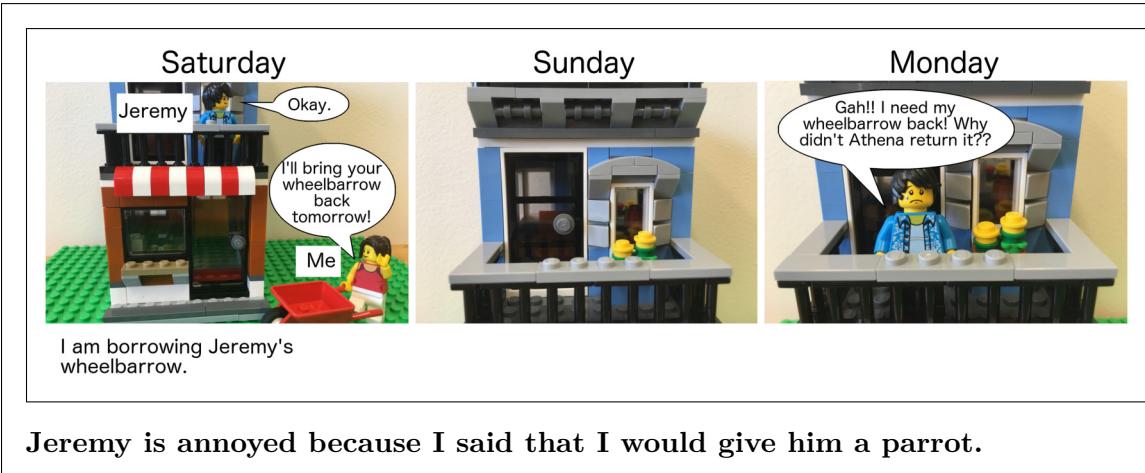
Kate is annoyed because I said that I would bring crackers.



**Jeremy is pleased because we finished his railing.**



**My boss is angry that I haven't finished my report.**

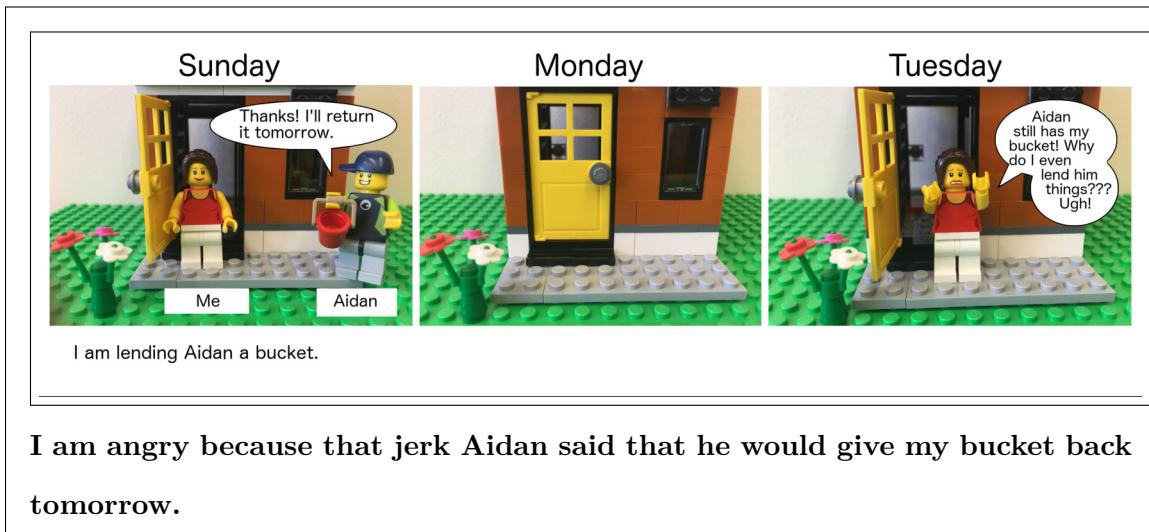
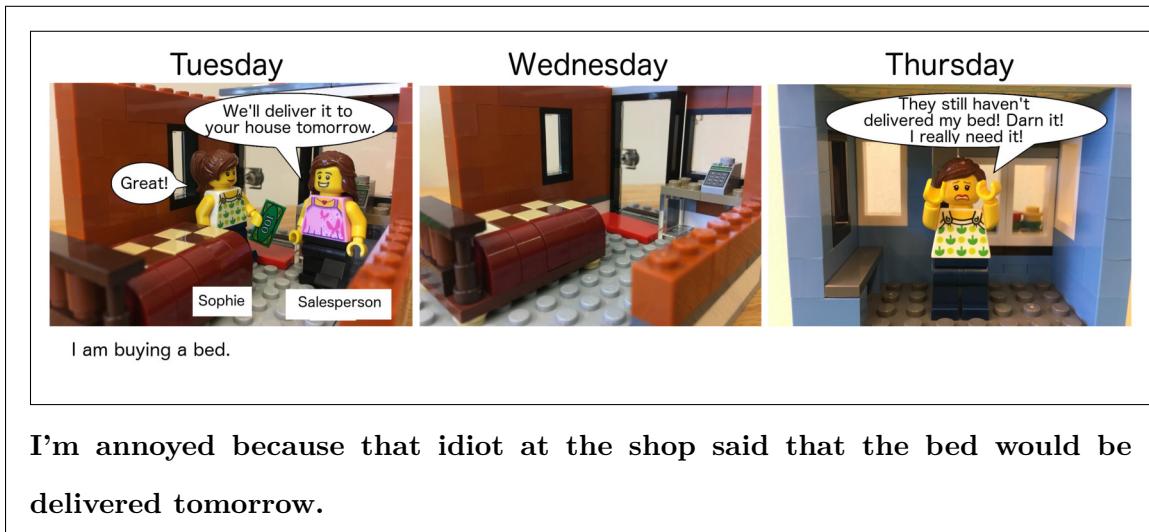


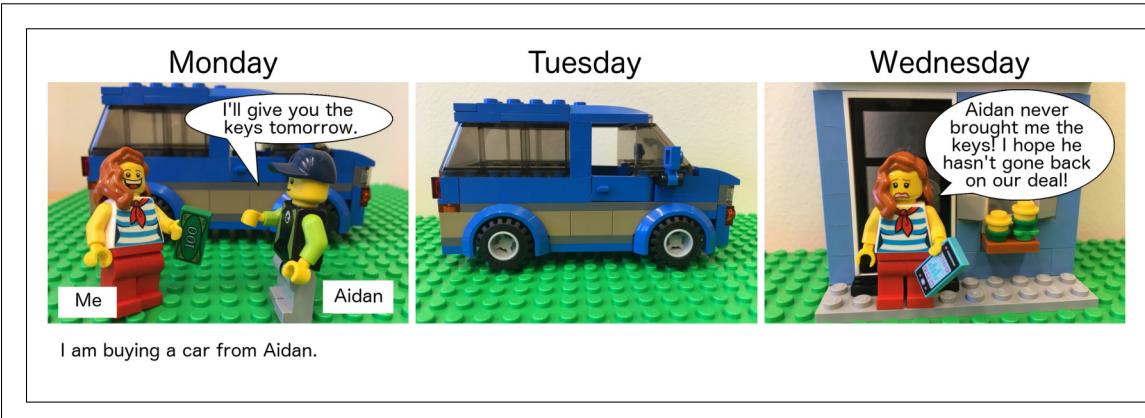
**Jeremy is annoyed because I said that I would give him a parrot.**

## APPENDIX I

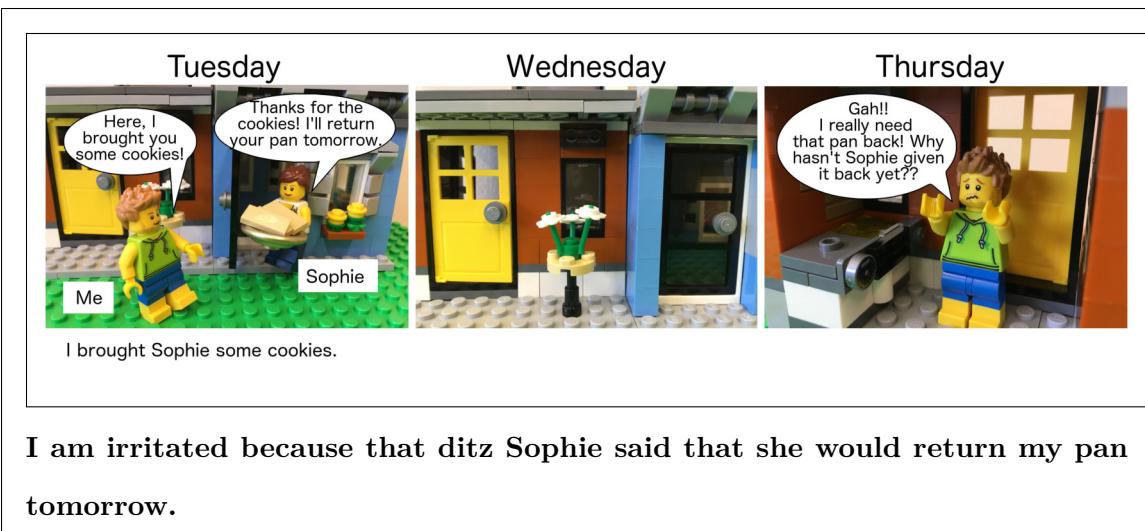
### EXPERIMENT 3B ITEMS

#### I.1 Main items

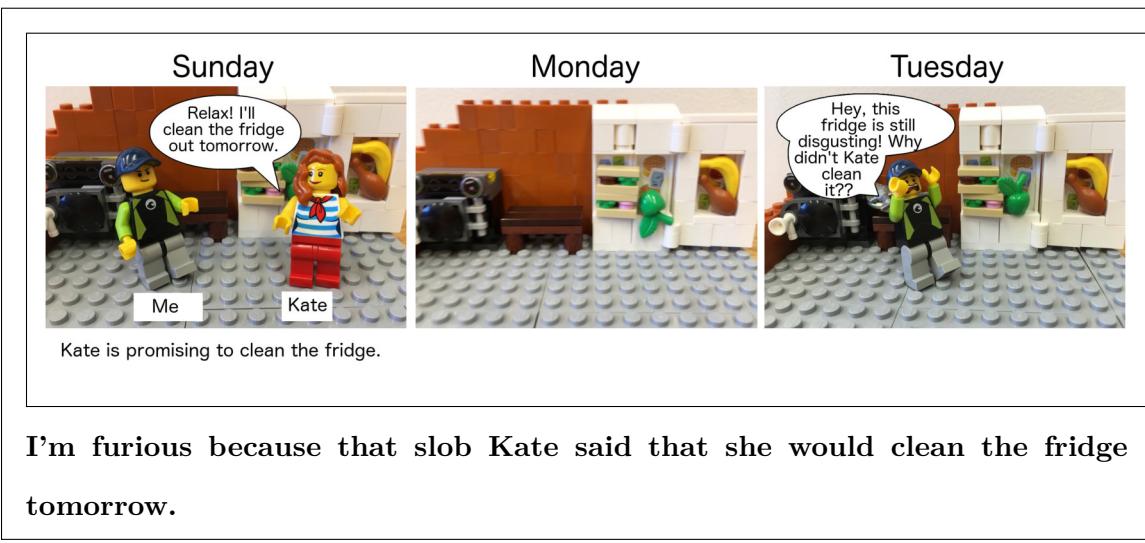




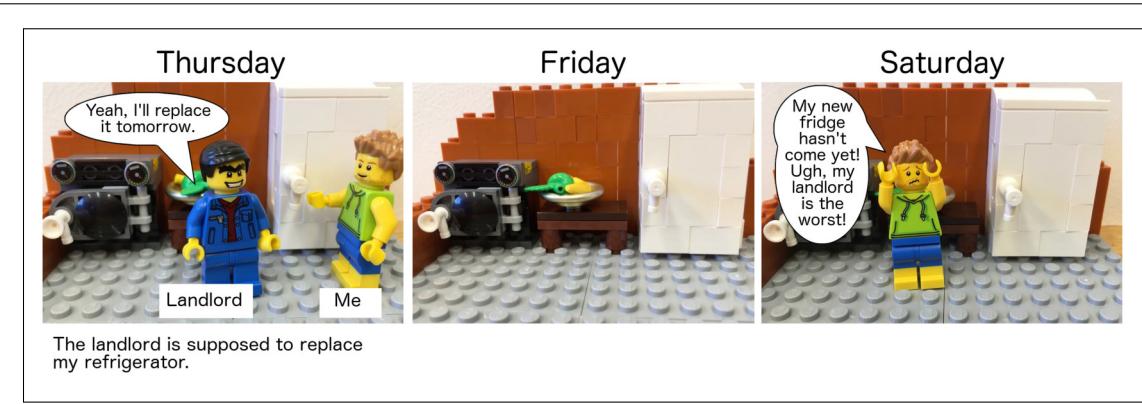
**I'm worried because that crook Aidan said he would give me the keys tomorrow.**



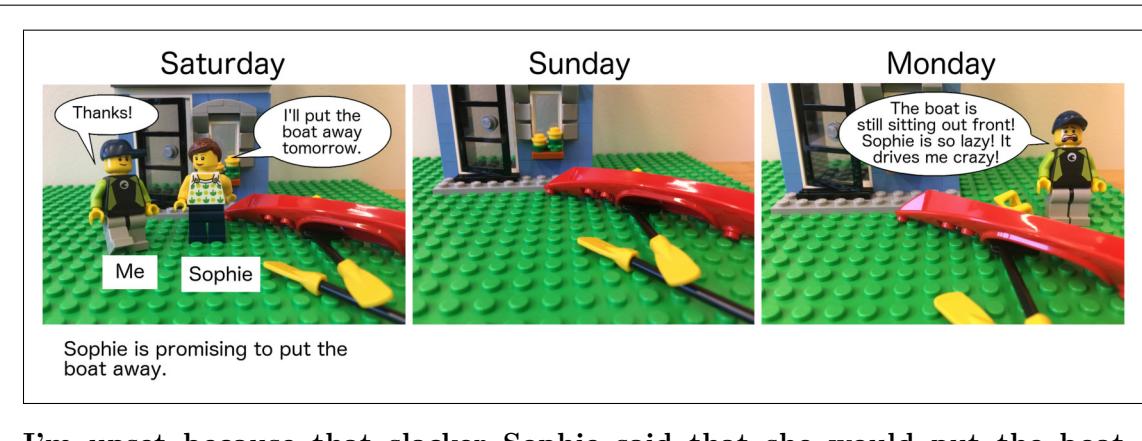
**I am irritated because that ditz Sophie said that she would return my pan tomorrow.**



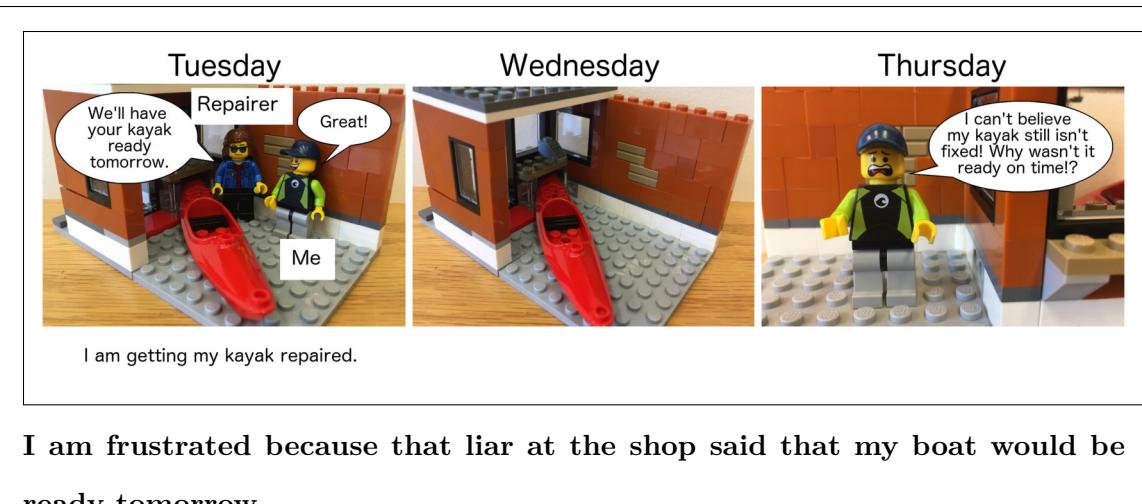
**I'm furious because that slob Kate said that she would clean the fridge tomorrow.**



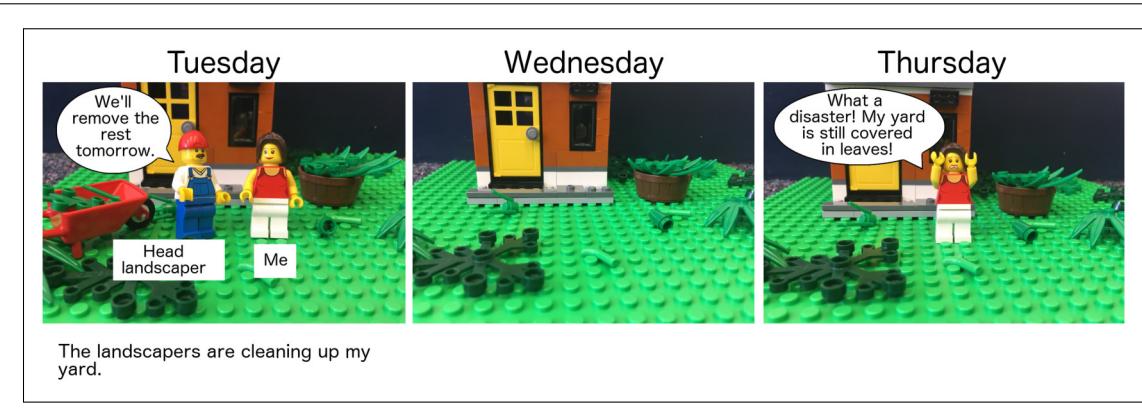
I am upset because that cheapskate landlord said that he would replace my fridge tomorrow.



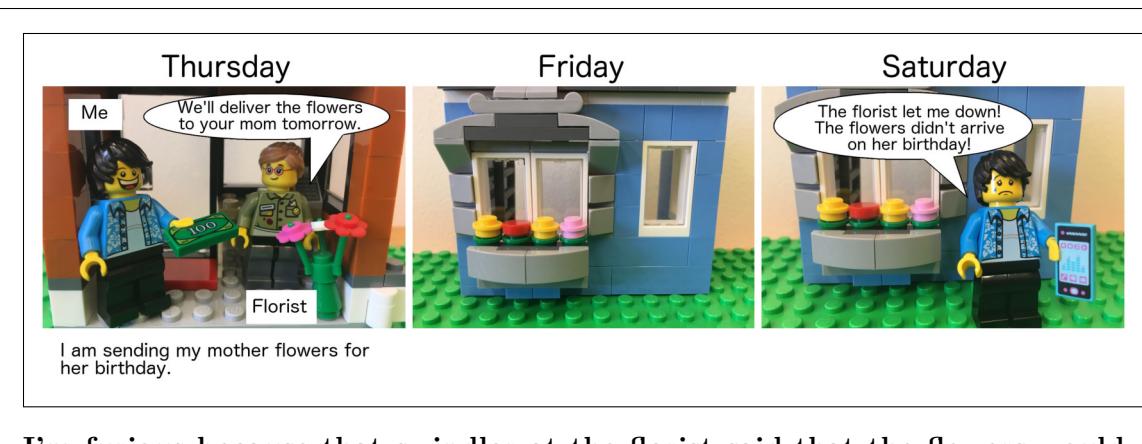
I'm upset because that slacker Sophie said that she would put the boat away tomorrow.



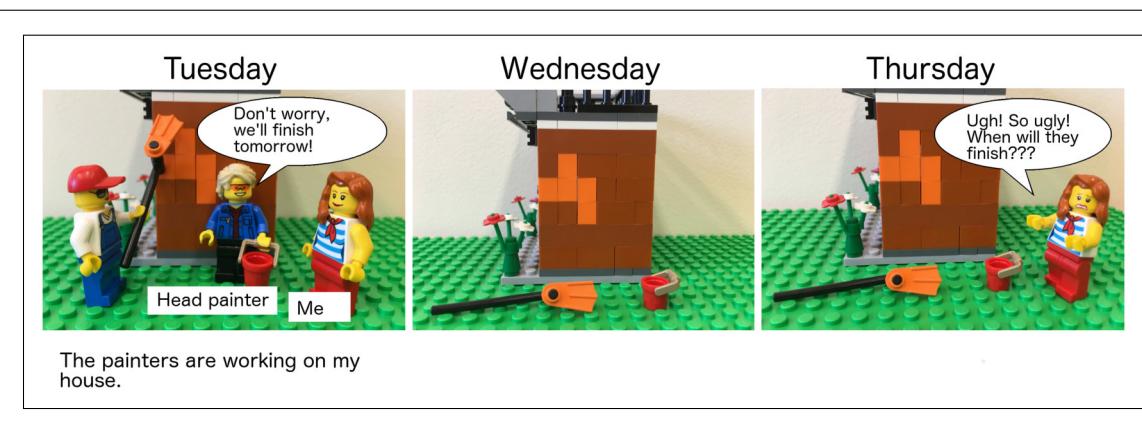
I am frustrated because that liar at the shop said that my boat would be ready tomorrow.



I'm frustrated because those slackers said that they would clean up this mess tomorrow.



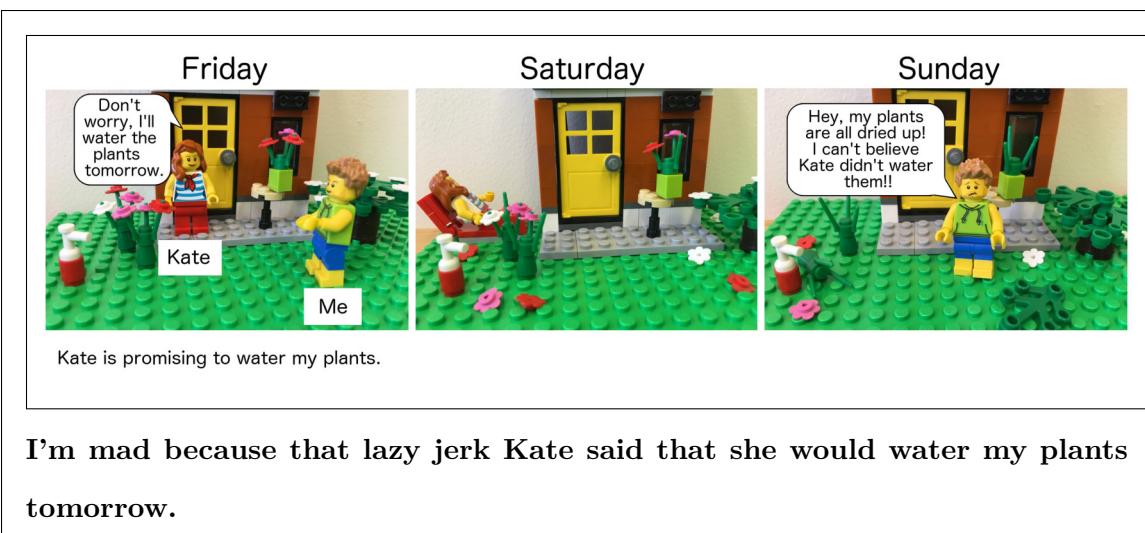
I'm furious because that swindler at the florist said that the flowers would be delivered tomorrow.



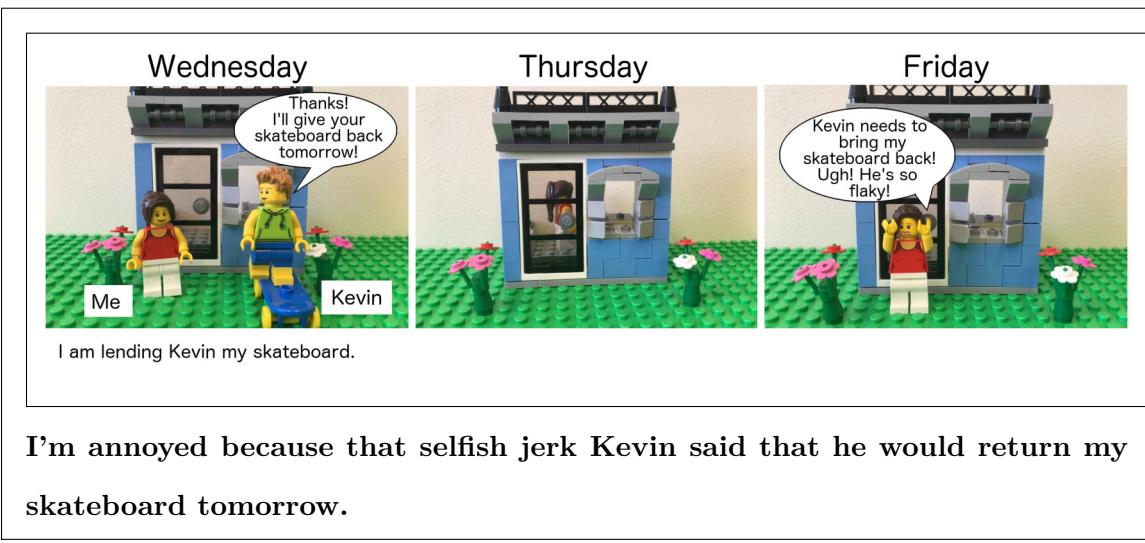
I am disappointed because those jerks said that they would finish the house tomorrow.



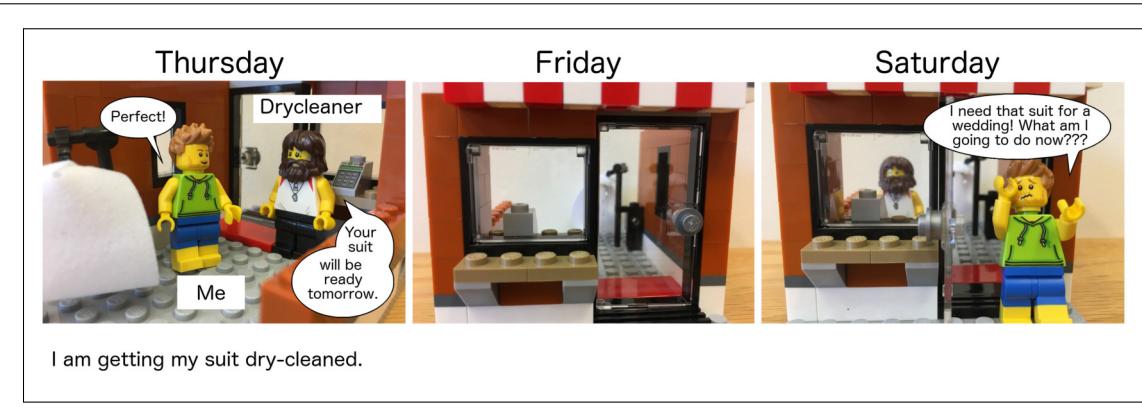
**I'm upset because that nitwit at the store said that they would have more iPhones tomorrow.**



**I'm mad because that lazy jerk Kate said that she would water my plants tomorrow.**



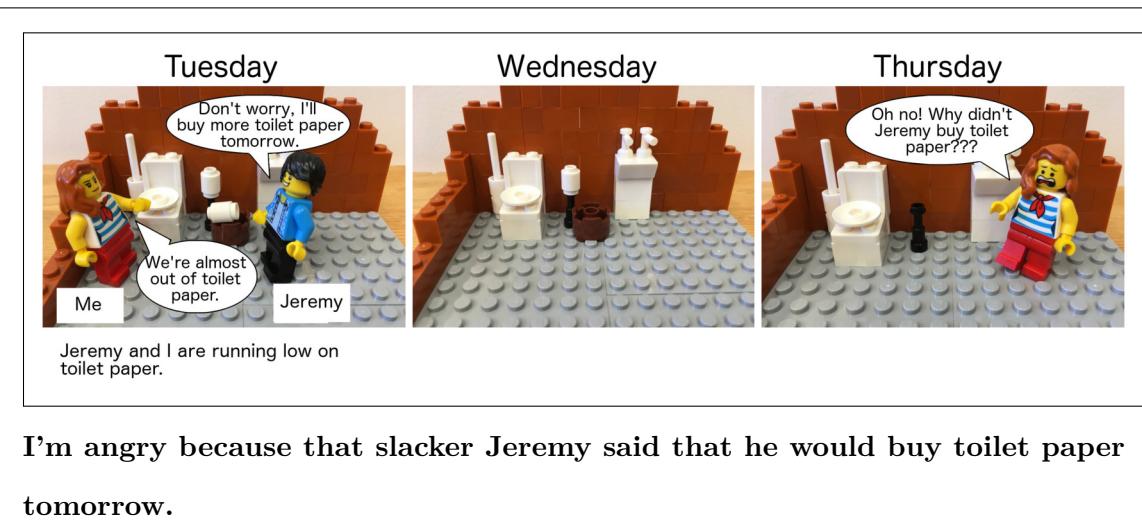
**I'm annoyed because that selfish jerk Kevin said that he would return my skateboard tomorrow.**



**I am angry because that liar at the cleaner's said that my suit would be ready tomorrow.**



**I'm disappointed because that doofus Aidan said that he would return my table tomorrow.**



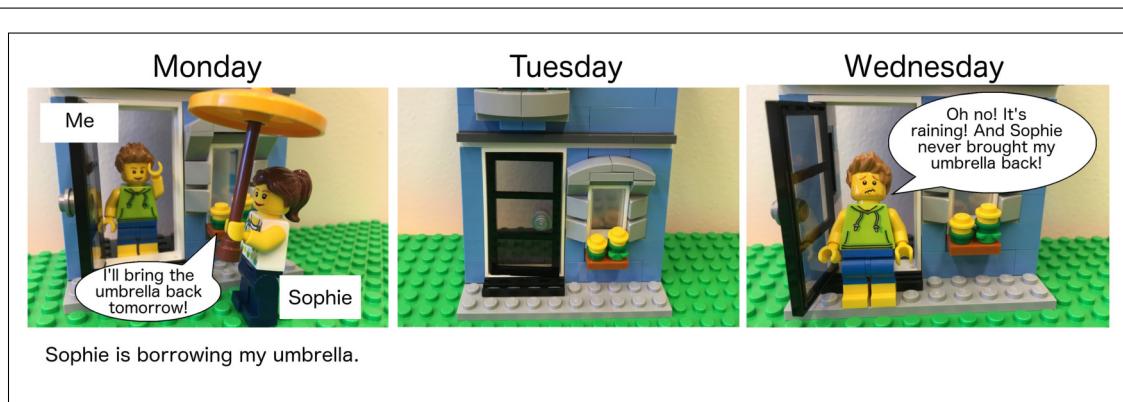


I am disappointed because those lazy landscapers said that they would finish the job tomorrow.



I am house-sitting for Sophie and Kevin.

I am worried because the dears said that they would be back tomorrow.



Sophie is borrowing my umbrella.

I'm upset because that airhead Sophie said that she would return my umbrella tomorrow.

## I.2 Fillers

**Sunday**

**Monday**

**Tuesday**

Jeremy and I are baking cookies.

**I'm sad because Jeremy lied.**

**Saturday**

**Sunday**

**Monday**

I am lending Sophie my binoculars.

**I am frustrated because I'm bored.**

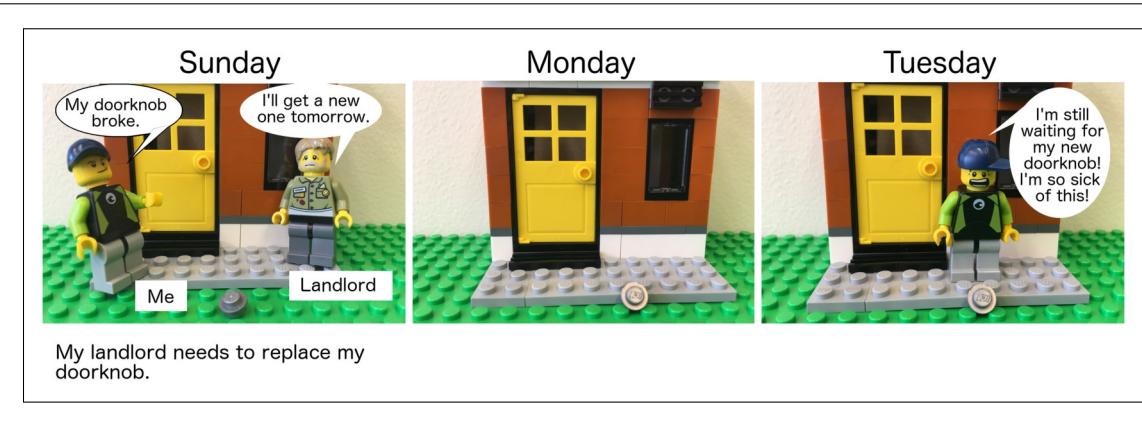
**Sunday**

**Monday**

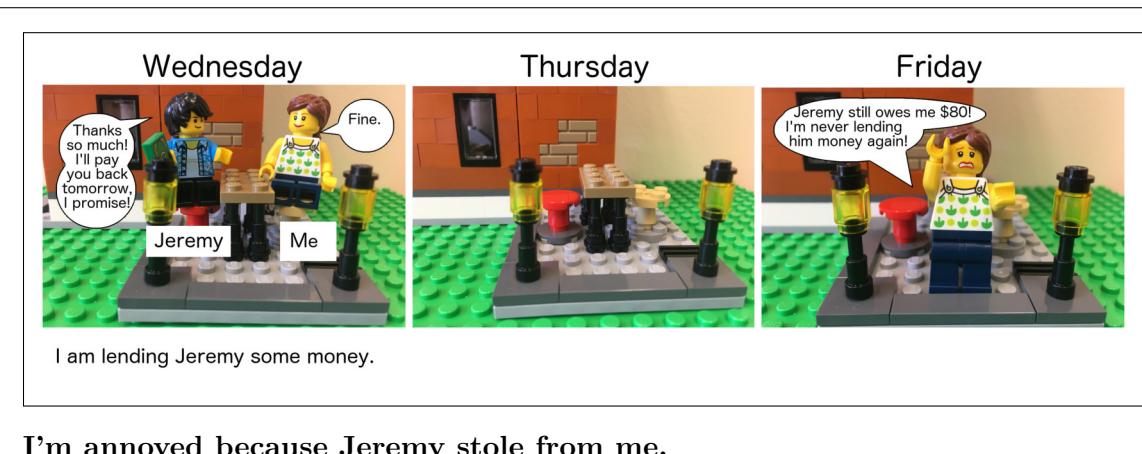
**Tuesday**

Kevin is promising to bring a cake for my birthday.

**I am happy because Kevin brought me a birthday cake.**



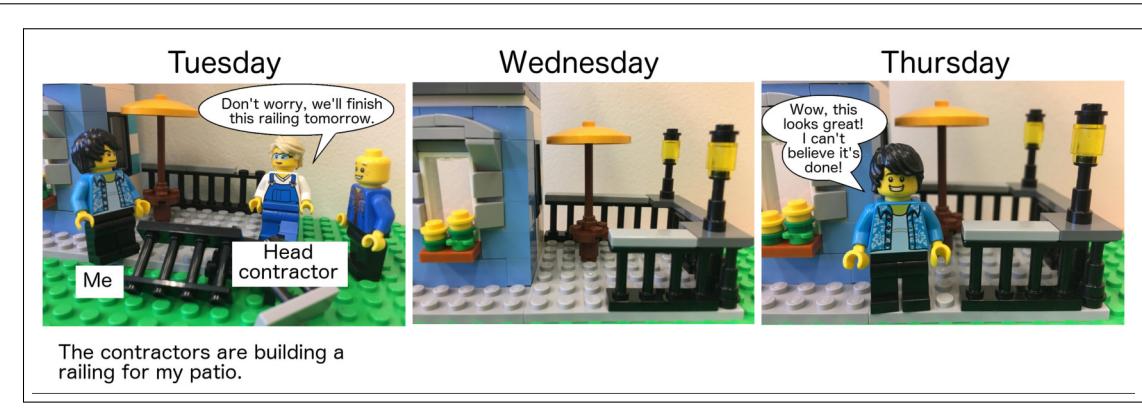
I am angry because the landlord didn't replace my door knob.



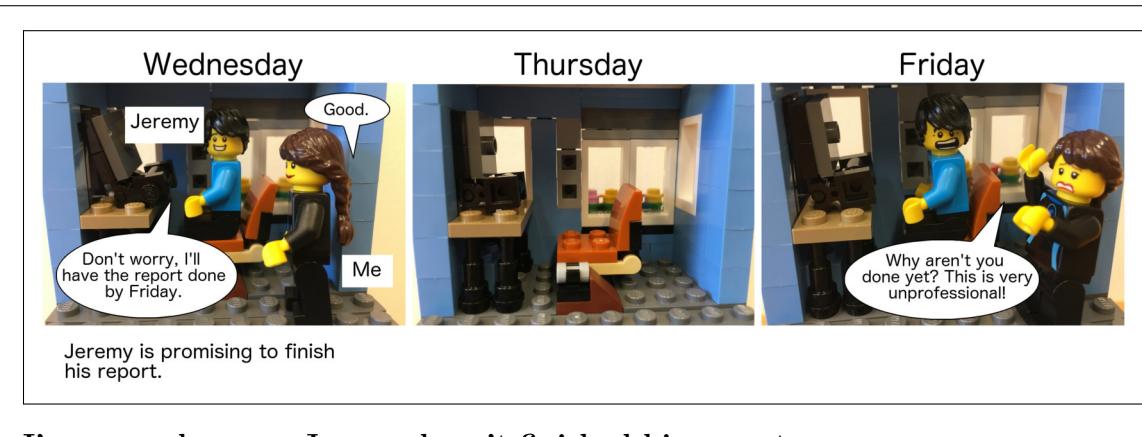
I'm annoyed because Jeremy stole from me.



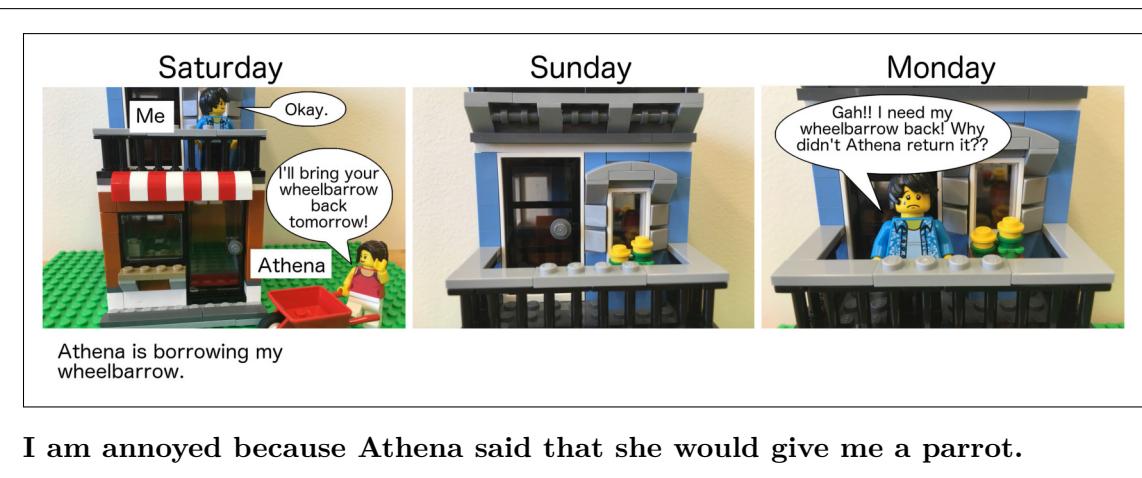
I'm annoyed because Sophie said that she would bring crackers.



I am pleased because the contractors finished my railing.



I'm angry because Jeremy hasn't finished his report.

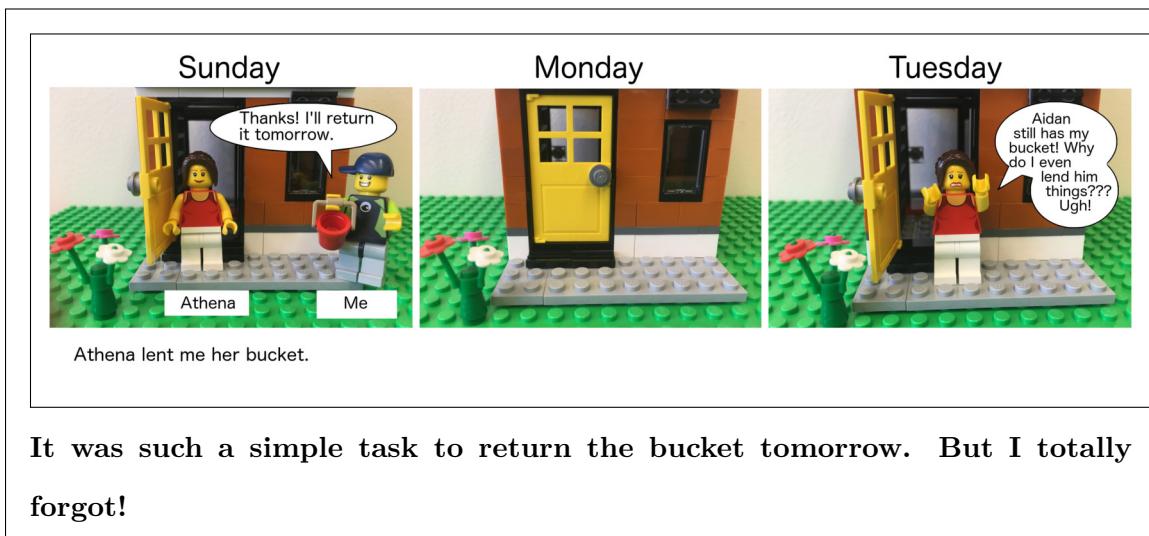
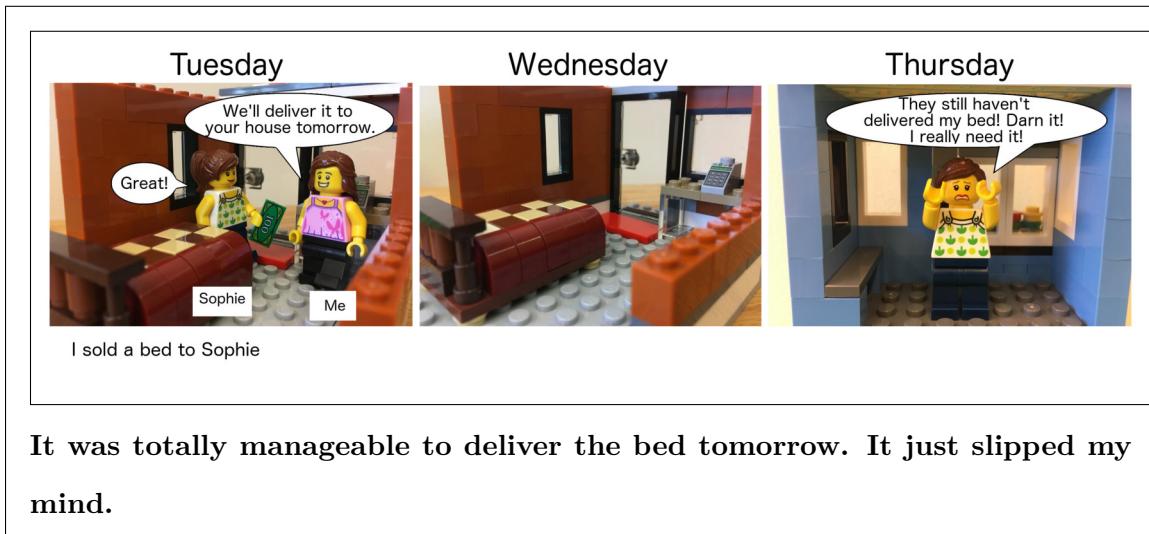


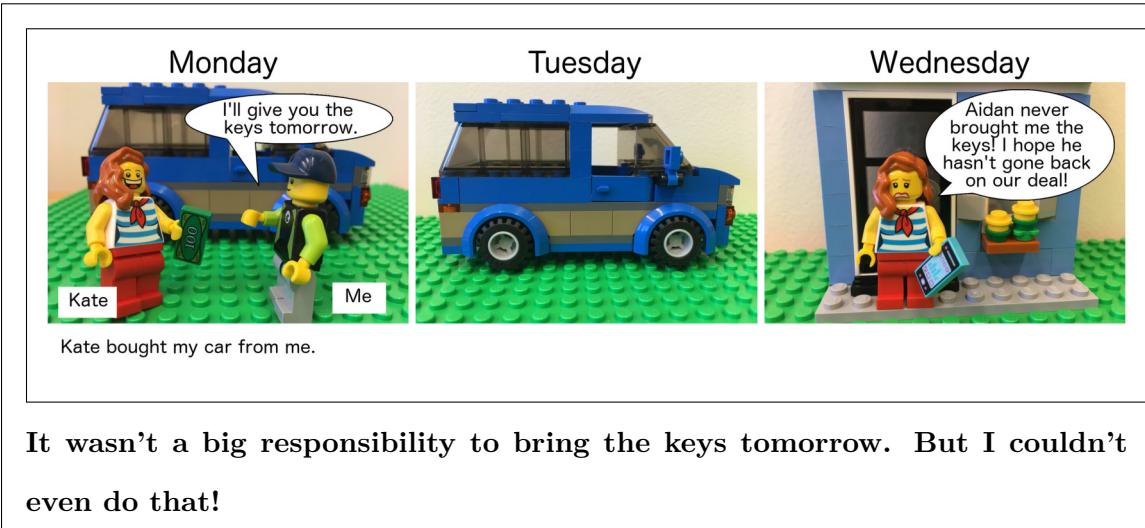
I am annoyed because Athena said that she would give me a parrot.

## APPENDIX J

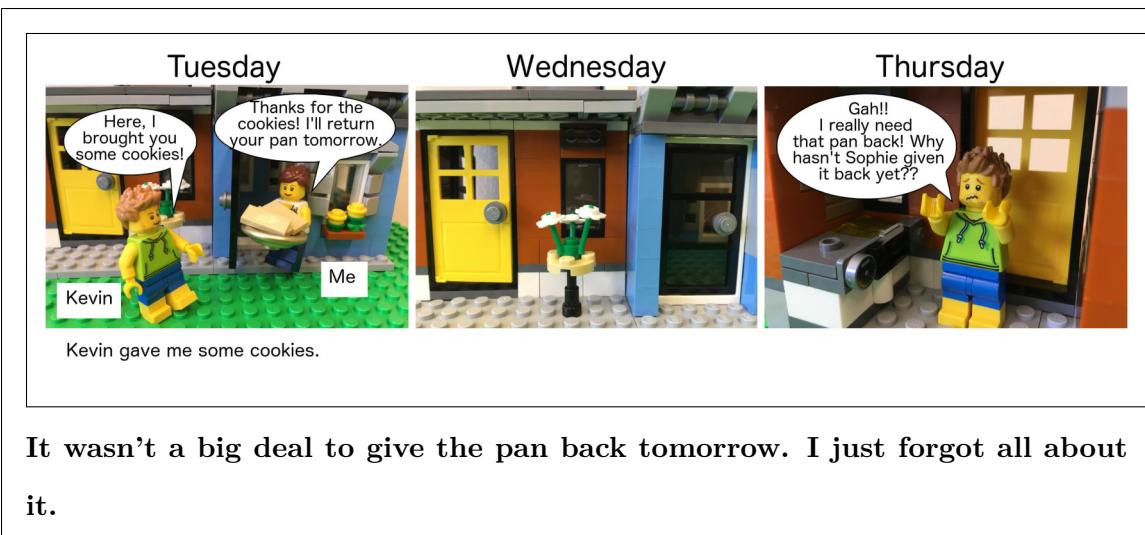
### EXPERIMENT 4 ITEMS

#### J.1 Main items

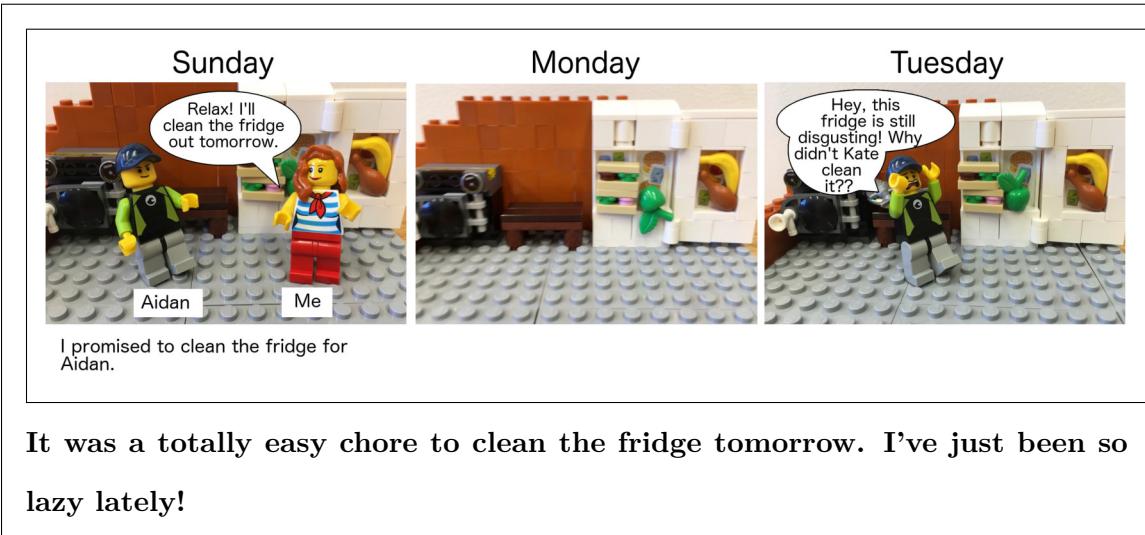




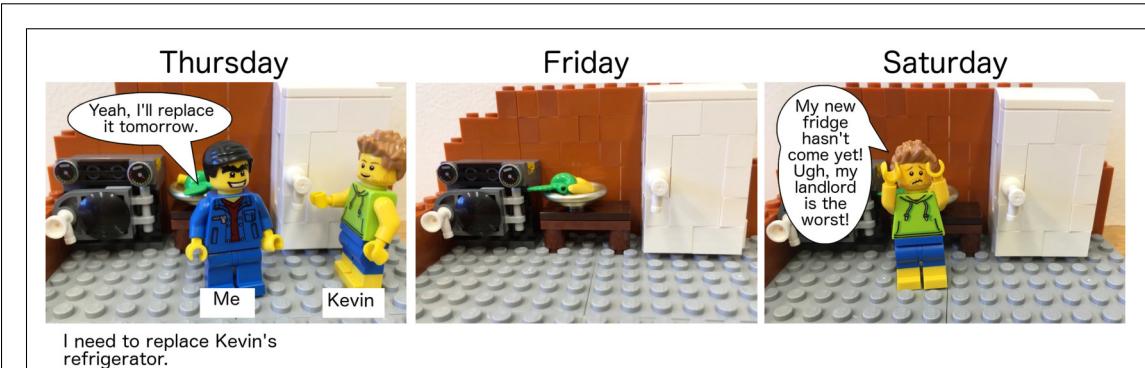
**It wasn't a big responsibility to bring the keys tomorrow. But I couldn't even do that!**



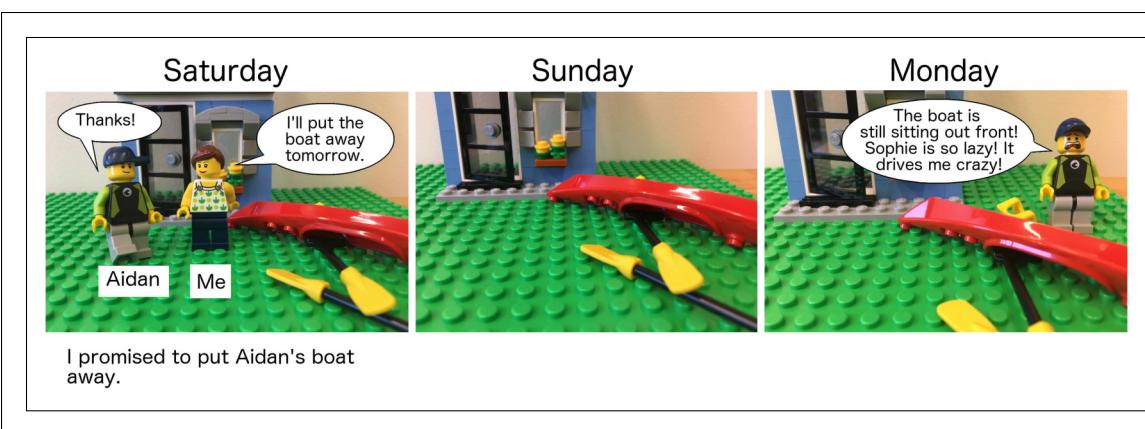
**It wasn't a big deal to give the pan back tomorrow. I just forgot all about it.**



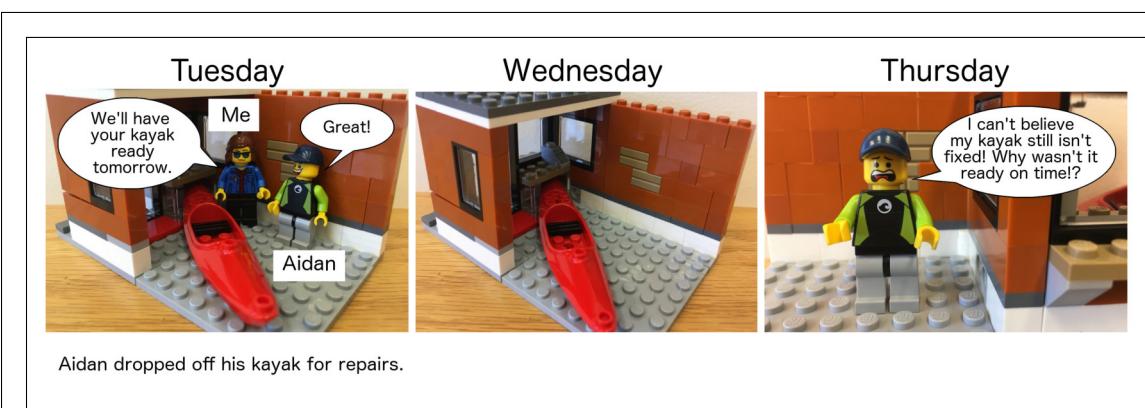
**It was a totally easy chore to clean the fridge tomorrow. I've just been so lazy lately!**



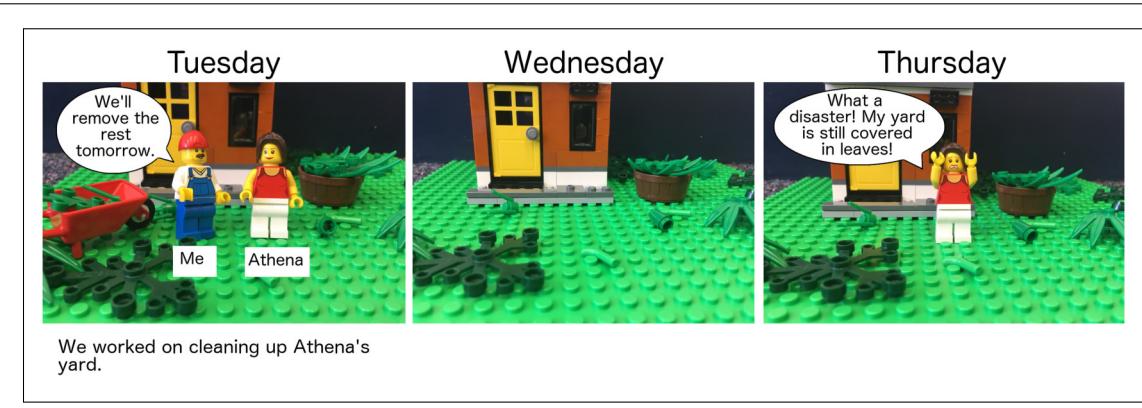
**It was a simple undertaking to replace the fridge tomorrow. I just didn't do it!**



**It wasn't a big deal to put the kayak away tomorrow. But I was so tired!**



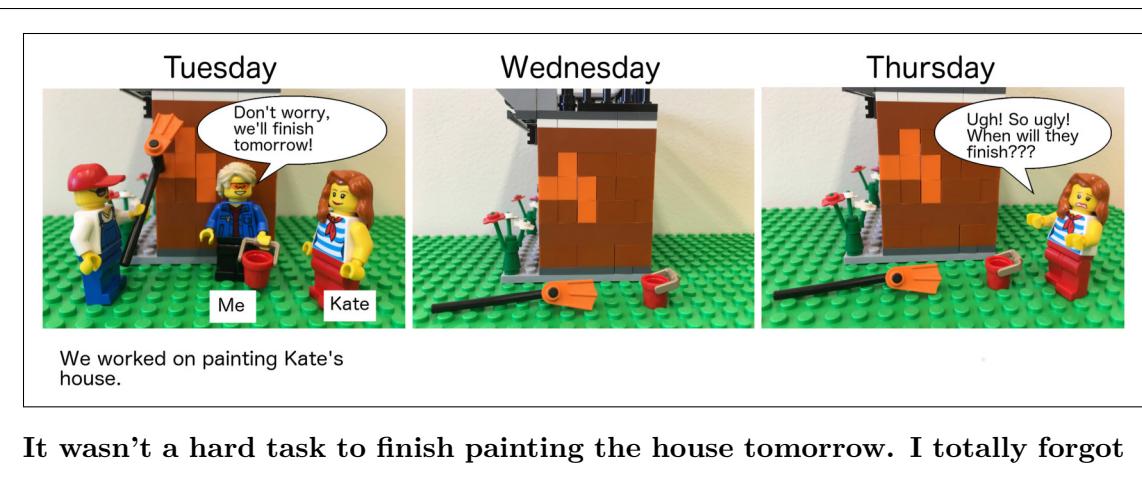
**It wasn't a big request to have the kayak ready tomorrow. But I just got so busy!**



**It was such a simple service to finish up the yard tomorrow. I can't believe I forgot about it!**



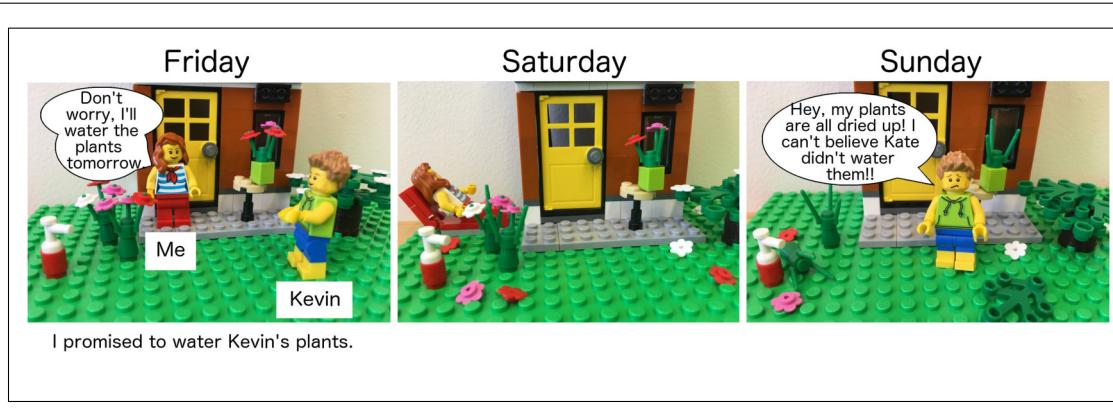
**It wasn't difficult to deliver the flowers tomorrow. We just lost the order slip.**



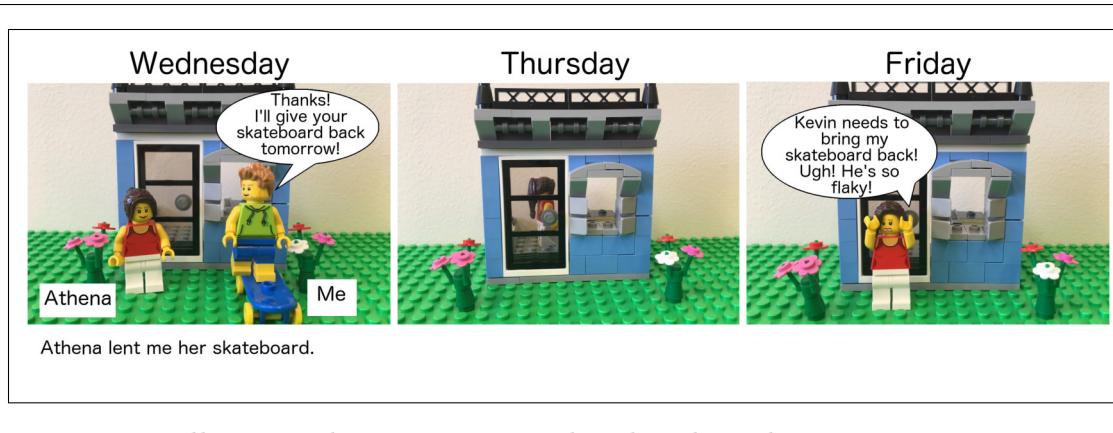
**It wasn't a hard task to finish painting the house tomorrow. I totally forgot that I had to be out of town!**



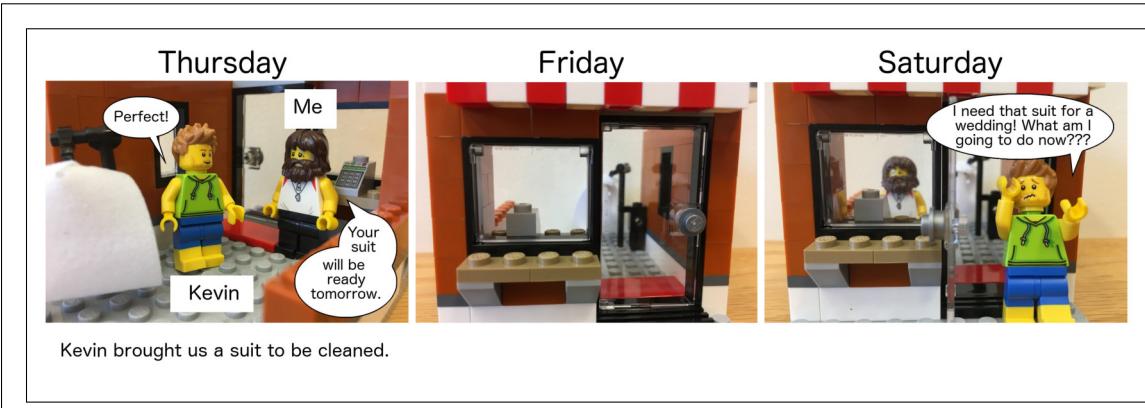
It wasn't a difficult thing to restock the iPhones tomorrow. But I just forgot to do it!



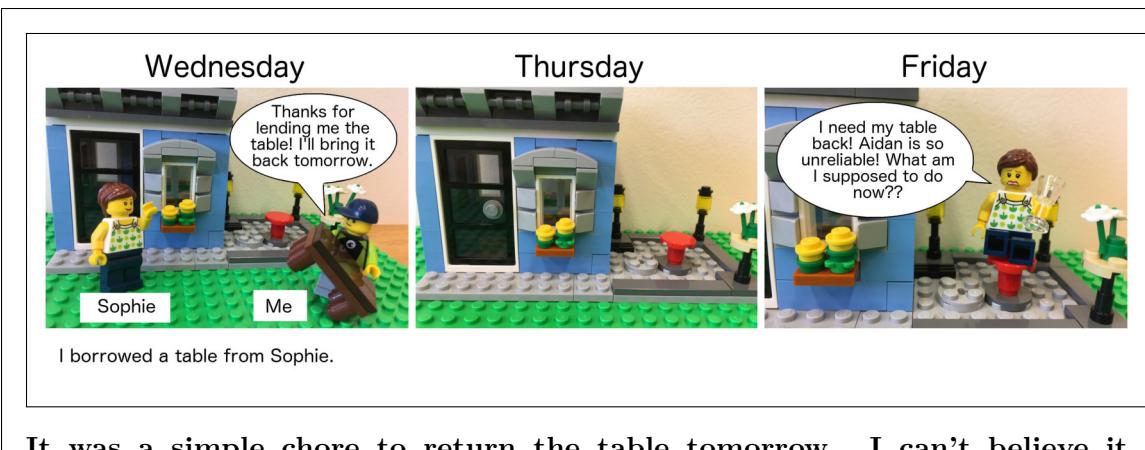
It wasn't a huge favor to water the plants tomorrow. I just had too many things to do.



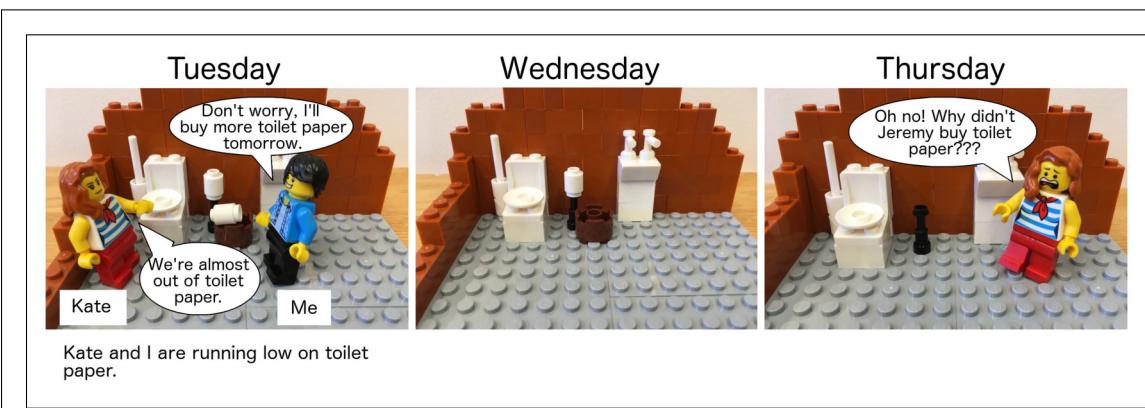
It was a really easy thing to return the skateboard tomorrow. I was just feeling so lazy!



**It was pretty easy to have the suit ready tomorrow. We just had too much work to do.**



**It was a simple chore to return the table tomorrow. I can't believe it slipped my mind!**



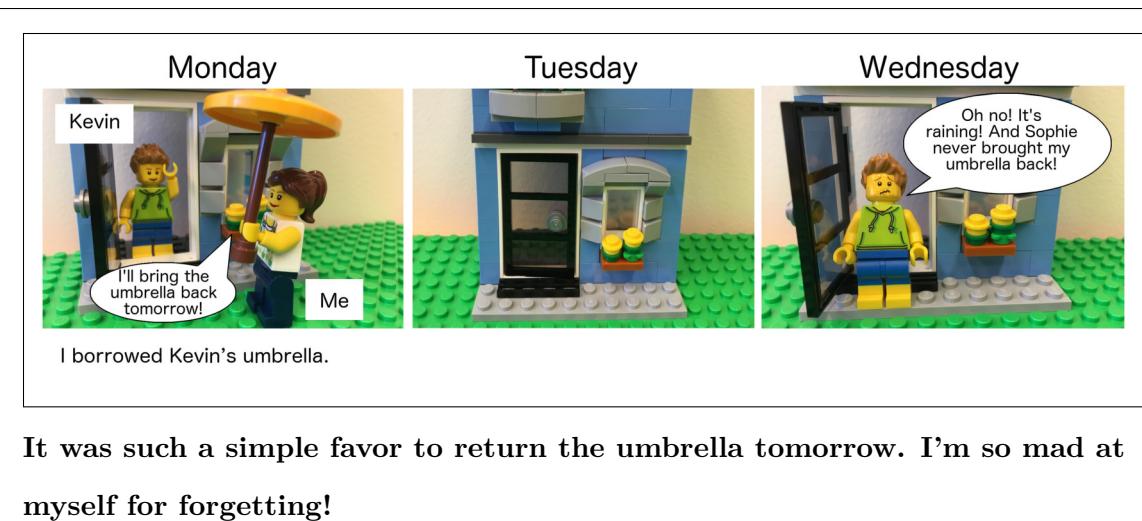
**It was a fairly routine chore to buy toilet paper tomorrow. I just lost track of all the things I had to do.**



**It was an easy job to finish hauling the tree tomorrow. We just got tied up with other jobs.**

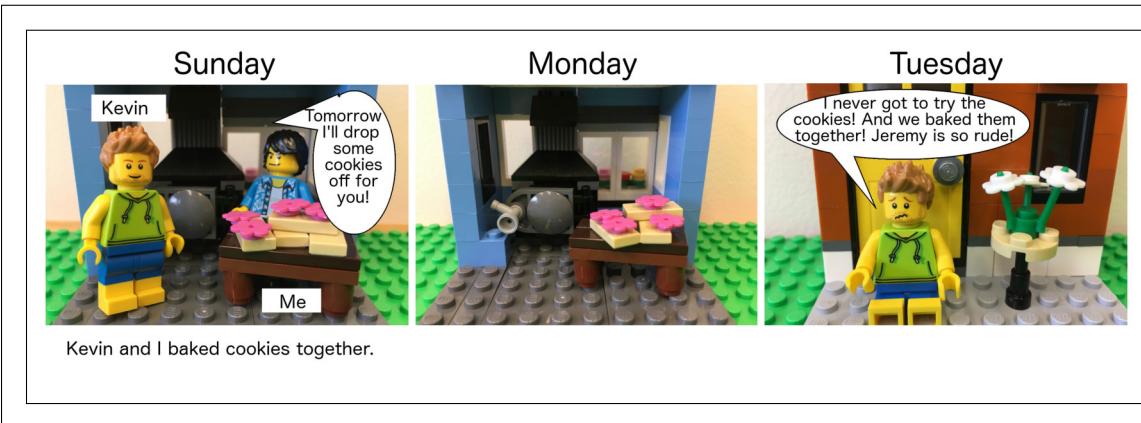


**It wasn't a big deal to drive back tomorrow. We just didn't think Kate would mind if we stayed longer.**

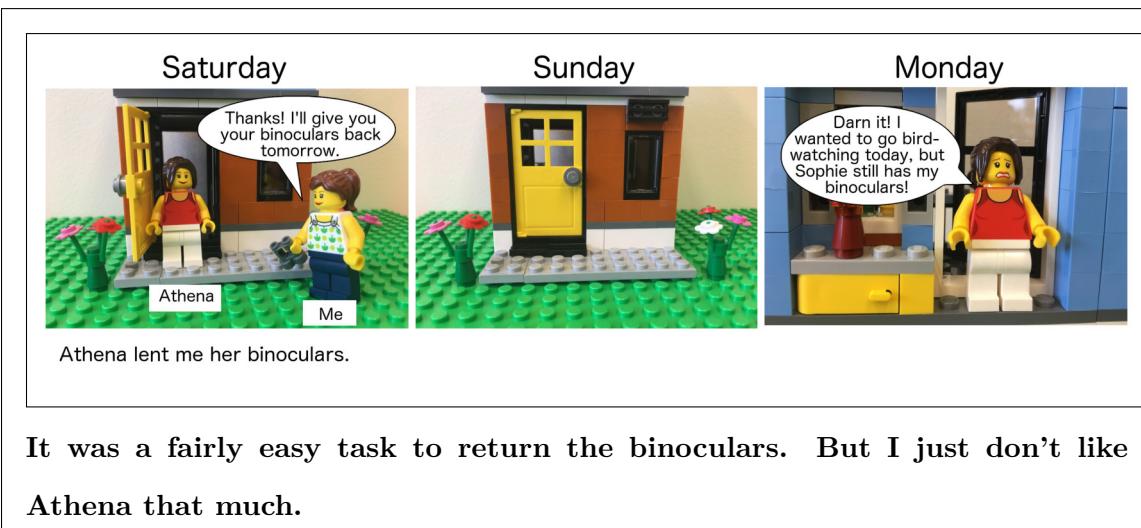


**It was such a simple favor to return the umbrella tomorrow. I'm so mad at myself for forgetting!**

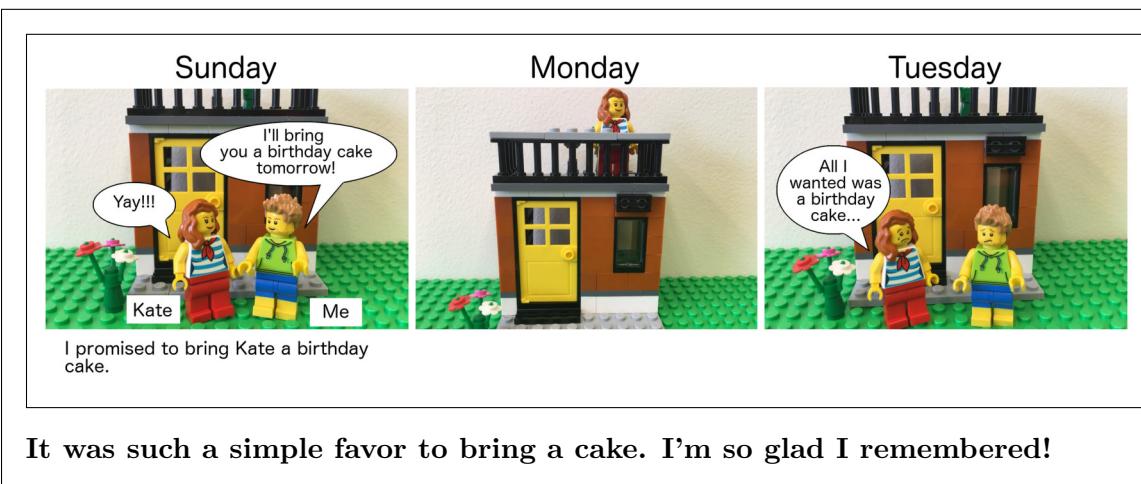
## J.2 Fillers

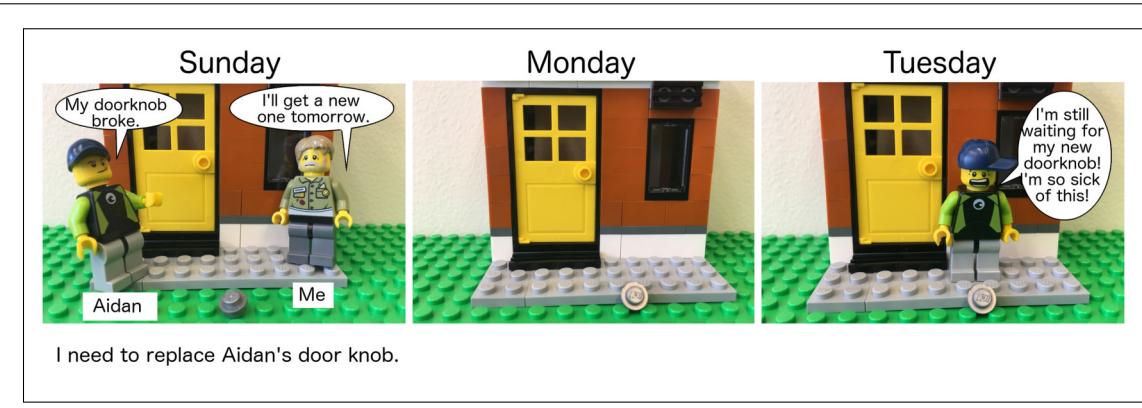


The nice thing to do was to give Kevin some of the cookies. But he wasn't very nice to me, so I didn't.



It was a fairly easy task to return the binoculars. But I just don't like Athena that much.





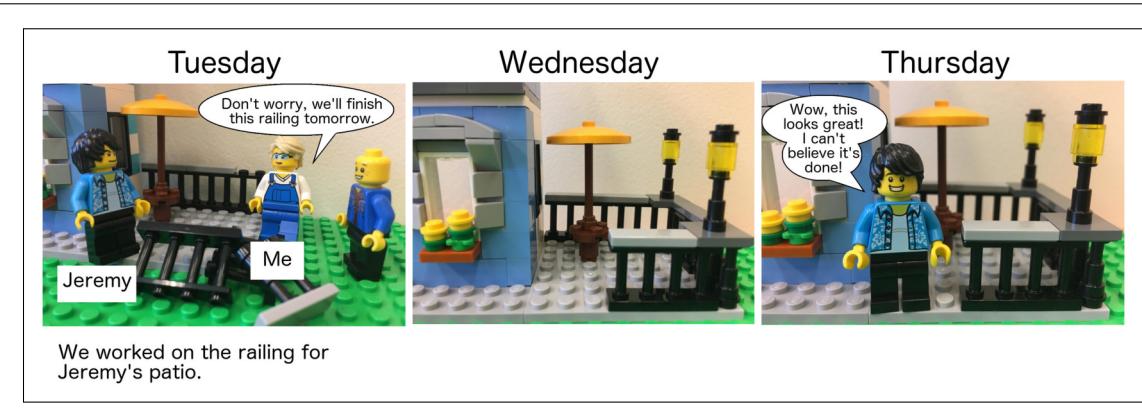
**It was such a straightforward job to replace the doorknob. I totally forgot to do it though.**



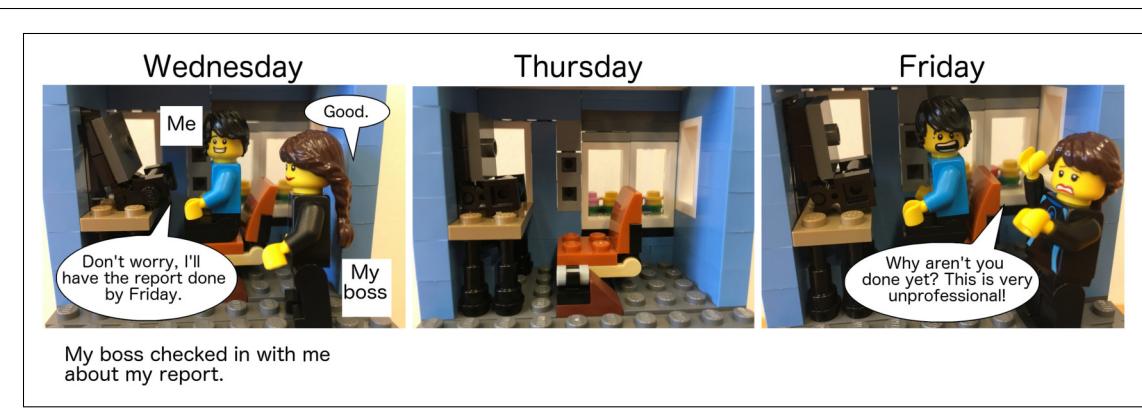
**It wasn't difficult to steal from Sophie. I don't feel bad about it at all!**



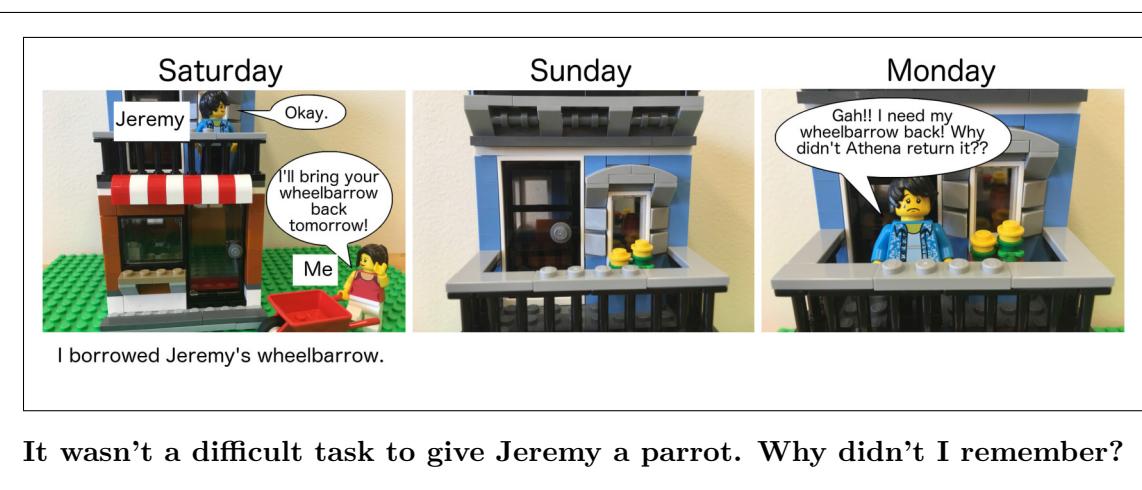
**It was an easy task to bring crackers to bring to the party. But I just forgot!**



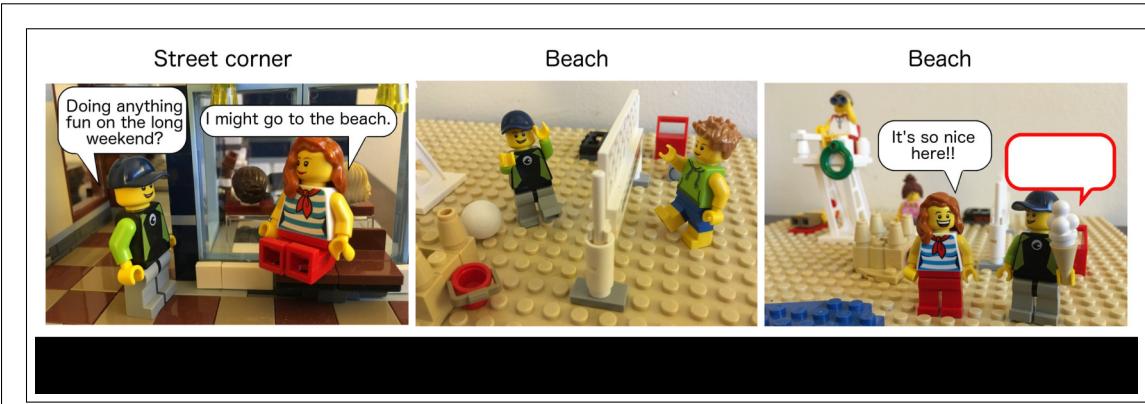
**It was an easy task to finish the railing. I'm glad we were able to get it done!**



**It was such an easy assignment to finish the report. I just didn't get to it!**



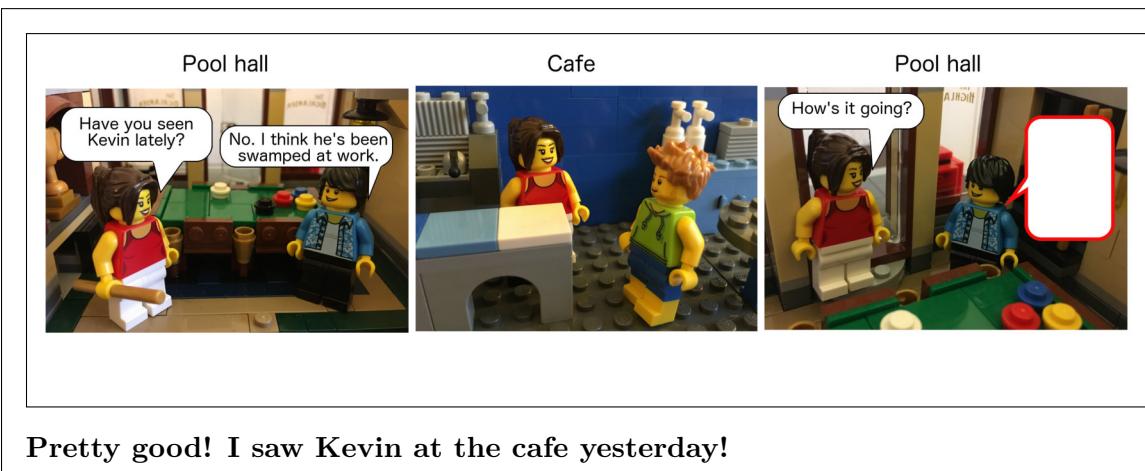
**It wasn't a difficult task to give Jeremy a parrot. Why didn't I remember?**



**Yeah! Kevin came with me yesterday. We had a blast!**



**This guy rode in on his bike! It was so rude!**



**Pretty good! I saw Kevin at the cafe yesterday!**

## APPENDIX K

### QUANTIFICATIONAL BINDING TASK 1 ITEMS

#### **K.1 Time quantification items**

1. Every time Kevin washes his car, it rains {tomorrow / the next day }.
2. Whenever Sophie drinks red wine, she oversleeps {tomorrow / the next day }.
3. Every time Jeremy has a loud party, his neighbors are grumpy {tomorrow / the next day }.
4. When the Seahawks win, Kate's coworkers all wear blue and green to work {tomorrow / the next day }.

#### **K.2 Speech report quantification items**

1. Every time Aidan tells his boss that he will work late, she comments on how tired he is tomorrow.
2. Every time Sophie says that she is going out for drinks after work, her boss tells her to come in early {tomorrow / the next day }.
3. Whenever Athena says that it is going to rain, it's {tomorrow / the next day }.
4. Every time Athena's neighbor says that he is going to have a party, she makes sure that she can sleep in {tomorrow / the next day }.
5. Whenever Jeremy says that it's going to snow, Athena wears her boots {tomorrow / the next day }.
6. Every time the office supervisor says there will be random inspections, Kevin makes sure his office is immaculate {tomorrow / the next day }.

7. Whenever Jeremy says that he is flying in late, his supervisor lets him take the morning off {tomorrow / the next day }.
8. Each time Kate tells her boss that her project will be finished today, she's still working on it {tomorrow / the next day }.
9. Each time Sophie says that she is going to take out the trash, Kevin ends up having to do it {tomorrow / the next day }.
10. Every time Kevin invites his coworkers for drinks after work, everyone shows up late {tomorrow / the next day }.
11. Whenever Aidan announces that he is throwing a party, everyone calls in sick {tomorrow / the next day }.
12. Every time the UPS person says that the package has been delivered, it doesn't show up until {tomorrow / the next day }.

## APPENDIX L

### QUANTIFICATIONAL BINDING TASK 2 ITEMS

#### L.1 Naturally occurring *tomorrow* items

1. I try to start winding down by 9pm so that I have time to tidy up the kitchen, think through what I'm going to wear tomorrow, and generally get my things together. That way I can sleep in as long as possible and still get to work on time!
2. One of my professors wears a different hat to work each day. He even has a website where you can pick a hat for him to wear tomorrow.
3. I set up an app on my phone that notifies me every time tomorrow's forecast calls for snow.

#### L.2 Perspectival *tomorrow* items

1. On Christmas Eve, every little girl stays awake for hours wondering what she will find under the Christmas tree tomorrow morning.
2. Every time you have to kick a drunk idiot out of the bar, you get to gloat about how hungover the jerk will be tomorrow.
3. My coworker is such a brat. Every time the jerk thinks it'll be sunny tomorrow, he calls in "sick" and I have to cover his shift.

#### L.3 Time quantification items

1. When the Seahawks win, our boss tells us that we can wear blue and green to work tomorrow.

2. Whenever Jeremy arrives late at night, his supervisor lets him take the morning off the next day.
3. Whenever Sophie drinks red wine, she oversleeps the next day.

#### **L.4 Speech report quantification items**

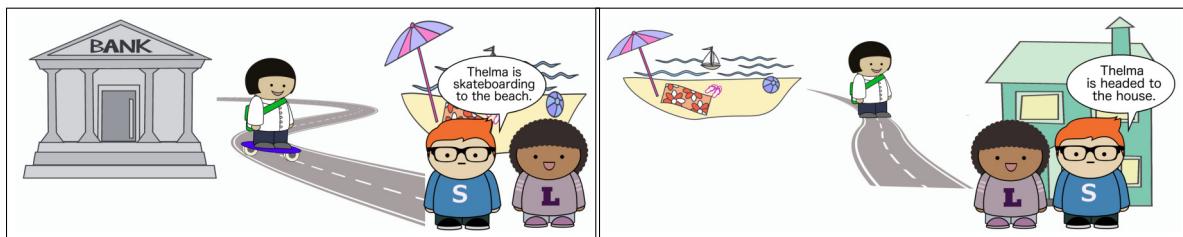
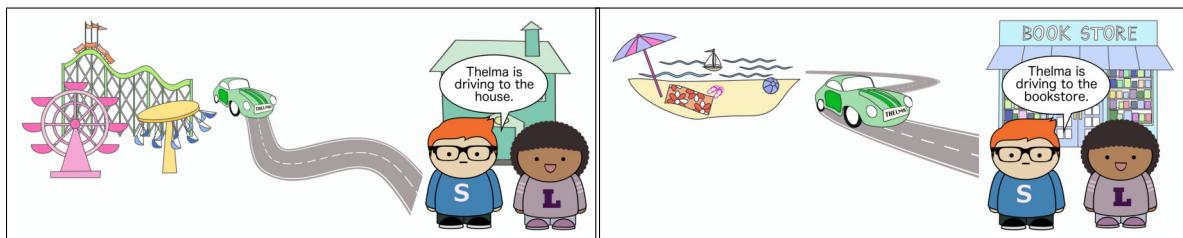
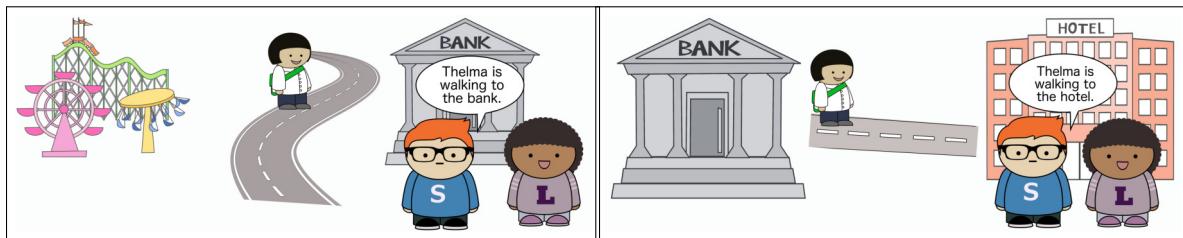
1. Every time my neighbor tells me he's hosting a party, I make sure I can go into work late tomorrow.
2. Each time Kate tells her boss that her project will be finished today, she's still working on it tomorrow.
3. Whenever Jeremy says that it's going to snow, we all wear our boots tomorrow.
4. Every time we say that we are going out for drinks after work, our boss tells us to come in early the next day.

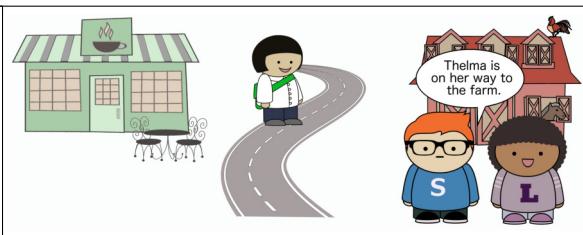
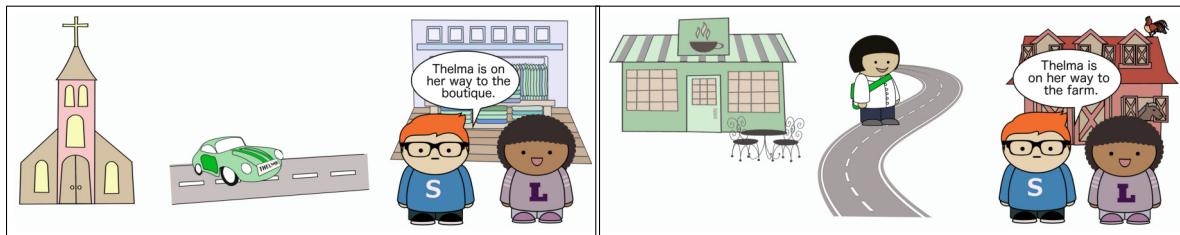
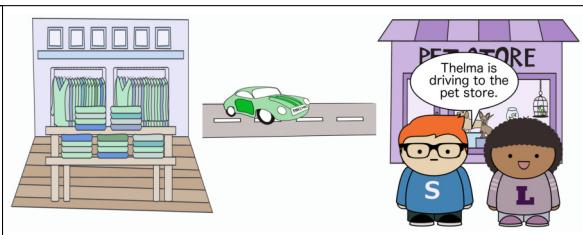
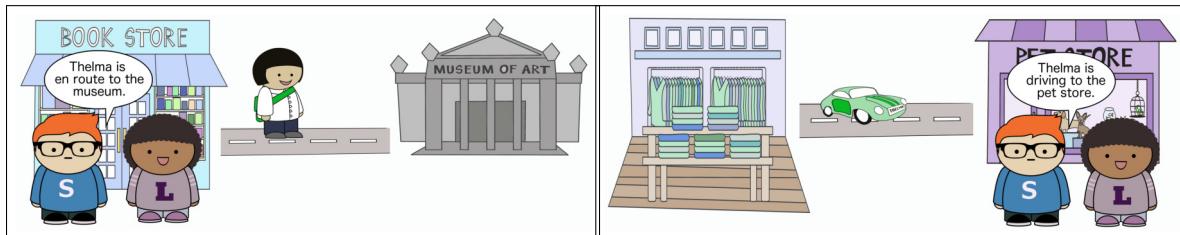
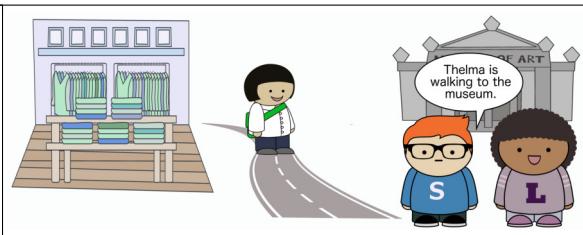
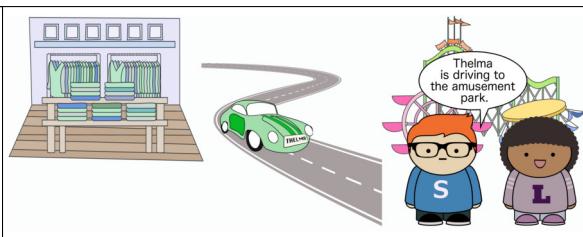
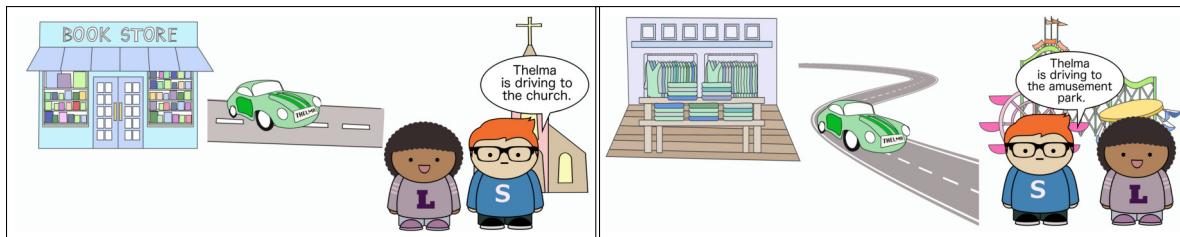
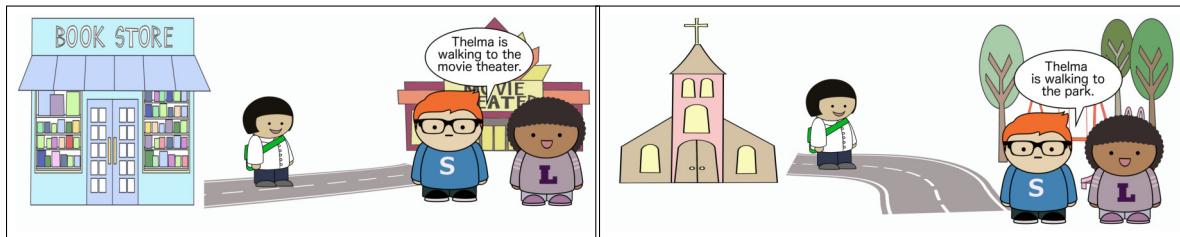
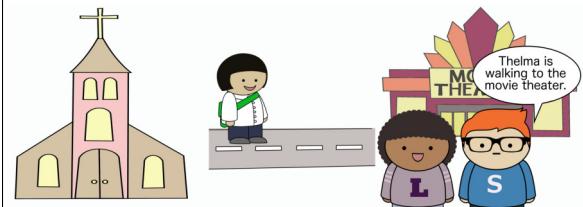
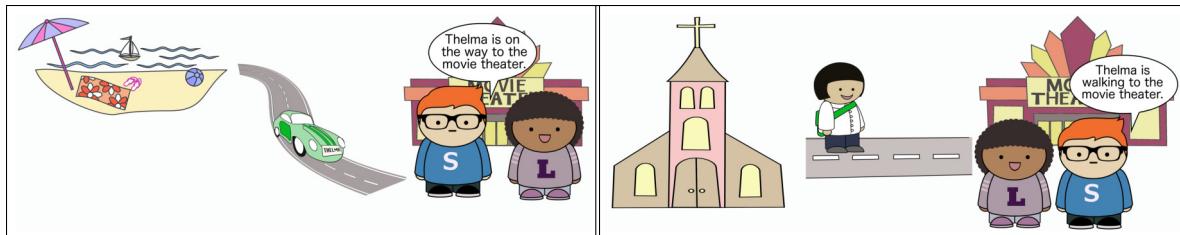
## APPENDIX M

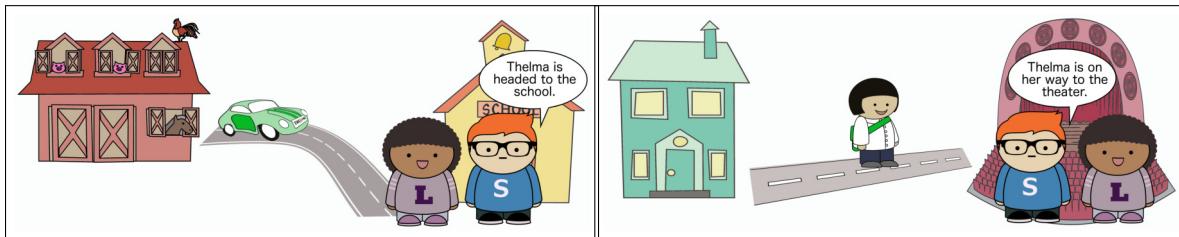
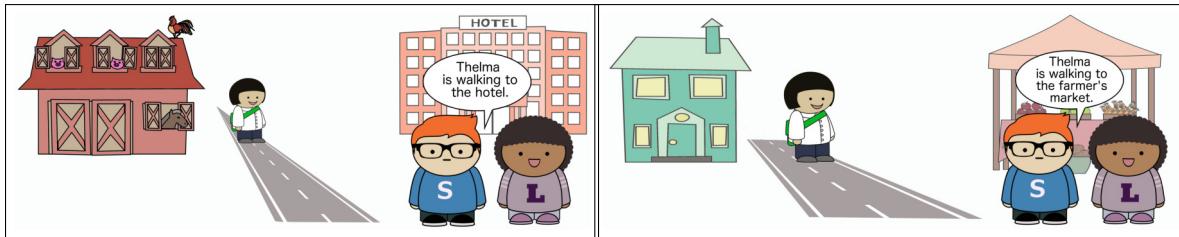
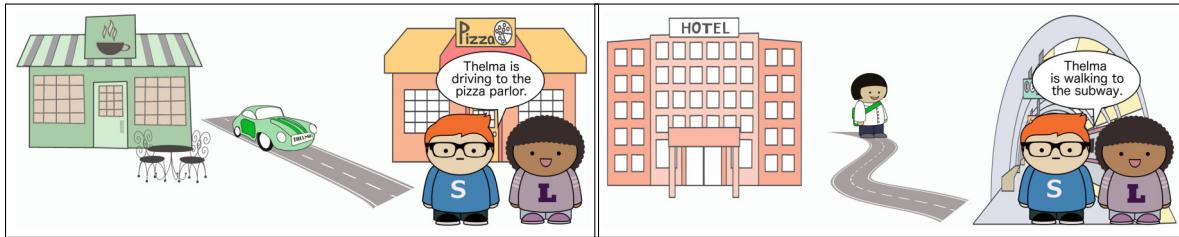
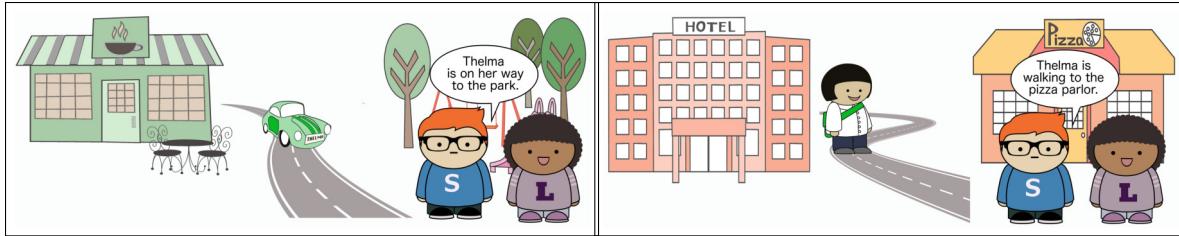
### EXPERIMENT 5A AND 5B ITEMS

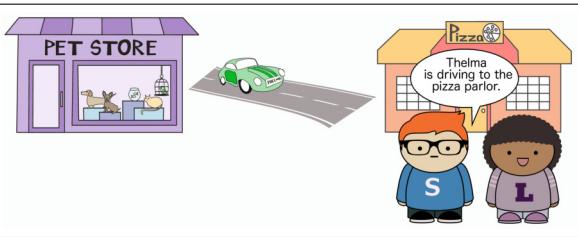
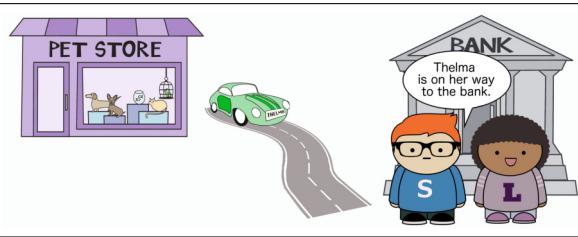
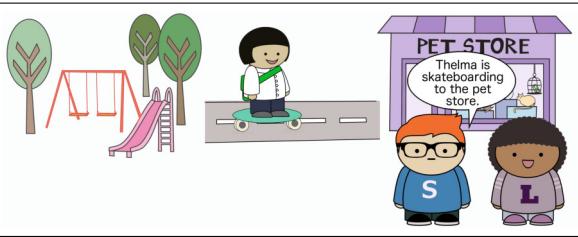
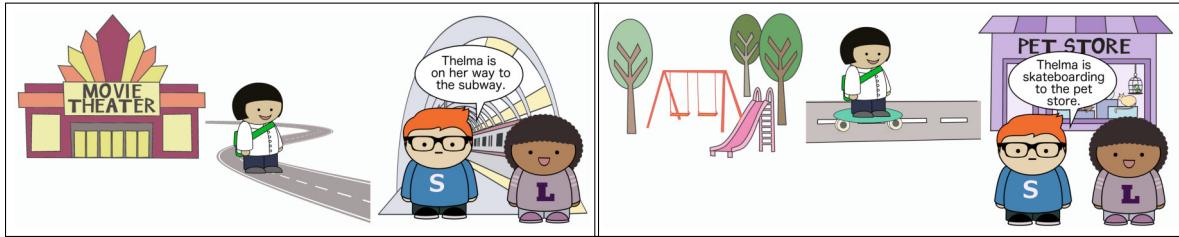
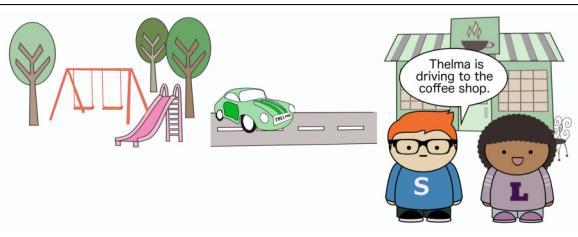
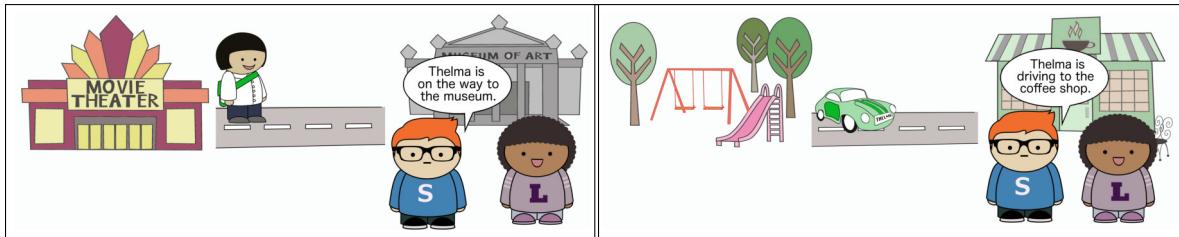
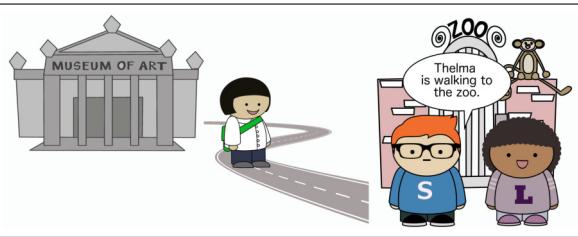
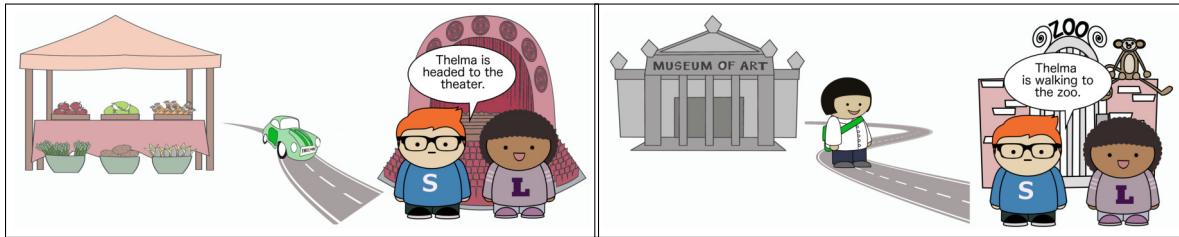
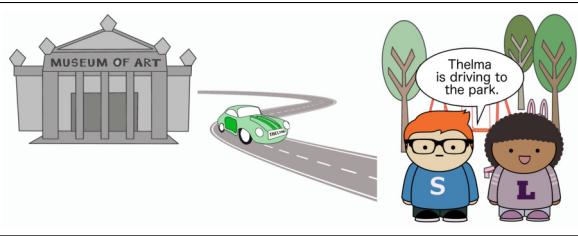
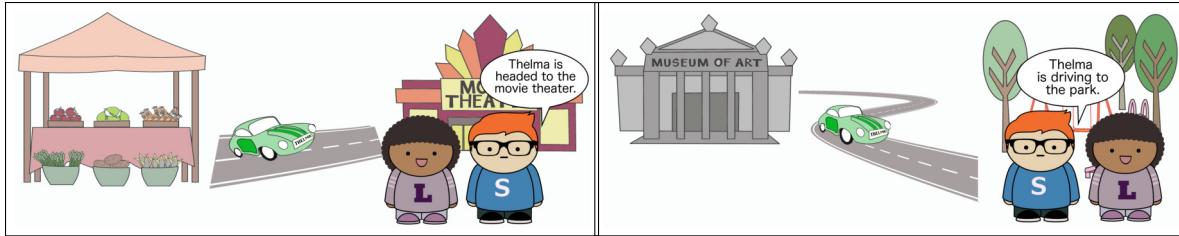
#### M.1 Main items

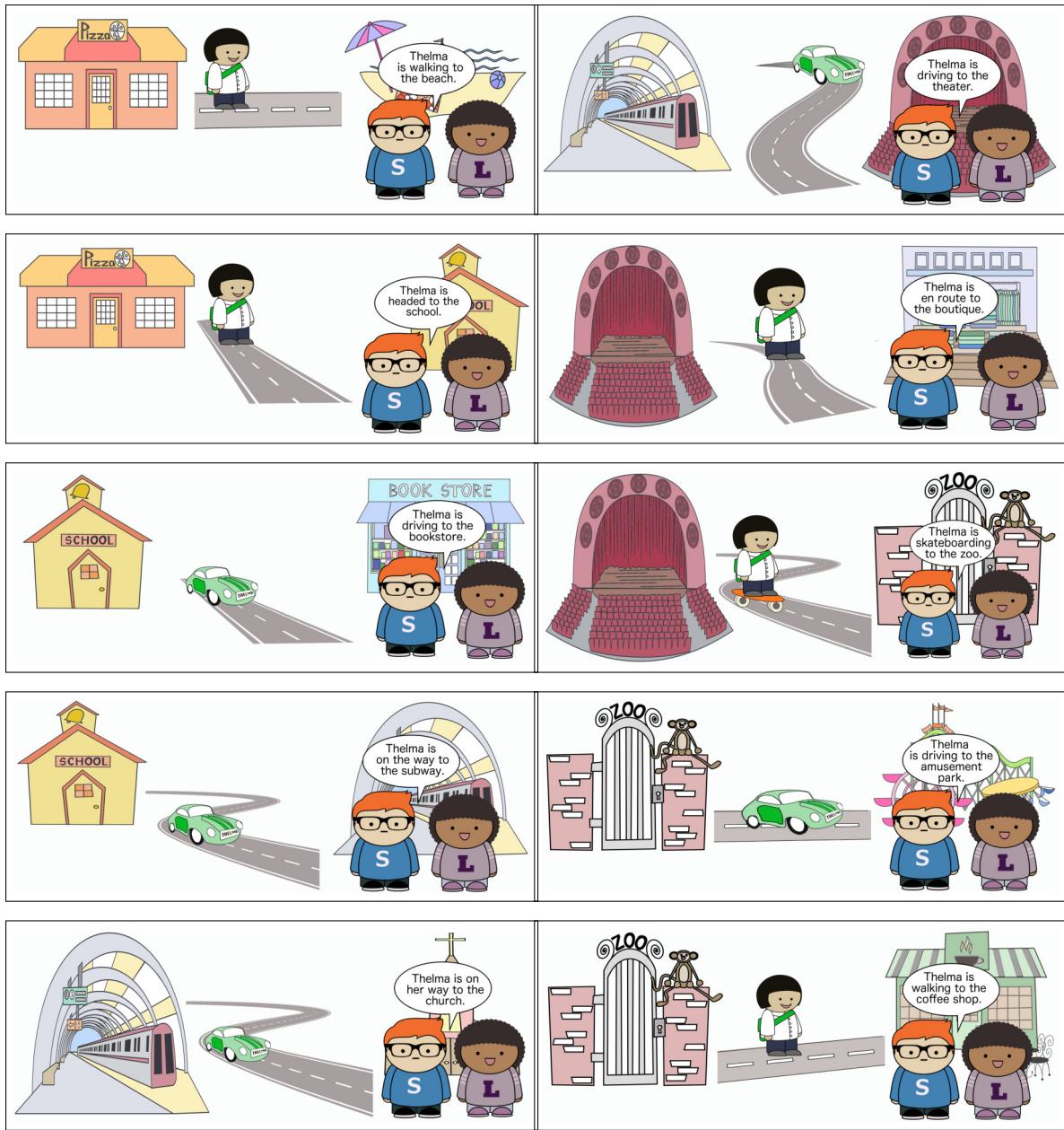
Each item below is shown in the Both condition. There were also Speaker, Listener, and None versions of each scene, as described in Chapter 7.



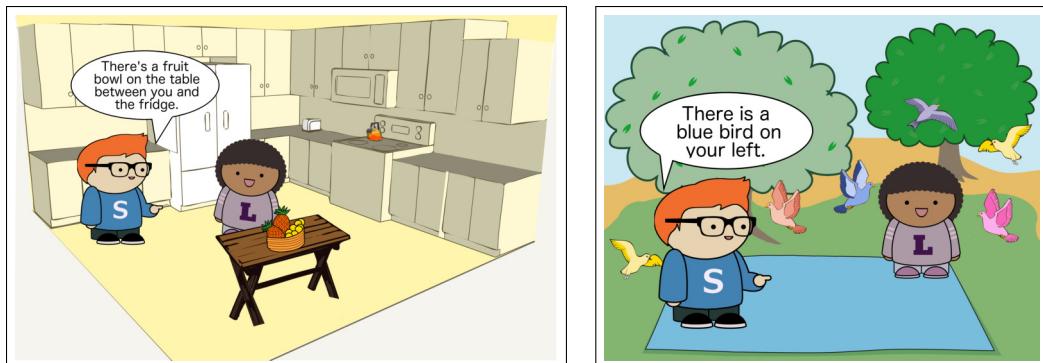


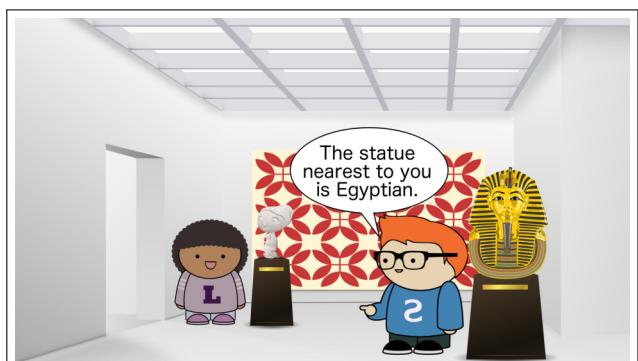
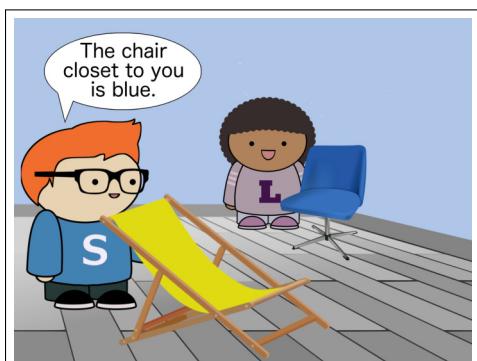
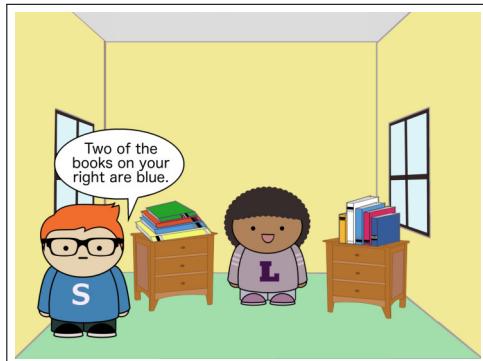
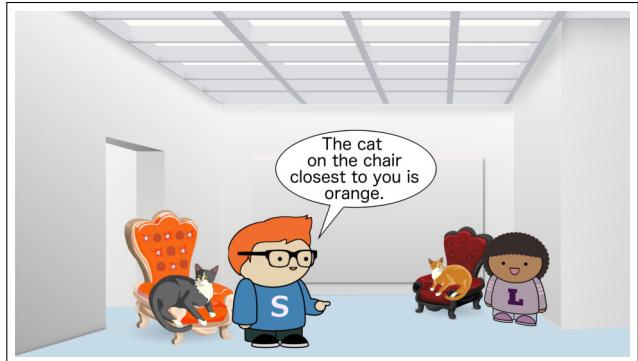


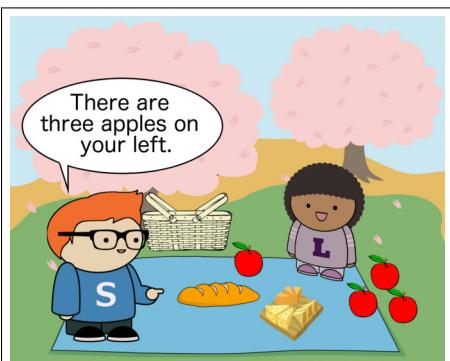
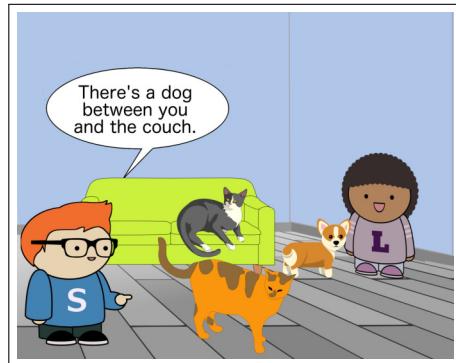
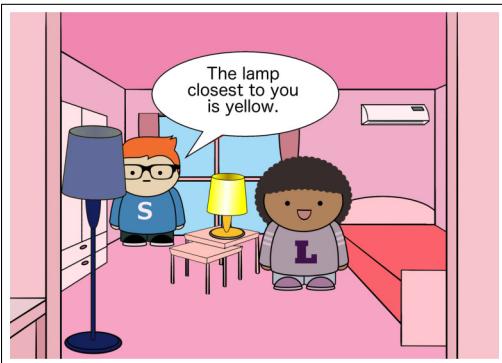
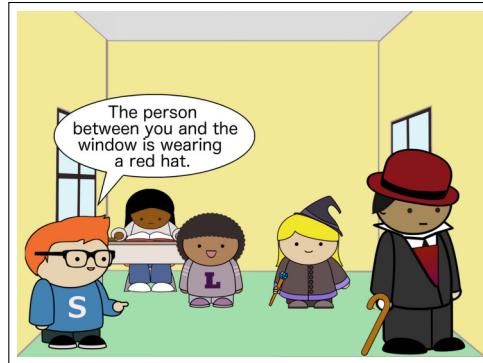
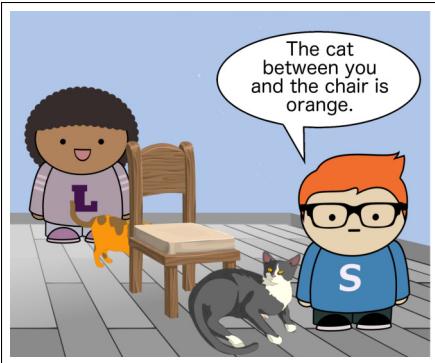


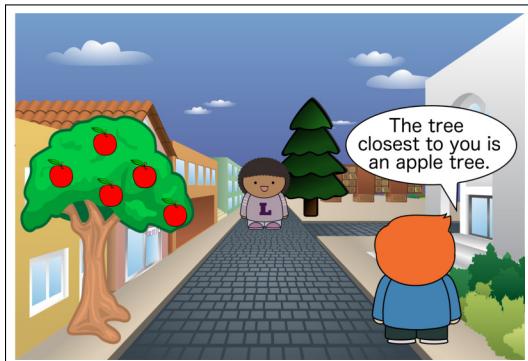
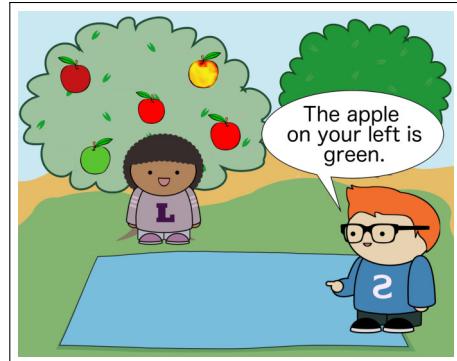


## M.2 Spatial task items

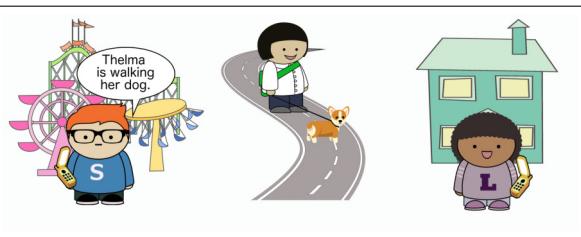
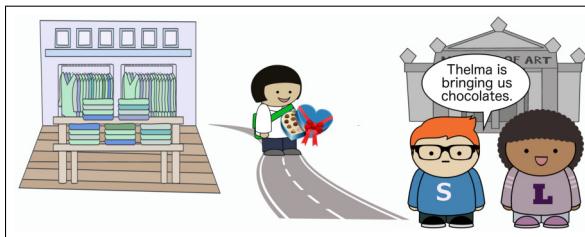
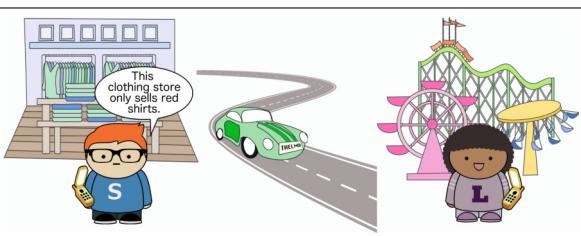
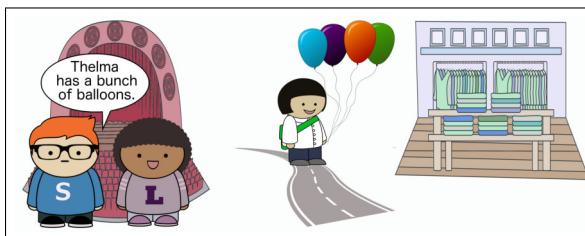
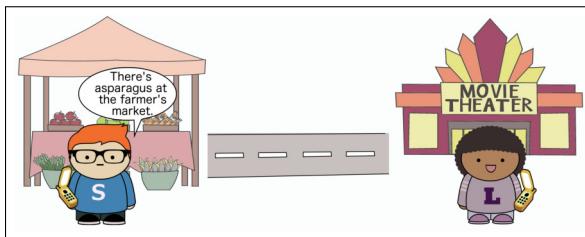


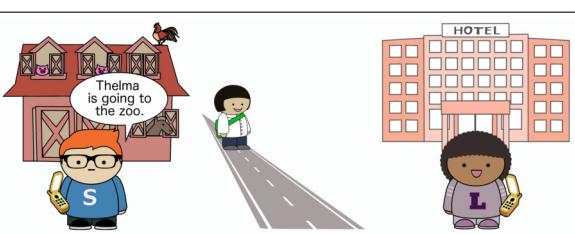
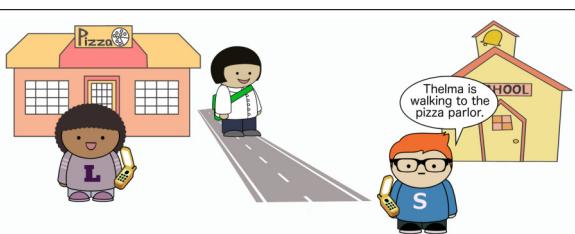
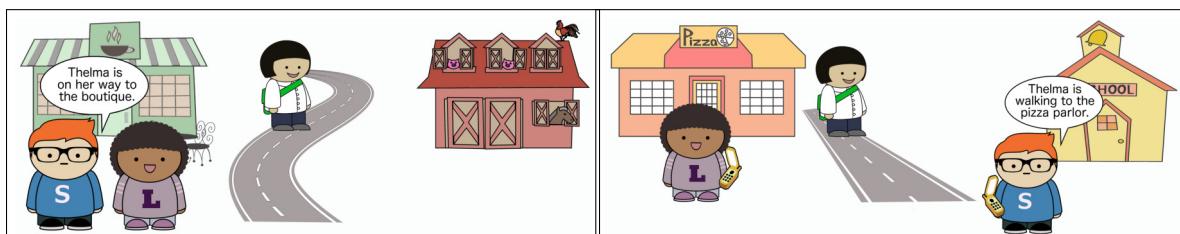
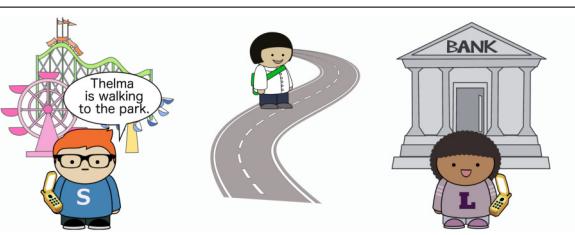
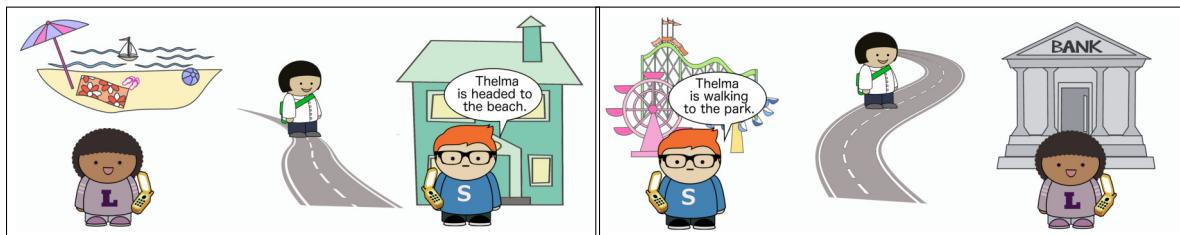
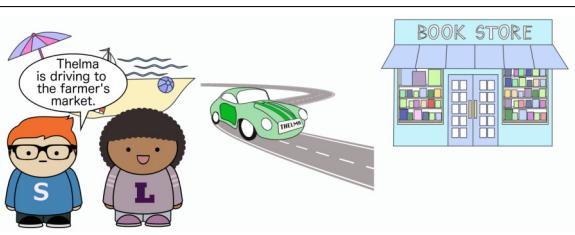
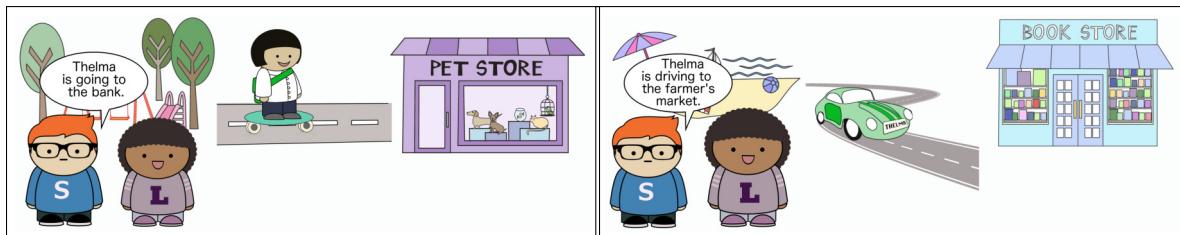
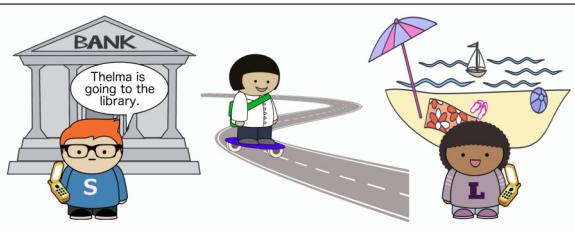
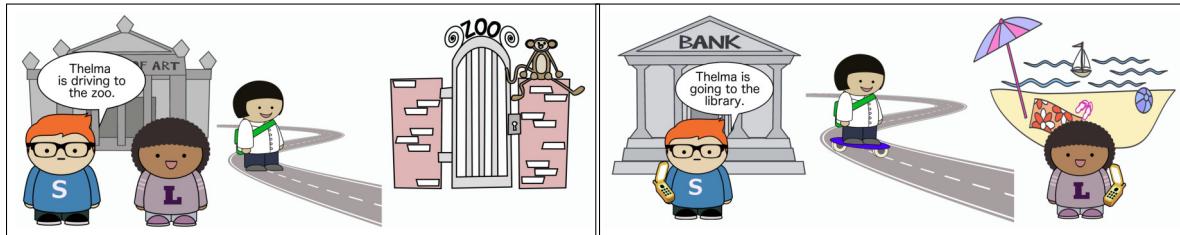
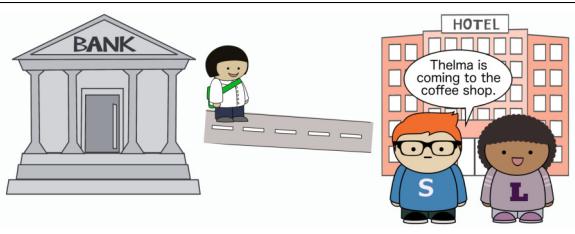
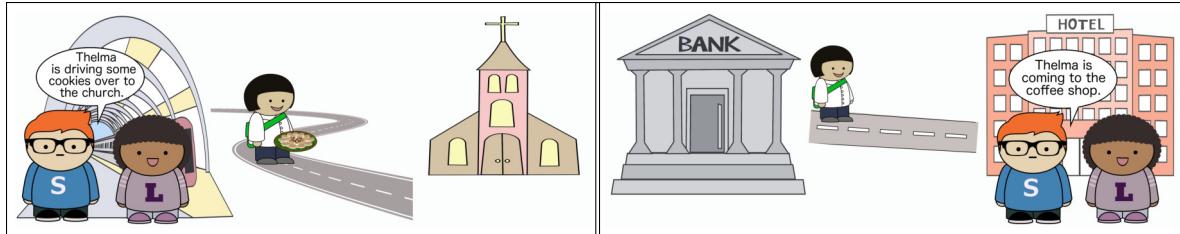


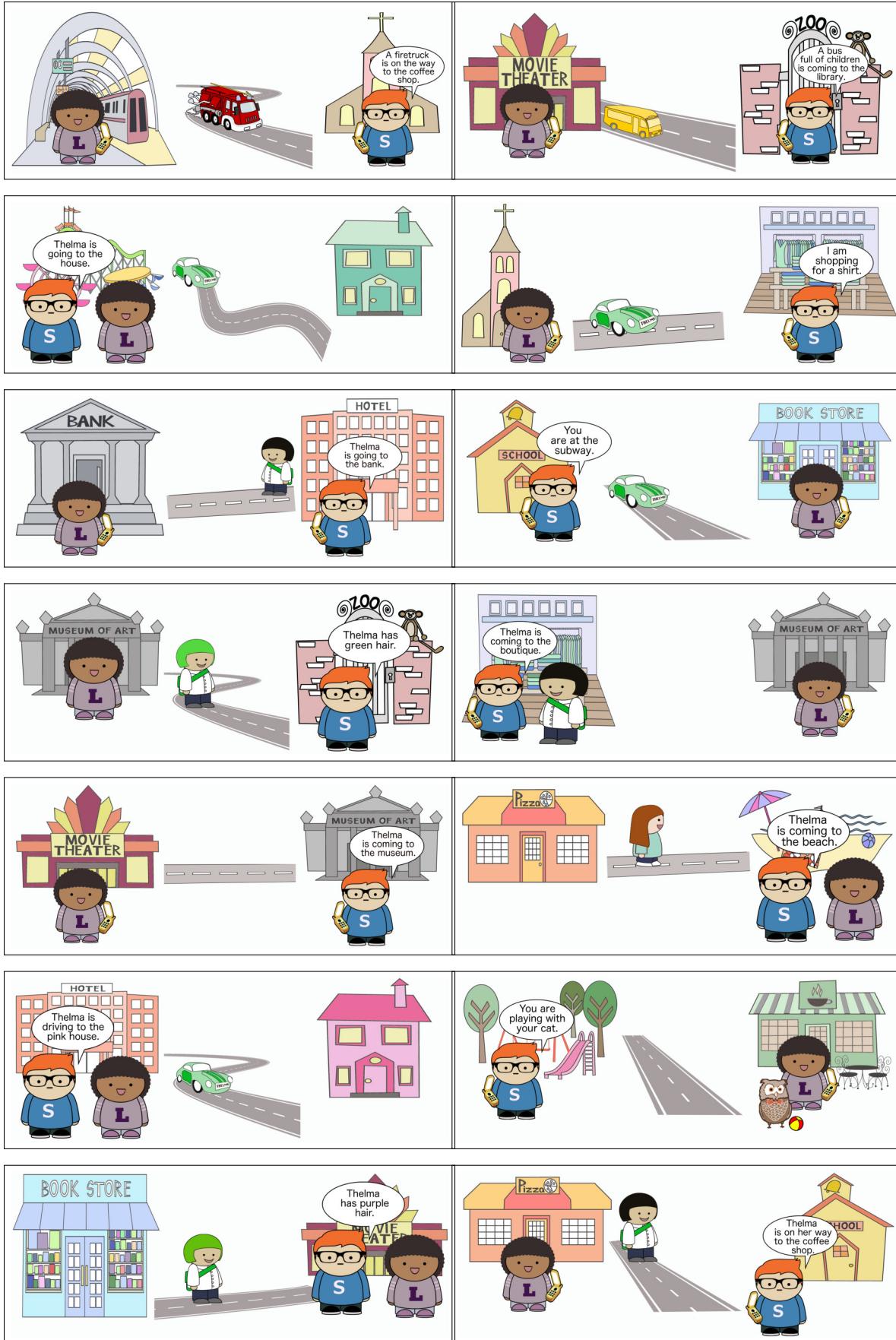




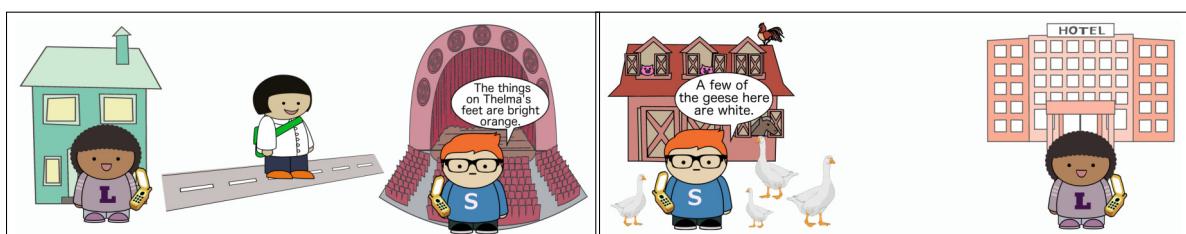
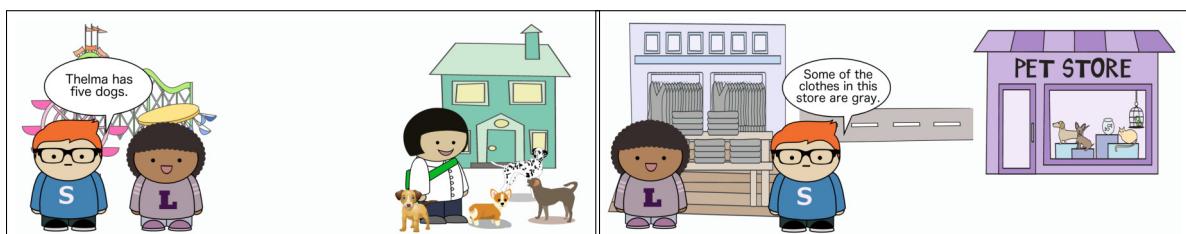
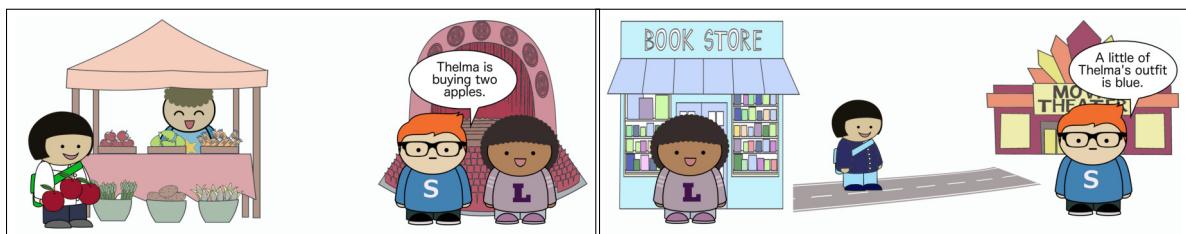
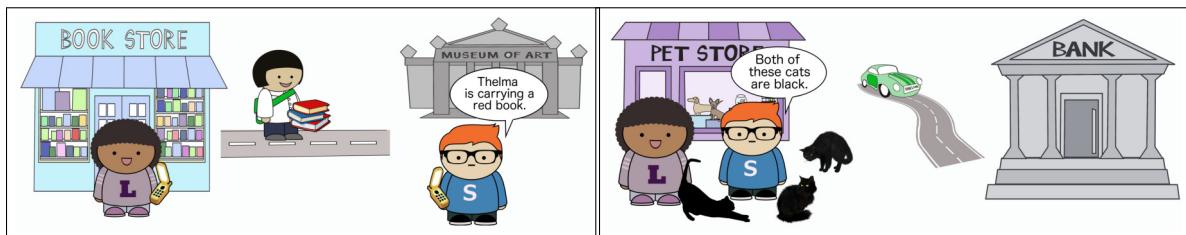
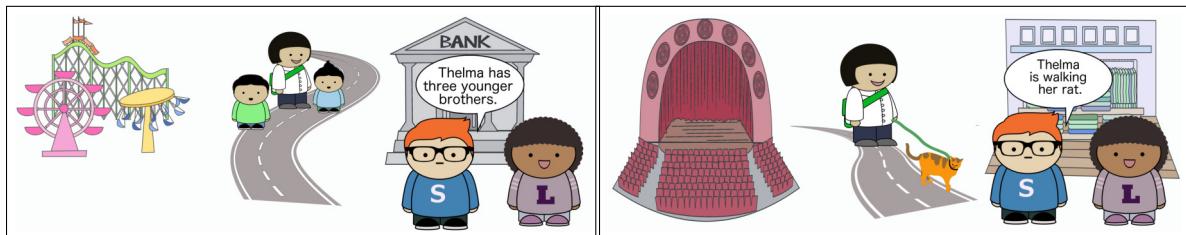
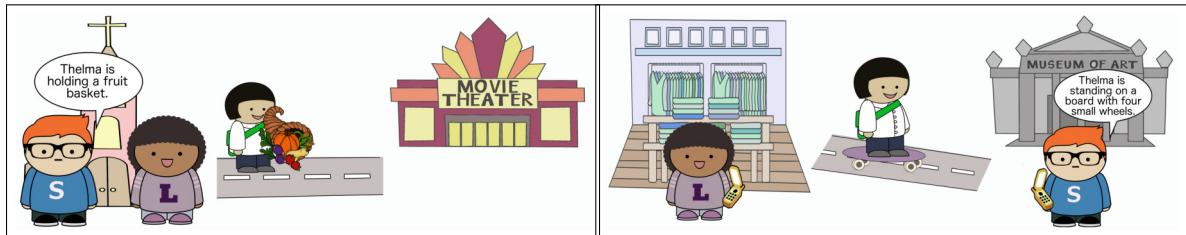
### M.3 Experiment 5a fillers

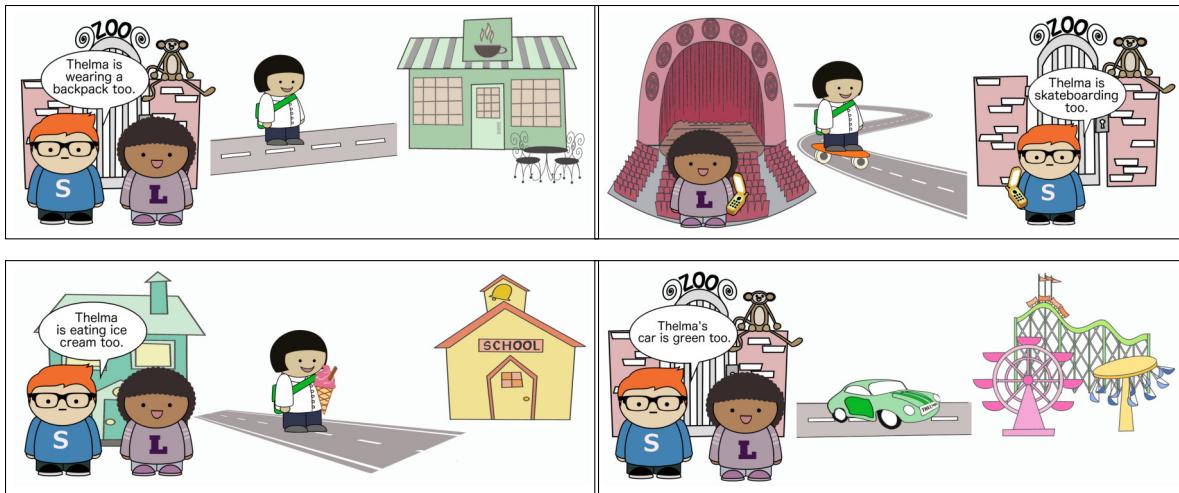






#### M.4 Additional fillers used in Experiment 5b



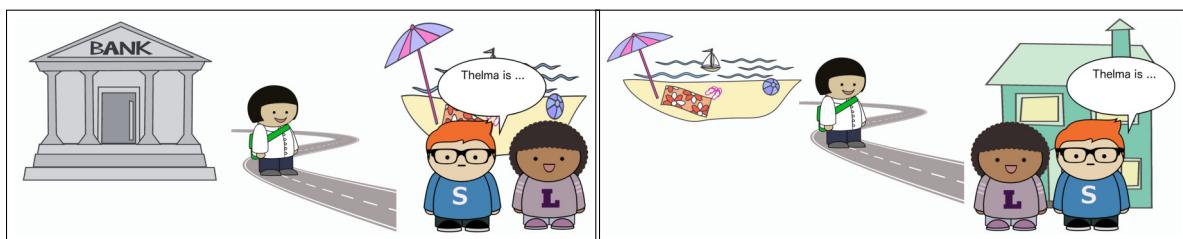
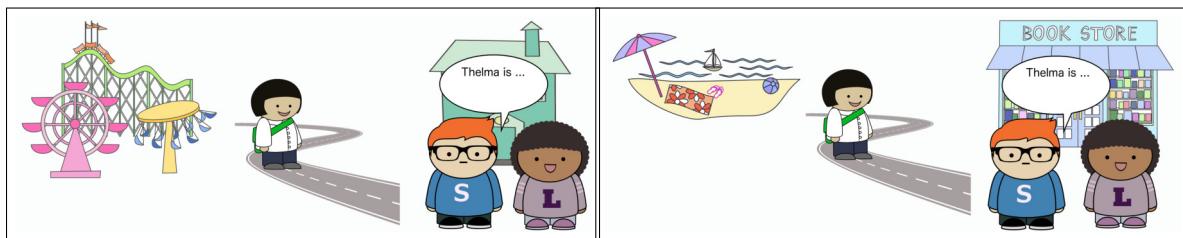
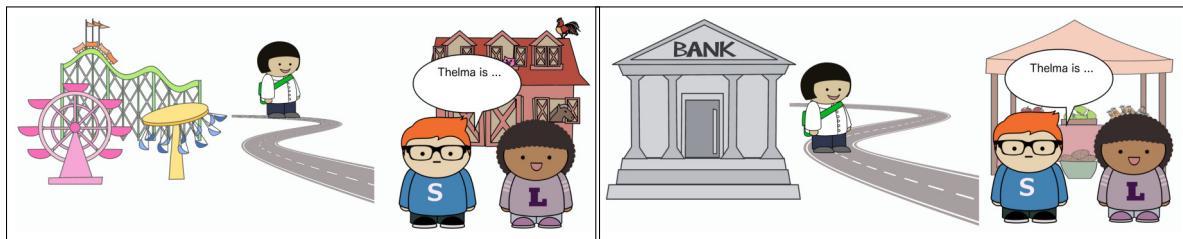
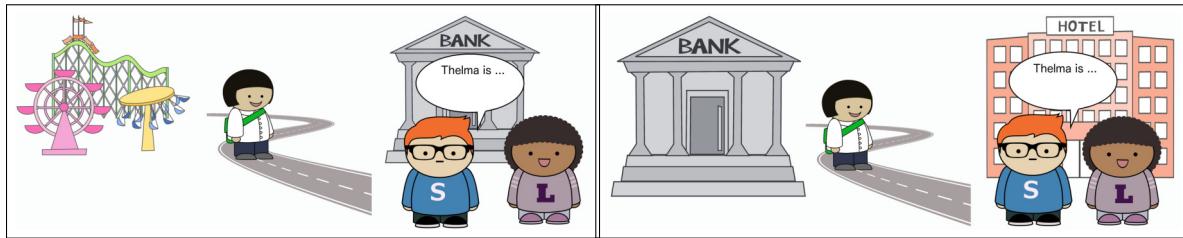


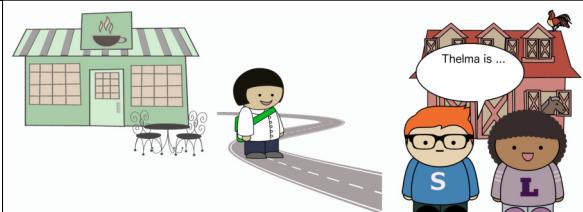
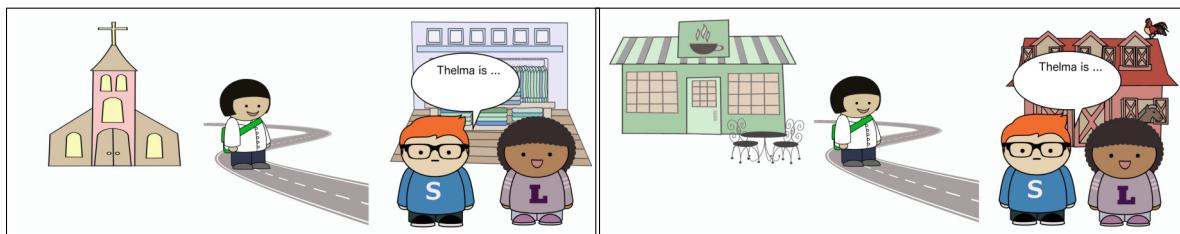
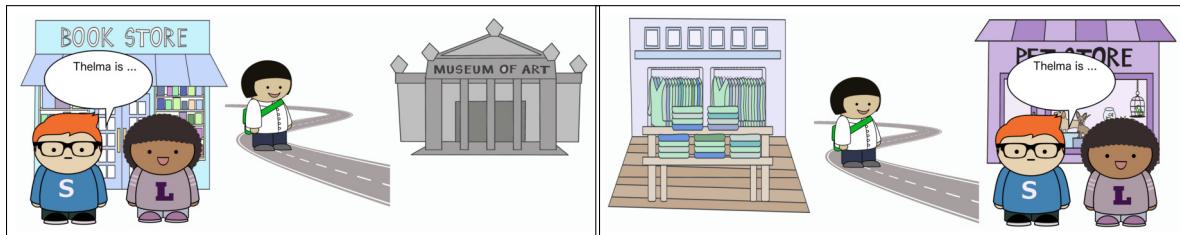
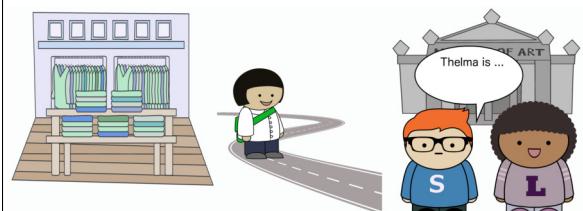
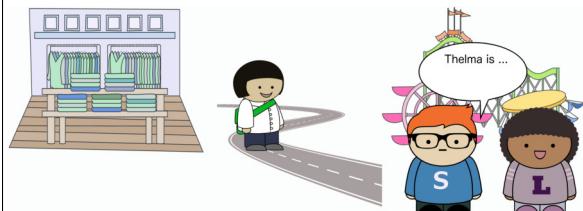
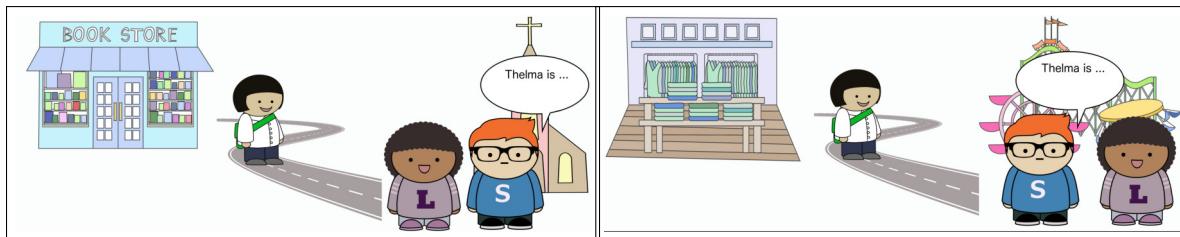
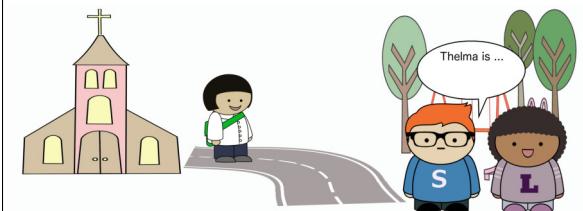
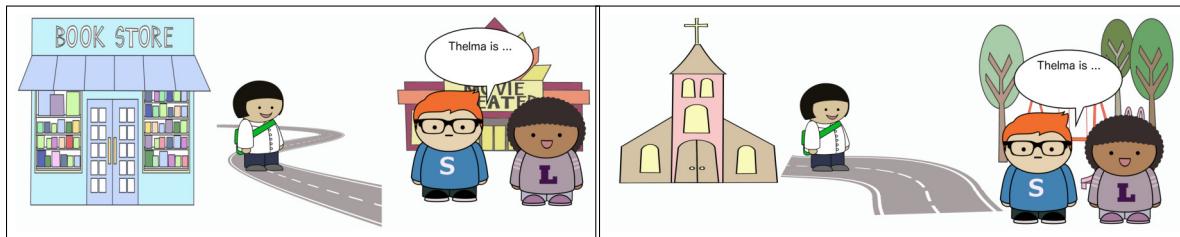
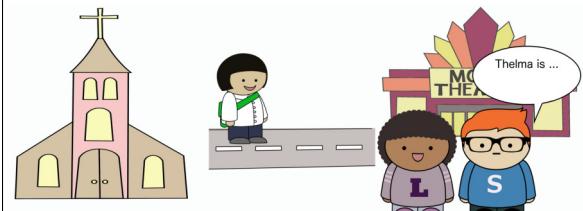
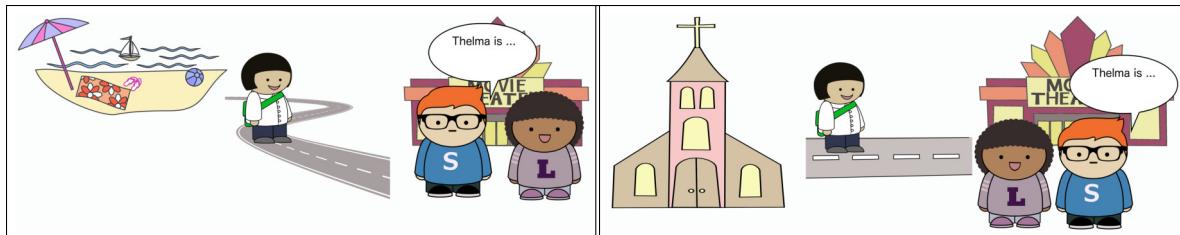
## APPENDIX N

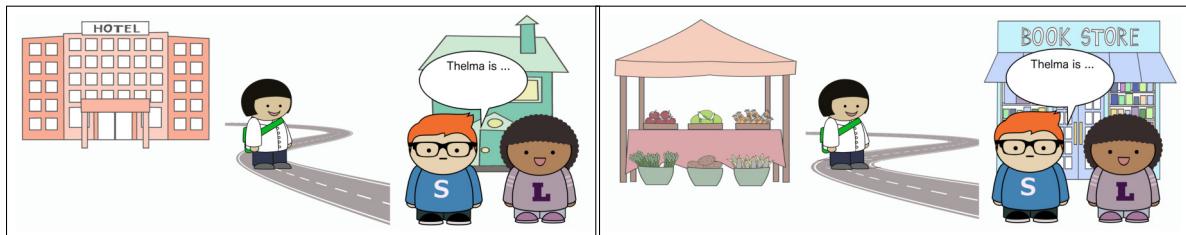
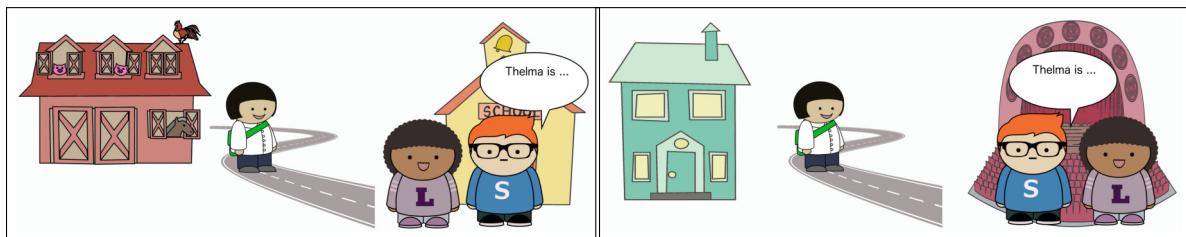
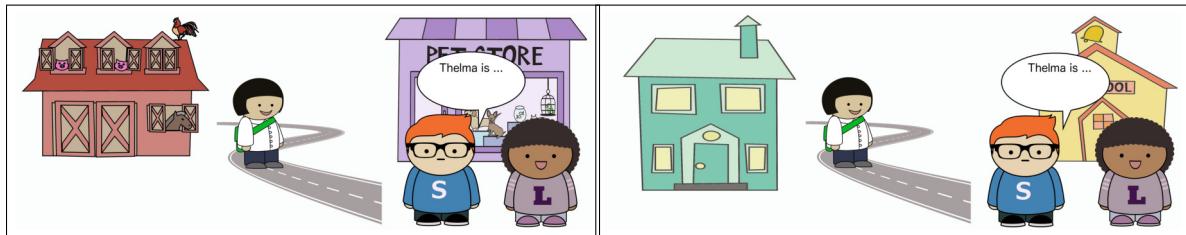
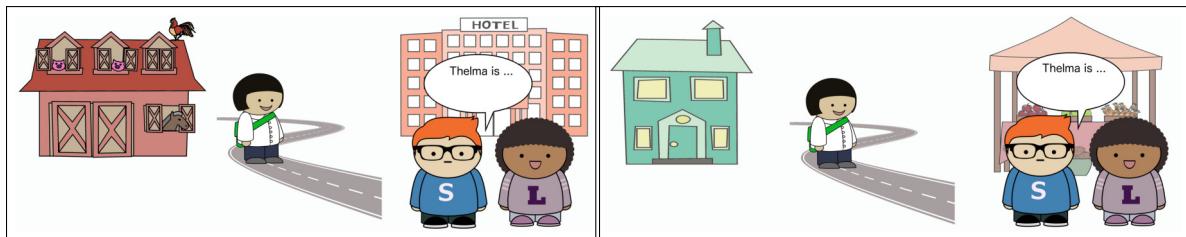
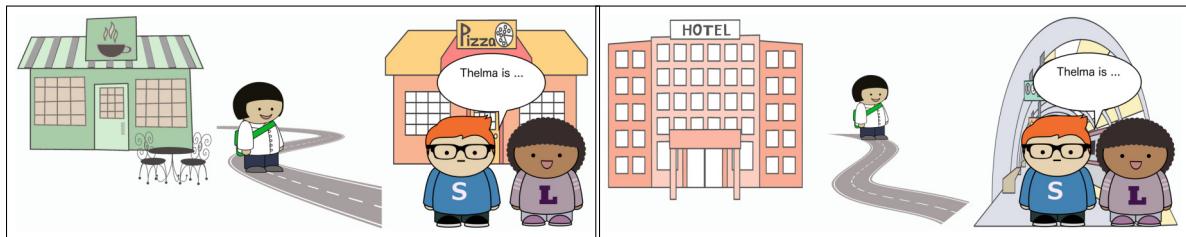
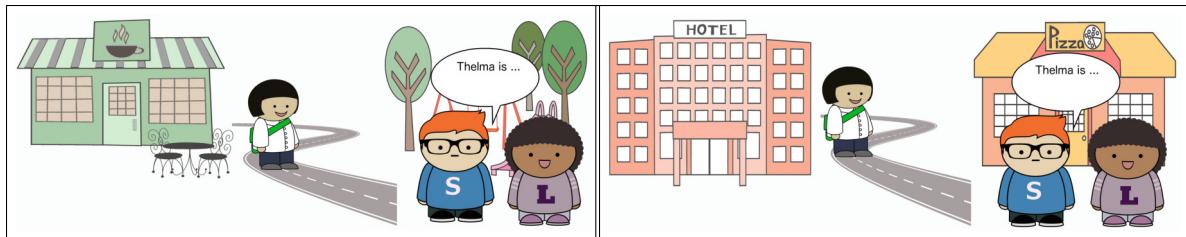
### EXPERIMENT 6 ITEMS

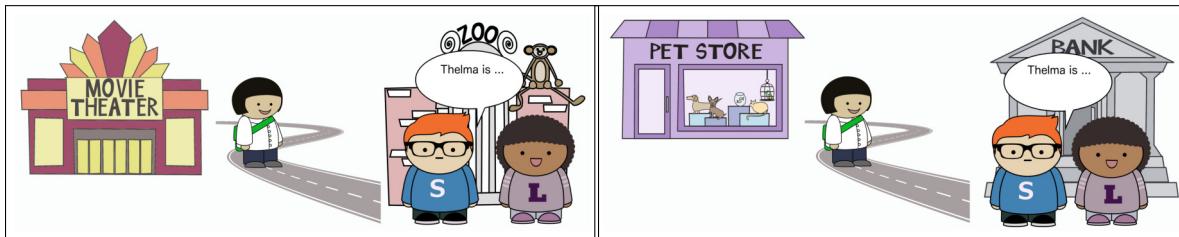
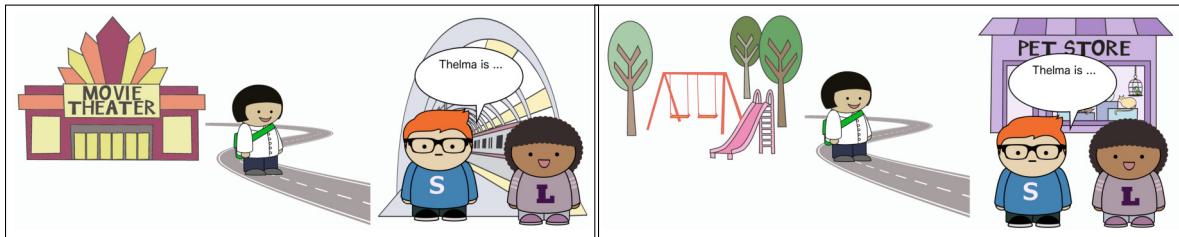
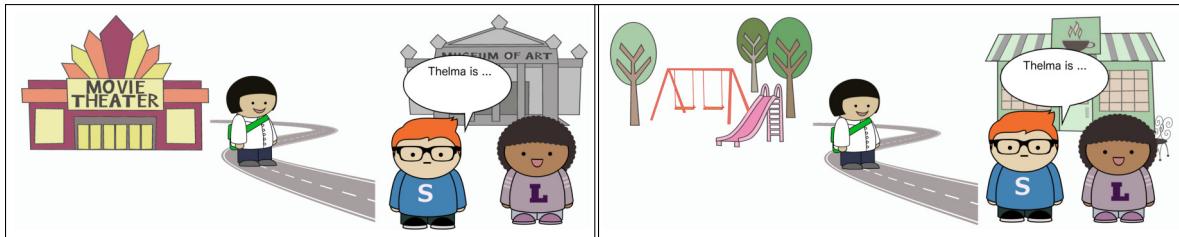
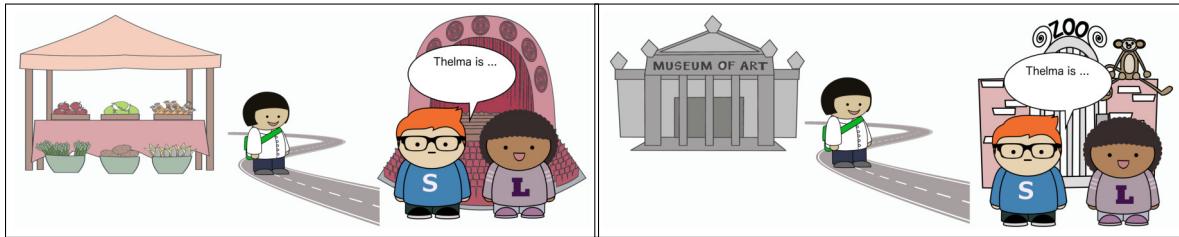
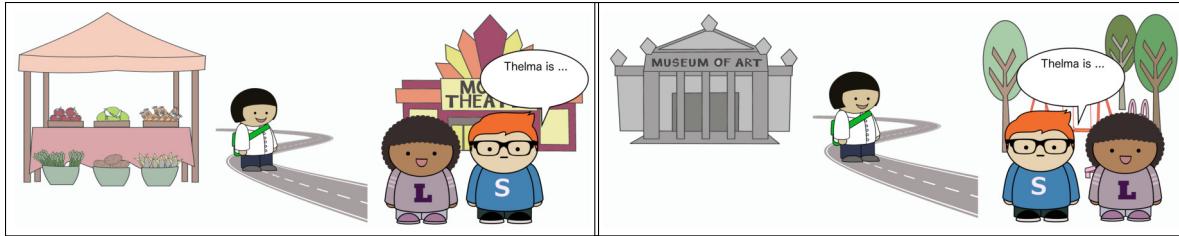
#### N.1 Main items

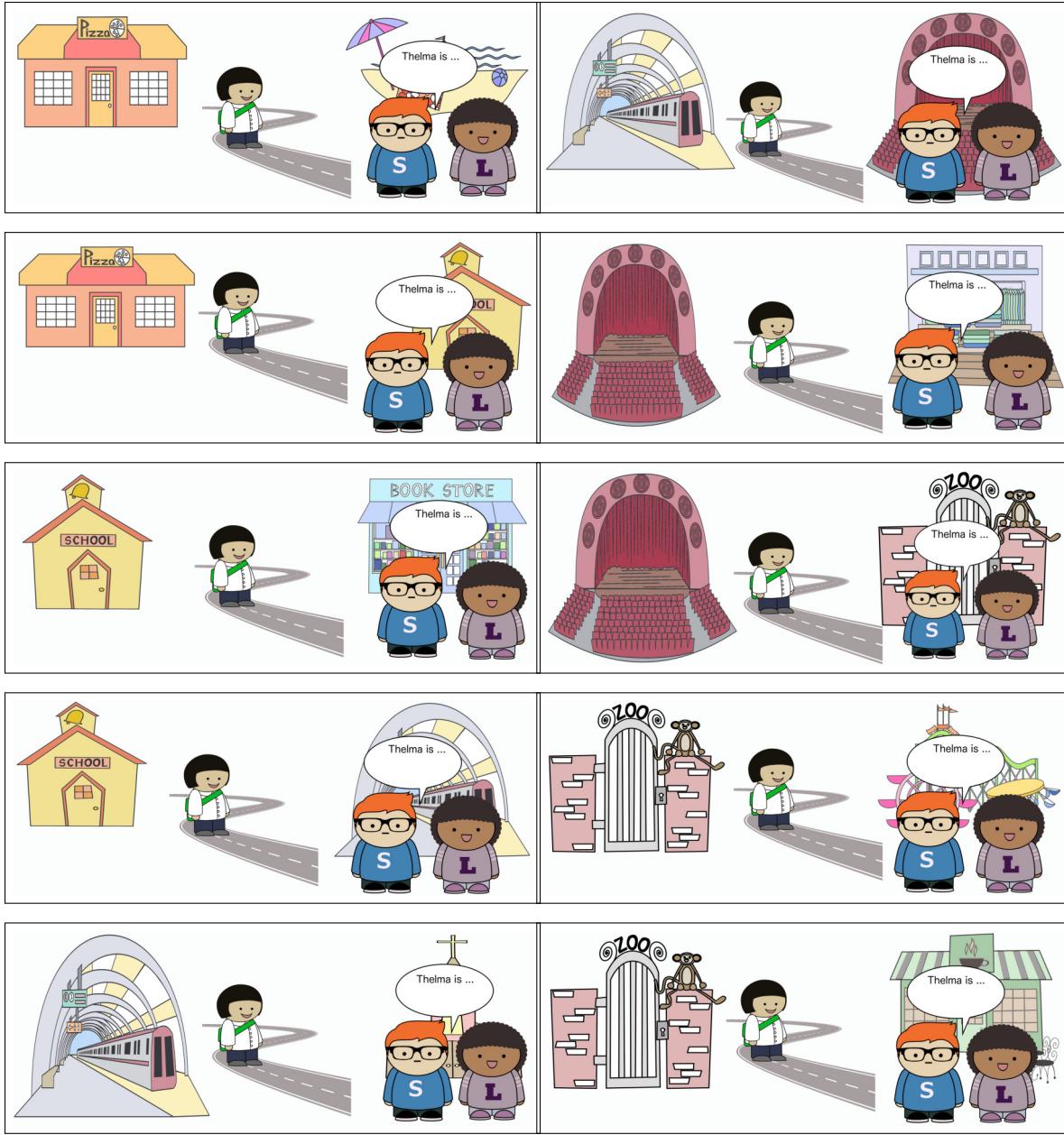
Each item below is shown in the Both condition. There were also Speaker, Listener, and None versions of each scene, as described in Chapter 8.





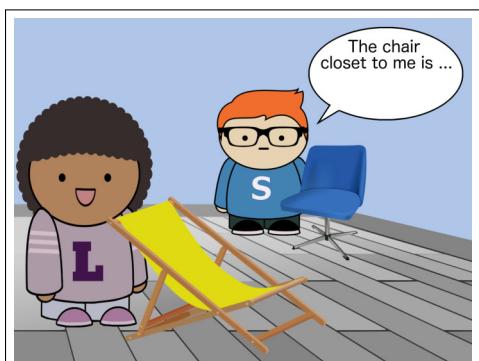
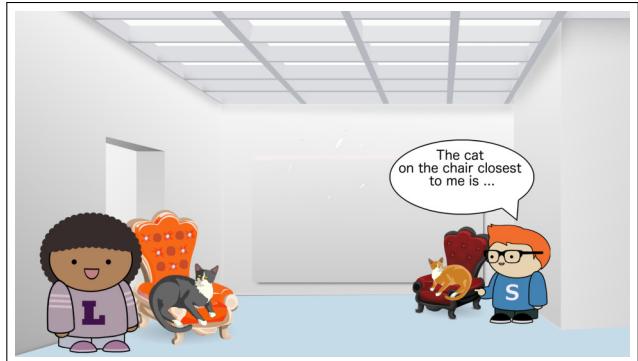
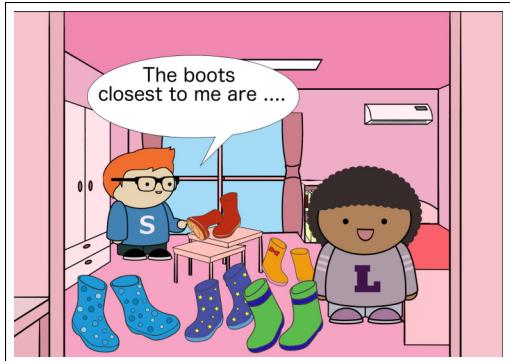


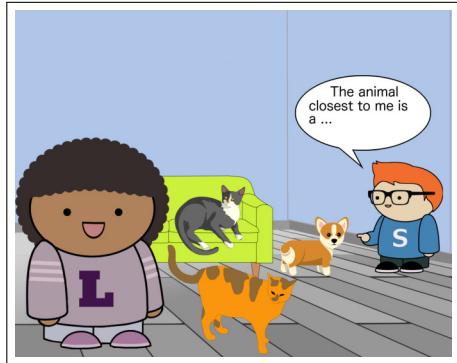
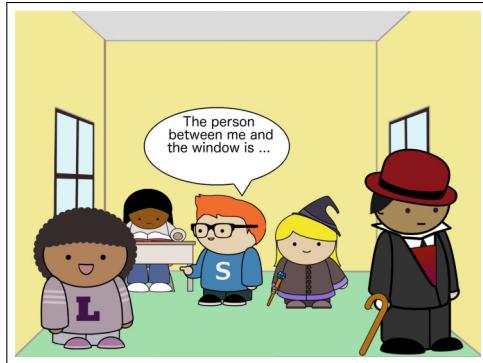
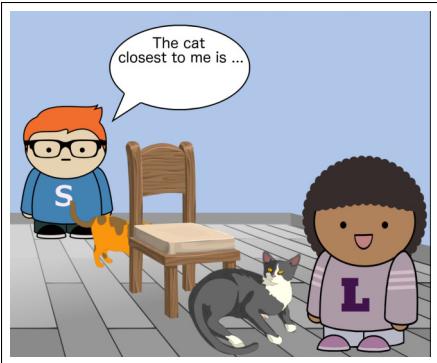


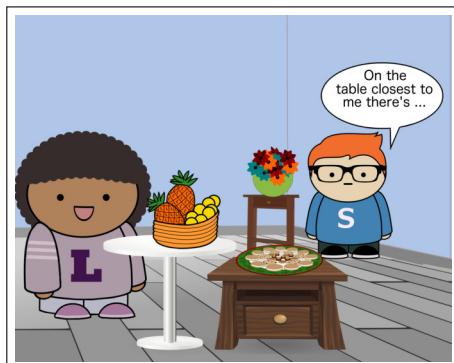
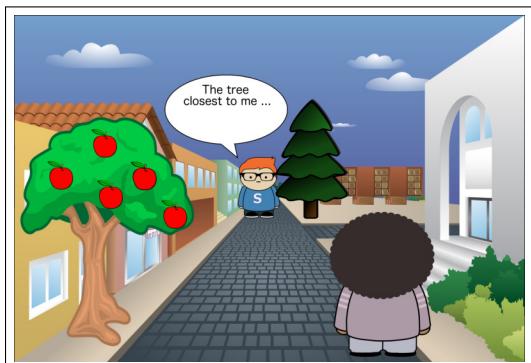
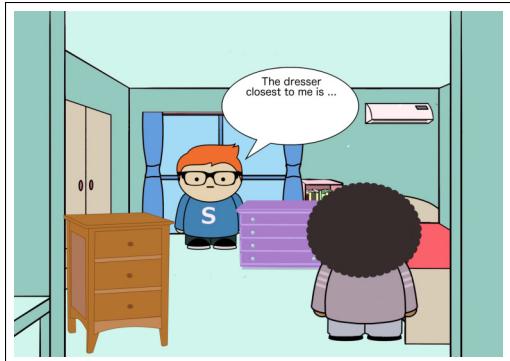


## N.2 Spatial task items

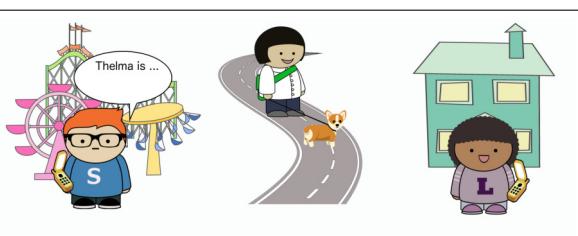
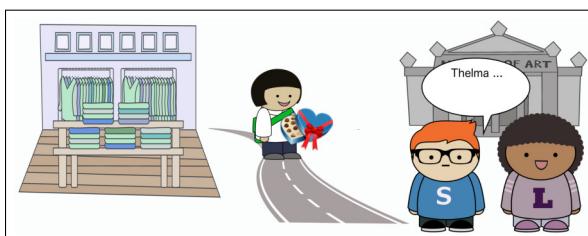
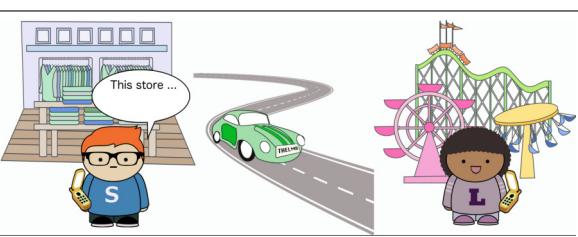
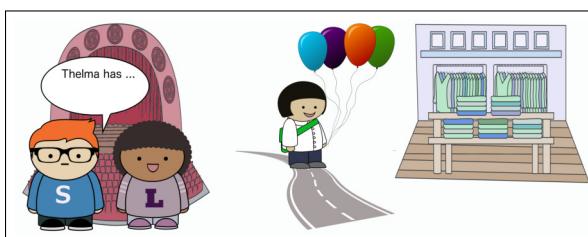
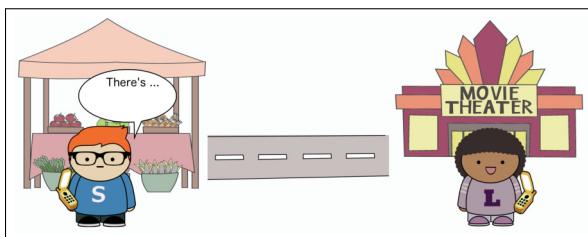


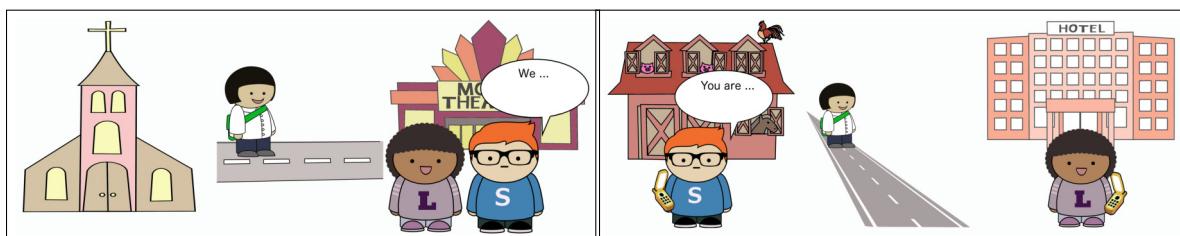
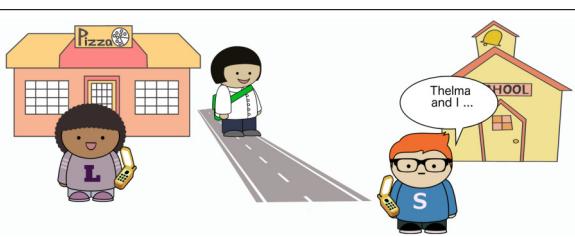
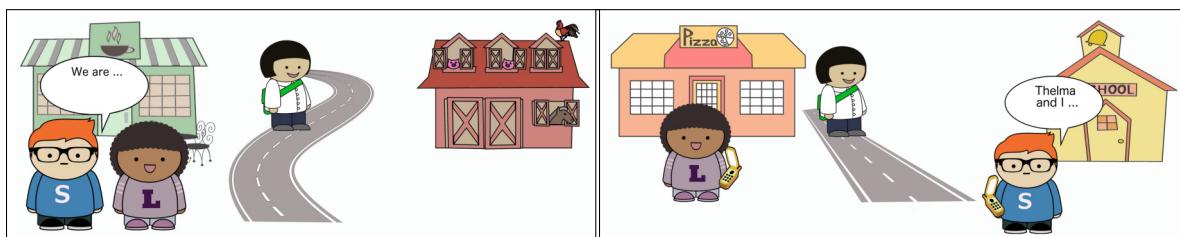
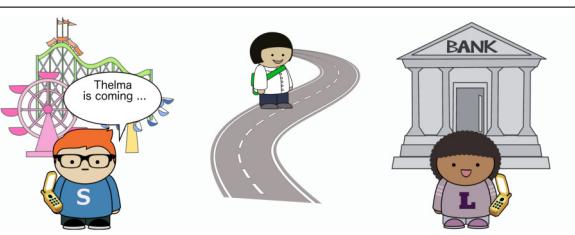
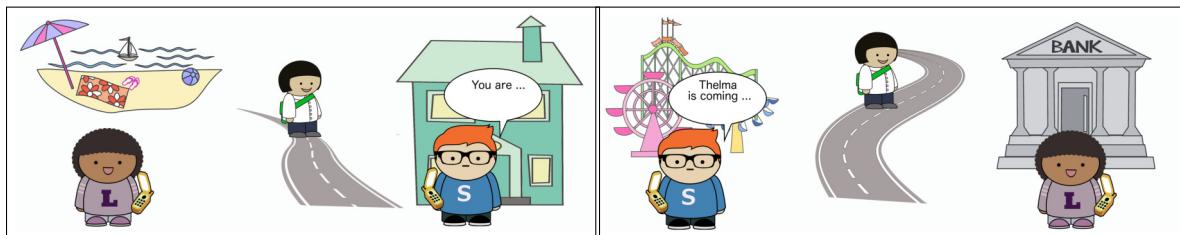
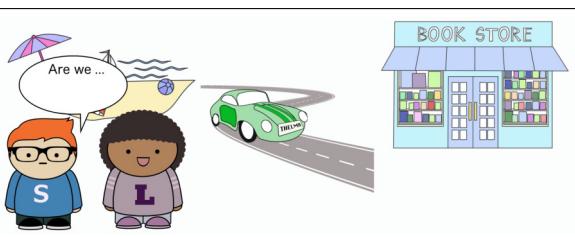
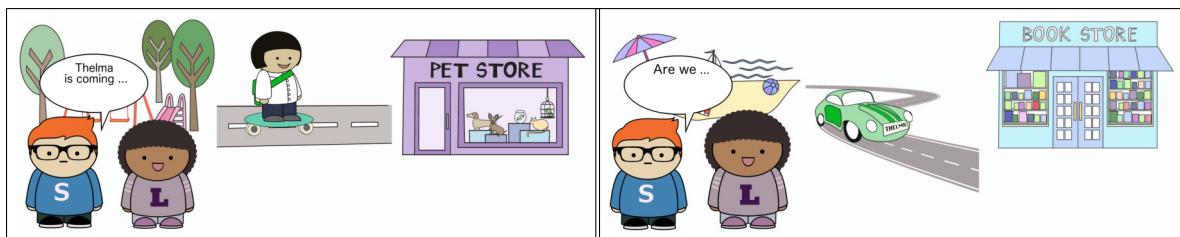
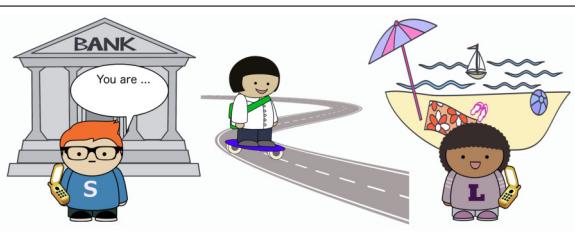
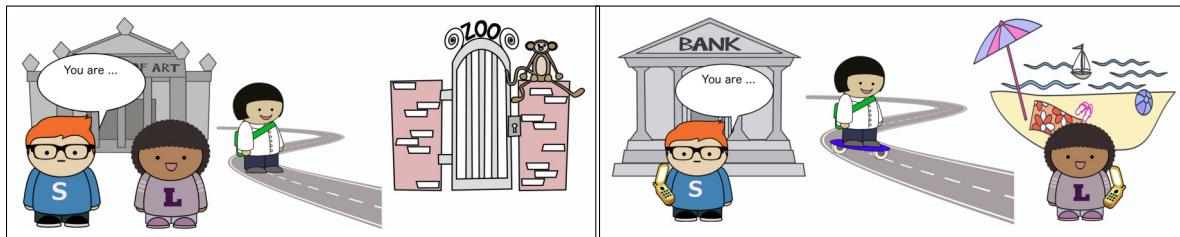
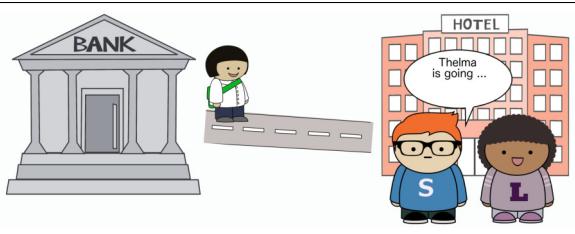
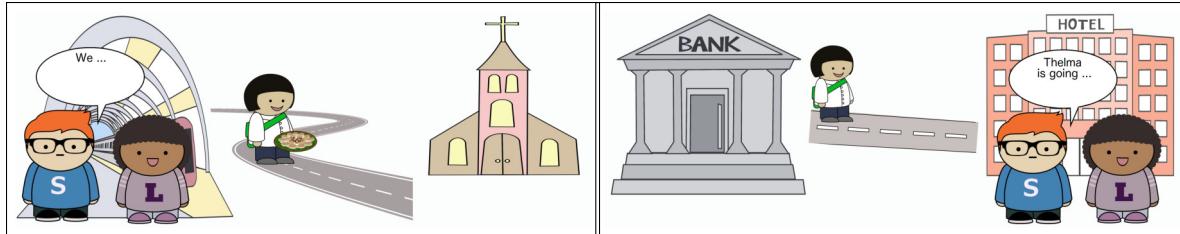


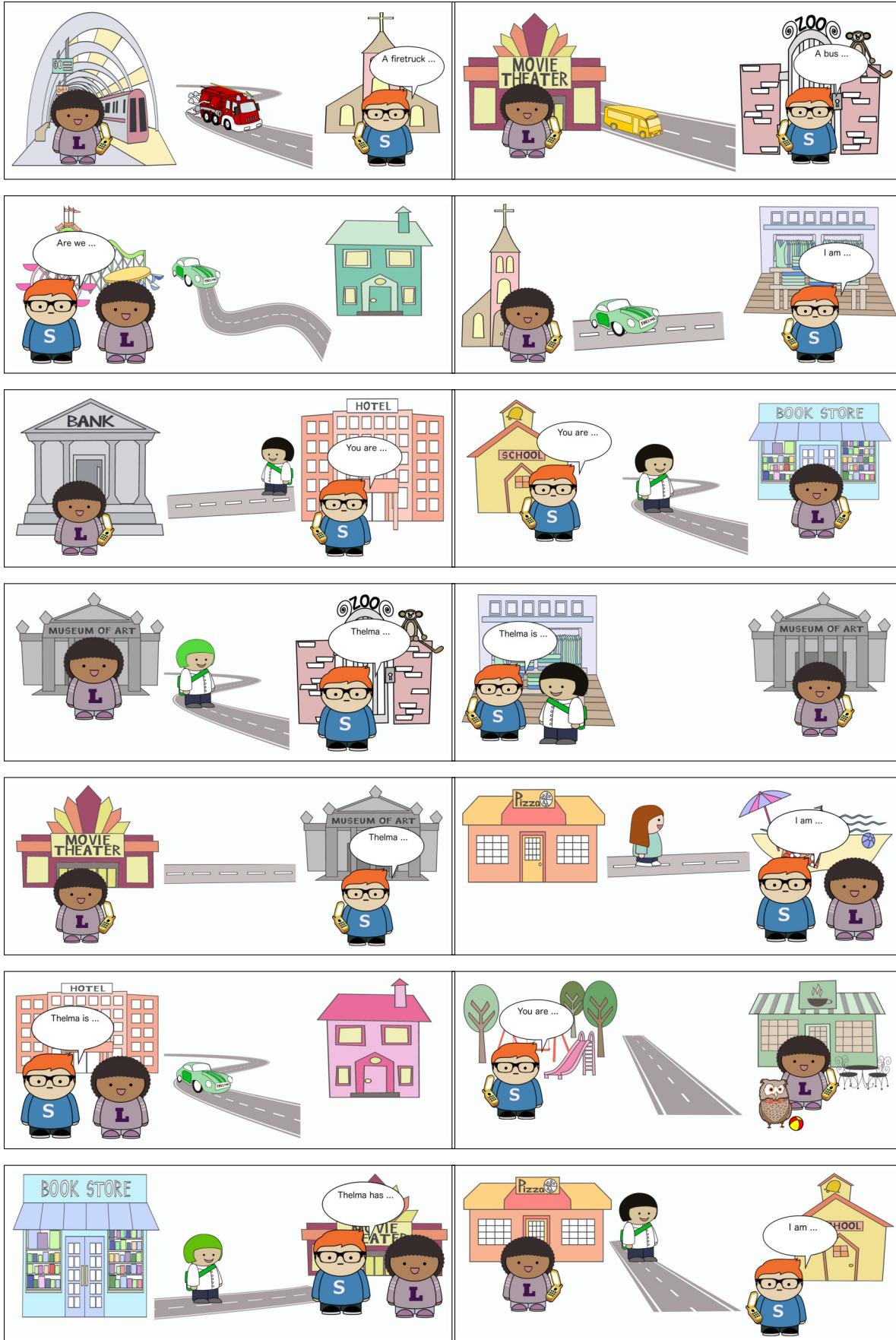




### N.3 Fillers







## BIBLIOGRAPHY

- Abbot-Smith, K., D. Williams, and D. Matthews (2020). What's new for you?: Interlocutor-specific perspective-taking and language interpretation in autistic and neuro-typical children. *Research in Autism Spectrum Disorders* 70.
- Abrusán, M. (2018). Protagonist Projection. In *Semantics and Linguistic Theory*.
- Abrusán, M. (2019). The Spectrum of Perspective Shift: Protagonist Projection vs. Free Indirect Discourse. *Linguistics and Philosophy*.
- Acheson, D., J. B. Wells, and M. C. MacDonald (2008). New and updated tests of print exposure and reading abilities in college students. *Behavioral Research Methods* 40, 278–289.
- Altshuler, D. (2016). *Events, States and Times*. De Gruyter.
- Altshuler, D. (2017). Tense and temporal adverbs (“I learned last week that there would now be an earthquake”). In L. Mattheuwson, C. Meier, H. Rullman, and E. Zimmerman (Eds.), *The Blackwell Companion to Semantics*. Wiley Blackwell.
- Altshuler, D. and U. Stojnić (2015). The attention-coherence model of prominence: A look at ‘now’.
- Amaral, P., C. Roberts, and E. A. Smith (2007). Review of *The Logic of Conventional Implicatures* by Chris Potts. *Linguistics and Philosophy* 30(6), 707–749.
- Anand, P. and F.-f. Hsieh (2005). Long-Distance Reflexives in Perspective. In J. Alderete (Ed.), *Proceedings of the West Coast Conference on Formal Linguistics*, Volume 24, pp. 43–51.
- Anand, P. and A. Nevins (2004). Shifty Operators in Changing Contexts. *Semantics and Linguistic Theory* 14.
- Banfield, A. (1982). *Unspeakable Sentences: Narration and Representation in the Language of Fiction*. Routledge.
- Barlew, J. (2015). ‘Coming’ to a doxastic agent. Volume 19.
- Barlew, J. (2017). *The semantics and pragmatics of perspectival expressions in English and Bulu: The case of deictic motion verbs*. Ph. D. thesis, The Ohio State University.
- Baron-Cohen, S., S. Wheelwright, R. Skinner, J. Martin, and E. Clubley (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders* 31, 5–17.
- Baroni, M., S. Bernardini, A. Ferraresi, and E. Zanchetta (2009). The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora. Volume 43(3), pp. 209–226.

- Bates, D., M. Mächler, B. Bolker, and S. Walker (2015). Fitting Linear Mixed-Effects Models Using lme4.
- Bergen, L., N. D. Goodman, and R. Levy (2012). That's what she (could have) said: how alternative utterances affect language use. In *Proceedings of the Annual Conference of the Cognitive Science Society*, Volume 34.
- Bergen, L., R. Levy, and N. D. Goodman (2016). Pragmatic Reasoning through Semantic Inference. *Semantics & Pragmatics* 9(20).
- Bilmes, L. (1995). The Grammaticalization of Thai ‘Come’ and ‘Go’. In *Berkeley Linguistics Society*, Volume 21, pp. 33–46.
- Bohn, M., M. H. Tessler, and M. C. Frank (2019). Integrating Common Ground and Informativeness in Pragmatic Word Learning.
- Brochhagen, T., M. Franke, and R. van Rooij (2016). Learning biases may prevent lexicalization of pragmatic inferences: a case study combining iterated (Bayesian) learning and functional selection.
- Brown-Schmidt, S. (2009). The role of executive function in perspective-taking during online language comprehension. *Psychonomic Bulletin and Review* 16, 893–900.
- Brown-Schmidt, S. and J. E. Hanna (2011). Talking in Another Person’s Shoes: Incremental Perspective-taking in Language Processing. *Dialogue and Discourse* 2, 11–33.
- Brown-Schmidt, S. and D. Heller (2018). Perspective-taking During Conversation. In G. Gaskell and S. A. Rueschemeyer (Eds.), *Oxford Handbook of Psycholinguistics*. Oxford University Press.
- Bylinina, L. (2014). *The grammar of standards*. LOT Dissertation Series.
- Bylinina, L., E. McCready, and Y. Sudo (2015). Notes on perspective sensitivity. In P. Arkadiev, I. Kapitonov, Y. Lander, E. Rakhilina, and S. Tatevosov (Eds.), *Donum Semanticum: Opera Linguistica Et Logica In Honorem Barbarae Partee A Disscipulis Amicisque Rossicis Oblata*.
- Cardinaletti, A. and G. Giusti (2001). ‘Semi-lexical’ motion verbs in Romance and Germanic. In M. Devos and J. van der Wal (Eds.), *Semi-lexical categories: the function of content words and the content of function words*.
- Carlson, R. (2014). The grammaticalization of ‘go’ as an intensifier in Supyire. In M. Devos and J. van der Wal (Eds.), *COME and GO off the Beaten Grammaticalization Path*.
- Carruthers, P. (2016). Two systems for mindreading? *Review of Philosophy and Psychology* 7(1), 141–162.
- Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, and topics. In C. Li (Ed.), *Subject and Topic*, pp. 25–76. New York: Academic Press.
- Charnavel, I. (2018). Deictic Perspective and Logophoric Exemption from Condition A. In *West Coast Conference on Formal Linguistics*, Volume 35.
- Charnavel, I. (2019). Logophoricity and Locality: a View from French Anaphors. *Linguistic Inquiry*.

- Child, S., J. Oakhill, and A. Garnham (2020). Tracking your Emotions: an Eye-Tracking Study on Reader's Engagement with Perspective during Text Comprehension. *Quarterly Journal of Experimental Psychology*.
- Christensen, R. H. B. (2019). Ordinal: Regression Models for Ordinal Data.
- Clark, E. V. and O. K. Garnica (1974). Is he coming or going? On the acquisition of deictic verbs. *Journal of Verbal Learning and Verbal Behavior* 13.
- Clements, G. N. (1975). The Logophoric Pronoun in Ewe: Its Role in Discourse. *Journal of West African Languages* 10, 141–177.
- Cohn-Gordon, R. and L. Bergen (2019). Verbal Irony, Pretense, and the Common Ground.
- Cohn-Gordon, R., N. D. Goodman, and C. Potts (2018). Pragmatically informative image captioning with character-level inference.
- Cohn-Gordon, R. and C. Qing (2018). Modeling “non-literal” social meaning with Bayesian pragmatics.
- Conan, N. (2008, July). America’s Romance with Rails May Rise.
- Cook, K. W. (1994). The temporal use of Hawaiian directional particles. In M. Pütz and R. Dirven (Eds.), *The Construal of Space in Language and Thought*, Number 8 in Cognitive Linguistics Research.
- Davidson, D. (1979). Quotation. *Theory and Decision* 11(1), 27–40.
- Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 600 million words, 1990-present*. Available online at [www.english-corpora.org/coca](http://www.english-corpora.org/coca).
- de Villiers, J. and P. A. de Villiers (2014). The Role of Language in Theory of Mind Development. *Topics in Language Disorders* 34(4), 313–328.
- de Vos, M. (2005). *The syntax of pseudo-coordination in English and Afrikaans*. Number 114 in Landelijke Onderzoekschool Taalwetenschap.
- Deal, A. R. (2014). Nez Perce embedded indexicals. *Semantics of Under-represented Languages of the Americas* 7.
- Deal, A. R. (2017). Shifty asymmetries: universals and variation in shifty indexicality.
- Degen, J., R. X. D. Hawkins, C. Graf, E. Kreiss, and N. D. Goodman (2019). When redundancy is useful: A Bayesian approach to ‘overinformative’ referring expressions. *Psychological Review in press*.
- Degen, J., M. H. Tessler, and N. D. Goodman (2015). Wonky worlds: Listeners revise world knowledge when utterances are odd.
- Doron, E. (1991). Point of view as a factor of content. *Semantics and Linguistic Theory* 1, 51–64.
- Dragomirescu, A. and A. Nicolae (2014). The multiple grammaticalization of Romanian *veni* ‘come.’ Focusing on the passive construction. In *COME and GO off the Beaten Grammaticalization Path*.
- Drummond, A. (2019). Ibex.

- Duff, J. (2018a). Says who?: The influence of speech and mental predicates on epithet interpretation.
- Duff, J. (2018b). *Somebody's Fool: Theory of Mind and the Interpretation of Epithets*. BA Thesis, University of Massachusetts, Amherst.
- Eckardt, R. (2014). *The Semantics of Free Indirect Discourse: How Texts Allow Us to Mind-read and Eavesdrop*.
- Epley, N., B. Keysar, L. Van Boven, and T. Gilovich (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology* 87(3), 327–339.
- Epley, N., C. K. Morewedge, and B. Keysar (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology* 40(6), 760–768.
- Faller, M. (2002). *Semantics and pragmatics of evidentials in Cuzco Quechua*. Ph. D. thesis, Stanford.
- Ferguson, H. J., I. Apperly, and J. E. Cane (2017). Eye tracking reveals the cost of switching between self and other perspectives in a visual perspective-taking task. *The Quarterly Journal of Experimental Psychology* 70(8), 1646–1660.
- Fillmore, C. (1966). Deictic Categories in the Semantics of Come. *Foundations of Language* 2, 219–227.
- Fludernik, M. (1995). *The Fictions of Language and the Languages of Fiction: the linguistic representation of speech and consciousness*. London: Routledge.
- Frank, M. C. and N. D. Goodman (2012). Predicting pragmatic reasoning in language games. *Science* 6084(336), 998–998.
- Frank, M. C. and N. D. Goodman (2014). Inferring word meanings by assuming that speakers are informative. *Cognitive Psychology* 75.
- Frazis, H. and J. C. Stewart (2014). Is the workweek really overestimated? *Monthly Labor Review June*.
- Garrett, E. J. (2001). *Evidentiality and Assertion in Tibetan*. Ph. D. thesis, UCLA.
- Garrod, S. C. and T. Sanford (1988). Thematic subjecthood and cognitive constraints on discourse structure. *Journal of Pragmatics* 12(5-6), 519–534.
- Gathercole, V. (1987). Towards a Universal for Deictic Verbs of Motion. In *Kansas Working Papers in Linguistics*, Volume 3.
- Gernsbacher, M. A. (2007). Presidential Column: On Not Being Human. *American Psychological Society Observer* 20(2).
- Geurts, B. and E. Maier (2005). Quotation in Context. *Belgian Journal of Linguistics* 17(1), 109–128.
- Goddard, C. (1997). The semantics of coming and going. *Pragmatics* 7(2).
- Goodman, N. D. and A. Stuhlmüller (2013). Knowledge and Implicature: Modeling Language Understanding as Social Cognition. *Topics in Cognitive Science* 5.

- Goodman, N. D. and A. Stuhlmüller (2019). The Design and Implementation of Probabilistic Programming Languages.
- Gordon, P. C., B. J. Grosz, and L. A. Gillion (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science* 17, 311–348.
- Grice, P. (1975). Logic and conversation. In P. Cole and J. Morgan (Eds.), *Syntax and semantics*, Number 3, pp. 41–58. New York: Academic Press.
- Grosz, B. J., A. K. Joshi, and S. Weinstein (1995). Centering: a framework for modeling the local coherence of discourse. *Association for Computational Linguistics*, 203–225.
- Gruber and G. Redeker (2014). *Pragmatics of Discourse Coherence: Theories and applications*. John Benjamins.
- Gunlogson, C. (2008). A question of commitment. *Belgian Journal of Linguistics* 22, 101–136.
- Harris, J. A. (2012). *Processing Perspectives*. Ph. D. thesis, University of Massachusetts, Amherst.
- Harris, J. A. and C. Potts (2009). Perspective-shifting with appositives and expressives. *Linguistics and Philosophy* 36(2), 523–552.
- Hassler, G. (1999). Crosslinguistic and diachronic remarks on the grammaticalization of aspect in Romance languages. In *Proceedings of the International Symposium on New Reflections on Grammaticalization*.
- Heller, D., K. S. Gorman, and M. K. Tanenhaus (2012). To name or to describe: Shared knowledge affects referential form. *Topics in Cognitive Science* 4(2), 290–305.
- Heller, D., D. Grodner, and M. K. Tanenhaus (2008). The role of perspective in identifying domains of reference. *Cognition* 108(3), 831–836.
- Hinterwimmer, S. (2017a). Prominent protagonists. *Journal of Pragmatics*.
- Hinterwimmer, S. (2017b). Two kinds of perspective taking in narrative texts. *Semantics and Linguistic Theory* 27, 282–301.
- Hobbs, J. R. (1979). Coherence and Coreference. *Cognitive Science* 3(1).
- Hooper, R. (2002). Deixis and aspect: The Tokelauan directional particles mai and atu. *Studies in Language* 26(2), 283–313.
- Horton, W. S. and B. Keysar (1996). When do speakers take into account common ground? *Cognition* 59(1), 91–117.
- Jacobson, P. (1999). Towards a Variable-Free Semantics. *Linguistics and Philosophy* 22(2), 117–184.
- John, O. and S. Srivastava (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin and O. P. John (Eds.), *Handbook of personality: Theory and research*, Volume 2, pp. 102–138. New York: Guilford Press.
- Josephs, L. S. (1975). *Palauan reference grammar*. Honolulu: University of Hawaii Press.

- Kaiser, E. (2015). Perspective-shifting and free indirect discourse: Experimental investigations. In *Semantics and Linguistic Theory*, Volume 25.
- Kaiser, E. and J. H. Lee (2017a). Experience matters: A psycholinguistic investigation of predicates of personal taste. *Semantics and Linguistic Theory* 27.
- Kaiser, E. and J. H. Lee (2017b). Predicates of personal taste and multidimensional adjectives: An experimental investigation. In *West Coast Conference on Formal Linguistics*, Volume 35.
- Kamp, H. (1971). Formal properties of ‘now’: to the memory of Arthur Prior. *Theoria* 37(3).
- Kao, J. T., J. Y. Wu, L. Bergen, and N. D. Goodman (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences* 111(33).
- Kaplan, D. (1989). Demonstratives: an essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals. In J. Almog, J. Perry, and H. Wettstein (Eds.), *Themes From Kaplan*, pp. 481–563. Oxford: Oxford University Press.
- Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford: CSLI Publications.
- Kehler, A. and H. Rohde (2013). A Probabilistic Reconciliation of Coherence-Driven and Centering-Driven Theories of Pronoun Interpretation. *Theoretical Linguistics* 39(1-2), 1–37.
- Keshet, E. (2005). A Matter of Taste.
- Keysar, B. and D. J. Barr (2003). Limits on theory of mind use in adults. *Cognition* 89, 25–41.
- Köder, F. and E. Maier (2016). When you isn’t you: the attraction of self-ascription in children’s interpretation of pronouns in reported speech. *Glossa* 2(1).
- Köder, F., E. Maier, and P. Hendriks (2015). Perspective shift increases processing effort of pronouns: a comparison between direct and indirect speech. *Language, Cognition and Neuroscience* 30(8).
- Korotkova, N. (2016). *Heterogeneity and uniformity in the evidential domain*. Ph. D. thesis, UCLA.
- Kratzer, A. (1996). Severing the External Argument from its Verb. In J. Rooryck and L. Zaring (Eds.), *Phrase Structure and the Lexicon*, pp. 109–137. Dordrecht: Kluwer.
- Kuno, S. and E. Kaburaki (1977). Empathy and Syntax. *Linguistic Inquiry* 8, 627–672.
- Lasersohn, P. (2005). Context dependence, disagreement, and predicates of personal taste. *Linguistics and Philosophy* 28, 643–686.
- Lawrence, D. H. (1920). *Women in Love*. Project Gutenberg.
- Lee, E. and J. Choi (2009). Two *nows* in Korean. *Journal of Semantics* 26(1).
- Lewis, D. (1979a). Attitudes De Dicto and De Se. *Philosophical Review* 88(513-543).
- Lewis, D. (1979b). Scorekeeping in a language game. *Journal of Philosophical Logic* 8, 339–359.

- Ligneau-Hervé, C. (2005). Perspective-Taking Judgments Among Young Adults, Middle-Aged, and Elderly People. *Journal of Experimental Psychology: Applied* 11(1), 53–60.
- Lin, S., B. Keysar, and N. Epley (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology* 46, 551–556.
- Long, M. R., W. S. Horton, H. Rohde, and A. Sorace (2018). Individual differences in switching and inhibition predict perspective-taking across the lifespan. *Cognition* 170.
- Loveland, K. A. (1984). Learning about points of view: spatial perspective and the acquisition of ‘I/you’. *Journal of Child Language* 11, 35–56.
- Maier, E. (2015). Quotation and unquotation in Free Indirect Discourse. *Mind and Language* 30(3), 345–373.
- Maier, E. (2016). A plea against monsters. *Grazer Philosophische Studien* 93, 363–395.
- Maier, E. (2017a). Mixed Quotation. In D. Gutzmann, L. Matthewson, C. Meier, H. Rullman, and E. Zimmerman (Eds.), *The Companion to Semantics*. Wiley Blackwell.
- Maier, E. (2017b). The pragmatics of attraction: explaining unquotation in Direct and Free Indirect Discourse. In P. Saka and M. Johnson (Eds.), *The Semantics and Pragmatics of Quotation*, Number 15 in Perspectives in Pragmatics. Springer.
- Maier, E. (2018). Quotes as complements. In P. Menéndez-Benito, I. Frana, and R. Bhatt (Eds.), *Making Worlds Accessible: Festschrift for Angelika Kratzer*.
- Mauri, C. and A. Sansó (2014). Go and come as sources of directive constructions. In Maud Devos and J. van der Wal (Eds.), *COME and GO off the Beaten Grammaticalization Path*.
- Miller, G. A. and P. Johnson-Laird (1976). *Language and Perception*. Cambridge: Cambridge University Press.
- Millis, K. K. (1995). Encoding discourse perspective during the reading of a literary text. *Poetics* 23(3), 235–253.
- Mitchell, J. E. (1986). *The formal semantics of point of view*. Ph. D. thesis, University of Massachusetts, Amherst.
- Mizuno, A., Y. Liu, D. L. Williams, T. A. Keller, N. J. Minshew, and M. A. Just (2011). The neural basis of deictic shifting in linguistic perspective-taking in high-functioning autism. *Brain* 134.
- Modiano, P. (2005). *Accident Nocturne*. Paris: Gallimard.
- Moltmann, F. (2010). Relative truth and the first person. *Philosophical Studies* 150(2), 187–220.
- Monroe, W. (2018). *Learning in the Rational Speech Acts Model*. Ph. D. thesis, Stanford.
- Monroe, W., R. X. D. Hawkins, N. D. Goodman, and C. Potts (2017). Colors in context: a pragmatic neural model for grounded language understanding. *Transactions of the Association for Computational Linguistics* 5, 325–338.

- Munro, P., F. H. Lopez, O. V. Méndez Martínez, Rodrigo Garcia, and M. R. Galant (1999). *Di'csyonaary X:tée'n Díi'zh Sah Sann Lu'uc / San Lucas Quiaviní Zapotec Dictionary / Diccionario Zapoteco de San Lucas Quiavini*. Los Angeles: UCLA Chicano Studies Research Center Publications.
- Murray, S. (2010). *Evidentiality and the Structure of Speech Acts*. Ph. D. thesis, Rutgers.
- Nakazawa, T. (2007). A typology of the ground of deictic motion verbs as path-conflating verbs: the speaker, addressee, and beyond. *Poznan Studies in Contemporary Linguistics* 43(2).
- Nakazawa, T. (2009). A typology of the ground of deictic motion verbs as path-conflating verbs: the entailment of arrival and the deictic center. *Poznan Studies in Contemporary Linguistics* 45(3).
- Nie, A., R. Cohn-Gordon, and C. Potts (2020). Pragmatic Issue-Sensitive Image Captioning.
- Nishigauchi, T. (2014). Reflexive binding: awareness and empathy from a syntactic point of view. *Journal of East Asian Linguistics* 23, 157–206.
- Oshima, D. (2006a). GO and COME revisited: What serves as a reference point? In *Berkeley Linguistics Society*, Volume 32.
- Oshima, D. (2006b). Motion deixis, indexicality, and presupposition. In M. Gibson and J. Howell (Eds.), *Proceedings of Semantics and Linguistic Theory*, Volume 16.
- Park, Y. (2014a). Indexical shift and the long-distance reflexive *caki* in Korean.
- Park, Y. (2014b). Indexicals and the long-distance reflexive *caki* in Korean. In *Generative Linguistics in the Old World*, Volume 37.
- Park, Y. (2017). *Attitudes de se and logophoricity*. Ph. D. thesis, University of Massachusetts, Amherst, Amherst, MA.
- Patel-Grosz, P. (2015). *Epithets at the Syntax-Semantics Interface*. Cambridge Scholars Publishing.
- Pearson, A., D. Ropar, and A. F. d. C. Hamilton (2013). A review of visual perspective taking in autism spectrum disorder. *Frontiers in Human Neuroscience* 7.
- Pearson, H. (2013). A Judge-Free Semantics for Predicates of Personal Taste. *Journal of Semantics* 30, 103–154.
- Poesio, M., R. Stevenson, B. Di Eugenio, and J. Hitzeman (2004). Centering: A Parametric Theory and Its Instantiations. *Computational Linguistics* 30(3), 309–363.
- Potts, C. (2005). *The logic of conventional implicatures*. Oxford Studies in Theoretical Linguistics. Oxford University Press.
- Potts, C. (2007a). The dimensions of quotation. In C. Barker and P. Jacobson (Eds.), *Direct Compositionality*, pp. 405–431. New York: Oxford University Press.
- Potts, C. (2007b). The expressive dimension. *Theoretical Linguistics* 33(2), 165–198.
- Potts, C., D. Lassiter, R. Levy, and M. C. Frank (2016). Embedded implicatures as pragmatic inferences undercompositional lexical uncertainty. *Journal of Semantics* 33(4).

- Pratt, M. W., R. Diessner, A. Pratt, B. Hunsberger, and M. S. Pancer (1996). Moral and social reasoning and perspective taking in later life: A longitudinal study. *Psychology and Aging* 11(1), 66–73.
- Pullum, G. K. (1990). Constraints on intransitive quasi-serial verb constructions in Modern Colloquial English. Volume 39.
- Qing, C. and R. Cohn-Gordon (2018). Use-conditional meaning in Rational Speech Act models.
- Qing, C., N. D. Goodman, and D. Lassiter (2016). A rational speech-act model of projective content. *Proceedings of the Cognitive Science Society* 38.
- Reboul, A., D. Delfitto, and G. Fiorin (2016). The Semantic Properties of Free Indirect Discourse. *Annual Review of Linguistics* 2, 255–271.
- Reinhart, T. and E. Reuland (1993). Reflexivity. *Linguistic Inquiry* 24, 657–720.
- Ricard, M., P. G. Girouard, and T. G. Décarie (1999). Personal pronouns and perspective taking in toddlers. *Journal of Child Language* 26(3).
- Rice, K. (1986). Some remarks on direct and indirect speech in Slave (Northern Athapaskan). In F. Coulmas (Ed.), *Direct and indirect speech*, pp. 47–76. Mouton de Gruyter.
- Richards (1976). Come and go reconsidered: Children's use of deictic verbs in contrived situations. *Journal of Verbal Learning and Verbal Behavior* 15(6), 655–665.
- Roberts, C. (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics & Pragmatics* 5, 1–69.
- Roberts, C. (2015). Indexicality: de se semantics and pragmatics.
- Ryskin, R. A., A. S. Benjamin, J. Tullis, and S. Brown-Schmidt (2015). Perspective-taking in Comprehension, Production, and Memory: An Individual Differences Approach. *Journal of Experimental Psychology: General* 144(5), 898–915.
- Sayers, D. (1926). *Clouds of Witness*. Distributed Proofreaders Canada.
- Sayers, D. (1935). *Gaudy Night*. Distributed Proofreaders Canada.
- Schlenker, P. (2003). A plea for monsters. *Linguistics and Philosophy* 26, 29–120.
- Schlenker, P. (2004). Context of thought and context of utterance. A note on free indirect discourse and the historical present. *Mind and Language* 19(3).
- Schwager, M. (2010). On what has been said in Tagalog: reportative ‘daw’. In T. Peterson and U. Sauerland (Eds.), *Evidence from evidentials*, pp. 221–246.
- Sells, P. (1987). Aspects of Logophoricity. *Linguistic Inquiry* 18, 445–79.
- Sharvit, Y. (2008). The Puzzle of Free Indirect Discourse. *Linguistics and Philosophy* 31(3), 353–395.
- Smith, C. S. (2009). Accounting for subjectivity (point of view). In R. P. Meier, H. Aristar-Dry, and E. Desрюs (Eds.), *Text, Time, and Context*, pp. 371–393. Springer.

- Smith, N. J., N. D. Goodman, and M. C. Frank (2013). Learning and using language via recursive pragmatic reasoning about other agents. In *Advances in Neural Information Processing Systems*, Volume 26.
- Speas, M. and C. Tenny (2003). Configurational properties of point of view roles. In A. M. Di Sciullo (Ed.), *Syntax and Semantics*, Volume 1 of *Asymmetry in Grammar*.
- Stalnaker, R. (2008). *Our knowledge of the internal world*. Oxford University Press.
- Stephenson, T. (2007). Judge dependence, epistemic modals, and predicates of personal taste. *Linguistics and Philosophy* 30(4), 487–525.
- Stojnić, U. and D. Altshuler (2019). Formal properties of ‘now’ revisited.
- Stokke, A. (2013). Protagonist Projection. *Mind and Language* 28(2), 204–232.
- Sudo, Y. (2018). Come vs. Go and Perspective Shift. Rutgers University.
- Sundaresan, S. (2018). Perspective is syntactic: evidence from anaphora. *Glossa*.
- Sundaresan, S. (2020a). A new theory of indexical shift.
- Sundaresan, S. (2020b). Shifty Attitudes: indexical-shift vs. perspectival anaphora.
- Taylor, K. A. (1988). We’ve got you coming and going. *Linguistics and Philosophy* 11(4).
- Ullman, T. D., Y. Xu, and N. D. Goodman (2016). The Pragmatics of Spatial Language.
- Volden, J., R. F. Mulcahy, and G. Holdgrafer (1997). Pragmatic language disorder and perspective taking in autistic speakers. *Applied Psycholinguistics* 18(2), 181–198.
- Wardlow, L. (2013). Individual Differences in Speaker’s Perspective Taking: The Roles of Executive Control and Working Memory. *Psychonomic Bulletin and Review* 20(4).
- Wechsler, S. (2010). What ‘you’ and ‘I’ mean to each other: Person indexicals, self-ascription, and theory of mind. *Language* 86(2).
- Wilkins, D. P. and D. Hill (1995). When ‘go’ means ‘come’: Questioning the basicness of basic motion verbs. *Cognitive Linguistics* 6, 209–259.
- Winston, M. (1988). *Deictic verbs: use and acquisition*. Ph. D. thesis, University of Connecticut.
- Woolf, V. (1925). *Mrs. Dalloway*. Project Gutenberg.
- Woolf, V. (1927). *To the Lighthouse*. Project Gutenberg.
- Yergeau, M. (2013). Clinically Significant Disturbance: On Theorists Who Theorize Theory of Mind. *Disability Studies Quarterly* 33(4).
- Yoon, E. J., M. H. Tessler, N. D. Goodman, and M. C. Frank (2016). Talking with tact: Polite language as a balance between kindness and informativity. In *Proceedings of the Cognitive Science Society*, Volume 38.
- Yoon, E. J., M. H. Tessler, N. D. Goodman, and M. C. Frank (2017). "I won’t lie, it wasn’t amazing": Modeling polite indirect speech.