

**GRADED MODALITY:  
QUALITATIVE AND QUANTITATIVE PERSPECTIVES**

**Daniel Lassiter**

**Department of Linguistics  
Stanford University  
danlassiter@stanford.edu**

Pre-final draft, as of December 17, 2018, of book to appear with Oxford University Press.

## Preface

This book is about the lexical semantics of scalar expressions, the lexical semantics of English modals, and what we can learn about each by studying them together. Both topics have been of great interest for linguists and philosophers of language, but work on modals has been especially intense in the last 50-odd years. Why is the lexical semantics of this particular corner of our vocabulary a topic of such abiding interest? Why have linguists, philosophers, and (increasingly) psychologists and computer scientists shown so much interest in the meaning of *should*, *must*, *likely*, and *ought*, leaving so few to obsess over the nuances of *grass*, *runway*, *scatter*, *lift*, and *above*?

One possible explanation is that the modals in question belong to the logical vocabulary of English, and that the others belong to the non-logical vocabulary. This may be correct, but I doubt that it is the only reason for their abiding interest. Modals provide a compelling object of study also because of their compelling subject matter: uncertainty, obligation, and goals are matters of deep practical significance to humans, playing a critical role our ability to talk and reason about our psychological mechanisms for navigating the complex and uncertain world we live in. They describe much of the rich social and psychological texture of our lives. What *should* we do? What are our friends and enemies *likely* to do? What do we *want* to accomplish, and how *must* we act in order to bring it about, in light of the information that we have about the world around us and about others' likely actions? Modal language allows us to communicate, inform, negotiate, and complain about topics which are not directly observable, but nevertheless intensely real to humans—and, presumably, to any beings who must act from within a thick fog of uncertainty, based on graded and sometimes conflicting values and with the constant possibility that their choices will engender social approval or reprimand.

It is no surprise, then, that thinkers who are interested in language and cognition have displayed considerable interest in modal language. The main body of modern work on this topic among formally-minded linguists and philosophers has been built on two important foundations: Kripke's (1963) possible-worlds semantics for modal logic, and Hintikka's (1962) logic of knowledge and belief. The tradition that emerged from these works—what I will refer to as the “quantificational” or “classical” semantics for modals—is to analyze modal language, wherever possible, in terms of first-order quantification over sets of possible ways the world might be, or “possible worlds”. *Necessary* and *obligatory* events are those that occur in **all** possible worlds within some appropriately restricted class; *possible* and *permissible* events are those that occur in **some** possible worlds within the relevant class; *impossible* and *forbidden* events are those that occur in **no** possible worlds within the relevant class; and so on.

The bulk of the attention in work on modal semantics has been paid to expressions for which this kind of treatment is most plausible, especially auxiliary verbs such as *must*, *might*, *may*, *should*, and quasi-auxiliaries such as *ought to* and *have to*. What is striking about these modals is the considerable degree to which they are grammatically inert: with the exception of certain limited scope ambiguities, they rarely interact with other expressions in an interesting way, so much so that many modals have been claimed (usually wrongly) to be unembeddable. While adjectives such as *possible*, *permissible*, *obligatory* and attitude verbs such as *think*, *believe*, *know*, and *want* have received a certain amount of attention in the formal semantics literature, their rather more intricate

grammatical behavior has largely gone unnoticed. In particular, their striking similarity to **gradable** expressions has not been analyzed in any detail until recently. Nor has the wealth of other modal expressions that are clearly scalar adjectives (*likely*, *plausible*, *probable*, *certain*, *good*, *desirable*, ...) received significant attention in the literature on the semantics of modality.

In fact, many expressions of modality in English are gradable: they accept degree modifiers, form comparatives and equatives, and pass other standard tests for gradability. For example, I might say that it is *very likely* or *almost certain* to rain; that it is *more likely to rain than to snow*; that I *need very much to sit down* or that I need it *more than I need to get where I'm going*; that it is *better to give money to charity than to gamble it on sports*; that Bill *ought very much* to leave the party, or that he *should leave more than Mary should*; and so on with many more examples. (The latter two are in the better-analyzed auxiliary class, but their gradability has been mostly unnoticed in the large body of previous work on these items. I will illustrate it with a number of naturalistic examples in §8.13.) Many modals form display other typical behaviors of gradable expressions as well, such as vagueness and sensitivity to contextually relevant alternatives.

The main contention of this book can be summed up briefly as follows: we can learn a lot about the lexical semantics of modals if we begin the inquiry afresh, focusing on a careful study of modals that are grammatically complex. We will learn more this way than by beginning with modals which are relatively inert grammatically, simply because the latter, if they do conceal intricate semantic structure, also provide few overt clues that would allow us to identify it. My strategy, then, is to develop a semantics for the complicated modals—in particular, the ones that are clearly gradable, such as *likely*, *certain*, and *good*—and then to use the conclusions reached in this way to infer the hidden structure behind the less revealing auxiliary modals, by examining their logical relationships with the gradable modals.<sup>1</sup>

Luckily, we do not have to start from scratch in analyzing gradable modals. There is a rich body of work on the lexical and compositional semantics of gradable expressions that we can build on. The book will thus begin with a detailed study of the semantics of scalar adjectives, the best-studied class of gradable expressions. In particular, in chapters 1 and 2 I will discuss both degree-based and qualitative semantics for gradable expressions, which provide two different perspectives on the semantics of scalar adjectives. These perspectives are complementary, rather than being in competition, and both are useful when we turn to the study of gradable modals in chapters 3-8.

There is a second, historical reason for beginning a book on modality by considering the theory of gradation in detail. While gradable modals (and, more generally, graded expressions of modality) have received very little attention in the literature until recently, there was some recognition of their existence in early work, with important efforts to integrate them into a possible-worlds semantics for modality in the Kripke/Hintikka tradition by Lewis (1973: §5) and Kratzer (1981, 1991). (Another early effort was Hamblin 1959, but this paper was, unfortunately, almost entirely ignored in subsequent work.) However, while the closely related theories of Lewis and Kratzer discussed modal comparatives, and Kratzer also analyzed some examples involving degree

---

<sup>1</sup> A terminological note: some theorists prefer to reserve the term “modal” for items in the syntactic category that linguists call “modal auxiliaries”: *must*, *may*, *might*, *should*, and so on. Nothing of theoretical significance seems to turn on this terminological issue, but I will follow the main body of work in linguistic semantics in using the term “modal” to refer to the broader, semantically defined category.

modification, neither drew out the connection to comparison and degree modification in other semantic domains or attempted to give a compositional semantics for complex modal expressions. These theories, while insightful, were thus radically incomplete. As we will see in chapter 3, it is technically straightforward to integrate the Lewis/Kratzer theory with modern theories of the lexical and compositional structure of gradable expressions using the Measurement-Theoretic apparatus introduced in chapter 2. However, the result turns out to make a range of incorrect predictions about the grammatical and inferential behavior of gradable modals.

These remarks suggest a second gloss on the primary message of this book: when we are theorizing about the lexical semantics of modals, the tradition has been to focus on the analogy with quantifiers such as *some*, *all*, and *none* and to try to graft on gradable modals as an afterthought. Instead, I suggest, our leading analogy should be to scalar adjectives such as *big*, *small*, *empty*, *full*, and *enormous*. I will try to see how far I can push this analogy. While it may not be possible to banish the specter of the classical semantics entirely, I will contend that even certain non-gradable modals may be fruitfully analyzed as scalar expressions whose thresholds do not, for grammatical reasons, interact with degree-binding operators. I will argue that this perspective provides a straightforward account of logical relations between gradable and non-gradable modals that are troubling for the classical semantics, and that it leads to many new ideas that have been obscured by the piecemeal approach to gradation in previous work.

A second key message of this book is that theories of modality have much to gain by engaging with the rich, sophisticated literature on related topics from neighboring fields—notably, cognitive science, formal epistemology, and decision theory. For example, modal semantics has developed in near-complete isolation from formal models of uncertain reasoning and decision-making, even when the theories were being developed by the same people (e.g., Bas van Fraassen and David Lewis). The scalar analyses of modal language developed here turn out to be closely related to the Bayesian analyses of uncertainty and value which occupy a prominent position in the literature on reasoning and decision-making, where they provide standard formalisms for the representation of uncertainty and value. In recent work in psychology and artificial intelligence, Bayesian treatments have become standard as well. Whatever the reason for the historical lack of engagement between theorists working on closely related topics in different fields, it is time to re-engage. This book can be seen as an argument that the Bayesian formalisms that are widespread elsewhere in philosophy and cognitive science can be put to good use in modal semantics. Their attraction comes not only from the attractive theoretical unification that they provide with other areas of inquiry, but also—and most importantly—from the fact that they allow us to make sense of many linguistic phenomena which are difficult or unnatural to analyze from within the classical perspective.

**History and caveats.** This book is a descendant of my 2011 NYU thesis *Measurement and Modality*. However, the present work differs in many respects from the thesis. I have reconsidered and revised most of the core theoretical claims of the dissertation, expanded the treatment of some items and suppressed others, and in one notable case—the interpretation of deontic *ought* and *should*—adopted a totally different and (I hope) improved analysis. Most other parts have been thoroughly rewritten as well, to improve presentation and content or in reaction to subsequent literature.

On the latter point, in particular, a disclaimer is in order. When I began to work on this topic

around 2009, the literature on gradable modals was tiny, consisting mainly of the work by Lewis and Kratzer mentioned above; Villalta's (2008) work on gradability and the Spanish subjunctive; a chapter of Eric Swanson's dissertation (2006); a brief but insightful section of Seth Yalcin's influential "Epistemic modals" (Yalcin 2007); and several pages in Portner's (2009) survey volume on modality which gave a very clear discussion of Yalcin's work and its relationship to Kratzer's and Kennedy's. The latter two sources, in particular, had a major impact on the development of the dissertation and eventually this book. In subsequent years, there has been an real explosion of work on the topic. In addition to my own work, the years since 2011 have seen a huge variety of relevant work, including a very insightful paper on epistemic comparison from a logical perspective by Holliday & Icard (2013) and an entire dissertation by Klecha (2014). Some have responded directly to the dissertation on which this book is based, and I have made a special effort to modify the content to reflect their insights as appropriate; in other cases, I have tried to incorporate or respond to these ideas as much as I am able.

The flood of new and exciting ideas, alternatives, and clever defenses of the classical semantics has been so strong in recent years that it has delayed publication of this book to a shameful extent. Frequently, as revisions seemed to be finished to a nearly acceptable level, new papers would appear in my news feed that would force me to rethink yet another part of the analysis, or to try to account for additional empirical phenomena that had not previously been on my radar. It is time to stop: unless I make a clean break now, this book will never be finished! I can only look forward to grappling with the many further ideas that will emerge from the literature on this rich topic. Where I should have discussed existing work but have failed, I can only offer my sincerest apologies, and a request: please do let me know so that I can do better next time.

There are many important aspects of modal semantics that I will not discuss here. For example, I will not attempt to account for a number of key deontic items—*must*, *may*, *permissible*, *obligatory*—simply because the book is already very long, and much of what I have to say is readily inferred from the treatment of *ought/should* and the usual assumptions about the semantic relations among these items. I also will not deal in any detail with conditionals or circumstantial modals. In the case of conditionals, while I touch on them in chapters 7-8 when discussing puzzles involving deontic conditionals, the treatment is fairly cursory: conditionals are very complex, and the book is already too long. In the case of circumstantial modals, it is because I do not have a clear picture of what I want to say. Neither will I discuss the recent methodological debate about whether we want the modal semantics to encode inferences at all, as opposed to leaving the semantics very weak and allowing some other form of reasoning to take over (Carr 2012; Charlow 2015). As I discuss in Lassiter (2015c), this debate is not really about modal semantics: it is a foundational issue for much of natural language semantics, which quite generally relies heavily on inference data to motivate proposals about the meanings of words and complex expressions. I do not see any reason to believe in the existence of a sharp distinction between meaning and inference, and indeed I think that there is much fruitful territory for work in the cognitive science of meaning which explicitly treats the former as a special case of the latter (cf. Goodman & Lassiter 2015). Along with most of the linguistic semantics literature, I will merrily conflate the two.

In addition, I will not talk about the debate among contextualists, relativists, and expressivists. For reasons articulated nicely by Yalcin (2011), I believe that this controversy, while fascinating,

is basically orthogonal to the issues discussed here. The key observation is that any time a formal semantics makes crucial reference to parameters—such as Kratzer’s modal base and ordering source, or scales/measure functions representing value and probability—the issue of what, if anything, determines the actual value of the parameter in question in a given conversational context is not something that the semantics can, or should, determine. Rather, this is a metasemantic question which is largely independent of the question of what to do with the interpretation once crucial parameter values have been filled in (at least, independent modulo certain details involving embedded modals which may provide insights into the syntax-semantics interface). Thus, while I describe my semantics throughout in terms that might appear more amenable to contextualists, I suspect that it would be relatively straightforward to interpret the parameters that I invoke in a way that would render the theory acceptable to relativists or expressivists.

**How to read this book.** The book is written to be a single extended argument, and the best way to read it is to follow the King of Hearts’ advice (Carroll 1866: ch.XII).

“Begin at the beginning,” the King said gravely, “and go on till you come to the end: then stop.”

While the book is unfortunately not very modular, not every reader will need to read every word. Here are some recommendations for abbreviation for specific groups of readers:

- A. Readers who have a thorough knowledge of **degree semantics** will be able to skim chapter 1.
- B. Readers who are primarily interested in **epistemic modality** should read chapters 1-6 (subject to the caveat in point A).
- C. Readers who are especially interested in **deontic modals** should read chapters 1-3 and 7-8.

# Table of contents

<b>Preface</b>	<b>i</b>
<b>Table of contents</b>	<b>vi</b>
<b>Acknowledgements</b>	<b>viii</b>
<b>Dedication</b>	<b>x</b>
<b>Chapter 1: Gradation, scales, and degree semantics</b>	<b>1</b>
1.1 Introduction and Motivation	1
1.2 Scalarity and Gradability	3
1.3 Gradable Adjectives and the Typology of Scales	7
1.4 Two Perspectives on Scales	15
1.5 Summary and Preview	16
<b>Chapter 2: Measurement theory and the typology of scales</b>	<b>18</b>
2.1 Introduction to Measurement Theory	19
2.2 Measure Functions, Interpretability, and the Typology of Scales in RTM	23
2.3 Compositional degree semantics without degrees	30
2.4 Measurement Theory and Natural Language Scales	33
2.5 Interval scales and intermediacy	43
2.6 Summary and Conclusion	47
2.7 Appendix	48
<b>Chapter 3: Previous work on graded modality: Lewis and Kratzer</b>	<b>50</b>
3.1 Modality and gradation	50
3.2 Classical and ordering-based semantics	51
3.3 Lewis on graded modality	52
3.4 Kratzer on graded modality	58
3.5 Conclusion	74
<b>Chapter 4: Epistemic adjectives: <i>Likely</i> and <i>probable</i></b>	<b>75</b>
4.1 Scale structure and modifiers	77
4.2 Boundedness and the relative/absolute distinction	90
4.3 Chapter summary	112
<b>Chapter 5: Certainty and possibility</b>	<b>114</b>
5.1 <i>Certain</i> and <i>sure</i>	114
5.2 Epistemic <i>possible</i>	126
<b>Chapter 6: Implications for the epistemic auxiliaries</b>	<b>136</b>
6.1 <i>Must</i> , <i>certain</i> , and (in)directness of evidence	136
6.2 Entailments among epistemic auxiliaries, <i>likely</i> , and <i>possible</i>	139
6.3 Theoretical accounts of <i>must</i> and <i>might</i>	141

6.4	Context-sensitivity of <i>might</i> and <i>must</i>	145
6.5	Alternative-sensitivity of <i>might</i> and <i>must</i>	149
6.6	Are <i>must</i> and <i>might</i> gradable?	152
6.7	A mixed theory, and the status of “epistemic <i>ought</i> ”	155
6.8	Conclusion	159
<b>Chapter 7:</b>	<b>Scalar goodness</b>	<b>162</b>
7.1	Minimum/maximum/relative classification	162
7.2	Aspects of scale structure	164
7.3	Is goodness positive, maximal, or intermediate?	165
7.4	Problems with maximality	167
7.5	The logic of scalar goodness	171
7.6	A possible-worlds semantics for scalar goodness	172
7.7	Applications	175
7.8	Varying probability and values in the Miners’ Puzzle	186
7.9	Positive form <i>good</i>	189
<b>Chapter 8:</b>	<b><i>Ought</i> and <i>should</i></b>	<b>194</b>
8.1	Basic connections among <i>ought</i> , <i>should</i> , and <i>better</i>	194
8.2	Defining <i>ought</i> and <i>should</i> in terms of <i>better</i> ?	195
8.3	Sloman’s Principle	200
8.4	Sloman’s Principle and supererogation	202
8.5	Uncertainty, choice, and information redux	205
8.6	Deontic detachment	209
8.7	Anti-coordination and agglomeration: An experiment	212
8.8	Modeling anti-coordination scenarios	217
8.9	The Smith argument	218
8.10	Weakening	221
8.11	Conflicting <i>oughts</i> and <i>shoulds</i>	225
8.12	Evaluating a revised classical theory of <i>ought</i>	229
8.13	Gradability of <i>ought</i> and <i>should</i>	232
8.14	Deriving the Smith and Weakening principles from the scalar semantics	237
8.15	Summary	239
<b>Chapter 9:</b>	<b>Concluding remarks</b>	<b>241</b>
<b>References</b>		<b>243</b>



## Acknowledgements

I am extremely grateful to the many people who gave me feedback on this work: first and foremost, Chris Barker, Seth Yalcin, Anna Szabolcsi, Chris Kennedy, and Philippe Schlenker. Each played a crucial role in helping these ideas take shape. Anna's work on weak islands first suggested to me that modals might not be quantifiers; having considered an implementation as degree expressions, I was primed for the probabilistic degree semantics for *probable* proposed in Yalcin 2007, which I learned of in Seth's 2009 seminar at NYU's Philosophy department. Philippe was the first to introduce me to measurement theory. Chris K. provided the critical semantic scaffolding, and posed clearly many questions about scale structure that inform key parts of the book. Chris B. gave detailed feedback on all parts, along with wisdom about linguistics and life, bulgogi, and much more.

This book does not adequately reflect Noah Goodman's influence on my thinking about language and cognition. Perhaps someday I will be able to write a book that does justice to the potential of computational cognitive science to inform modal semantics (and vice versa). But, for better or worse, this is not that book.

Discussions with Fabrizio Cariani on drafts of this book and related material had a considerable impact on chapters 7-8. Thanks, Fabrizio! Many thanks also to Stanley Peters and Alan Bale, who read drafts of several chapters and gave detailed feedback and discussion. I'm also grateful to Chris Potts, Beth Levin, Cleo Condoravdi, Annie Zaenen, Mike Frank, Lauri Karttunen, Tom Wasow, Judith Degen, Greg Scontras, Andreas Stuhlmüller, and Justine Kao for many conversations on topics related to this material and for making Stanford a wonderful place to live and learn; and to my students, especially those in my "Gradation & Modality" seminar in the winter of 2014 and my joint seminar with Cleo Condoravdi on "Modals & Conditionals", who suffered through drafts of several chapters and caught many errors and infelicities. In addition, I'm grateful to Wes Holliday and Thomas Icard, whose published and unpublished work inspired me and helped me to improve the discussion in numerous places; to Paul Portner, Aynat Rubinstein, and Angelika Kratzer for helpful and critical comments and for pointing out a blunder in the original version of my dissertation; to Peet Klecha, whose critical discussion on earlier versions of this work led to many improvements in the discussion in chapters 4-6; to Barry Smith for providing me with office space, philosophical stimulation, and excellent wine during my sojourn in London; and to Eric Swanson and Sarah Moss for their insights on modals, babies, and kayaking. I also owe thanks to Nate Charlow, Michael Glanzberg, Robert van Rooij, Wataru Uegaki, Elena Herburger, Graham Katz, Robert Shanklin, Simon Charlow, Daniel Rothschild, Bob Beddor, Adrian Brasoveanu, Pranav Anand, Valentine Hacquard, Salvador Mascarenhas, David Beaver, Louise McNally, Roger Levy, Eytan Zweig, Rick Nouwen, Larry Horn, Benjamin Spector, Julien Dutant, Napoleon Katsos, Florian Jaeger, Neil Myler, Chris Collins, Pauline Jacobson, Bob Frank, Michael Franke, Leon Bergen, Tom Leu, Tim Leffel, Mike Solomon, Tricia Irwin, Luka Crnic, and the many reviewers on this and related projects who I cannot thank by name.

I owe a great deal to Randall Nichols, Bert Vaux, Calvert Watkins, Warren Goldfarb, and Alan Musgrave, amazing teachers whose advice I did not always take (but should have), but who did succeed in instilling a love of language, reason, and the big picture. Finally, much love and many

thanks to Emma and Charlie, who have done everything they could to keep my feet on the ground while I was working on this.

## **Dedication**

To my mother Sally and my father Ike, for whose encouragement, love, and sacrifices I will always be grateful.

## CHAPTER 1

### Gradation, Scales, and Degree Semantics

#### 1.1 Introduction and Motivation

Recent work on modality in formal semantics has highlighted the fact that many modal expressions are GRADABLE; for example, they accept at least some degree modifiers and can take part in comparatives and equatives (Yalcin 2007, 2010; Portner 2009; Lassiter 2010, 2011a, 2015b). Although these papers have discussed epistemic modal adjectives for the most part, gradability in the modal domain goes beyond adjectives and beyond epistemic modals. Some examples are given in (1.1).

##### (1.1) Gradability among Modals

- a. How necessary is it to marinate meat before making jerkies? (Degree questions)
- b. Bill wants to leave as much as Sue wants to stay. (Equatives)
- c. I need to go on vacation more than I need to finish this work. (Comparatives)
- d. Concerns of autonomy ought very much to matter. (Degree modification)
- e. You are much more likely to roll an even number as a 5. (Modified comparatives)
- f. You are three times as likely to roll an even number as a 5. (Modified equatives)

The modal expressions in these examples cut across syntactic categories, including adjectives *necessary*, *certain*, main verbs *want*, *need*, and the (quasi-)auxiliary *ought*.

The examples in (1.1) closely resemble examples of gradability among the better-studied classes of non-modal gradable adjectives and verbs, for example:

##### (1.2) Gradability among Adjectives

- a. How angry can I make my teacher and still get an A? (Degree questions)
- b. The Gettysburg Memorial is as old as the Eiffel Tower. (Equatives)
- c. My child is cleverer than yours. (Comparatives)
- d. This carton of milk is almost empty. (Degree modification)
- e. Emma is much taller as Charlie. (Modified comparatives)
- f. Emma is three times as tall as Charlie. (Modified equatives)

##### (1.3) Gradability among Verb Phrases

- a. How much do you like chocolate? (Degree questions)
- b. Charlie likes chocolate as much as Emma does. (Equatives)
- c. I loathe *Battlefield Earth* more than any other movie. (Comparatives)
- d. Harriet has almost finished her art project. (Degree modification)
- e. I ran much more than you did. (Modified comparatives)
- f. I ran three times as much as you did. (Modified equatives)

In formal semantics, the most prominent approach to gradation adjectival and verbal domain is to tie these facts to *SCALES*—that is, to abstract representations of measurement to which gradable expressions relate their arguments. Scales are assumed to be composed of *DEGREES* which are partially or totally ordered. Roughly, then, *clever* is an expression which relates people to their degrees of cleverness; *loathe* is an expression which relates pairs of individuals  $(x,y)$  to the degree to which  $x$  loathes  $y$ ; and so on.

This book considers gradability, comparison, and other evidence for scalar semantics in the modal domain. I develop a semantics for a variety of epistemic and deontic expressions—those that talk about uncertainty, obligation, and neighboring concepts—which is very closely related to standard scale-based theories of the semantics of non-modal gradable expressions. In particular, modals such as *(un)likely*, *(un)certain*, *(im)possible*, *good/bad*, *ought*, and *should* are analyzed, like gradable adjectives, as expressions which relate their propositional arguments to points on a scale. The differences lie in the interpretation of the scale as likelihood or goodness, in the formal structure of the relevant scales, and in the way that individual lexical items’ interpretation depends on the scale in question.

Much of the book is dedicated to a detailed discussion of scalar structure and how it related to the logical properties of different kinds of modals. The usual assumption that modals are either *SOME*-, *ALL*-, or *NONE*-quantifiers over sets of possible worlds restricts the logical expressiveness of the theory of modality considerably: *SOME*- and *ALL*-quantifiers are upward monotone, and *NONE*-quantifiers are downward monotonic. However, as I will show, there is good reason to believe that at least some scalar deontic items are non-monotone. Since there do not seem to be any non-monotone quantifiers which we could plausibly invoke here, the existence of non-monotone modals calls into question the viability of a thoroughly quantificational semantics for modals. On the other hand, I show in chapter 2 that a scalar semantics can easily encode non-monotonicity, and that there are independent arguments that some English adjectives, such as *cold/hot* and temporal *long*, live on scales with a non-monotone (or “intermediate”) structure. In chapters 7-8, I show that we can account for a number of puzzles involving deontic items—especially *good*—if we suppose that they have a scalar semantics, and that their scales have the same formal structure as those of the intermediate adjectives.

Going further, I argue that some modal expressions which do not show evidence of gradability—notably the epistemic auxiliaries *might* and *must*—may have a semantics built around scales nonetheless. That is, the logical relations between gradable and non-gradable modals can be understood straightforwardly if both have a scalar semantics. Implementing this idea requires making a careful distinction between semantic **scalarity** and grammatical **gradability**, which is discussed in the next section of this introductory chapter. While the arguments for this final point are somewhat more tentative, if it is correct, we may be able to expunge direct quantification over possible worlds entirely from modal semantics, replacing it with propositional scales and thresholds, or (in some cases) quantification over sets of propositional alternatives.

Chapters 1 and 2 give the necessary theoretical and technical background on scalar semantics, both degree-based and qualitative. In chapters 3-8 I analyze the behavior of modals as scalar expressions, using entailment data, corpus data, and experimental results to assign these expressions to scale types which are independently attested in the semantics of gradable adjectives. I also show

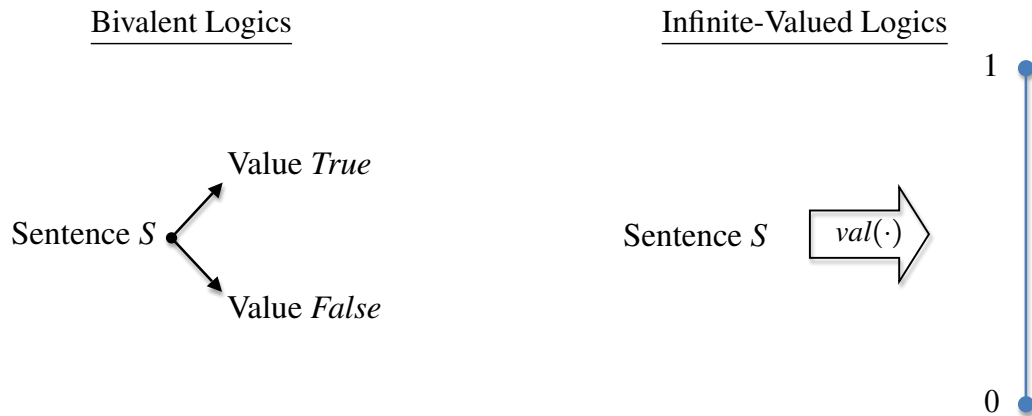
that, in a variety of domains, quantificational theories make incorrect predictions about the semantic behavior of epistemic and deontic modals, while the predictions of the scalar alternative that I develop are borne out.

## 1.2 Scalarity and Gradability

### 1.2.1 What is Scalar Semantics?

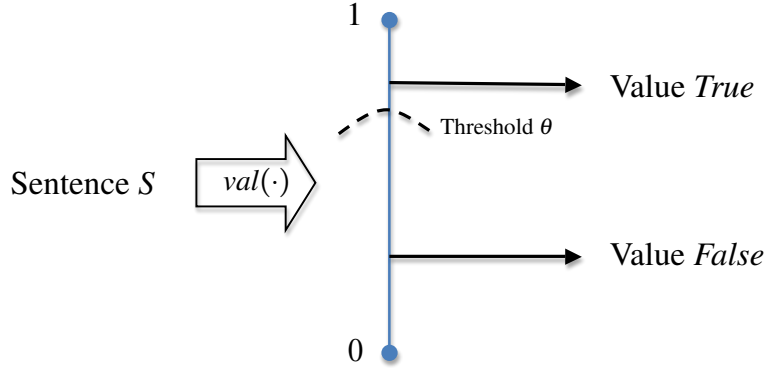
Scalar approaches to the semantics of various domains have gained increasing popularity in formal semantics in recent years, and have been employed in the analysis of gradable adjectives, telicity in verbs, prepositional phrases, common nouns, and elsewhere. Traditionally it has been assumed that the semantic effect of expressions is to partition their domain into two values, an EXTENSION and an ANTI-EXTENSION. In the simplest case, scalar representations can be seen as an enrichment of the classical assumption to allow for many values, often an infinite number; the scale is just the collection of all possible values of the representation, along with an ordering on the values. Two-valued representations can be re-captured from scalar representations, when appropriate, by the use of THRESHOLD VALUES.

For instance, in bivalent semantics the valuation function *val* maps well-formed and meaningful sentences to one of two values, True and False. A well-known (though controversial) proposal for a scalar enrichment of truth-values is fuzzy logic, where the range of *val* is not  $\{True, False\}$  but the infinite subset  $[0, 1]$  of the real numbers (Zadeh 1965, 1978). Fuzzy logic is just one of a variety of infinite-valued logics with this basic form.



Infinite-valued logics do not necessarily reject bivalence completely, though; it is always possible to recover bivalent truth-values from a linearly ordered infinite-valued representation by utilizing thresholds. A threshold is a distinguished value on the scale which is used to partition the domain of *val* into two subsets, which (in the simplest case) can be identified with the bivalent truth-values *True* and *False*. I will generally use  $\theta$  as a variable over threshold values. So, for example, if  $\theta$  is (arbitrarily) set at .7 then we have the following three-step mapping from sentences to bivalent truth-values:

### Three-Step Bivalent Logic



Here bivalent truth-values are determined by first mapping a sentence to a value in  $[0, 1]$ , then establishing a threshold value, and finally by comparing the value of the sentence to the threshold in order to determine whether it should count as True or False relative to that threshold.

This is not, as far as I know, a version of fuzzy logic that anyone has proposed seriously, but it serves to illustrate the two-step process for creating binary values that is used in the best-known variety of scalar semantics. Classically, properties have been thought of as sets of objects: “ $x$  has property  $P$ ” is true if and only if  $x \in A$ , where  $A$  is a set containing all and only the individuals or objects which have property  $P$ . Just as classical bivalent logic presupposes that all sentences can be assigned one of two values (True or False), the classical theory of properties presupposes that individuals are related to properties in one of two ways—they are either in the set or not:

$$(1.4) \quad \llbracket \text{tall} \rrbracket^{\mathcal{M}, w} = \lambda x_e [x \in \mathbf{tall}]$$

where  $\mathbf{tall} \subseteq \mathbf{D}_e$ , the domain of individuals.

Scalar semantics generalizes this approach by allowing that scalar expressions map objects to a possibly infinite set of values called **degrees**. In the simplest case, degrees are totally ordered, so that once a threshold has been chosen, the objects that map to a given degree are divided into two sets: those whose degree exceeds the threshold, and those for whom the threshold meets or exceeds their degree. (If the set is not *totally* ordered, there may be a third set of degrees which are not ordered relative to the threshold.) Unlike the contrived example of binarized fuzzy truth-values, this approach is widely thought to yield reasonable results in the case of gradable adjectives.

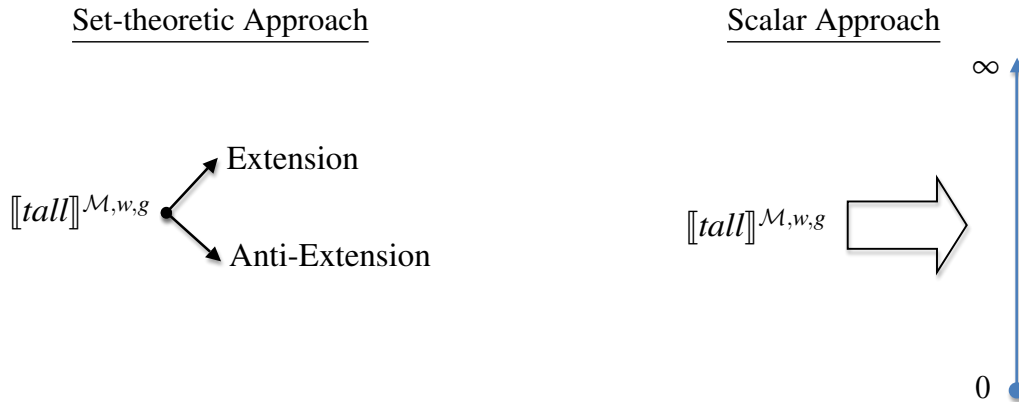
One well-analyzed example is the property of height, as instantiated e.g. by the adjectives *tall* and *short*. Individuals can be *tall* to various degrees; on the variety of scalar semantics that we will focus on throughout this book, this intuition is explained by treating *tall*, not as a function from individuals to truth-values as in (1.4), but as a function from individuals to degrees of height as in (1.5).

$$(1.5) \quad \llbracket \text{tall} \rrbracket^{\mathcal{M}, w} = \lambda x_e [\mathbf{height}(x)],$$

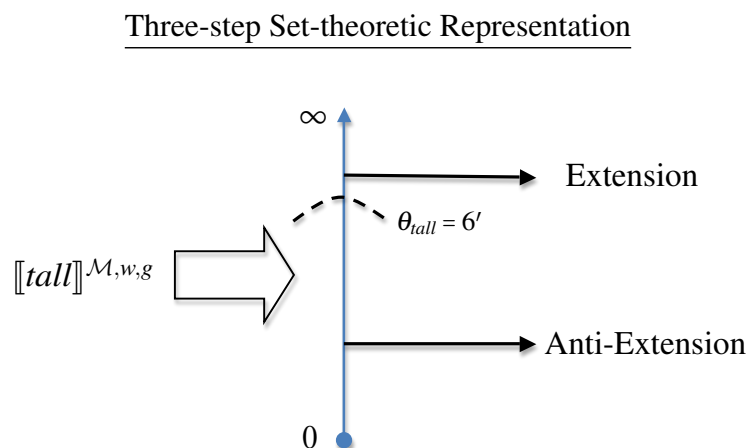
where  $\mathbf{height}(x)$  is a function from  $\mathbf{D}_{tall} \subseteq \mathbf{D}_e \rightarrow (0, \infty)$ .

This account—due in particular to Bartsch & Vennemann (1973); Kennedy (1997, 2007), and with close relatives in (e.g.) von Stechow 1984; Rullmann 1995; Schwarzschild & Wilkinson 2002

and many more—captures the basic features of gradability in a straightforward way. The essential idea is that gradable adjectives denote MEASURE FUNCTIONS, which map objects (etc.) to points on a scale. The difference between the set-theoretic and scalar conceptions of properties is essentially the same as the contrast between bivalent and infinite-valued approaches to truth-values:



Despite the fact that height is a property that comes in degrees, *tall* and *short* do sometimes function to partition their domains into (at least) two sets, as in the classical approach. The unmodified form *tall* in *Mary is tall* is called the POSITIVE FORM. In the measure function analysis, the denotation of the positive form is calculated by a three-step procedure. First, we find the height of the individual argument of the adjective; second, we establish a threshold value; and finally we compare the individual's height to the threshold value. Supposing (arbitrarily) that the threshold for counting as *tall*,  $\theta_{tall}$ , is 6 feet:



How the threshold value is determined for the positive form of gradable adjectives is a much-discussed and complex issue; at a minimum, discourse context, world knowledge, and lexical semantics play a role. There is also considerable debate about whether there is a unique threshold value operative in a given context. For simplicity's sake, I will often speak as if there is a fixed value of  $\theta_{tall}$  given each context. However, everything I say here could easily be generalized to allow for



indeterminate, fuzzy, or probabilistic threshold values. (For discussion of how these alternatives fit into the style of semantics used here, see Lassiter 2015a.)

A point that bears highlighting here is that there is nothing special about gradable adjectives with respect to this construction. Many natural language expressions are standardly analyzed as partitioning their domains into two sets: their extension and its complement, True and False, etc. In principle, the extension of any expression with these characteristics could turn out to be determined by means of a scalar representation along the lines just sketched. As long as there is some way of determining a scale, a measure function, and a threshold value, it is always possible to use scalar representations to determine classical set-theoretic/bivalent extensions. Of course, we have to ask on a case-by-case basis whether there is evidence for a scalar representation.

This book is essentially an extended argument for the usefulness of this approach in a domain to which it has rarely been applied, the analysis of modality. I will argue that many (perhaps all) epistemic and deontic modals have a semantics built that makes use of scales. Scalar modals denote (or have denotations which make crucial use of) measure functions on propositions, i.e. functions that map propositions to points on a scale. As in the case of scalar semantics for other expression-types, sentences containing scalar modals are assigned truth-values using an algorithm that makes crucial use of the two-step scalar construction described above. First, they map their propositional argument to a degree on the relevant scale. Then, they compare the result to a threshold value which is determined by some combination of lexical semantics, compositional semantics, and discourse context. A rough idea of the kind of truth-conditions of that modal sentences receive on this proposal can be seen in (1.6):

- (1.6) a.  $\llbracket \text{It is likely that Red will win} \rrbracket^{\mathcal{M},w} = 1$  if and only if  $\mu_{\text{likely}}(\llbracket \text{Red wins} \rrbracket^{\mathcal{M},w}) \geq \theta_{\text{likely}}$   
 b.  $\llbracket \text{Red ought to win} \rrbracket^{\mathcal{M},w} = 1$  if and only if  $\mu_{\text{ought}}(\llbracket \text{Red wins} \rrbracket^{\mathcal{M},w}) \geq \theta_{\text{ought}}$ ,

where  $\mu_{\text{likely}}$  and  $\mu_{\text{ought}}$  are the measure functions lexically associated with *likely* and *ought*, respectively.

This analysis still leaves many questions of interest unspecified, for example:

- What are the structural features of the scales which  $\mu_{\text{likely}}$  and  $\mu_{\text{ought}}$  map propositions to? What kind of logic do these structural features, along with other relevant constraints, induce?
- Do these scales differ in any semantically interesting way, and if so can any inferential differences between them be attributed to differences in scalar structure?
- Are there other items which map expressions to the same scale as *likely* and *ought*? If so, how and why do their meanings differ?
- How similar are *likely* and *ought* to (resp.) gradable adjectives like *tall*, and gradable verbs like *love*? Are there systematic differences in their semantics beyond the difference in the semantic type of their arguments?
- What kinds of operators can bind the threshold variables  $\theta_{\text{likely}}$  and  $\theta_{\text{ought}}$ ?

The bulk of this book is dedicated to a detailed investigation of these questions. I will argue that there are many interesting semantic connections between modals and non-modal adjectives in the

details of scalar structure, interaction with operators of various kinds, and effects of discourse context on interpretation.

### 1.2.2 What is Gradability?

What is gradability? What distinguishes the gradable expressions from the non-gradable ones, and how is it possible to be scalar without being gradable?

As I will use the term, an expression  $E$  is GRADABLE if it interacts grammatically with other expressions whose semantic function is to manipulate  $E$ 's threshold value. For example, *tall* is gradable because it interacts with the threshold-manipulating operators in (1.7): measure phrases, comparatives, and degree modifiers.

- (1.7) a. Joan is 5 feet tall.  
b. Harry is taller than Larry.  
c. Sam is very tall.

Although the method of deriving a classical extension for *tall* sketched in the last section looks rather roundabout, the intermediate steps come in handy when we are called upon to deal with threshold-manipulating operators like *very* and *5 feet*. By using a three-step process involving scales and threshold values to determine a classical extension for these expressions, we can account for the differences between *tall*, *very tall*, and *5 feet tall* in a straightforward way: *very* and *5 feet* temporarily change the threshold value which is used to determine a classical extension. So, for example, even if the global value of  $\theta_{tall}$  is 6 feet, (1.7a) will come out true if Joan's height is at least 5 feet, because  $\theta_{tall}$  has been reset to 5 feet for the purpose of evaluating this expression. There are numerous ways to implement this analysis compositionally, one of which will be described in the next section.

As this discussion implies, a precondition for an expression's being gradable is that it must be associated with a scale and a threshold value: otherwise the threshold-manipulating operators would have nothing to operate upon. However, the opposite implication does not necessarily hold—logically, there could be scalar expressions which are not gradable because they have a threshold which cannot be manipulated grammatically. If such expressions exist, they determine their extensions using scales as an intermediary, but have threshold values which are either fixed once and for all, or sensitive to contextual factors but not to grammatical manipulation.

This distinction between scalarity and gradability is important for the theory of modality, I will argue. Although evidence that an expression  $E$  is gradable is *ipso facto* evidence that  $E$  determines its extension using scales as an intermediary, a lack of evidence for gradability does not necessarily imply that scales are not implicated in  $E$ 's semantics. There is indirect but strong evidence that scales and threshold values are implicated in the semantics of epistemic and deontic, even those which do not combine with operators which manipulate the threshold value.

## 1.3 Gradable Adjectives and the Typology of Scales

This section described one way to implement the scalar analysis of gradable adjectives compositionally, based primarily on Kennedy (1997, 2007). For further discussion, see Lassiter 2015a;

Morzycki 2015. I also briefly discuss the semantics of the positive form, vagueness, the role of comparison classes, and the several types of scales which have been shown to be relevant in the semantics of gradable adjectives. Since adjectives provide the best-studied class of gradable expressions, this treatment is the gold standard for scalar semantics; many details, including the discussion of minimum/maximum/relative classification, scale type, and comparison classes, will play an important role in the scalar semantics of modality developed in later chapters.

### 1.3.1 Compositional Implementation of Scalar Semantics

The measure function analysis of gradable adjectives assumes that, in addition to the (at least) the types  $e$  for individuals,  $t$  for truth-values, and  $s$  for worlds—which have been standard in formal semantics since Montague 1973; Gallin 1975—there is also a basic type  $d$  (for “degree”). *Degrees*, in approaches of this type, are usually thought of as abstract representations of measurement organized into linearly ordered SCALES.

Formally, scales are structures at least as rich as  $\langle D, \leq \rangle$ , where  $D$  is a set of degrees and  $\leq$  is a reflexive, transitive, and antisymmetric binary order. (See §2.1 for definitions of these technical terms.) It is often assumed that the scale is connected, and that it is dense for at least some expressions, and possibly all (Fox & Hackl 2006; Nouwen 2008). When they are connected and dense, scales with this abstract structure can also be modeled as intervals on the real numbers  $\mathbb{R}$  (e.g., Kennedy & McNally 2005a).

While non-gradable adjectives like *British* and *geological* continue to be of type  $\langle e, t \rangle$  in degree semantics, gradable adjectives like *tall* and *happy* which take individual arguments are treated as functions of type  $\langle e, d \rangle$ , i.e. functions from individuals to degrees. The general semantic form of a gradable adjective  $A$  is

$$(1.8) \quad \llbracket A \rrbracket^{\mathcal{M}, w} = \lambda k_{\alpha} [\mu_A(k)]$$

where  $\mu_A$  is the measure function associated with the adjective  $A$  and  $k$  is a variable of type  $\alpha$ , as appropriate for the adjective in question. *Tall*, for example, expresses a function which takes an argument of type  $e$  and returns that individual’s degree of height.

$$(1.9) \quad \llbracket \text{tall} \rrbracket^{\mathcal{M}, w} = \lambda x_e [\mu_{\text{tall}}(x)]$$

Naturally, the interpretation of measure functions must be sensitive to the world parameter, since the heights of individuals can vary across possible worlds. So the official form of (1.8) is:

$$(1.10) \quad \llbracket A \rrbracket^{\mathcal{M}, w} = \lambda k_{\alpha} [\mu_A^w(k)],$$

where  $\mu_A^w$  is the function mapping objects  $k$  to objects the  $A$ -degree that they have world in  $w$ . However, we will generally elide the superscripted  $w$  except when it is specifically relevant.

The measure function analysis treats the comparative morpheme as a three-place relation between measure functions, degrees, and individuals:

$$(1.11) \quad \llbracket \text{more/-er} \rrbracket^{\mathcal{M}, w} = \lambda A_{\langle e, d \rangle} \lambda d_d \lambda x_e [A(x) \geq d]$$

(1.11) requires a syntactic structure where *-er* combines first with the main adjective and then with the comparative clause. The comparative clause itself denotes a definite description of a degree, and

is derived via ellipsis within the comparative clause and movement of a silent operator/*wh*-word (following Bresnan 1973; Chomsky 1977 a.o.). This movement triggers a further degree abstraction and a maximization operation (von Stechow 1984). (For simplicity I ignore phrasal comparatives.) For example:

(1.12) Mary is taller than Harry is.

a. LF: *Mary is*  $[[\text{more tall}] [\text{than } [Op_i [\text{Harry is } [\text{tall } t_i]]]]]$

b.  $\llbracket (1.12) \rrbracket^{\mathcal{M},w} = 1$  iff:  $\llbracket \text{more} \rrbracket^{\mathcal{M},w}(\llbracket \text{tall} \rrbracket^{\mathcal{M},w})(\llbracket \text{than } Op_i \text{ Harry is tall } t_i \rrbracket^{\mathcal{M},w})$

(1.13) a.  $\llbracket \text{than } Op_i \text{ Harry is tall } t_i \rrbracket^{\mathcal{M},w} = \mathbf{max}(\lambda d[\llbracket \text{tall} \rrbracket^{\mathcal{M},w}(\llbracket \text{Harry} \rrbracket^{\mathcal{M},w}) \geq d])$

b.  $= \mathbf{max}(\lambda d[\mu_{tall}(\mathbf{Harry}) \geq d])$

c.  $= \mu_{tall}(\mathbf{Harry})$

(1.14)  $\llbracket (1.12) \rrbracket^{\mathcal{M},w} = 1$  iff:  $[\lambda A \lambda d \lambda x[A(x) > d]](\lambda x_e[\mu_{tall}(x)])(\mu_{tall}(\mathbf{Harry}))$

(1.15)  $\llbracket (1.12) \rrbracket^{\mathcal{M},w} = 1$  iff:  $\mu_{tall}(\mathbf{Mary}) > \mu_{tall}(\mathbf{Harry})$

(Here and throughout the book, italicized words and phrases like *Harry* represent English expressions, while boldfaced expressions like **Harry** represent model-theoretic objects.)

Although the literature has generally concentrated on gradable adjectives of type  $\langle e, d \rangle$ , it is not difficult to adapt this analysis to gradability for arbitrary types; in general, for expressions of Boolean type  $\langle \alpha, t \rangle$ , the corresponding gradable type is  $\langle \alpha, d \rangle$ . For example, a proposition-embedding adjective such as *lucky*—as in *It's lucky that your parents didn't see us*—will fit the schema of (1.8) by setting  $\alpha = \langle s, t \rangle$ .

(1.16)  $\llbracket \text{lucky} \rrbracket^{\mathcal{M},w} = \lambda p_{\langle s, t \rangle}[\mu_{lucky}(p)]$

The comparative and other degree expressions can likewise be given type-polymorphic denotations which allow them to be applied to arbitrary gradable types as needed. So, for example, instead of treating *more/-er* as being of type  $\langle \langle e, d \rangle, \langle d, \langle e, t \rangle \rangle \rangle$ , we could write a type-polymorphic denotation which could equally well be applied to verbal comparatives:

(1.17)  $\llbracket \text{more/-er} \rrbracket^{\mathcal{M},w} = \lambda K_{\langle \alpha, d \rangle} \lambda d_d \lambda k_\alpha [K(k) > d]$

Similar type-polymorphic denotations could be constructed for *as*, *almost*, the positive morpheme **pos** to be introduced shortly, and other degree operators which can modify gradable expressions that take non-individual arguments.

### 1.3.2 The Positive Form, Vagueness, and Comparison Classes

Kennedy (2007) argues that the positive form (with no overt degree modification) is derived via a silent morpheme (or type-shifting operation) *pos*.

(1.18)  $\llbracket \text{pos} \rrbracket^{\mathcal{M},w} = \lambda A_{\langle e, d \rangle} \lambda x_e [A(x) > \theta_A]$

$\theta_A$  is just a free variable here; how its value is determined in context is a complex and controversial question which touches on issues relating to vagueness, the semantics of comparison classes, and minimum/maximum/relative status discussed in this and the next subsection.

The literature on vagueness is vast and contentious, and I will avoid the topic as much as possible here this book (though the interested reader may consult Lassiter 2011b; Lassiter & Goodman 2013, 2015a for my favored approach). For the purposes of this work, we can think of vagueness as a kind of pervasive context-sensitivity of the threshold value, as argued for example by e.g. Fara (2000); Barker (2002); Kennedy (2007). Roughly, certain adjectives in the positive form and with some degree modifiers determine their threshold value by reference to a “norm” or “standard value” whose value is constrained (but perhaps not fully determined) by features of the semantics of the expression, its grammatical environment, and the discourse context.

For example, the sentences in (1.19) can both be true, even if the elephant is much bigger than the flea:

- (1.19) a. This flea is big.  
b. This elephant is not big.

This difference is plausibly explained by assuming that *big* is interpreted with respect to a standard value which is sensitive to features of the discourse (in some way which remains to be elucidated). On this account, then, (1.19a) means something like “This flea is big relative to the relevant norm *N*”, where *N* is given by context.

According to many authors, the threshold value is constrained in part by reference to an implicit or explicit COMPARISON CLASS. If so, the context- and norm-sensitivity of (1.19a) comes down to roughly “This flea is big relative to the expected value for comparison class *C*”, where *C* is some set of objects of which the flea in question is a member. On the plausible assumption that the implicit comparison class relevant to evaluating (1.19a) is the set of fleas, while the relevant comparison class for (1.19b) is the set of elephants, we have the beginnings of an explanation of how these two sentences can be simultaneously true.

APs with explicit comparison classes are typically of the form *A for a NP*, as in (1.20).

- (1.20) a. Harry is heavy for a jockey.  
b. Harry is heavy for a sumo wrestler.

The use of an explicit comparison class generally seems to require that the individual to which the adjective is applied is a member of the comparison class: thus (1.20a) is infelicitous unless Harry is a jockey, and (1.20b) is infelicitous unless he is a sumo wrestler (Kennedy 2007). In Kennedy’s analysis, this is evidence that comparison classes exert their semantic effect by restricting the domain of the measure function. For concreteness’ sake I will make this assumption when we discuss comparison classes below, although the details of this analysis are not vital here. (See for example Bale 2011; Solt 2011 for further discussion and alternatives.)

### 1.3.3 The minimum/maximum/relative typology

Recent work has emphasized the importance of ADJECTIVE TYPE and SCALE TYPE in the semantics of gradable adjectives (Hay, Kennedy & Levin 1999; Rotstein & Winter 2004; Kennedy & McNally 2005a; Kennedy 2007). The fact that adjectives and scales come in a variety of forms will be reflected in our treatment of modality as well.

Unger (1971) seems to have been the first to notice that gradable adjectives come in two types, which he called ABSOLUTE and RELATIVE. Relative adjectives like *tall* and *expensive* have been a primary focus on work on vagueness. It is generally unclear just how tall someone has to be in order to count as tall, or what dollar amount makes an item expensive, even if we have a fully specified comparison class. Relatedly, relative adjectives tend to have context-dependent interpretations: what counts as “expensive” will depend on a number of factors, such as what kinds of objects are under discussion (lunch options or houses), how much money we have available, and perhaps more.

Closely related to the relative adjectives are EXTREME ADJECTIVES, exemplified by *huge*, *tiny*, and *ecstatic*. Extreme adjectives also combine with comparison classes, and in the positive form mean roughly that their argument has a much greater degree of size/happiness than the norm or expected value (subject to the same caveats as relative adjectives about persistent vagueness, and questions about how exactly the comparison class achieves its semantic effect). They do show a number of interesting semantic and distributional differences from typical relative adjectives, though: see Morzycki 2012 for a detailed discussion.

However, Unger points out that vagueness and context-sensitive meaning is not necessarily present in all types of adjectives: for example, whether or not an object is *flat* is plausibly an all-or-nothing affair. If a road has any bumps in it, it is not “flat” but at best “almost flat” or “approximately flat”. This differs considerably from heights and costs: a road must be maximally flat if it is to be flat at all, while someone can be tall without being maximally tall (whatever this would even mean). Adjectives like “flat” are those which Unger calls **absolute**.

The absolute adjectives cleave further into two groups, the MAXIMUM ( “maximum-standard”, “total”) and MINIMUM ( “minimum-standard”, “partial”) adjectives (Cruse 1986; Yoon 1996; Rotstein & Winter 2004; Kennedy & McNally 2005a; Kennedy 2007). Maximum adjectives like *flat*, *full*, *straight*, and *safe* require that an object have a maximal degree of the property in question in order to count as instances of the concept. So, for example, if I tell you that my beer glass is full, it is strange to continue by asserting that it could be fuller (Kennedy 2007). In contrast, minimum adjectives like *bent* and *dangerous* require only that an object have a non-minimal degree of the property in question; for example, an antenna is bent if it has any amount of bend in it. Unlike maximum adjectives, there is generally no oddity in saying that something is bent, but also that it could be more bent.

As Kennedy discusses in detail, absolute (minimum and maximum) adjectives share a number of properties. For instance, the positive form of both types of adjective typically has sharp boundaries and thus an apparent lack of vagueness. Absolute adjectives are also much less sensitive to comparison classes: for example, *This is bent for an antenna* seems strange, at best a funny way to say *This antenna is bent*. Kennedy argues that we can account for all of these properties if absolute adjectives require that their threshold be an extreme scalar value. On this analysis, if  $A$  is a minimum adjective, then  $\theta_A$  must be the minimum point on  $S_A$ , the scale associated with  $A$ : an antenna is bent just in case it has a non-zero degree of bend. Likewise, if  $A$  is a maximum adjective, then  $\theta_A$  must be the maximum point on  $S_A$ : a glass is full just in case it has a maximal degree of fullness. In other words, if  $A$  is maximum, then  $x$  is **pos**  $A$  is true if and only if  $A(x) = \mathbf{max}(S_A)$ . If  $A$  is minimum-standard, then  $x$  is **pos**  $A$  is true if and only if  $A(x) > \mathbf{min}(S_A)$ . (See, however, Lassiter & Goodman 2013 for arguments that this gloss is only approximately true, and that the

difference in vagueness is one of degree rather than of kind.)

### 1.3.4 Scale Structure

If absolute adjectives in the positive form constrain their thresholds to be at the minimum or maximum point of the scale, it follows that an adjective can only be absolute if it is associated with a scale which has a minimum or maximum as appropriate. For this reason scale structure places constraints on minimum/maximum/relative classification: we cannot speak meaningfully of maximum and minimum values with all types of adjectives. Recently a theory of scale types focusing on BOUNDEDNESS properties has been developed by Rotstein & Winter (2004); Kennedy & McNally (2005a); Kennedy (2007). The crucial observation is that scales may vary in the presence or absence of a lower bound, and independently in the presence or absence of an upper bound. Furthermore, this variation can be related to a number of linguistically interesting properties in addition to the relative/absolute distinction.

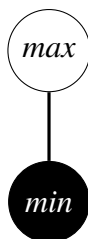
For instance, *danger* is a property which is presumably associated with a scale that has no maximum value: after all, there is no upper limit on how dangerous something can be, at least in principle. A number of theorists have suggested that this fact about *tall* and similar adjectives has linguistic consequences. von Stechow (1984); Rullmann (1995) argue that (1.21) is semantically ill-formed because there is no unique or maximal degree *d* such that San Francisco is not dangerous to degree *d*. As a result, the denotation of the comparative clause (cf. §1.3.1) makes reference to a degree that does not exist.

(1.21) # Los Angeles is more dangerous than San Francisco isn't.

The ill-formedness of (1.21) is not a special fact about degrees of danger, though: sentences of this form are infelicitous with various other adjectives, such as *rich*.

(1.22) # Mary is richer than Sam isn't.

*Rich*, too, is intuitively associated with a scale with a lower bound (\$0) but no upper bound. You can keep getting richer forever if you have enough time, energy, and luck.



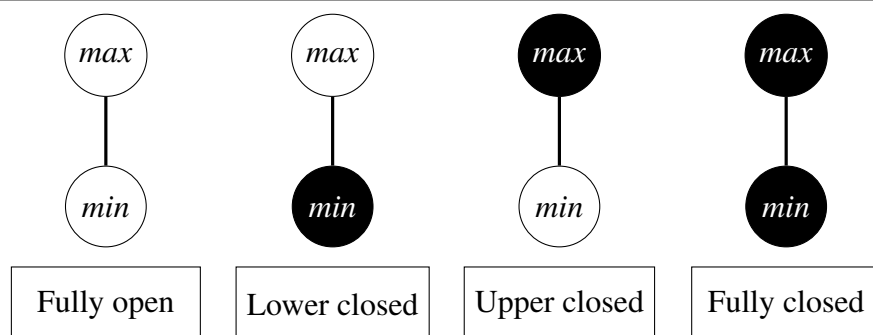
**Figure 1.1** A lower-closed scale.

---

We can represent what *dangerous* and *rich* share in terms of a LOWER-CLOSED scale (Figure 1.1). Here, a filled (black) circle represents an endpoint that is included in the scale, while an

unfilled (white) circle represents an endpoint that is excluded. Note that we use *min* and *max* in the general case, since we cannot assume that every scale will be readily related to numerical values as *rich* and *tall* are, or that 0 will be the minimum for all scales that are.

Let us restrict attention for the moment to scales that are connected: any two points on the scale are ordered. If we allow for all logically possible variations with respect to boundedness properties, the typology of possible scales with respect to boundedness is as Figure 1.2. Kennedy & McNally (2005a) discuss this typology and show that all four of these possibilities are instantiated among gradable adjectives in English.



**Figure 1.2** Possible scale types with respect to boundedness properties.

---

There is a clear connection between the minimum/maximum/relative typology discussed above and the scale types just given. In order to get the interpretation that we ascribed to them, maximum adjectives like *full* and *flat* must be associated with a scale which has a maximum element. This limits them to upper-closed and fully closed scales. Likewise, minimum adjectives like *bent* and *dangerous* cannot be associated with a scale which lacks a minimum element, which limits them to either lower-closed or fully closed scales.

Rotstein & Winter (2004); Kennedy & McNally (2005a); Kennedy (2007) give a number of empirical tests for minimum/maximum/relative status and boundedness properties. As an example, consider the degree modifier *slightly*. *x is slightly A* is true, roughly, just in case *x* has the property picked out by *A* to a small but non-zero degree. If we want to cash out this intuition more precisely, note that we have to assume that it makes sense to talk about a zero degree of the property denoted by *A*—that is, that *A* can sensibly be associated with a scale with a minimum element. If *A*'s scale does not have a minimum, we expect semantic anomaly. The presence or absence of a minimum element on the relevant scales, then, can be invoked to explain why the sentences in (1.23) are acceptable while those in (1.24) are not:

- (1.23) a. This neighborhood is slightly dangerous.  
           b. This antenna is slightly bent.
- (1.24) a. # This neighborhood is slightly safe.  
           b. # This antenna is slightly straight.



If the sentences in (1.24) can be interpreted at all, they must be taken to describe e.g. *how much* of the neighborhood is safe, rather than *the degree to which* the neighborhood is safe.

The explanation of (1.23)-(1.24) given by the authors cited is that *dangerous* is associated with a lower-closed scale. Intuitively this corresponds to the observation that a neighborhood can get more and more dangerous ad infinitum; however, there is a minimum amount of danger that it can have, namely complete safety. As a result modification by *slightly*, which is restricted to adjectives whose scale has a minimum element, is acceptable. However, *safe* has a scale which is the inverse of the scale of *dangerous*, and so is upper-closed. As a result, modification by *slightly* is not permitted because the scale has no minimum. Kennedy & McNally (2005a); Kennedy (2007) give a number of other arguments which converge on these conclusions, some of which will be reviewed in chapter 4.

Further examples of tests for minimum/maximum/relative status and scale type are *almost*, *completely*, and proportional modifiers. Rotstein & Winter (2004) show that, if an adjective can be modified by *almost*, its scale has a maximum:

- (1.25) a. This neighborhood is almost safe/#dangerous.  
b. This antenna is almost straight/#bent.

Likewise, Kennedy & McNally show that *completely*-modification is acceptable with a degree-modifying meaning only when an adjective has a scale with a maximum element:

- (1.26) a. This neighborhood is completely safe/#dangerous.  
b. This antenna is completely straight/#bent.

Note that *completely* is sometimes possible with scales with no maximum, but in these cases it indicates emphasis, correction, or high speaker confidence rather than maximization.

- (1.27) **Mary:** The president is not tall.  
**Sue:** Uh-uh! He is completely tall.

On the other hand, *proportional* modifiers like *half*, *90%*, and *mostly* require comparing an object's distance from the minimum point to the distance between the maximum and minimum points. Therefore, they can only modify adjectives which are associated with a scale that has maximum and minimum points. (See ch.2, §2.4.4 for a refinement of this generalization.) Neither *safe* nor *dangerous* is associated with a scale that has the requisite structure, but e.g. *full/empty* and *open/closed* are. This can be used to explain the data in (1.28).

- (1.28) a. # This neighborhood is half/90%/mostly dangerous/safe.  
b. This glass is half/90%/mostly full/empty.  
c. This window is half/90%/mostly open/closed.

Again, the examples marked as infelicitous can be given an interpretation on which the adverb quantifies a different property—the proportion of the spatial area of the neighborhood which is safe/dangerous, where the latter is a fully closed scale ranging from proportion 0 to 1. This is not the directly degree-modifying reading that we are interested in, though.

Boundedness and the minimum/maximum/relative distinction will crop up repeatedly in the discussion of modality in later chapters, particularly when we discuss the epistemic adjectives

*possible, probable, likely, and certain* in chapters 4-5. We will consider the relationship between the two in more detail there.

## 1.4 Two Perspectives on Scales

As I have presented it, following in particular Kennedy (1997, 2007), scales are ordered sets of degrees. What are degrees? According to Kennedy and others (e.g., von Stechow 1984; Bierwisch 1989; Heim 2001), degrees are abstract representations of measurement. On this account, degrees of height or happiness exist, and these scales have the structure that they do, independent of whether any objects in the world actually possess those degrees of height or happiness.

This is probably the mainstream perspective in formal semantics, but it is not universal. Cresswell (1976); Klein (1991); Sassoon (2010a); van Rooij (2011); Bale (2011) and others have argued that degrees should be thought of as **equivalence classes**: sets of objects all of which bear the “exactly as *P* as” relation to each other, for the relevant property *P*. These authors take their inspiration in this regard from the Representational Theory of Measurement, an algebraic approach to measurement which has been highly influential in psychology, philosophy, and economics. For measurement theorists, degrees do exist, but they exist as an abstraction from the real-world objects which instantiate them and the qualitative relations that these objects bear to each other. Fundamental to this approach are BINARY ORDERS with varying amounts of structure, and CONCATENATION OPERATIONS which relate simple and compound objects.

Measure functions mapping objects to real numbers are employed in measurement theory as well, but care is taken to ensure that the numerical representations do not carry any information that is not already inherent in the qualitative structures underlying them. So, we can speak freely of objects such as “the degree to which Sam is tall”, but the existence of this degree is dependent on the prior existence of a qualitative structure representing heights, containing among other things a set of individuals who bear the “exactly as tall as” relation to Sam.

The degree-based and measurement-theoretic perspectives are sometimes thought to be in competition, and they may well carry different philosophical commitments; for example, the measurement-theoretic perspective might be more attractive to someone who wishes to avoid ontological commitment to abstracta. As far as formal semantics is concerned, though, there is nothing to choose here: any degree semantics can be translated into an equivalent measurement-theoretic implementation, as we will discuss in some detail in the next chapter. As a result, the choice of whether to include degrees in our ontology is largely a matter of convenience, philosophical proclivity, or the usefulness of the representation in clarifying properties of the formal theory.

Although the versions of the degree-based and measurement-theoretic analyses that are considered in this book are formally equivalent, the algebraic perspective of measurement theory is very useful to adopt here, and will be the subject of some detailed formal discussion in chapter 2. The process of constructing scales using measurement theory—rather than simply treating them as unanalyzed primitives—will suggest new possible parameters of variation in scale type, several of which, I will argue, are in fact instantiated in natural language scales, and vital for the understanding of modality in natural language. Needless to say, since the two approaches are equivalent, nothing that I will propose makes the use of measurement theory obligatory. The situation is comparable

to the relationship between formal logic and its algebraic treatment: even though an algebraic re-formulation of (say) propositional logic is provably equivalent to the more familiar style of presentation, certain aspects of the theory become clearer from an algebraic perspective, and certain methods of proof become available which were previously obscured (Halmos & Givant 1998). Similarly, measurement theory as I use it here does not add anything vital to standard degree semantics, but it allows us as theorists to adopt a different perspective on our familiar degree semantics which suggests new ways of viewing problems and new connections.

For these reasons, chapter 2 is dedicated to a presentation of the aspects of the Representational Theory of Measurement that are most relevant for this book, in particular the method of constructing measure functions from qualitative orderings. I will also use chapter 2 to argue for an expanded range of scale types which can be given a natural formulation in measurement-theoretic terms, which will play an important role in the scalar semantics for modals developed in chapters 3-8.

## 1.5 Summary and Preview

This chapter gave an overview of scalar semantics and summarized a compositional semantics of gradability and comparison which treats scalar expressions as measure functions. Some of the high points of this discussion are that

- Scalar expressions determine a classical (two-valued) extension by a three-step process of mapping their argument to a value on a scale, establishing a threshold value, and comparing the value of their argument to the threshold value.
- In degree-based approaches, scales are partially or totally ordered sets of degrees which vary at least according to the presence or absence of minimum and maximum elements.
- Gradable adjectives come in at least four types:
  - Relative adjectives like *tall* and *rich*;
  - Extreme adjectives like *huge* and *ecstatic*;
  - Minimum adjectives like *bent* and *dangerous*;
  - Maximum adjectives like *full*, *safe* and *straight*.
- Despite the grammatical similarities between complex expressions of modality and gradable expressions more generally, only very recently has there been any attempt to give them a unified semantics.

The similarities between gradable and modal expressions suggest that we need to develop a theory which explains their commonalities. Furthermore, it is clear that this theory should be compositional and, where appropriate, should use the resources of existing theories of gradability. It remains to be seen just how closely related gradable expressions and modals are, however.

In the remainder of this book I will argue that we should give up the venerable assumption that modals are a complicated sort of quantifiers. Instead, I argue, the commonalities between gradable and modal expressions are due to the fact that both have a semantics built around scales.

We should build a detailed semantics for modal expressions in the same way that we do for gradable adjectives: by considering the formal structure of the relevant scales and where the threshold value falls for various simple and complex scalar expressions. I will also pose a number of problems for quantificational semantics for modality and show that they are better accounted for from scalar perspective.

The next chapter lays the groundwork for this argument. I present an algebraic approach to the construction of scales building on Measurement Theory (Krantz, Luce, Suppes & Tversky 1971). The degree-based theories outlined in this chapter can be recast straightforwardly in this framework, and the use of Measurement Theory clarifies a number of issues with respect to the analysis of modality and suggests new avenues of inquiry for both degree semantics and modal semantics. In addition, chapter 2 will propose further distinctions of scale type, motivated by data involving gradable adjectives, and this refined typology will be crucial in the analysis of gradable modals in later chapters.

## CHAPTER 2

### Measurement Theory and the Typology of Scales

This chapter discusses the issue of scale type in more detail, using the formal tools of the Representational Theory of Measurement (RTM, also known simply as “measurement theory”, and not to be confused with a very different enterprise called “measure theory”). RTM allows us to use qualitative (algebraic) and quantitative (measure function) characterizations of scales interchangeably, using what is effectively a type of supervaluation semantics for measure functions.

A number of linguists have suggested accounting for gradability and comparison in various domains using the resources of measurement theory rather than an apparatus taking degrees and scales composed of degrees as primitive (Cresswell 1976; Klein 1980, 1982, 1991; Krifka 1989, 1990, 1998; Nerbonne 1995; Sassoon 2007, 2010a; Bale 2006, 2008, 2011; van Rooij 2009, 2011). This approach is sometimes presented as a competitor of degree-based theories (e.g., by Klein (1980, 1982), but see Klein 1991 for a different perspective). However, the measurement-theoretic approach is not really an *alternative* to anyone’s semantics of gradation; rather, it is a framework of considerable expressive power into which apparently diverse semantic proposals can be translated and compared. For example, both degree-based and delineation-based semantics for gradable adjectives can be expressed using measurement theory. However, degree-based approaches, and particularly measure function-based approaches, are the most straightforward to state using RTM tools, and I will focus on them here for this reason. See, however, Burnett 2016 for a detailed treatment of related issues around scalar structure within a delineation approach.

Using RTM as the basis for a degree semantics carries a number of advantages for our purposes in this book. First, most previous theories of modality that acknowledge gradation (e.g., those of Lewis (1973) and Kratzer (1981)) are couched in qualitative terms. As a result, qualitative semantics for gradation provides a useful perspective for the project of unifying modal semantics and degree semantics, allowing us to study the strengths and weaknesses of previous theories and make comparisons with the theory developed in later chapters.

Second, the process of constructing a degree semantics using RTM force us to be explicit about the mathematical properties attributed to the scales that we associate with gradable expressions in natural languages. The enriched typology of scales that I will motivate linguistically in this chapter is naturally expressed using RTM. Several aspects of the scalar typology that is developed in this chapter will, in turn, form crucial components of the scalar semantics developed later in the book for epistemic and deontic modals. The formal properties of scales that are studies here will be used to explain a number of puzzling phenomena involving the interaction of modals with degree modification, disjunction, conditionals, and other interactions.

Third, RTM has interesting points of contact with algebraic semantics, both in the Boolean semantics tradition (Keenan & Faltz 1985; Winter 2001) and structured-domain semantics for the semantics for plurals and events (as in Link 1983, 1998; Bach 1986; Krifka 1989, a.o.). In particular, I will suggest in this chapter that the concatenation operation in measurement theory can be treated as a restricted version of the operation of algebraic join (sum formation) that is familiar from work in the semantics of plurals. This identification has important ramifications for the semantics of modality, since concatenation plays a role in the construction of many scales, and join realizes

natural language disjunction in algebraic semantics. The behavior of disjunction under modals will, in turn, provide a crucial source of data for the theoretical developments presented later in the book.

## 2.1 Introduction to Measurement Theory

Measurement Theory was developed beginning in the late 19th century as a mathematical foundation for measurement in the physical and psychological sciences (e.g., Helmholtz 1887; Hölder 1901). Its modern incarnation, the Representational Theory of Measurement, stems from the foundational work on scale types by Stevens (1946), written with a strongly psychological focus: “Is it possible to measure human sensation?” Prior to this work it was often assumed that measurement had an intrinsic connection to numbers per se, and that it was senseless to speak of measurement unless, for example, some content could be given to the notion of addition applied to measurements in the domain in question. Since it was not obvious how to map measurement of sensations to an addition operation in many domains, quite a few psychologists had concluded that the concept of measurement did not make sense when applied to human perception.

Stevens’ approach was essentially to work from the other direction: instead of starting with criteria for which operations are necessary for something to count as “measurement” and trying to impose these on the data of psychology, Stevens suggested looking at the kinds of data that were available and seeing what mathematical operations they could support. Formally, this means treating scales as algebraic structures consisting of one or more sets and an  $n$ -ary relation, and (optionally) some further relations and operations. These algebraic structures and the qualitative relations that they encode are taken as basic, and numerical measurement involves asking which kinds of numerical representations faithfully preserve the structure of various types of scales; that is, what is the class of homomorphisms from a scale into the natural numbers, integers, rationals, or real numbers as the case may be. RTM was extended and further formalized by Scott & Suppes (1958); Suppes (1959); Suppes & Zinnes (1963); Narens (1985) and others, most authoritatively in Krantz et al. (1971). A good introduction to RTM is Roberts 1979.

The usefulness of this approach to measurement is illustrated by the contrast between measurements of length and width on the one hand and clock time on the other. Measurements of time can extend as far back as you like, and the choice of zero point is arbitrary; in contrast, all reasonable measurements systems for length have a fixed minimum (zero length) and will share a fixed zero point (the same). Differences of this type crop up routinely in physical and psychological measurement, and are reflected in intuitive judgments of the felicity of certain kinds of statements. It is unremarkable to say of one building that it is twice as wide or tall as another one, but difficult to make sense of the bare claim that one event occurred twice as late as another. To interpret this, we would have to supply a third time with respect to which we are implicitly measuring lateness. In RTM, this difference is traced back to a qualitative difference between the scales which determines which kinds of quantitative statements are interpretable, as I will explain in §2.2.1.

So: Imagine that you had to construct a measurement system from scratch, without the help of a system of numbers. How would you go about doing this?

The first thing to decide, obviously, is what sort of property  $P$  you are measuring. The second step is to consider the relative ordering of pairs of objects of interest with respect to the ordering

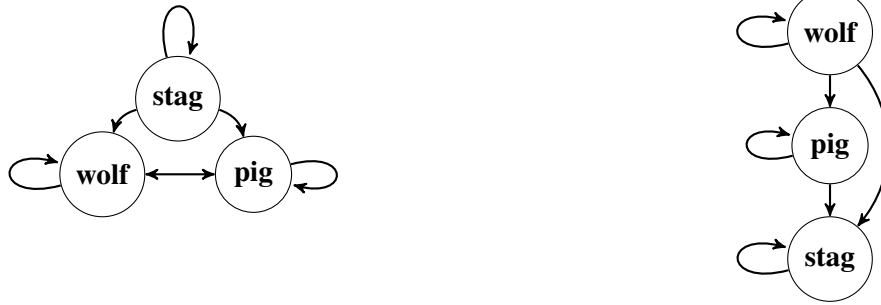
that you are trying to create. Given a domain of objects  $X$  which we are interested in, you can ask of each  $x, y \in X$ : Does  $x$  outrank  $y$  with respect to property  $P$ ? Does  $y$  outrank  $x$ ? Are they equal in  $P$ -ness? Are they incomparable?

This, at least, is the start of a rational reconstruction of the systems of measurement that human languages and human societies utilize. Starting with a **domain**  $X$ , we construct a **binary relation**  $\succsim_P$  using a comparison procedure like the one just outlined.  $\succsim_P$  is just the set of all pairs  $(x, y)$  such that  $x$  is equal to or greater than  $y$  with respect to property  $P$ . So, for example, with the tiny domain  $X = \{\text{stag}, \text{wolf}, \text{pig}\}$  we might have:

$$(2.1) \quad \succsim_{\text{size}} = \{(\text{stag}, \text{stag}), (\text{stag}, \text{wolf}), (\text{stag}, \text{pig}), (\text{pig}, \text{pig}), (\text{pig}, \text{wolf}), (\text{wolf}, \text{wolf}), (\text{wolf}, \text{pig})\}$$

$$(2.2) \quad \succsim_{\text{loudness}} = \{(\text{wolf}, \text{wolf}), (\text{wolf}, \text{pig}), (\text{wolf}, \text{stag}), (\text{stag}, \text{stag}), (\text{pig}, \text{pig}), (\text{pig}, \text{stag})\}$$

This can be represented more clearly as a directed graph with arrows playing the role of  $\succsim$ :

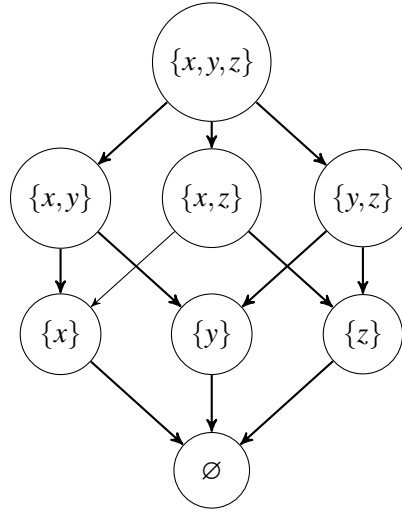


**Figure 2.1** Graphical depictions of the binary relations corresponding to graded properties *size* (left) and *loudness* (right).

We can then inspect these relations to see if they have any other properties of interest. For example, the two binary relations in (2.1) are REFLEXIVE—everything is at least as great as itself with respect to size and loudness—and TRANSITIVE—if  $(x, y) \in \succsim_P$  and  $(y, z) \in \succsim_P$ , then  $(x, z) \in \succsim_P$ . (I will mostly write  $x \succsim_P y$  as an abbreviation for  $(x, y) \in \succsim_P$ .)

The relations in (2.1) and (2.2) are also both CONNECTED (a.k.a. COMPLETE). This means that any two objects in the domain  $X$  are comparable:  $\forall x \forall y [x \succsim_P y \vee y \succsim_P x]$ . Many binary relations are not connected, however, for instance the familiar subset relation depicted in Figure 2.2: neither  $\{x, y\} \subseteq \{y, z\}$  nor  $\{y, z\} \subseteq \{x, y\}$ . (To avoid clutter I suppress the reflexive and transitive arrows in this and later figures.)

Another property of interest is ANTISYMMETRY, which is satisfied by a binary relation  $\succsim_P$  if and only if  $x \succsim_P y \wedge y \succsim_P x$  implies  $x = y$ . In the small domain of (2.1),  $\succsim_{\text{loudness}}$  is antisymmetric (by accident of the small size of  $X$ , though).  $\succsim_{\text{size}}$  is not, though, since  $\text{pig} \succsim_{\text{size}} \text{wolf}$  and  $\text{wolf} \succsim_{\text{size}} \text{pig}$ —that is, these objects are judged to be equivalent in size—but  $\text{pig} \neq \text{wolf}$ . I will often use  $x \approx_P y$  as an abbreviation for  $(x \succsim_P y) \wedge (y \succsim_P x)$ , and I will use  $x >_P y$  to abbreviate  $(x \succsim_P y) \wedge \neg(y \succsim_P x)$ .



**Figure 2.2** Subset relation  $\subseteq$  with domain  $\{x, y, z\}$ .

Another important concept that can be derived from a binary order is the EQUIVALENCE CLASS:

(2.3) The equivalence class of  $x$  relative to a relation  $\succsim_P$ , written  $[x]_P$ , is the set  $\{y \mid y \approx_P x\}$ .

Some useful terms for frequently occurring types of relations are:

- (2.4) a. **Preorder** (a.k.a. **quasi-order**): transitive and reflexive.  
 b. **Partial order**: transitive, reflexive, and antisymmetric.  
 c. **Weak order**: transitive, reflexive, and connected.  
 d. **Total order**: transitive, reflexive, connected, and antisymmetric.

Note that a total order  $\succsim_P$  is a special case of a weak order where, since  $\succsim_P$  is antisymmetric,  $[x]_P$  is the unit set  $\{x\}$  for any  $x$ . Similarly, a partial order is a special case of a preorder where  $[x]_P = \{x\}$  for every  $x$ .

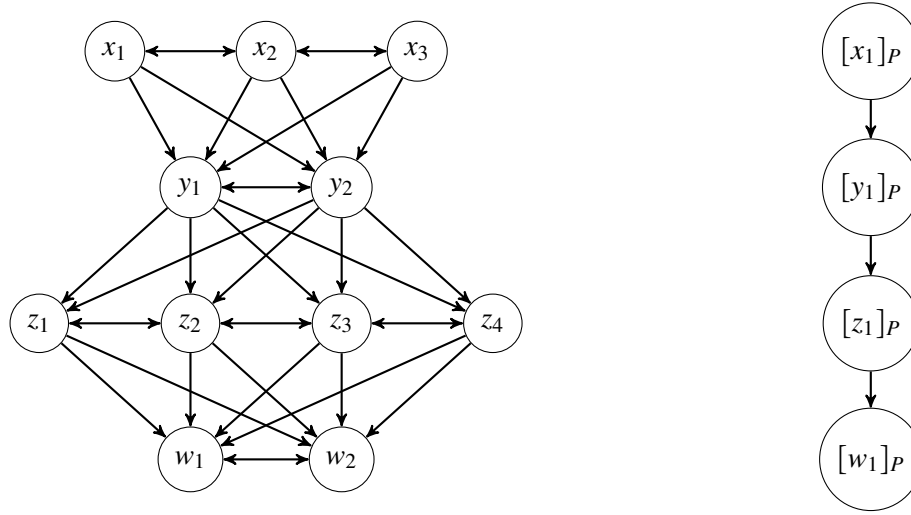
In fact, every weak order is systematically related to a total order in the following way. Let  $(X/\approx_P)$  be the set of equivalence classes under the relation  $\succsim_P$  with domain  $X$ , i.e.  $\{Y \mid \exists x \in X : Y = [x]_P\}$ . Then the relation  $\succsim_P^*$  is the REDUCTION of  $\succsim_P$  if and only if

- (2.5) a.  $\succsim_P^*$  is a binary relation on  $(X/\approx_P)$ ; and  
 b.  $x \succsim_P y$  if and only if  $[x]_P \succsim_P^* [y]_P$ .

That is, if  $\succsim_P$  is a binary relation on  $X$ , then its reduction  $\succsim_P^*$  is the corresponding binary relation on the set of equivalence classes of  $X$  with respect to  $\approx_P$ . If  $\succsim_P$  is a weak order, then  $\succsim_P^*$  will be a total order, as depicted in Figure 2.3.

Similarly, the relation  $\succsim_Q$  depicted on the left side of Figure 2.4 is a preorder, since neither  $y_1, y_2 \succsim_Q z_1, z_2$  nor  $z_1, z_2 \succsim_Q y_1, y_2$ . Its reduction  $\succsim_Q^*$ , pictured on the right side of Figure 2.4, is a partial order.



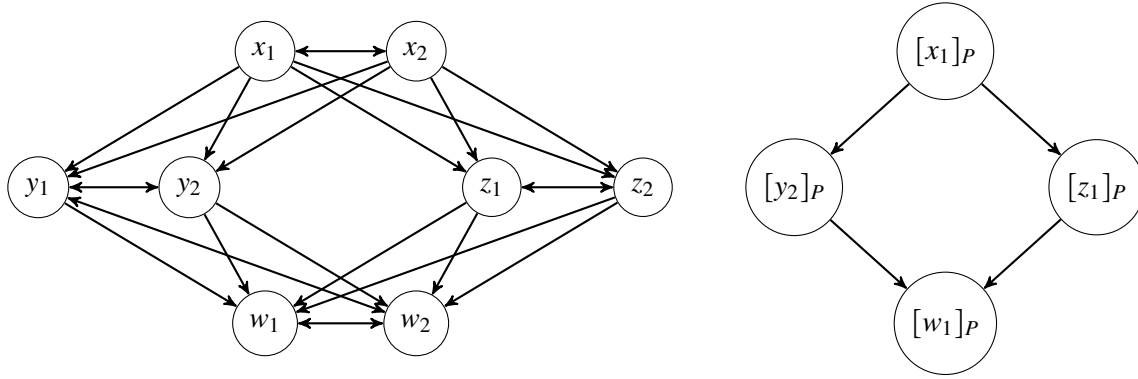


**Figure 2.3** A weak order (left) and total order (right) representing its reduction to equivalence classes.

The binary relations  $\succsim_{size}$  and  $\succsim_{loudness}$  can also be thought of as part of algebraic STRUCTURES based on a domain  $X$  and a binary relation. For present purposes, we can think of a structure as a tuple  $\langle X, \succsim_P, \dots \rangle$ , where  $X$  is any set,  $\succsim_P$  is a  $n$ -ary relation on  $X$  (so, in the binary case, for all  $(x, y) \in \succsim_P, x \in X$  and  $y \in X$ ), and further members of the tuple represent other sets, relations on  $X$ , or (possibly partial) operations on  $X$ . For example, the property **size** can be thought of extensionally as a structure  $\langle X, \succsim_{size}, \dots \rangle$ , where  $X$  is some set of objects of which it makes sense to talk about their size, and the ellipsis represents additional constraints on this property that we may choose to add later.

With a larger domains of objects, we can start to ask more interesting questions about the structure  $\langle X, \succsim_P, \dots \rangle$ . One question is whether we can fill in the ellipsis with a CONCATENATION operation on  $X$  with useful properties. Concatenation is a function  $\circ \subseteq (X \times X) \rightarrow X$ , which I will usually write in infix form:  $x \circ y = z$ . Intuitively, concatenation can be thought of taking two objects and forming a complex object from them (although there is no need for them to be physical objects, and concatenation need not involve any procedure as in this metaphor). For example, if we are measuring the lengths of a number of rods  $x, y$ , and  $z$  we might ask not only whether  $x \succsim_{length} y$ ,  $x \succsim_{length} z$ , and  $y \succsim_{length} z$ , but also how  $x$  compares in length to  $y$  and  $z$  placed end-to-end lengthwise, i.e. the concatenation  $y \circ z$ . Similarly if we are measuring the weights of the same objects using a balance, we can find out whether  $x \succsim_{weight} y$  by considering whether the pan drops to the right when  $x$  is placed on the left pan and  $y$  is placed on the right; if it does not, then  $x \succsim_{weight} y$ . Equally, though, we can compare the weight of  $x$  to the concatenation of  $y$  and  $z$  by placing  $x$  on the left-hand side of the balance and placing  $y$  and  $z$  on the right-hand side at the same time. If the pan drops right, then  $(y \circ z) > x$ .

Concatenation is important in a theory of the expression of measurement in natural language because semantic domains have a part-whole structure. In this case, we would like to know not



**Figure 2.4** A preorder (left) and a partial order (right) representing its reduction to equivalence classes.

only how atomic objects  $x$ ,  $y$ , and  $z$  compare to each other relative to graded property  $P$ , but also how these compare to various complex objects which contain these as parts. If we know that  $x \geq_P y$  and  $x \geq_P z$ , this information does not yet determine whether  $x \geq_P (y \circ z)$ . As I will show later in this chapter and also in the discussion of modality in subsequent chapters, properties vary in whether this inference is valid: INTERMEDIATE graded properties such as temperature are associated with scales for which the inference

$$x \geq_P y \wedge x \geq_P z \rightarrow x \geq_P (y \circ z)$$

is valid, while ADDITIVE properties such as weight and length are associated with scales for which the same inference always fails except in trivial cases.

## 2.2 Measure Functions, Interpretability, and the Typology of Scales in RTM

In the previous subsection we managed to introduce some basic concepts of measurement without mentioning numbers or degrees. Even though the idea of measurement without numbers or degrees may seem odd, the goal of RTM is to justify the use of these constructs in scientific practice, and for this purpose it would obviously be unwise to use them in the definitions. Rather, measurements using numbers or degrees are justified by showing the existence of a homomorphism  $\mu$  from a qualitative structure  $\langle X, \geq_P, \dots \rangle$  into a structure making use of numbers such as  $\langle \mathbb{R}, \geq, \dots \rangle$ —or, if you like, into  $\langle D, \geq, \dots \rangle$  for some specified set of degrees.

(2.6) A function  $\mu$  is a **homomorphism** from a qualitative structure  $\mathcal{S}_P = \langle X, \geq_P \rangle$  into a numerical structure  $\langle \mathbb{R}, \geq \rangle$  if and only if, for all  $x, y \in X$ ,

- $\mu(x) \in \mathbb{R}$ ,  $\mu(y) \in \mathbb{R}$ , and
- If  $x \geq_P y$ , then  $\mu(x) \geq \mu(y)$ .

If  $\mathcal{S}_P$  also contains further relations or operations, then similar conditions apply. For instance, if  $\mathcal{S}'_P = \langle X, \geq, \circ \rangle$ , where  $\circ$  is a binary operation on  $X$ , then  $\mu$  is a homomorphism from  $\mathcal{S}'_P$  into  $\langle \mathbb{R}, \geq, + \rangle$  if and only if in addition, for all  $x, y, z \in X$ ,

- If  $x \circ y = z$ , then  $\mu(x) + \mu(y) = \mu(z)$ .

Intuitively, the requirement that  $\mu$  be a homomorphism limits us to candidate  $\mu$  which preserve all of the information contained in the source structure, while possibly adding more due to the structure inherent in the real numbers.

The danger of using real numbers to represent the simple qualitative ordering  $\langle X, \succsim \rangle$  is that the reals have additional structure beyond what is contained in  $\langle X, \succsim \rangle$ . For example, if  $\mu(x) = 8$  and  $\mu(y) = 3$ , the numerical representation would allow us to ask whether  $x$ 's measure is  $1/2$  of  $y$ 's (no) and whether it is the base-2 logarithm of  $y$ 's (yes). But these questions do not make sense with respect to the source structure  $\langle X, \succsim \rangle$ . So, we need some way to ensure that we do not, in the process of developing a numerical representation of this qualitative ordering, trick ourselves into thinking that it makes sense to talk about logarithms.

We can eliminate the extra structure by considering not just the information contained in one particular homomorphism  $\mu$ , but the information that is common to all homomorphisms. That is, we will recapture the fact that some of our scales have a less rich structure than  $\mathbb{R}$  by universally quantifying over homomorphisms from  $\langle X, \succsim_P, \dots \rangle$  into  $\langle \mathbb{R}, \geq, \dots \rangle$ . Statements that are true in all such homomorphisms are made true by properties of the qualitative source structure; similarly for statements that are false in all such homomorphisms. When all homomorphisms agree on the truth-value of a statement  $S$ , we will say that  $S$  is INTERPRETABLE. (This property will be discussed and illustrated in more detail below).

The empirical payoff is that we are able to represent contrasts between scales like temperature and clock time (where interval comparisons makes sense, but ratio comparisons do not) and scales like height (where both types of statements make sense).

(2.7) Sam grew from 2 feet to 3 feet, and Harry grew from 4 feet to 6 feet.

- ✓ So, Harry grew twice as much as Sam did.
- ✓ So, Harry is now twice as tall as Sam is.

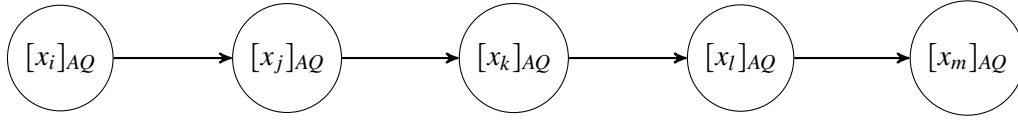
(2.8) I ran from 2PM to 3PM, and you ran from 4PM to 6PM.

- ✓ So, you ran for twice as long as I did.
- # So, you started running twice as late as I did.

As we will see, this difference in interpretability can be attributed to a qualitative difference in the underlying structure of the scales: amount of growth is measured on a RATIO SCALE, while length of time is measured on an INTERVAL SCALE.

### 2.2.1 Ordinal Scales, Admissibility, and Interpretability

To see how interpretability works, consider a structure  $\langle X, \succsim_{AQ} \rangle$  where  $X$  is the set of cities in the United States with population over 1,000,000 and  $x \succsim_{AQ} y$  is interpreted as “ $x$  has air quality as least as good as  $y$ ”. (The example is from Roberts 1979; this is apparently a measurement system which was actually employed in some locales in the 1970's.) Assume that  $\succsim_{AQ}$  is a weak order whose reduction  $\succsim_{AQ}^*$  is a total order on the equivalence classes under the  $\succsim_{AQ}$  relation, as depicted in Figure 2.5.



**Figure 2.5** Equivalence classes of cities under the relation “has air at least as good as”.

A homomorphism from  $\langle X, \succsim_{AQ} \rangle$  into  $\langle \mathbb{R}, \geq \rangle$  is any function  $\mu$  with domain  $X$  and range  $\mathbb{R}$  where  $x \succsim_{AQ} y$  only if  $\mu(x) \geq \mu(y)$ . So, for example, the following are all homomorphisms, as long as all cities in an equivalence class are mapped to the same number.

$$\mu_1 = \begin{pmatrix} x_i \longrightarrow 5 \\ x_j \longrightarrow 4 \\ x_k \longrightarrow 3 \\ x_l \longrightarrow 2 \\ x_m \longrightarrow 1 \\ \dots \end{pmatrix} \quad \mu_2 = \begin{pmatrix} x_i \longrightarrow 10 \\ x_j \longrightarrow 6 \\ x_k \longrightarrow 2 \\ x_l \longrightarrow 1 \\ x_m \longrightarrow 0 \\ \dots \end{pmatrix} \quad \mu_3 = \begin{pmatrix} x_i \longrightarrow 2,048,348 \\ x_j \longrightarrow 194 \\ x_k \longrightarrow 193 \\ x_l \longrightarrow 22 \\ x_m \longrightarrow -438 \\ \dots \end{pmatrix}$$

If  $\mu$  is a homomorphism from a structure  $\langle X, \succsim, \dots \rangle$  into  $\langle \mathbb{R}, \geq, \dots \rangle$ , we will say that  $\mu$  is an **admissible measure function** on  $\langle X, \succsim, \dots \rangle$ . In general, a measure function  $\mu : X \rightarrow \mathbb{R}$  is admissible relative to a qualitative scale  $\mathcal{S}$  just in case  $\mu$  faithfully preserves all of the qualitative structure in  $\mathcal{S}$ , according to some intended interpretation of the relations and operations.  $\mu$  must be a homomorphism from  $\mathcal{S} = \langle X, \succsim, \dots \rangle$  into a target structure  $\langle \mathbb{R}, \geq, \dots \rangle$ . For example, if  $\mathcal{S} = \langle X, \succsim, \circ \rangle$  and the target is  $\langle \mathbb{R}, \geq, + \rangle$ , then  $\mu$  is admissible just in case, for all  $x$  and  $y$ ,  $x \succsim y \rightarrow \mu(x) \geq \mu(y)$  and  $\mu(x \circ y) = \mu(x) + \mu(y)$ . We will frequently leave the specification of the source and target structures implicit when it is clear from context.

The concept of admissibility is crucial in measurement-theoretic semantics, and will appear frequently in the rest of the book.

$\langle X, \succsim_{AQ} \rangle$  is an example of an ORDINAL SCALE, one of the weakest scale types standardly employed in RTM:

(2.9) If a structure  $\langle X, \succsim_P \rangle$  is an **ordinal scale** then, for all admissible  $\mu$ ,  $x \succsim_P y \Leftrightarrow \mu(x) \geq \mu(y)$ .

Every weak order has at least as much structure as an ordinal scale. In the case of the weak order  $\succsim_{AQ}$ , for example, (2.9) is clearly satisfied: for example, since  $x_j \succ_{AQ} x_k$  we have  $\mu_1(x_j) = 4 > \mu_1(x_l) = 2$ ,  $\mu_3(x_j) = 194 > \mu_3(x_l) = 22$ , etc.<sup>1</sup>

<sup>1</sup> Note that stronger scale types, such as ratio scales and interval scales to be defined below, also have this property. I use an “if ... then” statement here in order to avoid overlap between scale types, but this is just a matter of definition: we could equally well define the scale types so that all interval and ratio scales are also ordinal scales, for example.

Also, I use the term “scale” to designate the qualitative source structures alone. This is slightly different from typical usage in RTM, where this term is used to designate a qualitative structure and a target structure together with some specific homomorphism. In the latter case, we would talk about classes of scales that are related by some transformation rather than classes of measure functions that are so related. Hopefully this terminological difference will not introduce any confusion.

Another way to characterize an ordinal scale is in terms of transformations among the admissible measure functions: any monotone increasing transformation of an admissible  $\mu$  is also an admissible  $\mu$ .

- (2.10) If a structure  $\langle X, \succsim_P \rangle$  is an **ordinal scale** then, for all admissible measure functions  $\mu$  and all order-preserving (monotone increasing) functions  $f: \mathbb{R} \rightarrow \mathbb{R}$ :  
 $\mu'(x) = f(\mu(x))$  is also an admissible measure function.

(2.10) captures the fact, already apparent from the air quality example, that the relative distance between the measures of objects is not important in an ordinal scale, but only the relative ordering of the measures assigned to objects. (See Krantz et al. 1971: §2.1 for a proof of the equivalence of the conditions in (2.9) and (2.10).)

One consequence of the relatively weak structure of an ordinal scale is that many quantitative statements that can be framed using the numbers assigned to objects do not have a stable truth-value across admissible  $\mu$ . For example, consider the statements:

- (2.11) a.  $x_j$  has better air than  $x_k$  does.  
 b.  $x_m$  has better air than  $x_i$  does.  
 c.  $x_j$  has air twice as good as  $x_l$  does.

Are these statements true or false? Well, if we looked only at  $\mu_1$ , the first would appear to be true, the second false, and the third true. If we consider  $\mu_2$  and  $\mu_3$ , however, (2.11a) and (2.11b) remain true and false respectively, but (2.11c) comes out false. Since all three measure functions are of equal status, a natural move is to declare (2.11a) to be true and (2.11b) false, but to conclude that (2.11c) does not have a truth-value relative to this structure.

In RTM this type of situation is usually described by saying that (2.11a) and (2.11b) are “meaningful” while (2.11c) is “meaningless”. Since this term is already overloaded both in ordinary usage and in formal semantics, I will formulate this concept instead as a constraint on semantic “interpretability”:

- (2.12) A statement  $S$  whose meaning makes reference to a measure  $\mu$  is **interpretable** only if, for all admissible transformations  $\mu'$  of  $\mu$ , the truth-value of  $S$  remains the same if  $\mu'$  is used instead of  $\mu$ .

Since all homomorphisms from  $\langle X, \succsim_P \rangle$  into  $\langle \mathbb{R}, \geq \rangle$  are admissible, the effect of (2.12) is that statements are undefined if they make reference to quantitative features of the numerical representation that are not contained in the qualitative source structure  $\langle X, \succsim_P \rangle$ . That is, if a measure function shows some patterned behavior in  $\mathbb{R}$ , this is ignored unless the same pattern is also observed in the underlying qualitative structure. This makes it possible to use real numbers, which have a very complex structure, to represent simpler structures accurately. In the case of an ordinal scale  $\langle X, \succsim_P \rangle$ , only simple comparisons between values and statements that can be built using Boolean combinations of such statements will be interpretable, for example the statements in (2.13).

(2.13)

$$a \text{ is } \left\{ \begin{array}{l} \text{at least as } P \text{ as} \\ \text{more } P \text{ than} \\ \text{at most as } P \text{ as} \\ \text{less } P \text{ than} \\ \text{exactly as } P \text{ as} \end{array} \right\} b.$$

## 2.2.2 Ratio Scales

Treating (2.11c) as undefined seems fine, but there are other similar statements which should clearly get truth-values, for example:

- (2.14) a. Sam is twice as tall as his little brother.  
 b. I ran 4.3 times as far as you did.  
 c. It is three times as likely to rain as it is to snow.

What sorts of structures are needed to make these statements come out as interpretable? It turns out that a sufficient condition for statements like *x is n times as P as y* to have stable truth-conditions across all admissible  $\mu$  is that the property *P* be associated with a RATIO SCALE. Ratio scales can be characterized easily in terms of admissible measure functions:

- (2.15) If a structure  $\langle X, \succcurlyeq_P, \circ, \dots \rangle$  is a **ratio scale** then, for all admissible  $\mu$  and all  $x \in X$ :  
 $\mu'(x) = n \times \mu(x)$  is also admissible, where  $n \in \mathbb{R}^+$  (the positive real numbers).

Ratio scales are those where all and only admissible transformations are those which involve multiplying each value by the same positive real number. Familiar examples of ratio scales include measurements of extent (length, width, height, etc.), mass, and weight. For example, measurements of extent in feet and meters can be converted by using the transformation

$$\text{length in feet} = 0.3048 \times \text{length in meters.}$$

and its converse

$$\text{length in meters} = \frac{1}{0.3048} \times \text{length in feet} \approx 3.2808 \times \text{length in feet.}$$

Similarly, in ordinary usage measurements in pounds and kilograms can be converted without loss of information by the transformations

$$\text{weight in kilograms} = 2.2 \times \text{weight in pounds.}$$

and

$$\text{weight in pounds} = \frac{1}{2.2} \times \text{weight in kilograms} \approx 0.455 \times \text{weight in kilograms.}^2$$

<sup>2</sup> Of course, in scientific usage kilograms are a measure of mass rather than weight, and so the transformation would have to take into account gravity; but this is probably not linguistically relevant.

If  $\mu$  and  $\mu'$  are both admissible measure functions for some ratio scale  $\langle X, \succsim_P, \circ, \dots \rangle$ , then the ratio  $\frac{\mu(x)}{\mu(y)}$  is guaranteed to be equal to the ratio  $\frac{\mu'(x)}{\mu'(y)}$ . This is because  $\mu'(x) = n \times \mu(x)$  for some  $n > 0$ , and so

$$\frac{\mu'(x)}{\mu'(y)} = \frac{n \times \mu(x)}{n \times \mu(y)} = \frac{\mu(x)}{\mu(y)}$$

As a result, statements involving ratios like (2.14) are predicted to be interpretable for ratio scales, since they will be true or false in every admissible  $\mu$ .

Another important property of ratio scales is that they are additive with respect to concatenation: that is, as long as two objects do not overlap, then the measure assigned to their concatenation is the sum of their individual measures.

(2.16) A scale  $\langle X, \succsim_P, \circ, \dots \rangle$  is **additive** with respect to concatenations iff, for all non-overlapping  $x, y$  and all admissible  $\mu$ ,  $\mu(x \circ y) = \mu(x) + \mu(y)$ .

Two further properties of ratio scales are worth noting here. First, it follows from additivity that  $\mu(x \circ y) > \mu(x)$  and  $\mu(x \circ y) > \mu(y)$ , and so that  $(x \circ y) \succ_P x$  and  $(x \circ y) \succ_P y$ . This contrasts with some familiar types of interval scales, as we will see in a moment. Second, ratio scales have a fixed minimum corresponding to  $\mu(x) = 0$  under all admissible  $\mu$ . There may not, however, be any element in the domain which has measure zero.

The characterization of ratio scales as being “stronger” than ordinal scales, in the sense that they “contain more information”, can now be made precise as follows: scale  $\mathcal{S}$  is stronger than scale  $\mathcal{S}'$  just in  $\mathcal{S}$  permits a smaller class of transformations. For example, ordinal scales have the property that any monotone increasing transformation of an admissible measure function is also admissible. Ratio scales permit a subset of the monotone increasing functions: those for which, for all  $x$ ,  $\mu'(x) = n \times \mu(x)$ , with  $n > 0$ . However, there are many other monotone increasing transformations that are not permissible for ratio scales, and so that ratio scales are a stronger scale type.

In order to ensure that a qualitative scale  $\langle X, \succsim_P, \circ, \dots \rangle$  is indeed a ratio scale, we must check that it has a number of formal properties. Most of the work in a measurement-theoretic characterization of ratio scales is in stating these properties and proving that they do indeed characterize the class of scales for which all admissible measure functions are additive and related by the transformation in (2.15). The practical usefulness of this exercise is that it allows us to check whether the individual axioms describe assumptions that are reasonable to make about the domain in question, and so to decide whether a certain measurement procedure can support talk about addition, subtraction, ratios, etc. However, there are many different ways to characterize a ratio scale, and the choice of axioms depends to a certain extent on our assumptions about the domain—its cardinality, boundedness, and whether it has a part-whole structure—and also on our assumptions about the nature of the concatenation relation. The usual assumptions from RTM will not suffice for us, since they require an unstructured and infinite domain and the possibility of concatenating any two arbitrary elements. In contrast, it is much more natural for us to assume here that many domains (especially,  $D_e$  and  $D_{\langle s, t \rangle}$ ) have a part-whole structure, and that the concatenation operation is significantly restricted. Therefore, I will postpone discussion of the axiomatic characterization of ratio (and interval) scales until after our assumptions about semantic domains have been laid out in section 2.4.

### 2.2.3 Interval Scales

There are certain properties of interest whose scales seem to be stronger than an ordinal scale, but weaker than a ratio scale. Temperature and clock time are two familiar examples: if it is 10 degrees Celsius in Boston and 30 degrees Celsius in Atlanta, it is not natural to describe this situation using (2.17).

(2.17) # It is three times as hot in Atlanta as it is in Boston.

The formal explanation offered by RTM of the oddity of (2.17) is that translating the temperature measurements into another equally valid measurement system for temperature—Fahrenheit—could make (2.17) false, violating the condition on interpretability in (2.12). Specifically, Fahrenheit and Celsius measurements are related by the transformation

$$\text{temperature in Fahrenheit} = \frac{9}{5} \times \text{temperature in Celsius} + 32.$$

Under this transformation, the temperatures in Atlanta and Boston do not have the ratio  $\frac{30}{10} = 3$ , but  $\frac{86}{50} = 1.72$ . Since the truth-value of (2.17) is not stable under this transformation, we conclude that this statement and other statements involving non-trivial ratios are uninterpretable.

However, we do not want temperature to have as little structure as an ordinal scale. Not just any monotone increasing transformation of Celsius would give us a usable measurement system for temperature; we need one that preserves information about differences, as in:

(2.18) It is 30 Celsius in Atlanta, 10 Celsius in Boston, 35 Celsius in Rio de Janeiro, and 25 Celsius in Rome. So, Atlanta is hotter than Boston by twice as much as Rio is hotter than Rome.

Even though the absolute ratio statement (2.17) does not keep its truth-value in the transformation from one admissible measure function (Celsius) to another (Fahrenheit), the **difference ratio** in (2.18) does. This is because the Celsius ratio  $(30 - 10)/(35 - 25) = 2$  is equal to the Fahrenheit ratio  $(86 - 50)/(95 - 77) = 2$ .

We cannot associate temperature with a mere ordinal scale if we want to explain the fact that the relative size of intervals on the temperature scale are stable quantities across admissible  $\mu$ : allowing all monotone increasing transformations would destroy this information. Similar considerations hold, for example, for clock time, where the numbers assigned to points in time are not interpretable, nor are their ratios, but the relative sizes of intervals are interpretable quantities:

(2.19) I ran from 3PM to 4PM, and you ran from 6PM to 8PM.

- a. # So, you started running twice as late as I did.
- b. So, you ran for twice as long as I did.

In order to capture these features using RTM, temperature and time are standardly associated with INTERVAL SCALES, which I will also characterize using the class of admissible transformations.

(2.20) If  $\mathcal{S}_P$  is an **interval scale** then, whenever  $\mu$  is admissible for  $\mathcal{S}_P$ , for all  $\alpha \in \mathbb{R}^+$  and  $\beta \in \mathbb{R}$ :  $\mu'(x) = \alpha \times \mu(x) + \beta$  is also admissible for  $\mathcal{S}_P$ .



The conversion from Celsius to Fahrenheit given above is an example of such a transformation, setting  $\alpha = \frac{9}{5}$  and  $\beta = 32$ . Another example of an interval scale which we will make considerable use of in later chapters is **expected value** (in the utility-theoretic sense). With this class of transformations, ratios of differences will always be interpretable with interval scales, but absolute ratios will be interpretable only in the trivial case where  $\mu(x) = \mu(y)$ . See section 2.4 below for discussion of the qualitative characterization of interval scales that is appropriate for our purposes.

These three scale types—ordinal, ratio, and interval—are the three standard scale types in RTM that are most relevant for us here. In the next section we will sketch a way to integrate qualitative scales into a degree-based compositional semantics. In the rest of the chapter will use measurement-theoretic considerations to try to refine our picture of the kinds of scales that are needed to represent in a theory of the lexical semantics of English adjectives.

### 2.3 Compositional degree semantics without degrees

With basic concepts of RTM in place, it is relatively straightforward to define a compositional semantics for degree expressions which is build on the kinds of qualitative structures just described. I will go through some details of this derivation here. However, it is worth emphasizing up front that continuing to use degrees, as I will here, is merely a convenience. Like quantitative systems of measurement in general, it is always possible to construct an equivalent, purely qualitative representation. However, quantitative measurements are frequently much simpler to represent and reason about, especially as the underlying scales grow in complexity (Krantz et al. 1971; Suppes 2002). In addition, continuing to use degree semantics here allows us to keep the compositional assumptions simple and fairly standard, making it easier to focus on the issue of primary interest: the lexical semantics of scalar expressions.

In the degree semantics discussed in chapter 1, we interpreted adjectives  $A$  by reference to a measure function  $\mu_A^w$  which was designated as the interpretation of the scalar property  $A$  in world  $w$ . In the present context, we will interpret such sentences instead by looking to a qualitative scale  $\mathcal{S}_A^w$ , which is associated by the model with adjective  $A$  in world  $w$ . Qualitative scales determine a class of admissible measure functions which are referred to in the compositional semantic derivation. The latter proceeds almost exactly as in the Kennedy-style semantics introduced in chapter 1, using the degrees that these measure functions assign to objects.

There are two key differences, though. First, doing degree semantics without degrees means dropping the assumption that the interpretation procedure supplies a privileged measure function for each scalar expression—for example, that interpreting height adjectives like *tall*, *long*, *heavy* always requires us to choose a privileged unit of measurement (feet, miles, meters, pounds, newtons, etc.). Instead, models provide qualitative scales, and the qualitative scales are used to constrain the degree semantics via the concept of admissibility.

Second, the fact that the interpretation procedure does not hand us a single measure function per scalar expression means that we need a procedure for adjudicating disagreement among admissible measure functions. The proposed solution is to utilize a supervaluation technique which implements the “interpretability” concept described informally above. That is, a sentence will receive a truth-value *simpliciter* if and only if its truth-value is the same relative to every admissible measure

function.

Like the degree semantics in chapter 1, the qualitative semantics for gradation is able to provide interpretations for comparatives, equatives, and more complex expressions of degree such as ratio and proportional modifiers. It also accounts for the intuition that scale type restricts which degree expressions can be considered, as in the case of minimum-deviation (*slightly*) or maximizing (*totally*) degree expressions on scales that lack the necessary endpoints. However, the present account also explains why certain complex degree expressions that are acceptable with some adjectives—say, *Jacksonville is twice as large as Chicago*—are unacceptable with other adjectives—?? *Jacksonville is twice as hot as Chicago*. The latter is neither true nor false, because *hot* evokes an interval scale  $\mathcal{S}_{hot}$  for which there will always be admissible measures that disagree on the truth-value of the sentence.

Here are details of one possible compositional implementation of this strategy. Consider the schema given in chapter 1 for interpreting scalar expressions, and the type-polymorphic definition of *more/-er*:

$$(2.21) \quad \llbracket A \rrbracket^{\mathcal{M},w} = \lambda k_{\alpha} [\mu_A^w(k)],$$

$$(2.22) \quad \llbracket \text{more/-er} \rrbracket^{\mathcal{M},w} = \lambda A_{\langle e,d \rangle} \lambda d_d \lambda x_e [\mu_A^w(x) \geq d]$$

We will keep these definitions unchanged, except that—since we are now assuming that the models do not contain any measures—we must parameterize the interpretation function somehow in order to provide an interpretation for the various  $\mu$ 's. I will do this by adding as a semantic parameter a function  $\mathbf{m} : (W \times \mathcal{A}) \rightarrow \mathbf{M}$ . Here,  $W$  is the set of possible worlds,  $\mathcal{A}$  is the set of all scalar expressions in the language, and  $\mathbf{M}$  is the set of all possible measure functions. For any world  $w \in W$  and scalar expression  $A \in \mathcal{A}$ ,  $\mathbf{m}(w, A)$  yields the measure function that we have until now been writing as  $\mu_A^w$ —or, when ignoring intensionality, simply as  $\mu_A$ .

Now, the official denotation of an arbitrary scalar expression  $A$  is given in (2.23).

$$(2.23) \quad \llbracket A \rrbracket_{\mathbf{m}}^{\mathcal{M},w} = \lambda k_{\alpha} [\mathbf{m}(w, A)(k)],$$

An equivalent form that makes use of the convention  $\mu_A^w := \mathbf{m}(w, A)$  is:

$$(2.24) \quad \llbracket A \rrbracket_{\mathbf{m}}^{\mathcal{M},w} = \lambda k_{\alpha} [\mu_A^w(k)],$$

which is just the original definition from chapter 1, explicitly parametrized by  $\mathbf{m}$ .

Measure functions serve in this semantics as a convenient fiction that allows the compositional interpretation to proceed smoothly. By carefully defining truth using supervaluations – on which see, for example, Suppes 1959; van Fraassen 1966, 1968; Kamp 1975; Fine 1975; Kamp & Partee 1995—we can interpret simple and complex expressions that make explicit or implicit reference to degrees while avoiding reference to degrees in the metalanguage.

Specifically, we assume that contexts of utterance do not provide a value for the semantic parameter  $\mathbf{m}$ . We then define simple truth for an expression of type  $t$  in terms of all of the possible truth-values that any parametrized interpretation function could yield, as long as the associated  $\mathbf{m}$  assigns admissible measure functions  $\mu_A$  to all scalar expressions  $A \in \mathcal{A}$ .

$$(2.25) \quad \textbf{Generalized admissibility: A function } \mathbf{m} : (W \times \mathcal{A}) \rightarrow \mathbf{M} \textbf{ is } \textbf{gen-admissible} \textbf{ if and only if, for all } w \in W \textbf{ and } A \in \mathcal{A}, \mathbf{m}(w, A) \textbf{ is admissible for } \mathcal{S}_A^w, \textbf{ the qualitative scale associated with } A \textbf{ at } w.$$

$$(2.26) \quad \text{For any } E \text{ of type } t: \llbracket E \rrbracket^{\mathcal{M},w} = \begin{cases} 1 \text{ if, for all gen-admissible } \mathbf{m}, \llbracket E \rrbracket_{\mathbf{m}}^{\mathcal{M},w} = 1; \\ 0 \text{ if, for no gen-admissible } \mathbf{m}, \llbracket E \rrbracket_{\mathbf{m}}^{\mathcal{M},w} = 1; \\ \text{undefined otherwise.} \end{cases}$$

This is just another way to state the basic RTM invariance concept: no additional information should be imported into the interpretation beyond what actually resides in the qualitative source structures. This is ensured if we decline to assign a truth-value to any statement for which the choice among admissible measure functions could make a difference to the truth-value. These statements are **uninterpretable**.

Here are a few examples.

$$\begin{aligned} (2.27) \quad & \llbracket \text{Mary is taller than Harry is} \rrbracket^{\mathcal{M},w} \\ &= 1/0 \text{ if for all/no gen-admissible } \mathbf{m}, \llbracket \text{Mary is taller than Harry is} \rrbracket_{\mathbf{m}}^{\mathcal{M},w} = 1 \\ &= 1/0 \text{ if for all/no gen-admissible } \mathbf{m}, \mu_{tall}^w(\mathbf{Mary}) > \mu_{tall}^w(\mathbf{Harry}) \\ &= 1/0 \text{ if for all/no } \mathcal{S}_{tall}^w\text{-admissible } \mu_{tall}^w, \mu_{tall}^w(\mathbf{Mary}) > \mu_{tall}^w(\mathbf{Harry}) \end{aligned}$$

Since admissible measure  $\mu_{tall}^w$  are, by definition, required to agree with  $\succsim_{tall}^w$  on the ordering of individuals, this definition renders (2.27) equivalent to the purely qualitative ordering statement (2.28), as desired.

$$(2.28) \quad \llbracket \text{Mary is taller than Harry is} \rrbracket^{\mathcal{M},w} = 1 \text{ iff } \mathbf{Mary} \succ_{tall}^w \mathbf{Harry}.$$

Second example, with a ratio modifier:

$$\begin{aligned} (2.29) \quad & \llbracket \text{Mary is exactly twice as tall as Harry is} \rrbracket^{\mathcal{M},w} \\ &= 1/0 \text{ if for all/no gen-admissible } \mathbf{m}, \llbracket \text{Mary is exactly twice as tall as Harry is} \rrbracket_{\mathbf{m}}^{\mathcal{M},w} = 1 \\ &= 1/0 \text{ if for all/no } \mathcal{S}_{tall}^w\text{-admissible } \mu_{tall}^w, \mu_{tall}^w(\mathbf{Mary}) = 2 \times \mu_{tall}^w(\mathbf{Harry}) \\ &= 1/0 \text{ if for all/no } \mathcal{S}_{tall}^w\text{-admissible } \mu_{tall}^w, \mu_{tall}^w(\mathbf{Mary})/\mu_{tall}^w(\mathbf{Harry}) = 2 \end{aligned}$$

As discussed in section 2.2.2, if  $\mathcal{S}_{tall}^w$  is a ratio scale, then the ratio  $\mu_{tall}^w(x)/\mu_{tall}^w(y)$  will be the same for all admissible  $\mu_{tall}^w$ , for any  $x$  and  $y$  for which these measure functions are defined. (2.29) is thus guaranteed to be either true or false.

Third example, in which things go haywire due to the use of a ratio modifier with an adjective whose scale is merely interval.

$$\begin{aligned} (2.30) \quad & \llbracket \text{Jacksonville is exactly twice as hot as Chicago is} \rrbracket^{\mathcal{M},w} \\ &= 1/0 \text{ if for all/no gen-admissible } \mathbf{m}, \llbracket \text{J'ville is exactly twice as hot as Ch. is} \rrbracket_{\mathbf{m}}^{\mathcal{M},w} = 1 \\ &= 1/0 \text{ if for all/no } \mathcal{S}_{hot}^w\text{-admissible } \mu_{hot}^w, \mu_{hot}^w(\mathbf{J'ville}) = 2 \times \mu_{hot}^w(\mathbf{Ch.}) \\ &= 1/0 \text{ if for all/no } \mathcal{S}_{hot}^w\text{-admissible } \mu_{hot}^w, \mu_{hot}^w(\mathbf{J'ville})/\mu_{hot}^w(\mathbf{Ch.}) = 2 \end{aligned}$$

As discussed in §2.2.3, if  $\mathcal{S}_{hot}^w$  is an interval scale, then the ratio  $\mu_{hot}^w(x)/\mu_{hot}^w(y)$  will not generally be the same for all admissible  $\mu_{hot}^w$ . Normally, there will be admissible measures that make less or

more than 2. The exception is when  $x \approx_{hot} y$ , in which case  $\mu_{hot}^w(x) = \mu_{hot}^w(y)$  and the ratio will be 1. (2.30) is thus undefined unless Jacksonville and Chicago have the same temperature, in which case it is false.

No doubt there are other, more revisionary ways to construct a workable semantics for gradable expressions without taking degrees as basic objects in the ontology (e.g., Burnett 2014). However, the supervaluation method just sketched has the advantage of allowing us to port over without modification the insights of degree semantics as developed by many previous theorists and sketched in chapter 1. For the rest of this chapter, I will continue to develop features of the RTM perspective on scales, assuming as background the mapping into degree semantics discussed in this section.

When discussing modal semantics in subsequent chapters, we will make use of the duality between degree scales and qualitative scales by using whichever representation is convenient. As long as we are clear about the formal properties of the representation being used, we can be assured that an equivalent representation is available using the other format. This will be useful, in particular, because it allows us to consider qualitative ordering-based semantics for modality proposed in previous literature and immediately convert it into an equivalent degree-based theory, from which various grammatical and logical consequences can immediately be seen to follow. The most important are boundedness properties and, as we will now consider, the differential behavior of the concatenation/join operation in different kinds of scales.

## 2.4 Measurement Theory and Natural Language Scales

The interpretation of concatenation appropriate for natural language semantics is rather different from the typical use in RTM. Concatenation—interpreted as a restricted variant of the semantic operation of sum formation, or join—is relevant to the interpretation of a variety of scale types, including interval scales. Variability in how scales respond to concatenation stems from what further axioms a structure obeys, and in particular whether the structure is positive or intermediate with respect to concatenation. This results in variable interactions between scalar properties and the join operation, notably non-Boolean *and* (in the domain of individuals) and Boolean *or* (in the domain of propositions).

### 2.4.1 Concatenation as restricted join

Many domains standardly used in natural language semantics are Boolean, i.e. have a type ending in *t*. Boolean domains have the structure of a **BOOLEAN ALGEBRA**:

(2.31) A **Boolean algebra** is a structure  $\langle X, \vee, \wedge, \neg, \perp, \top \rangle$ , which obeys the following axioms as well as their inverses:  $\top \in X, \perp \in X$ , and for all  $a, b, c \in X$ ,

- a. **Associativity**:  $a \vee (b \vee c) = (a \vee b) \vee c$
- b. **Commutativity**:  $a \vee b = b \vee a$
- c. **Absorption**:  $a \vee (a \wedge b) = a$
- d. **Distributivity**:  $a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$
- e. **Complements**:  $a \vee \neg a = \top$

(The “inverses” of these axioms are the formulas that you get if you interchange  $\vee$  and  $\wedge$  throughout and change  $\top$  to  $\perp$  in (2.31e).)

The fact that Boolean domains have this structure is closely related to the fact that expressions whose type ends in  $t$  are in a 1-to-1 relationship with sets: e.g., the type  $\langle e, t \rangle$  function  $\lambda x_e[P(x)]$  is the characteristic function of the set  $\{x \mid P(x)\}$ . Relative to some fixed domain of individuals  $\mathbf{D}_e$ , the set of all type  $\langle e, t \rangle$  meanings forms an algebra isomorphic to the powerset lattice  $\langle \mathcal{P}(\mathbf{D}_e), \cup, \cap, -, \emptyset, \mathbf{D}_e \rangle$ . This is a Boolean algebra, where meet  $\wedge$  is identified with set intersection  $\cap$ , join  $\vee$  with set union  $\cup$ , and complement  $-$  with set complement  $-$ .

This structure may not be general enough for other semantic domains, though. For example, type  $e$  has no inherent structure in Montague’s (1973) theory. In the Boolean semantics of Keenan & Faltz (1985), this is taken as a point in favor of eliminating type  $e$  from the object language altogether. Although there is a set of individuals  $X$  whose powerset algebra forms the domain of type  $\langle e, t \rangle$  expressions, expressions usually assigned type  $e$  such as *Barack Obama* or *the Queen of England* do not receive interpretations in type  $e$ , but in the Boolean domain  $\langle \langle e, t \rangle, t \rangle$ . For instance,  $\llbracket \text{Barack} \rrbracket^{\mathcal{M}, w} = \lambda P.P(\mathbf{Barack}) = \{P \mid P(\mathbf{Barack})\}$ , the (characteristic function of) the set of properties that the real-world individual associated with the name “Barack” has.

Once of the nice consequences of this approach is that *or* can be interpreted as the join operation, and *and* with the meet operation, in any Boolean type, including the type  $\langle \langle e, t \rangle, t \rangle$  in which individuals denote in Keenan & Faltz’s (1985) theory.

- (2.32) a.  $\llbracket \text{run and jump} \rrbracket^{\mathcal{M}, w} = \{x \mid \mathbf{run}(x)\} \cap \{y \mid \mathbf{jump}(y)\}$   
 b.  $\llbracket \text{Barack and Michelle} \rrbracket^{\mathcal{M}, w} = \{P \mid P(\mathbf{Barack})\} \cap \{Q \mid Q(\mathbf{Michelle})\}$   
 c.  $\llbracket \text{Barack or Michelle} \rrbracket^{\mathcal{M}, w} = \{P \mid P(\mathbf{Barack})\} \cup \{Q \mid Q(\mathbf{Michelle})\}$

From here we get the equivalence *Barack and Michelle like carrots*  $\leftrightarrow$  *Barack likes carrots and Michelle likes carrots*: the property denoted by *like carrots* just has to be in the intersection of the set of properties that Barack has and the set of properties that Michelle has.

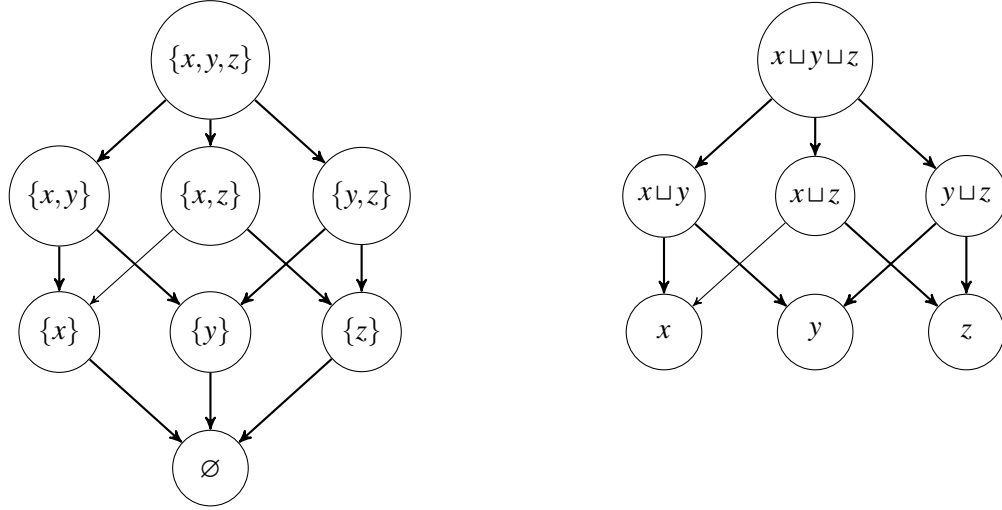
However, Link (1983) points out that treating *and* as denoting the meet operation in all domains makes it difficult to account for collective predicates such as intransitive *meet*. In this case, the equivalence between  $P(x \text{ and } y)$  and  $P(x) \text{ and } P(y)$  does not hold.

- (2.33) a. Barack and Michelle met in 1989.  
 b. Barack met in 1989 and Michelle met in 1989.

Not only does (2.33b) not mean the same as (2.33a), it’s not clear that it is even intelligible.

Link argues that we should abandon the assumption that the domain of individuals  $D_e$  is an unstructured set. Instead, he assigns it a MEREOLOGICAL (part-whole) structure: the individuals **Barack** and **Michelle** are proper parts of the compound individual (**Barack**  $\sqcup$  **Michelle**). The latter is known as the INDIVIDUAL JOIN (*i*-join) or SUM of **Barack** and **Michelle**. We can then translate (2.33a) as saying that the complex individual (**Barack**  $\sqcup$  **Michelle**) participated in a meeting event which occurred in 1989. Distributive predicates such as *like carrots*, on the other hand, have the special property that they hold of a compound individual if and only they hold of all of the atoms which make up the compound individual, so that the equivalence  $P(x \text{ and } y) \text{ iff } P(x) \text{ and } P(y)$  holds only in this special case.

The structure that Link ascribes to the domain of individuals treats it as an UNBOUNDED JOIN SEMILATTICE. An unbounded join semilattice is essentially a Boolean algebra minus the bottom element  $\perp$  (i.e.  $\emptyset$ , in the case of a powerset algebra). See Figure 2.6. As the name suggests, a join semilattice is closed under the join operation, but is not generally closed under meet and complement because of the absence of a bottom element.<sup>3</sup>



**Figure 2.6** Left: Boolean algebra  $\mathcal{P}(\{x,y,z\})$ , with arrows representing the proper subset relation. Right: join semilattice with atoms  $\{x,y,z\}$ . On a mereological interpretation, arrows represent the proper part relation.

Formally, an unbounded join semilattice is characterized as follows. (All quantifiers in the definition are implicitly restricted to  $X$ .)

- (2.34) An **unbounded join semilattice** is a structure  $\langle X, \sqcup \rangle$ , where  $X$  is a non-empty set, and  $\sqcup$  is a binary operation on  $X$  that obeys
- a. **Closure:**  $\forall x \forall y \exists z : x \sqcup y = z$
  - b. **Commutativity:**  $\forall x \forall y : x \sqcup y = y \sqcup x$
  - c. **Associativity:**  $\forall x \forall y \forall z : x \sqcup (y \sqcup z) = (x \sqcup y) \sqcup z$
  - d. **Idempotency:**  $\forall x : x \sqcup x = x$
  - e. **No Bottom Element:**  $\neg \exists x \forall y : x \sqcup y = y$

Note that the last axiom (2.34e) can be written in a more intuitive way if we define an auxiliary relation  $\sqsubseteq$  “part of”, where

$$x \sqsubseteq y \Leftrightarrow x \sqcup y = y.$$

<sup>3</sup> Link’s proposal is widely but not universally accepted, and there have been concerted efforts to maintain a uniformly Boolean interpretation of *and* and *or* (Winter 2001; Champollion 2013). This would clearly be theoretically desirable, but is not clear yet whether these efforts will be successful: see Scha & Winter 2015 for a recent survey and discussion of prospects.

We can then rewrite (2.34e) as

$$\neg \exists x \forall y : x \sqsubseteq y,$$

i.e., “There is nothing in  $X$  that is a part of everything in  $X$ ”.

What is interesting about Link’s proposal for the purposes of a measurement-theoretic semantics of degree is this: the domain of individuals, like the domain of propositions, is internally structured via a join operation. Furthermore, there are close intuitive and formal correspondences between the join operation in structured domains and the concatenation operation in RTM. On the intuitive level, it is obvious that we want the weight of John and the weight of Mary to have a systematic connection with the weight of the complex individual (**John**  $\sqcup$  **Mary**). The latter quantity should be the sum of their individual weights. This is, of course, exactly what we expect if weight is a ratio scale, and the join operation gives us the interpretation of the concatenation operation in weight measurement.

The intuitive connection between concatenation and join was noted by Krifka (1989: 79), who suggests in his RTM-inspired treatment of amount expressions that “concatenation ... can be defined as join ... restricted to non-overlapping individuals”. In fact, it is not possible to draw this equation using the standard axiomatizations of ratio scales. However, Krantz et al. (1971: §3.4) describe an alternative axiomatization of ratio scales that does not have these issues, and is compatible with Krifka’s suggestion that we use a restricted version of the join operation to define ratio scales.

In light of the discussion of temperature and danger in §2.5.2, it is also desirable to generalize the definition of concatenation so that it is not only applicable to additive structures such as ratio scales. We will deal with ratio scales first, and then consider the extension to interval scales.

First we define “overlap”: two objects overlap if they share any part (other than the bottom element, if one exists). Recall from above that  $x \sqsubseteq y$  means “ $x$  is a part of  $y$ ”, i.e.,  $x \sqcup y = y$ .

(2.35) Relative to structured domain  $X$ ,  $x$  and  $y$  **overlap** if and only if there is some  $z$  which is part of both  $x$  and  $y$ , and  $z$  is not an intrinsic bottom element (e.g., the empty set). More formally,  $x$  overlaps  $y$  iff  $x, y \in X$  and there is a  $z$  such that

- $z \sqsubseteq x$  and  $z \sqsubseteq y$ ;
- There is some  $z' \in X$  such that  $z \not\sqsubseteq z'$ .

With this definition in hand, we can state the general claim about the interpretation of concatenation that is relevant for a qualitative scalar semantics for natural language.

(2.36) The concatenation operation  $x \circ y$  is defined if and only if

- $x$  and  $y$  have the same semantic type  $\alpha$ , and
- $x$  and  $y$  do not overlap.

When it is defined,  $x \circ y = x \sqcup y$ , where  $\sqcup$  is the join operation in the domain  $D_\alpha$  of semantic objects of type  $\alpha$ .

## 2.4.2 What makes a ratio scale?

The standard axiomatizations of ratio scales cannot be extended directly to mereologically structured domains due to their requirement of unrestricted concatenation and non-existence of an inherent

maximum. Neither of these assumptions is viable for linguistic applications. However, Krantz et al. (1971: 84) give an alternative axiomatization which makes room for restricted concatenation and inherent maxima. Here I sketch a variant of this axiomatization, somewhat simplified as a result of our interpretation of concatenation. An appendix to this chapter (§2.7) discusses the relationship between the axioms in (2.37) and the characterization given by Krantz et al.

In the following, recall that we are **not** assuming that the base set  $X$  in a structure  $\langle X, \succsim, \circ \rangle$  is the join semilattice consisting of the entire domain of individuals  $D_e$ . Rather, we include only individuals for which it makes sense to talk about the extent to which they have the relevant scalar property. For example, when we are dealing with qualitative scales representing properties of physical objects—height, weight, etc.—the base set  $X$  will contain only individuals for whom such comparisons make sense. The number seven, dignity, and other abstract objects are excluded along with anything that contains them as a part.

(2.37) Consider the structure  $\langle X, \succsim, \circ \rangle$ , where  $X$  is a join semilattice,  $\succsim$  is a weak order on  $X$ , and  $\circ$  is disjoint join as defined in (2.36). Suppose that the following conditions are also satisfied:

- a. **Monotonicity:** If neither  $a$  nor  $b$  overlaps  $c$ , then  $a \succsim b$  iff  $a \circ c \succsim b \circ c$ .
- b. **Positivity:**  $a \circ b \succ a$  for all  $a, b \in X$  such that  $a$  and  $b$  do not overlap.
- c. **Solvability:** If  $a \succ b$ , then there is some  $d \in X$  such that  $a \succsim b \circ d$ .
- d. **Archimedean:** for all  $x \in X$ , there is no infinite sequence of non-overlapping  $y_1, y_2, \dots$  such that
  - for all  $j$  and  $k$ ,  $y_j \approx y_k$ ;
  - for all  $j$ ,  $x \succ_P y_1 \circ y_2 \circ \dots \circ y_j$ .

If all of these conditions are satisfied, then  $\langle X, \succsim, \circ \rangle$  is a ratio scale, and there is a measure function  $\mu : X \rightarrow \mathbb{R}$  such that

- $a \succsim b$  if and only if  $\mu(a) \geq \mu(b)$ ;
- If  $a$  and  $b$  do not overlap, then  $\mu(a \circ b) = \mu(a) + \mu(b)$ ;
- For any  $\mu'$  satisfying these two conditions and any  $x \in X$ , there is a positive  $n$  such that  $\mu'(x) = n \times \mu(x)$ .

(Proof: see §2.7 for discussion and references.)

The key assumptions that are required to ensure that the qualitative scale  $\langle X, \succsim, \circ \rangle$  satisfies the requirements in (2.37)—and so constitutes a ratio scale. Here is some discussion of the meaning of these axioms:

- **Weak order:**  $\succsim$  is reflexive and transitive, and any two objects can be compared.
- **Monotonicity:** A truck weighs more than a car if and only if the truck *with Rover sitting on top* weighs more than the car *with Rover sitting on top*. Similarly, a book costs more than a candy bar if and only if the book and a parking ticket cost more together than the candy bar and the parking ticket together. This axiom is highly intuitive, at least for the main candidates for constituting ratio scales that we have discussed.



- **Positivity:** For any two disjoint things for which we can speak sensibly of their degree of the property in question, their join has a greater degree of the property than either. A consequence is that, if  $x$  is a proper part of  $y$ , then either  $x < y$  or it does not make sense to talk about both individuals' degrees of the property. The former seems to be a reasonable assumption concerning the weight of two physical objects. The latter is a plausible analysis of how we count objects such as people, whose proper parts are not themselves people. (This axiom makes sense for scales such as weight and height, where it is not possible to have a zero degree of the property, but it needs a slight modification for scales such as cost, since some things are free. See §2.4.3 below.)
- **Solvability:** This axiom requires ratio scales to have a fairly rich domain. Consider, for example, a domain with just three atomic individuals— $X = \{a, b, c, a \circ b, a \circ c, b \circ c, a \circ b \circ c\}$ —where  $a$ ,  $b$ , and  $c$  have weights 10, 15, and 20 pounds respectively. The structure  $\langle X, \succ_{weight}, \circ \rangle$  cannot be a ratio scale because  $a \circ b \succ_{weight} c$ , but there is no individual  $d$  such that  $a \circ b \succ_{weight} c \circ d$ . (This problem does not arise in ordinary degree semantics, because the domain of degrees is simply assumed to be given in advance and to have whatever mathematical structure is required by a given application, regardless of the number of individuals that are actually being measured.)
- **Archimedean:** this axiom serves to rule out infinitesimal measures. If it were not applicable to weights, for example, it would be possible that a compound object formed of an infinite number of objects that have non-zero weight could still weigh less than some finite object. This axiom is intuitive in the sense that, for properties that are plausibly modeled using ratio scales, it is difficult to conceive of a scenario in which its failure would make sense.

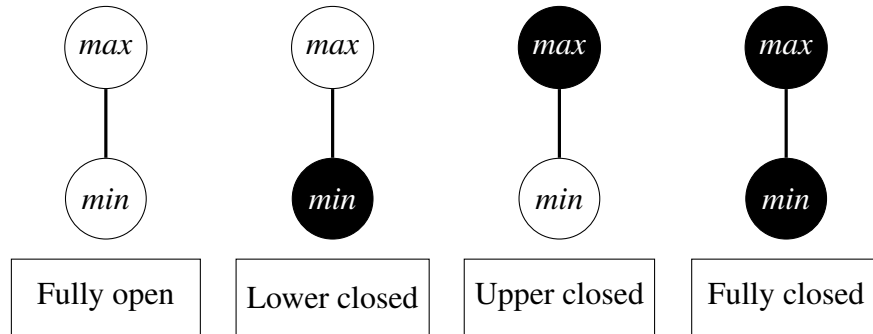
The upshot is that, if these axioms represent reasonable assumptions to make about the domain of some scalar property  $P$ , then the scale  $\mathcal{S}_P$  is ratio and all admissible measure functions  $\mu_P$  are additive.

Solvability will frequently fail in small domains, and in some cases it may be necessary to fill out the domain with “pseudo-objects” (e.g., a set of weights) in order to maintain the formal structure of a ratio scale. However, for the purpose of a general theory of natural language semantics, it seems reasonable to suppose that the domains (of individuals, events, propositions, etc.) are always very rich. Certainly, if we consider natural language semantics to be part of cognitive science, the domains of scales are not limited by comparisons among *actually* existing individuals, but should also include comparisons with and among *imaginable* individuals—all those that are relevant to the mental representation of a scale by an agent who has the relevant concept.

Additional candidates for failure in particular cases might include connectedness, monotonicity, and positivity. Connectedness probably fails for many scales that are lexicalized in natural languages, for example, beauty and cleverness. An even more interesting class, for our purposes, are the “intermediate” scales discussed below, for which positivity and monotonicity are clearly inappropriate: see §2.5.3.

### 2.4.3 Boundedness in ratio scales

The boundedness of scales has not traditionally been a topic of major interest in RTM, but, as we saw already in chapter 1, it is an issue of considerable import in the semantics of gradability (Hay et al. 1999; Rotstein & Winter 2004; Kennedy & McNally 2005a; Kennedy 2007). Figure 2.7 shows the well-known typology of scales from this work, repeated from chapter 1, §1.3.4.



**Figure 2.7** Possible scale types with respect to boundedness properties.

---

To get a usable measurement-theoretic characterization of boundedness properties, we need to make several distinctions. First, separate INHERENT boundedness from ACCIDENTAL boundedness. One prominent kind of accidental boundedness is the upper- and lower-boundedness exhibited by any property in a finite domain. In a finite model, the ratio scale  $\mathcal{S}_{tall} = \langle X, \succsim_{tall}, \circ \rangle$  will have an upper bound (the height of the tallest object in  $X$ ) and a non-zero lower bound (the height of the shortest object in  $X$ ). This is not enough to make *tall* behave as if it had a fully closed scale, judging by the usual tests.

- (2.38) a. This glass is almost full.  
       b. # Sam is almost tall.
- (2.39) a. This glass is half full.  
       b. # Sam is half tall.

It is not possible, for example, to interpret (2.39b) as meaning “Sam is half as tall as the tallest person around”. Accidental boundedness is not very interesting for our purposes; the interpretation and acceptability of degree modifiers and the like does not appear to be responsive to properties that hold of domains simply due to contingent features of a particular model.

Inherent boundedness, the kind that is linguistically interesting, is associated with some structural feature of a scale. For example, a Boolean algebra is inherently upper- and lower-bounded, due to the presence of a bottom element  $\perp$  and a top element  $\top$  in its structural description (cf. (2.31)).

In considering bounds, it is important to distinguish between asymptotic and non-asymptotic boundedness. For example, all admissible measure functions on ratio scales as defined in (2.37)—such as height and weight—map their domain into the positive reals  $(0, \infty)$ . Nothing has measure

0, and it does not make sense to talk about the weight of a non-physical object. All ratio scales are at least asymptotically bounded below by zero.

However, some intuitive examples of ratio scales are truly lower-bounded. Cost is an example: the zero clearly exists—some things have zero cost, and nothing can have negative cost—so, the cost scale seems to have an inherent minimum which is actually occupied by some things in the domain of the scale. In addition, statements about ratios of cost are interpretable.

- (2.40) a. The house cost seven times as much as the car.  
b. The house was seven times as expensive as the car.

(Other examples of lower-bounded ratio scales include speed, wealth, frequency, proportion, and probability.) However, the ratio scale axioms given above exclude the possibility of zero-measure objects. We can generalize the definitions to deal with this possibility of ratio scales by treating inherent minima as a special case in two places.

(2.41)  $\langle X, \succsim_P, \circ, \perp_P \rangle$  is a **lower-closed ratio scale** if it satisfies the Monotonicity and Solvability axioms in (2.37) and in addition:

- a. **Bottom:**  $\perp_P \in X$  and, for all  $a \in X$ ,  $a \succsim_P \perp_P$ .
- b. **Positivity** (modified):  $a \circ b \succ_P a$  for all  $a, b \in X$  if  $a$  and  $b$  do not overlap and  $b \not\approx \perp$ .
- c. **Archimedean** (modified): for all  $x \in X$ , there is no infinite sequence of non-overlapping  $y_1, y_2, \dots$  such that
  - for all  $i$ ,  $y_i \not\approx_P \perp_P$ ;
  - for all  $j$  and  $k$ ,  $y_j \approx_P y_k$ ;
  - for all  $j$ ,  $x \succsim_P y_1 \circ y_2 \circ \dots \circ y_j$ .

If these conditions are met, then there is an admissible measure function  $\mu$  such that

- $\mu(\perp_P) = 0$  and, for all  $x$  and  $y$  in  $X$ ,  $\mu(x \circ y) = \mu(x) + \mu(y)$ ;
- For any admissible  $\mu'$  meeting these conditions, there is some positive  $n$  such that  $\mu'(x) = n \times \mu(x)$ , for all  $x \in X$ .

(Sketch of proof technique and references: see the chapter appendix, §2.7.)

Cost is a scale that clearly has no inherent maximum: there is no principled limit to how much something can cost. However, as we saw in chapter 1, many scalar properties of interest for natural language semantics do have inherent maxima. As Krantz et al. (1971: 85) discuss, the axiomatization of ratio scales given above generalizes directly to scales that have an inherent maximum as long as the maximum element meets a “divisibility” condition.

(2.42)  $\langle X, \succsim_P, \circ, \top_P \rangle$  is an **upper-closed ratio scale** if

- a.  $\langle X, \succsim_P, \circ \rangle$  is a ratio scale;
- b.  $\top_P \in X$  and, for all  $a \in X$ ,  $a \leq_P \top$ ;
- c. **Divisibility:** There are  $a$  and  $b$  in  $X$  such that  $a \circ b \approx_P \top_P$ .

All admissible measure functions  $\mu$  on an upper-bounded ratio scale are additive and have a range that is bounded above by the measure of the intrinsic maximum,  $\mu(\top_P)$ . Upper-closed ratio scales,

like all ratio scales, are at least asymptotically bounded below; so the range of an admissible  $\mu$  is either  $(0, \mu(\top_P)]$  or  $[0, \mu(\top_P)]$  depending on whether there is also an intrinsic minimum.

Ratio scales with both an intrinsic maximum and minimum are fully closed:

- (2.43)  $\langle X, \succsim_P, \circ, \perp_P, \top_P \rangle$  is an **fully-closed ratio scale** if
- a.  $\langle X, \succsim_P, \circ, \perp_P \rangle$  is a lower-closed ratio scale;
  - b.  $\langle X, \succsim_P, \circ, \top_P \rangle$  is an upper-closed ratio scale.

If we analyze a scalar property using any type of ratio scale, we are predicting that absolute ratio statements such as *x is n times as P as y* are interpretable. In addition, if the scale is upper- or fully-closed, we also predict that references to a maximum (*almost/totally P*) are meaningful, as well as proportional statements (*x is half/n% P*). This is useful because it allows us to propose tentative classifications for some of the adjectives that were discussed with reference to the degree-based framework in chapter 1.

- (2.44) a. This glass of beer is twice as full as that one.  
b. This glass of beer is twice as heavy as that one.

These statements are both clearly interpretable, and so we can conclude that both *full* and *heavy* are ratio-scale properties.

#### 2.4.4 Proportional modifiers, scale type, and boundedness

A contrast between *full* and *heavy* emerges, however, when we compare the ratio modifiers in (2.44) to the proportional modifiers in (2.45).

- (2.45) a. This glass of beer is totally/half/30% full.  
b. # This glass of beer is totally/half/30% heavy.

Adapting the reasoning in Kennedy & McNally 2005a, *totally P* makes reference to an intrinsic maximum, and so it can occur in an interpretable statement only if  $S_P$  has an intrinsic maximum. This is satisfied in the case of *full*, but not of *tall*. On the other hand, *half* and *n%* appear to have the more demanding interpretations in (2.46)-(2.47).

$$(2.46) \quad \llbracket x \text{ is half } P \rrbracket^{\mathcal{M}, w} = 1 \text{ iff } \frac{\mu_P(x)}{\mu_P(\top_P)} = .5.$$

$$(2.47) \quad \llbracket x \text{ is } n\% P \rrbracket^{\mathcal{M}, w} = 1 \text{ iff } \frac{\mu_P(x)}{\mu_P(\top_P)} = n/100.$$

(To keep things simple, I assume here an “exactly” interpretation for *half* and *n%*.)

The definitions in (2.46) and (2.47) obviously require that the scale be upper-bounded by a maximum element  $\top_P$ . In addition, the ratios referred to these definitions will not have stable values across admissible  $\mu_P$  if the scale is interval: ratios of values in interval scales are disrupted by the possibility of linear transformation (adding/subtracting a constant to all values). However, ratios are preserved if the scale is ratio scale, since ratio scales only permit multiplicative transformations (multiplication by a positive value), which will not affect the ratio of  $\mu_P(x)$  to  $\mu_P(\top_P)$  for any  $x$ .

As a result, these definitions generate empirical predictions about the kinds of scalar predicates that should accept modification by *half* and *n%*:

- the scale must have an inherent maximum *and* minimum (though the minimum may be merely asymptotic);
- the scale must have enough formal structure to render interpretable (stable across admissible  $\mu$ ) the ratio between an arbitrary degree and the degree associated with the scalar maximum.

This means that the scale will be an upper-bounded ratio scale. Note that the requirement is that *only* predicates on upper-bounded ratio scales will accept modification by *n%* and *half*. Nothing requires that *all* such predicates will do so.

As suggested above, since the definitions in (2.46) and (2.47) reference  $\mu_P(\tau)$ , they require that the upper bound must be on the scale in both cases—in the sense that it is ordered by  $\geq_P$  and that it is in the domain of  $\mu_P$ . However, the definitions leave room for the possibility that the lower bound may be merely asymptotic. (As a reminder, height and weight are non-upper-bounded example of ratio scales with merely asymptotic lower bounds.) This fact may help to explain why proportional modifiers are acceptable in complex constructions involving equatives and comparatives, even when they are not acceptable with the adjectives from which they are formed. Compare the sentences in (2.48):

(2.48) # Mary is 50% tall.

“ $\mu_{\text{height}}(\mathbf{Mary})$  is 50% of the upper bound of the interval  $(0, \tau_{\text{height}}]$ .”

(2.49) ✓ Mary is 50% as tall as Bill.

“ $\mu_{\text{height}}(\mathbf{Mary})$  is 50% of the upper bound of the interval  $(0, \mu_{\text{height}}(\mathbf{Bill})]$ .”

(2.50) ✓ Mary is 50% taller than Bill.

“ $\mu_{\text{height}}(\mathbf{Mary})$  is  $\mu_{\text{height}}(\mathbf{Bill})$  plus 50% of the upper bound of  $(0, \mu_{\text{height}}(\mathbf{Bill})]$ .”

(2.48) is ill-formed because its interpretation makes reference to a scalar maximum which does not exist. In contrast, (2.49) and (2.50) succeed because the interpretation of equatives and comparatives involves a restriction to a subscale of heights which is upper-bounded: here, the interval extending from 0 to Bill’s height.

An obvious next question is whether proportional modification would succeed on a scale where the upper bound exists but is “off the scale”, i.e., merely asymptotic. The definitions in (2.46) and (2.47) predict that it should not, but a more complex definition in terms of limits would make the same predictions in these cases while also leaving for merely asymptotic upper bounds. I do not know of any way to test this prediction empirically: we would need an analogous case where the upper bound is merely asymptotic, and I do not know of any clear candidates. (For example, comparatives and equatives with *short* are not relevant since  $\mathcal{S}_{\text{short}}$  is not a ratio scale, and its lower bound is infinite height.) Chapter 4 will discuss a proposal according to which likelihood is such a scale (§4.2.6), but I will argue that there is little independent motivation for analyzing likelihood in this way. If so, the jury must remain out on whether proportional modifiers require a true maximum, or indeed if proportional modifiers are homogeneous in this respect.

## 2.5 Interval scales and intermediacy

To round out the refined typology of scale types that we are developing in this chapter, we turn to interval scales. After giving the formal characterization of interval scales, I will consider the relationship between these scales and the interpretation of concatenation as join suggested in the previous section. The most important observation is that there is a class of scales (temperature, danger, expected value, ...) which are **intermediate** with respect to concatenations. This stands in contrast to the **positivity** of ratio scales.

### 2.5.1 Formal characterization of interval scales

Interval scales such as temperature and clock time have the interesting property that there is no fixed zero point, but the size of *gaps* between points is stable. As we saw above, this means that interval scales are unique up to positive affine transformation: for any admissible  $\mu$  and  $\mu'$ , there is some positive  $\alpha$  and some  $\beta$  such that  $\mu'(x) = \alpha \times \mu(x) + \beta$  for any  $x$  in the domain.

The general strategy for axiomatizing this scale type is to have the binary relation  $\succsim$  order not individual objects, but pairs of objects representing gaps. That is, interval scales are structures  $\langle X, Y, \succsim_P \rangle$ , where  $Y \subseteq X \times X$  is a set of pairs of objects in  $X$ , and  $\succsim_P$  is a binary relation on  $Y$ . The relation

$$(a, b) \succsim_P (c, d)$$

can be read “ $a$  exceeds  $b$  with respect to property  $P$  by more than  $c$  exceeds  $d$ ”. From this, we can define an ordering  $\succsim_P^i$  directly on individuals: for all  $a, b \in X$ ,

$$a \succsim_P^i b \text{ if and only if } (a, b) \succsim_P (b, b).$$

For example,  $a$  is at least as hot as  $b$  if and only if the difference in temperature between  $a$  and  $b$  is at least as great as zero (the difference in temperature between  $b$  and itself). This is the qualitative equivalent of the claim that  $\mu(a) \geq \mu(b)$  for all  $S_P$ -admissible  $\mu$ .

In contrast to ratio scales, I will not appeal to the axiomatic characterization of interval scales in detail elsewhere in the book. However, for completeness I give the axiomatization briefly here. (The reader can skip to subsection 2.5.2 without loss of content.)

(2.51)  $\langle X, Y, \succsim_P \rangle$  is an **interval scale** if and only if

- a.  $Y \subseteq X \times X$  and  $\langle Y, \succsim \rangle$  is a weak order;
- b. **Positive-negative:** For all  $a, b, c, d$  in  $X$  if  $(a, b) \succsim_P (c, d)$  then  $(d, c) \succsim_P (b, a)$ ;
- c. **Monotonicity:** For all  $a, b, c, a', b', c'$  in  $X$ : if  $(a, b) \succsim_P (a', b')$  and  $(b, c) \succsim_P (b', c')$ , then  $(a, c) \succsim_P (a', c')$ ;
- d. **Associativity:** For all  $a, b, c, d$  in  $X$ : if  $(a, b) \succsim_P (c, d)$  and  $(c, d) \succsim_P (x, x)$ , then there are  $e, f \in X$  such that  $(a, e) \succsim_P (c, d)$  and  $(f, b) \succsim_P (c, d)$ .
- e. **Archimedean:** Every strictly bounded standard sequence is finite, where a “standard sequence” is a sequence of individuals  $a_1, a_2, \dots$  in  $X$  such that,  $(a_1, a_2) \not\succsim_P (a_1, a_1)$  and, for every  $i$ ,  $(a_i, a_{i+1}) \approx_P (a_{i+1}, a_{i+2})$ .

If these conditions hold, then there is an admissible measure function  $\mu$  such that

$$(a, b) \succsim_P (c, d) \text{ if and only if } \mu(a) - \mu(b) \geq \mu(c) - \mu(d),$$

and, for all  $\mu'$  meeting this description, there is some  $\alpha \in \mathbb{R}^+$  and  $\beta \in \mathbb{R}$  such that  $\mu'(x) = \alpha \times \mu(x)$ . (Proof: Krantz et al. 1971: 158.)

### 2.5.2 Concatenation-as-join, positivity, and intermediacy

With some notable exceptions (especially, Luce & Narens (1985)), it is usual in RTM to suppose that concatenations are defined only on ratio scales. However, this interpretation is not available to us once we have adopted the interpretation of concatenation as restricted join. For any property whose domain has internal structure (e.g., that of a Boolean algebra or a join semilattice), concatenations should be available and potentially relevant.

This prediction seems to be correct: there are scalar properties for which concatenation is intuitively a meaningful operation, but the additivity and positivity properties of ratio scales are clearly inappropriate. Suppose you have two bowls of soup  $x$  and  $y$ , and you pour the contents of  $x$  and  $y$  into a larger bowl  $z$ . If we ask about the volume of soup in  $z$ , the answer is straightforward:  $z$ 's volume is just the sum of the volumes that  $x$  and  $y$  had before the pouring began. In RTM terms, this makes sense because  $z = x \circ y$ , and given additivity this implies  $\mu_{\text{volume}}(z) = \mu_{\text{volume}}(x) + \mu_{\text{volume}}(y)$ . Obviously, a ratio classification requires positivity: as long as neither  $x$  nor  $y$  has volume 0, the volume of  $z$  is greater than the volume of either  $x$  or  $y$ . This means that the following should be a valid inference as indeed it is.

- (2.52) a. This bowl of soup has volume  $v$ , and that one has volume  $v' > v$ .  
b. So, if we pour both into a larger bowl, the result will be a bowl with a volume greater than  $v'$ .

If we ask about the temperature of  $z$  instead, however, things are more difficult. If temperature were positive with respect to concatenations, then we would expect the following inference to be valid.

- (2.53) a. This bowl of soup is 40 degrees Celsius. That one is 20 degrees Celsius.  
b. # So, if we pour both into a larger bowl, the result will be a bowl which is more than 40 degrees Celsius.

Since this argument is specious, we can conclude that temperature is not positive.

Another non-positive property is danger. Suppose that some object  $x$ —say, a U.S. state—is covered by  $n$  non-overlapping parts (say, its counties), where  $x = (y_1 \circ y_2 \circ \dots \circ y_n)$ . For any property  $P$  for which concatenation is positive, we can validly infer that  $x$  is  $P$  to a greater degree than any of its proper parts:  $(x \succ_P y_1) \wedge (x \succ_P y_2) \wedge \dots \wedge (x \succ_P y_n)$ . For additive properties such as size, this inference seems extremely trivial. Consider, for example, what you could infer from a list of the areas of all of the counties in the state of Georgia: the state obviously has an area greater than any of the counties that make it up.

- (2.54) a. Fulton County, Georgia has an area of 535 square miles.

- b. Cobb County, Georgia has an area of 345 square miles.
- c. So, the state of Georgia has an area greater than 535 square miles.

However, the corresponding inference is specious when we consider danger rather than size.

- (2.55)
- a. Fulton County is extremely dangerous.
  - b. Cobb County is quite safe (only slightly dangerous).
  - c. # So, the state of Georgia is extremely dangerous or worse.

We cannot deal with these facts by refusing to countenance concatenations when reasoning about temperature and danger: after all, there *is* a systematic correspondence between the temperature or danger of an object and the temperature or danger of its proper parts. It could not be, for example, that all parts of Georgia are extremely dangerous, but that the state as a whole is very safe. By virtue of our understanding of the properties **temperature** and **danger**, we know how the temperature or danger of a complex object relates to the temperature or danger of its component parts, and we know that this relationship is not additive. Instead, concatenation seems to produce an object that has an INTERMEDIATE degree of the property in question.

- (2.56)
- a. This bowl of soup is 40 degrees Celsius. That one is 20 degrees Celsius.
  - b. So, if we pour one into the other the result will be a bowl which is somewhere between 20 and 40 degrees Celsius.
- (2.57)
- a. Fulton County is extremely dangerous.
  - b. Cobb County is quite safe (only slightly dangerous).
  - c. So, Fulton and Cobb Counties taken together are somewhere between quite safe and extremely dangerous (e.g., moderately dangerous).

Basically, it looks as if instead of obeying the positivity assumption, these properties respond to concatenation as in (2.58).

- (2.58) **Intermediacy of  $\succsim$  with respect to  $\circ$ :** Suppose that  $x \succ_P y$  and  $x \circ y$  is defined. Then  $x \succ_P (x \circ y) \succ_P y$ .

In chapter 7 we will return to this definition in the context of comparative goodness, which I will argue is also an intermediate scalar property.

If this is the right characterization of concatenation for the temperature and danger scales, we can add to our typology of scale types the *intermediate interval scales*:

- (2.59)  $\langle X, Y, \succsim_P \rangle$  is an INTERMEDIATE INTERVAL SCALE if and only if
- a.  $\langle X, Y, \succsim_P \rangle$  is an interval scale;
  - b.  $\succsim_P^i$  is intermediate with respect to  $\circ$ .

(Where  $\succsim_P^i$  is the individual comparison associated with  $\succsim_P$ . See §2.5.1 for the formal definition.)

An interesting question which I will not try to resolve here is whether natural languages lexicalize other kinds of scales that are intermediate with respect to concatenations.



### 2.5.3 Other alternatives to positivity

The obvious next question is whether, in addition to intermediacy and positivity, there are other ways that scalar properties might distribute over concatenation/join.

Here are some options, all of which have either been proposed in previous linguistic, psychological, or philosophical literature, or will be discussed in this book.

- (2.60) a. **Additive:**  $\mu(x \circ y) = \mu(x) + \mu(y)$   
 b. **Superadditive:**  $\mu(x \circ y) > \mu(x) + \mu(y)$   
 c. **Subadditive:**  $\mu(x) + \mu(y) > \mu(x \circ y) \geq \mu(x)$   
 d. **Maximal:** If  $x \geq y$ , then  $(x \circ y) \approx x$   
 e. **Intermediate:** If  $x \geq y$ , then  $x \geq (x \circ y) \geq y$   
 f. **Minimal:** If  $x \geq y$ , then  $(x \circ y) \approx y$   
 g. **Subtractive:** If  $x \geq y$ , then  $\mu_P(x \circ y) = \mu_P(x) - \mu_{P-1}(y)$ , where  $\mu_{P-1}$  is a measure function associated with the antonym of  $\mu_P$  with shared units.  
 h. **Atomic Only:** The domain of the relation  $\geq_P$  contains no concatenations, i.e.  $x \geq_P y$  implies that  $x, y$  are atomic.

Additive scales have already gotten a good deal of discussion, and are clearly relevant for many purposes. Super- and sub-additivity have been discussed in some detail in the psychological literature on probability judgment (e.g., Tversky & Koehler 1994; Macchi, Osherson & Krantz 1999). Subtractive scales have not been noticed previously, but they appear to be the correct characterization of the scales associated with the antonyms of additive adjectives, such as *short* and *light*. Also, I argued in the previous section that temperature and danger are intermediate with respect to concatenation; in chapter 7 I will add goodness to the list.

Maximality will be relevant throughout the chapters on modals, since it is a property of Lewis's (1973) semantics for gradable modals, and also (modulo connectedness) of the semantics of Kratzer (1981, 1991). If we use the Lewis-Kratzer semantics to provide qualitative scales for epistemic and deontic adjectives, then these adjectives are predicted to be maximal with respect to concatenation (disjunction), and their antonyms (*unlikely*, *bad*) should be minimal. See chapter 3 for extensive discussion.

Finally, many gradable properties seem to be Atomic Only, i.e. distributive: if John and Mary are happier than Sue, this cannot mean, for instance, that the sum or the average of their happiness exceeds Sue's, but only that each of them is individually happier than Sue.<sup>4</sup>

### 2.5.4 Boundedness in interval scales

Another question that arises here is whether interval scales are ever inherently upper- or lower-bounded. This issue deserves a more detailed formal discussion than I am prepared to give here. Nevertheless, I will venture a few linguistically-motivated speculations.

<sup>4</sup> Incidentally, this point problematizes Sassoon's (2010a) argument that *happy* is a ratio scale adjective. I suspect that the data points that Sassoon adduces in support of this hypothesis have an alternative explanation, along the lines of §4.1.3.

In general, there does not seem to be any formal barrier to allowing that some interval scales may have lower and/or upper bounds. For example, *danger* is a plausible candidate for being an interval scale that is both intermediate and lower-bounded. Its intermediacy was motivated above (§2.5.2); lower-boundedness is indicated by linguistic tests, for example, the acceptability of *slightly* and the upper-boundedness of its antonym *safe*.

- (2.61) a. This neighborhood is slightly dangerous. (*slightly*-modification)  
 b. This neighborhood is completely/almost safe.

Conversely, if *dangerous* is associated with a lower-bounded interval scale, its antonym *safe* must be associated with an upper-bounded interval scale.

There are also some candidates for fully closed interval scales. For example, the pair *similar/different* passes the usual tests for double-boundedness (2.62); but this is probably not a ratio scale, given that positivity is not very plausible (2.63).

- (2.62) My book is  $\left\{ \begin{array}{l} \text{completely} \\ \text{slightly} \end{array} \right\} \left\{ \begin{array}{l} \text{similar to} \\ \text{different from} \end{array} \right\}$  yours.

- (2.63) ?? Bill is similar to Obama. So, Bill and his extended family are even more similar to Obama.

We might think to make similarity a fully closed **ordinal** scale, but this would be too weak: then we would not be able to interpret even extremely weak and intuitively meaningful quantitative comparisons such as (2.64).

- (2.64) Bill is much more similar to Obama than Jim is.

These considerations might be taken to motivate the idea that *similar/different* are on a fully closed interval scale (cf. Tversky 1977). This is an interesting possibility, in particular, because the differences between such a scale and a fully closed ratio scale are very subtle: both would render statements about ratios and proportions interpretable, and they would differ only in that the ratio scale is positive over concatenations.

Since nothing in the rest of the book will rely on the answer to these questions, I will not pursue them further here. However, a full typology of the scales invoked by natural languages will need to provide an answer.

## 2.6 Summary and Conclusion

This chapter gave an overview of the Representational Theory of Measurement, which provides a method for constructing measure functions from ordered qualitative structures. We discussed several of the standard RTM scale types—ordinal, interval, and ratio—and proposed a method of doing “degree semantics without degrees” by constructing degree scales from an underlying qualitative semantics using supervaluations and the RTM concept of admissibility. Next, I suggested a way to integrate the RTM apparatus with the structured-domains required for Boolean and plural semantics by treating concatenation as restricted join; proposed an expanded range of scale types; and situated the familiar boundedness properties within this range. In addition to the boundedness properties

familiar from Kennedy & McNally (2005a), there is a distinction between ordinal, interval, and ratio scales, and a distinction between scales which are additive with respect to concatenation and those which are intermediate.

For reference, here is a brief summary of the partial typology of scales that was developed in this chapter.

- (2.65)
- a. All scales are assumed to be transitive.
  - b. Scales may be **connected** (complete) or not.
  - c. Connected scales may be
    - **ordinal, interval, or ratio;**
    - **upper- and/or lower-bounded;**
    - **positive or intermediate** with respect to concatenations.

This classification includes only the distinctions that I have been able to motivate in chapters 1-2 using English data. No doubt there are further scale types waiting to be discovered, both in English and other languages.

In the remainder of the book I will use measurement-theoretic reasoning and the expanded range of scales discussed here to investigate the structure of epistemic and deontic modals. In chapter 3 I will use the tools developed in this chapter to clarify and render compositional previous proposals for the interpretation of graded modals, in the process revealing some important limitations relating to the structures of the qualitative scales assumed. Building on this discussion, chapters 4-5 will use data involving degree modification of epistemic expressions and their inferential behavior to motivate assigning an important class of epistemic adjectives to fully closed ratio scales—like the one associated with the adjective pair *full/empty*. Chapter 6 then discusses implications of this conclusion for the semantics of non-gradable epistemic modals such as *must* and *might*. In chapters 7-8 I will turn to deontic modals, arguing that one of the scale types identified here—the intermediate interval scale—plays a crucial role in their semantics.

## 2.7 Appendix

The interpretation of concatenation as restricted join allows us to simplify the axiomatic characterization of ratio scales considerably relative to that given by Krantz et al. (1971: 84), along the lines described in §2.4.2. Krantz et al.'s formulation is given in (2.66).

- (2.66) Consider a structure  $\langle X, \succsim, B, \circ \rangle$ , where  $X$  is non-empty;  $\succsim$  is a weak order on  $X$ ;  $B$  is a subset of  $X \times X$  representing pairs of objects that can be concatenated; and  $\circ : B \rightarrow X$  is a function taking the concatenable pairs to their concatenation. Suppose that the following conditions are also satisfied:
- a. **Associativity:** If  $(a, b) \in B$  and  $(a \circ b, c) \in B$ , then  $(b, c) \in B$ ,  $(a, b \circ c) \in B$ , and  $((a \circ b) \circ c) \succsim (a \circ (b \circ c))$ .
  - b. **Monotonicity:** If  $(a, c) \in B$  and  $a \succsim b$ , then  $(c, b) \in B$  and  $a \circ c \succsim c \circ b$ .
  - c. **Solvability:** If  $a \succ b$ , then there is some  $d \in X$  such that  $(b, d) \in B$  and  $a \succsim b \circ d$ .
  - d. **Restricted positivity:** If  $(a, b) \in B$ , then  $a \circ b \succ a$ .

- e. **Archimedean:** for all  $x \in X$ , there is no infinite sequence  $y_1, y_2, \dots$  such that,
- for all  $j$  and  $k$ ,  $y_j \approx y_k$ ;
  - for all  $j$ ,  $y_1 \circ y_2 \circ \dots \circ y_j$  is defined, and  $x \succsim_P y_1 \circ y_2 \circ \dots \circ y_j$ .

If all of these conditions are satisfied, then  $\langle X, \succsim, B, \circ \rangle$  is a ratio scale, and there is a measure function  $\mu : X \rightarrow \mathbb{R}$  such that

- $a \succsim b$  if and only if  $\mu(a) > \mu(b)$ ;
- If  $(a, b) \in B$ , then  $\mu(a \circ b) = \mu(a) + \mu(b)$ ;
- For any  $\mu'$  satisfying these two conditions,  $\mu'(x) = n \times \mu(x)$  for all  $x \in X$ .

**Proof:** see Krantz et al. 1971: 87.

The Link-style semantics for plurals and our interpretation of concatenation allow us to re-interpret the axiomatization in a way that may be more intuitive, and also to simplify it slightly: the discussion in §2.4.2 does not include an axiom corresponding to (2.66a) because these requirements are enforced by the interpretation of concatenation as restricted join. Rationale: If  $a$  and  $b$  do not overlap, and  $a \sqcup b$  does not overlap with  $c$ , then clearly  $b$  and  $c$  do not overlap, nor does  $a$  with  $b \sqcup c$ . So,  $b$  and  $c$  are concatenable, as are  $a$  and  $b \sqcup c$ . Furthermore, since join is associative,  $(a \sqcup b) \sqcup c = a \sqcup (b \sqcup c)$ . The condition that  $((a \circ b) \circ c) \succsim (a \circ (b \circ c))$  thus follows trivially from the reflexivity of  $\succsim$ .

The variant discussed in the main text also differs in the formulation of axiom (2.66b), because this axiom is unsatisfiable given our interpretation of concatenation as restricted join. Our version corresponds to that given by Krantz et al. (1971: 204-205); see also Narens 2007: 32-35.

The discussion in §2.4.3 also motivates weakening the positivity and Archimedean axioms (66d-66e) to make room for the existence of objects that have a zero degree of the property in question (e.g., cost is a ratio scale even though some things are free). We can extend the proof technique to such structures by reasoning about the substructure which excludes the equivalence class containing the intrinsic minimum: see Narens 2007: 33-34.

## CHAPTER 3

### Previous Work on Graded Modality: Lewis and Kratzer

#### 3.1 Modality and gradation

The term “modal” is used in at least two different ways. Sometimes it is used to pick out a syntactic category, the MODAL AUXILIARIES *may, might, can, could, should, would, must*, and the quasi-auxiliary *ought*. I will use “modal” in a more expansive way to refer to expressions which have a particular semantic flavor. As Portner (2009: 1) puts it:

[M]odality is the linguistic phenomenon whereby grammar allows one to say things about, or on the basis of, situations which need not be real.

This is more of a pointer than a definition – Portner precedes it with the proviso “I am not too comfortable trying to define modality” – but it provides a reasonable characterization of modality as a semantic phenomenon.

Construed this way, a wide variety of natural language expressions have (or have been claimed to have) modal semantics, going well beyond the small set of modal auxiliaries: conditionals, imperfective verbs, the future tense, expressions of mood, evidentials, many attitude verbs, perhaps certain nominal quantifiers (e.g., *many*), and probably much more. I will not use the term “modal” this broadly here, though. I am primarily interested in modal expressions that are able to take propositions as arguments and fall into one of four syntactic categories: auxiliaries, verbs, adjectives, and sentential adverbs.

##### (3.1) AUXILIARIES

- a. Harry should be in Sacramento by now.
- b. My brother can bench press 250 pounds.
- c. All cameras must be checked at the door.

##### (3.2) VERBS

- a. I need to go to Sacramento.
- b. My mother wants to be on television.
- c. You are required to wait behind the line.

##### (3.3) ADJECTIVES

- a. We are unable to fulfill your request.
- b. It is likely that we have missed our train.
- c. It is impermissible to fake illness to get out of work.

##### (3.4) ADVERBS

- a. Evidently, we have missed our train.
- b. We will possibly be in Houston next week.

- c. Obligatorily, children are picked up by 3PM.

Since these expressions come in several syntactic categories, we might expect that their semantics will vary somewhat as well. Nevertheless, the modals in (3.1)-(3.4) are often analyzed as having a common semantics built on some variant of standard modal logic—the semantic framework associated with Kripke (1963) and much following work in logic, philosophy, and computer science.

Modals are traditionally thought to come in several semantic types (or “flavors”), and certain of the auxiliary modals are ambiguous between two or more of these types. For example, *must* can be interpreted **epistemically** (“It *must* be, given what is known”, **deontically** (“You *must* do this, according to the law/morality”), **teleologically** (“You *must* do this in order to accomplish some goal”), or **bouletically** (“I *must* have this if I am to get what I want”). Another important modal type is **dynamic** or **circumstantial** modality, which talks about abilities and potentials.

Modal adjectives, adverbs, and verbs are generally pickier about flavor than the auxiliaries. For instance, *want* is restricted to bouletic modality, *permissible* to deontic modality, and *likely* to epistemic modality. Note in connection with the latter that, although epistemic modality is traditionally contrasted with **doxastic** modality (“given what is believed”), the term “epistemic modal” is widely used to describe modals for which it is unclear whether knowledge or mere belief is the best gloss. In keeping with common practice, I will not distinguish the two.

## 3.2 Classical and ordering-based semantics

### 3.2.1 Classical Modal Logic

The classic analysis of modality treats modal expressions essentially as restricted quantifiers over possible worlds. The restriction is provided by an accessibility relation  $R$ , which comes in various types associated with the modal flavors just discussed. Certain expressions, e.g. *want*, are lexically associated with one or several accessibility relations – in the case of *want*,  $R$  must be bouletic – while others are freer, e.g. *can*, which can be associated with (at least) epistemic, deontic, or dynamic  $R$ .

Many modal expressions receive a plausible interpretation in classical modal logic as (implicit) existential or universal quantifiers over accessible worlds. Letting  $R(w)$  be the worlds that are accessible from  $w$ , we define:

- (3.5) a.  $\llbracket \text{necessarily/must } \phi \rrbracket^{\mathcal{M},w} = 1$  iff for all  $w' \in R(w)$ :  $\llbracket \phi \rrbracket^{\mathcal{M},w'} = 1$ .  
 b.  $\llbracket \text{possibly/might } \phi \rrbracket^{\mathcal{M},w} = 1$  iff for some  $w' \in R(w)$ :  $\llbracket \phi \rrbracket^{\mathcal{M},w'} = 1$ .  
 c.  $\llbracket \phi \text{ is impossible} \rrbracket^{\mathcal{M},w} = 1$  iff for no  $w' \in R(w)$ :  $\llbracket \phi \rrbracket^{\mathcal{M},w'} = 1$ .

Although this approach works reasonably well for expressions with modal force at the “extremes”, such as *must*, *might*, *possible*, and *impossible*, it is more difficult to apply to intermediate grades of modality or to comparative modalities (Kratzer 1991). For example, consider the intermediate modality *probable*, and comparative modalities such as *It is (morally) better that  $\phi$  than it is that  $\psi$* . Since the set of accessible worlds is unordered, the best we seem to be able to do in classical modal logic is:

- (3.6)  $\llbracket \phi \text{ is probable} \rrbracket^{\mathcal{M},w} = 1$  iff, for the relevant epistemic accessibility relation  $R$ , there are more worlds  $w'$  such that  $wRw'$  and  $\llbracket \phi \rrbracket^{\mathcal{M},w'} = 1$  than there are worlds  $w''$  such that  $wRw''$  and  $\llbracket \neg\phi \rrbracket^{\mathcal{M},w''} = 1$ .
- (3.7)  $\llbracket \phi \text{ is better than } \psi \rrbracket^{\mathcal{M},w} = 1$  iff, for the relevant deontic accessibility relation  $R$ , there are more worlds  $w'$  such that  $wRw'$  and  $\llbracket \phi \rrbracket^{\mathcal{M},w'} = 1$  than there are worlds  $w''$  such that  $wRw''$  and  $\llbracket \psi \rrbracket^{\mathcal{M},w''} = 1$ .

These interpretations are not very promising, though. If there happen to be an infinite number of  $\phi$ - and  $\psi$ -worlds, both of these sentence-types will come out as trivially false no matter what. Even putting this technical issue to the side, it seems unlikely that counting worlds would give us the right truth-conditions: for instance, the question of whether  $\phi$  is morally better than  $\psi$  has no obvious connection to the *number* of worlds that instantiate these two propositions.

A further problem with the use of standard tools of modal logic here is that the truth-conditions of complex expressions are generally stipulated in the meta-language rather than being derived compositionally. This is particularly damaging in the case of  $\phi$  is *better than*  $\psi$ , a comparative sentence: presumably, the truth-conditions of this sentence ought to be derived using the same formal apparatus that we used to treat non-modal comparatives such as *John is happier than Mary* above. A related problem involves intermediate grades of modality with degree modifiers such as *It is somewhat probable that*  $\phi$  and  $\phi$  is *much better than*  $\psi$ . Such sentences are transparently related to complex degree expressions such as *Mary is somewhat happy* and *Sue is much funnier than Bill*, which have a compositional interpretation in a degree- or delineation-based theory of gradability and comparison. Presumably a compositional interpretation is needed for their modal counterparts as well.

### 3.3 Lewis on graded modality

#### 3.3.1 The basic proposal

In his book *Counterfactuals*, Lewis (1973) presents a semantics for counterfactual sentences and describes a number of extensions, including an extension to graded modality. This effort is historically important, since it provides the first extended effort to give a formal semantics for graded modals. It has also been very influential on subsequent theorizing about modals, particularly because of its influence on the later work of Kratzer (1981, 1991).

Lewis gives several equivalent formulations of his basic semantics for counterfactuals. One of these (see his §2.3) involves a qualitative ordering  $\succsim_w$  of some set  $S$  of accessible worlds in terms of “comparative similarity” to the evaluation world  $w$ , where  $w$  is in  $S$ . (Lewis’ symbol was actually closer to  $\leq_w$ , but the symbol used here is more intuitive for the purposes that we will put it to later on.) Lewis assumes that the similarity ordering is a **weak order**: that is, it must be

- reflexive: for all  $w' \in S$ ,  $w' \succsim_w w'$ ;
- transitive: for all  $w_1, w_2, w_3 \in S$ , if  $w_1 \succsim_w w_2$  and  $w_2 \succsim_w w_3$ , then  $w_1 \succsim_w w_3$ ;
- connected: for all  $w_1, w_2 \in S$ , either  $w_1 \succsim_w w_2$  or  $w_2 \succsim_w w_1$ .

In addition, Lewis requires that the ordering be centered, so that every world is more similar to itself than any other world is: for any distinct  $w, w'$ ,  $w >_w w'$ .

Lewis also presents an equivalent semantics in terms of a world-relative “Comparative Possibility” relation  $\geq^{sw}$  on propositions, rather than worlds. The key assumptions are that  $\geq^{sw}$  is a weak order, and also that it obeys the following conditions. (The first two labels are my inventions.)

- **Maximality:** If  $A$  is a proposition and  $\mathcal{B}$  is a set of propositions, then  $A \geq^{sw} \bigcup \mathcal{B}$  if and only if  $A \geq^{sw} B$  for all  $B \in \mathcal{B}$ . (Lewis: “The union of a set of propositions is the greatest [upper] bound of the set.”)
- **Size-insensitivity:** if  $\{w_1\} >^{sw} A$  for all  $A$  in a set of propositions  $\mathcal{A}$ , then  $\{w_1\} >^{sw} \bigcup \mathcal{A}$ .
- **Centering:** if  $w \in A$ ,  $w \in B$ , and  $w \notin C$ , then  $A \approx^{sw} B$  and  $A >^{sw} C$ . (Lewis: “All and only truths are maximally possible.”)

Lewis’ specific proposals about how to utilize these structural assumptions in the analysis of counterfactuals will not concern us here. However, the structures themselves are of considerable interest because they recur in a nearly-identical form in subsequent work, including Kratzer’s account of modals of all stripes. The story begins in chapter 5 of Lewis 1973, where Lewis discusses an application of the structures just described in the analysis of certain deontic conditionals—those of the form *If  $\phi$ , ought  $\psi$* . Deontic conditionals have long been discussed as a challenge for theories of modality. Lewis points out that some important problems in this literature could be alleviated if we interpret them relative to an ordering which is closely related to his semantics for counterfactuals. First, we re-interpret  $\geq^{sw}$  as “at least as good as”, and we drop the centering axiom (since it would be absurd to suppose that the actual world is always deontically optimal). Second, we assign *ought*-conditionals the following truth-conditions:

$$(3.8) \quad \llbracket \text{If } \phi, \text{ ought } \psi \rrbracket^{\mathcal{M}, w} = 1 \text{ iff } \{w' \mid \llbracket \phi \wedge \psi \rrbracket^{\mathcal{M}, w'} = 1\} >^{sw} \{w' \mid \llbracket \phi \wedge \neg \psi \rrbracket^{\mathcal{M}, w'} = 1\}.$$

For example, the proposed analysis would have it that *If Mary robs a bank, she ought to go to jail* is true just in case it is better that Mary robs a bank and goes to jail than it is that Mary robs a bank and does not go to jail. More eloquently, on the supposition that Mary does indeed rob a bank, it is better that she go to jail than not. The main attraction of this analysis is that the sentence can be true even when *Mary ought to rob a bank* is false: the conditional version only compares alternative propositions which make the antecedent true, without evaluating the goodness of the antecedent directly.

We will have more occasion to discuss deontic conditionals in chapters 7-8. For the moment, it will be more useful to focus on another, more general application of Comparative Possibility to which Lewis’ proposal for deontic conditionals points: deontic comparatives. In the section where he introduces a semantics for deontic conditionals (§5.1), Lewis points out that the same Comparative Possibility relation could be used to give a semantics for deontic comparatives (again, with the centering axiom removed).

$$(3.9) \quad \begin{array}{ll} \text{a. } \llbracket \phi \text{ is as good as } \psi \rrbracket^{\mathcal{M}, w} = 1 & \text{iff } \{w' \mid \llbracket \phi \rrbracket^{\mathcal{M}, w'} = 1\} \geq^{sw} \{w' \mid \llbracket \psi \rrbracket^{\mathcal{M}, w'} = 1\}. \\ \text{b. } \llbracket \phi \text{ is better than } \psi \rrbracket^{\mathcal{M}, w} = 1 & \text{iff } \{w' \mid \llbracket \phi \rrbracket^{\mathcal{M}, w'} = 1\} >^{sw} \{w' \mid \llbracket \psi \rrbracket^{\mathcal{M}, w'} = 1\}. \end{array}$$



### 3.3.2 Aside: Convention for interpretation of Greek sentence letters

Since the sets of worlds referred to in (3.9) are predictable in form and cumbersome to write out, let's define an abbreviatory convention for use here and in the rest of the book.

- (3.10) In contexts in which a proposition (a semantic object) is clearly called for,  $\phi$ —which technically serves as a placeholder for sentences *qua* expressions—can also be used to denote the set of worlds in which a sentence is true relative to model  $\mathcal{M}$ .

$$“\phi” \rightsquigarrow \left\{ \begin{array}{l} \text{the expression } \phi, \text{ or} \\ \text{the proposition } \{w \mid \llbracket \phi \rrbracket^{\mathcal{M},w} = 1\} \end{array} \right\}, \text{ whichever makes sense in context.}$$

A similar convention holds for other Greek letters  $\psi, \psi', \dots, \psi_i, \dots, \chi, \chi', \dots, \chi_i, \dots$

I will use this convention throughout the rest of the book; hopefully no confusion will result. It allows us to write out Lewis' proposal for deontic comparatives in a more perspicuous form:

- (3.11) a.  $\llbracket \phi \text{ is as good as } \psi \rrbracket^{\mathcal{M},w} = 1$  iff  $\phi \geq^{s_w} \psi$ .  
b.  $\llbracket \phi \text{ is better than } \psi \rrbracket^{\mathcal{M},w} = 1$  iff  $\phi >^{s_w} \psi$ .

### 3.3.3 Measurement-theoretic considerations

Measurement Theory provides a useful way to analyze the suitability of Lewis' proposal as a general semantics for deontic comparatives.  $\mathcal{S}_{good}$ , the goodness scale, is an ordinal scale whose basic binary relation is  $\geq_{good}$  (i.e., Lewis'  $\geq^{s_w}$ ). The latter orders (some subset of)  $\mathcal{P}(W)$ , the set of all propositions, in terms of their relative goodness. Since  $\geq_{good}$  is a weak order, a goodness scale  $\mathcal{S}_{good}$  with this structure is an **ordinal scale** (§2.2.1). The measure functions  $\mu_{good}$  that are admissible for this scale are all and only those which agree with  $\geq_{good}$  on the relative ordering of propositions:  $\mu_{good}$  is admissible only if  $\mu_{good}(\phi) \geq \mu_{good}(\psi)$  whenever  $\phi \geq_{good} \psi$ .

Using these familiar tools from chapter 2, Lewis' proposal can be converted into a compositional degree semantics using the supervaluation method described in §2.3.

$$(3.12) \quad \llbracket \phi \text{ is as good as } \psi \rrbracket^{\mathcal{M},w} \left\{ \begin{array}{l} = 1 \text{ if } \mu_{good}(\phi) \geq \mu_{good}(\psi) \text{ for all } \mathcal{S}_{good}\text{-admissible } \mu_{good}; \\ = 0 \text{ if } \mu_{good}(\phi) \geq \mu_{good}(\psi) \text{ for no } \mathcal{S}_{good}\text{-admissible } \mu_{good}; \\ \text{is undefined otherwise.} \end{array} \right.$$

$$(3.13) \quad \llbracket \phi \text{ is better than } \psi \rrbracket^{\mathcal{M},w} \left\{ \begin{array}{l} = 1 \text{ if } \mu_{good}(\phi) > \mu_{good}(\psi) \text{ for all } \mathcal{S}_{good}\text{-admissible } \mu_{good}; \\ = 0 \text{ if } \mu_{good}(\phi) > \mu_{good}(\psi) \text{ for no } \mathcal{S}_{good}\text{-admissible } \mu_{good}; \\ \text{is undefined otherwise.} \end{array} \right.$$

For now-familiar reasons, this interpretive procedure yields the same result as the direct interpretation that Lewis gives:  $\llbracket \phi \text{ is better than } \psi \rrbracket^{\mathcal{M},w}$  will hold iff  $\phi >_{good} \psi$ . Since Lewis assumes that comparative goodness is a connected scale, *good*-comparatives and -equatives will never be undefined. Monotonic transformations are required to preserve the basic ordering, and so all admissible measure functions will agree on the relative ordering of any two propositions' degrees of goodness.

However, as discussed in §2.2.1, ordinal scales do not contain any quantitative information. The only quantitative relationships between objects that are invariant with respect to an ordinal scale are weak comparisons (“at least as”) and relations which can be built up using Boolean combinations of them, such as equivalence relations (“exactly as”) and strict comparisons (“more than”). The choice to associate comparative goodness with an ordinal scale thus leads to a non-trivial empirical prediction: deontic comparatives will be uninterpretable if they make reference to any quantitative information, no matter how weak. (Lewis himself makes the same point regarding his formally closely related proposal for counterfactuals: see his §2.4.) Whatever the situation is for counterfactuals, this is enough to show that Lewis’ proposal is not quite right for deontic comparatives. After all, (3.14b) is clearly meaningful, and it does not mean the same as (3.14a).

- (3.14) a.  $\phi$  is better than  $\psi$ .  
b.  $\phi$  is much better than  $\psi$ .

Here are some naturalistic examples:

- (3.15) a. I try that old wheeze where you say to yourself something like: “It is much better to have done something and regretted it than not done something and then wondered for the rest of your life, blah, blah, blah.” You know what? That saying is a crock ...  
b. It is far better to grasp the universe as it really is than to persist in delusion, however satisfying and reassuring. (Carl Sagan)

We saw already in chapter 2 (§2.2.3) that statements about differences are uninterpretable in the RTM sense if the underlying scale is ordinal. The truth-conditions of differentials refer to differences between degrees, and these are stable quantities only if the scale is at least as strong as an interval scale. This means that differential comparatives involving *good* should be uninterpretable if the relevant scales are ordinal or weaker. This is clearly incorrect: the differential comparative in (3.14b) requires a certain (vague, context-dependent) minimal difference in goodness between  $\phi$  and  $\psi$ , and Lewis’ semantics is not able to guarantee this.

In order to render (3.14b) and other comparisons of difference interpretable, Lewis’ assumptions about scale type must be strengthened. As we saw in chapter 2, the minimal assumption that is needed for difference statements like (3.14b) to come out as interpretable is that the underlying scale be an interval scale. For example, the interval scale of temperature supports comparisons of difference. *Atlanta is 10 degrees Celcius hotter than Chicago* is an interpretable statement which is true if and only if

$$[\mu_{hot}(\mathbf{Atlanta}) - \mu_{hot}(\mathbf{Chicago})] \geq \mathbf{10\ degrees\ Celcius}.$$

This statement is the same as a claim about ratios:

$$\frac{\mu_{hot}(\mathbf{Atlanta}) - \mu_{hot}(\mathbf{Chicago})}{\mathbf{10\ degrees\ Celcius}} \geq 1.$$

When we convert from Fahrenheit to Celsius, or employ any other positive affine transformation  $\mu'(x) = \alpha \times \mu(x) + \beta$ , the ratio on the left-hand side of this inequality will remain the same. As a result, the statement maintains its truth-value under all admissible transformations.

For similar reasons, *Atlanta is much hotter than Chicago* is interpretable because, however **much** is resolved, the ratio

$$\frac{\mu_{hot}(\text{Atlanta}) - \mu_{hot}(\text{Chicago})}{\text{much}}$$

is the same under all positive affine transformations, as long as the quantity picked out by **much** is transformed in the same way. The truth-condition—that this ratio is at least 1—is unaffected by the transformation. This fact about admissibility reflects the underlying model-theoretic fact that interval scales, unlike ordinal scales, contain information about the relative size of differences.

More generally, interval scales have a structure for which differences between degrees (e.g.,  $\mu_{good}(\phi) - \mu_{good}(\psi)$ ) are stable quantities, in the sense that the relative magnitude of two differences is not affected by scale transformation. Given that  *$\phi$  is much better than  $\psi$*  is clearly meaningful, then, it seems that the minimal assumption that we can make is that goodness is an interval scale. Of course, it is always possible that the goodness scale is even stronger, for example, that it is a ratio scale. In chapter 7 I will show that comparative goodness is indeed associated with an interval scale.

### 3.3.4 Maximality and disjunction

Lewis’ semantics for comparative goodness encodes a second set of problematic assumptions that go beyond the claim that the scale is merely ordinal, and shed light on the general issue of concatenation/join that was discussed in detail in chapter 2. Recall that Lewis crucially requires that the scale be **Maximal** and **Size-insensitive** as discussed in §3.3.1 above. The motivation for introducing these properties is not any deep commitment on Lewis’ part to their correctness; instead, the primary goal seems to be to keep the semantics of for comparative goodness discussed in chapter 5 of *Counterfactuals* as close as possible in logical terms to the similarity-based semantics for counterfactuals discussed in chapter 1-2 of Lewis’ book. Indeed, while introducing his semantics for comparative goodness, Lewis notes that the scale that he has just axiomatized probably does not correspond to any intuitive notion of goodness.

This is not instrumental or intrinsic betterness of any familiar form, but rather *maximax betterness*. Roughly, we are comparing  $\phi$ -at-its-best with  $\psi$ -at-its-best, and ignoring the non-best ways for  $\phi$  and  $\psi$  to hold. (Lewis 1973: 101)

Despite Lewis’ qualms, his proposal has (with variations) become the go-to semantics for graded deontics and indeed a variety of gradable and non-gradable modalities in subsequent literature. Why does Lewis express discomfort with its predictions, then?

The “maximax” character of this semantics is a direct effect of the **Maximality** and **Size-insensitivity** requirements. Lewis’ hesitation about the generality of these conditions suggests that he may have realized that they cause the semantics to encode some fairly strange inferential properties. Specifically, Lewis’ semantics has the feature that the position of a proposition in the ordering is determined exclusively by the positions of the highest-ranked *singletons* formed from worlds in the proposition. If we are considering whether it is better than not that Nader runs for president—which is, incidentally, equivalent to *Nader ought to run* according to many proposals—the only thing that is relevant in Lewis’ semantics is whether the set containing the *best*

worlds where Nader runs strictly outranks the set containing the *best* worlds where he does not. Nothing else matters, even intuitively relevant information such as the distribution of non-optimal worlds (are they almost all good, or almost all bad?) and the relative likelihood of ending up in a very good world vs. a very bad one. For example, if it is possible that Nader will win and it would be the best of all possible outcomes if he did, then *It is better than not for Nader to run* will come out true even if we know that we will probably end up in a nightmarish situation if he runs. (As we did.) Lewis' caveat is, I suggest, an oblique acknowledgement that **Maximality** and **Size-insensitivity** are not in general reasonable conditions to impose on a semantics for deontic comparisons: information about likelihoods and about non-maximal worlds should not be ignored.

Another gloss on this theory's insensitivity to non-maximal possibilities is the following: Lewis is concerned to ensure that the following inference comes out valid. (Recall that  $\vee$  is equivalent in this setting to the operation of set union. This, in turn, is the realization of the concatenation operation in the Boolean domain of propositions, as long as the disjuncts are mutually exclusive: see §2.4.1.)

- (3.16)    a.  $\phi \approx^{sw} \psi$ .  
             b.  $\phi >^{sw} \chi$ .  
             c. Therefore,  $\phi \approx^{sw} (\psi \vee \chi)$ .

Intuitively, the goal is to ensure that the expansion of  $\psi$  to include some relatively remote possibilities (the  $\chi$ -worlds) cannot influence the outcome of the comparison.

Lewis apparently thought that (3.16) was a desirable inference pattern to enforce in a semantics for counterfactuals. Perhaps so, but it is far from obvious that this inference pattern is reasonable for comparative goodness: is the degree of goodness of a proposition, or an action, really fully determined by the *best* possible way that things could turn out if it holds/is performed? And should the inference in (3.16) hold regardless of the relative likelihood of  $\phi$ ,  $\psi$ , and  $\chi$ ? Despite the formal assumptions of his official proposal, Lewis seems to think not: at the very least, it should not be a logical impossibility for non-maximal possibilities to affect the result of a comparison. I agree, and I will argue in chapters 7-8 that various problems that have plagued theories of deontic semantics can be attributed to a widespread, mostly unnoticed acceptance of this validity. Sometimes, I will show, the expansion of a proposition to include additional non-maximal worlds can lead it to have a lower degree of goodness.

The problematic nature of the **Maximality** and **Size-insensitivity** principles becomes even clearer when we consider a hypothetical extension to epistemics. (Note: Lewis does not propose such an extension.) The validity of the inference in (3.16) on Lewis' proposal leads straight to a validity that I will call the **Disjunctive Inference**.

- (3.17)    a.  $\phi \geq^{sw} \psi$ .  
             b.  $\phi \geq^{sw} \chi$ .  
             c. Therefore,  $\phi \geq^{sw} (\psi \vee \chi)$ .

The intuition behind the validity is as follows: since only maximal possibilities are relevant to the comparison, the position of  $\psi \vee \chi$  in  $\geq^{sw}$  will be fully determined by either the position of  $\psi$  or the position of  $\chi$ , whichever is greater. But since  $\phi$  is ranked at least as high as both, it will also

be ranked at least as high as their disjunction, since the disjunction has a rank equal to that of the higher-ranked disjunct.

In this case, the deontic version of the inference seems unproblematic. (Assume here that *A*, *B*, and *C* are mutually exclusive options.)

- (3.18) a. It's as good for Bill to perform action *A* is as it is for him to perform *B*.  
b. It's as good for Bill to perform *A* is as it is for him to perform *C*.  
c. Therefore, it's as good for Bill to perform *A* is as it is for him to either perform *B* or perform *C*.

However, if we were to extend Lewis' semantics to epistemic comparisons, the result is clearly not a valid inference:

- (3.19) a. It's as likely that Bill will perform action *A* is as it is that he will perform *B*.  
b. It's as likely that Bill will perform *A* is as it is that he will perform *C*.  
c. Therefore, it's as likely that Bill will perform action *A* as it is that he will perform either *B* or perform *C*.

After all, if *A* is just a bit more likely than *B*, and it's just a bit more likely than *C*, it could easily be *less* likely that Bill will perform *A* will hold than it is either that he'll perform *B* or that he'll perform *C*. Even worse, if *A*, *B*, and *C* are all equally likely, then *A* will *always* be less likely than the disjunction of *B* and *C*. For example, a single roll of a fair die is as likely to come up 1 as it is to come up 2, and it is as likely to come up 1 as it is to come up 3. However, it is clearly *not* as likely to come up 1 as it is to come up either 2 or 3: the latter event is *exactly twice as likely*.

In this hypothetical extensions of Lewis' semantics to epistemic comparisons, the validity of (3.19) is a consequence of the **Maximality** and **Size-insensitivity** assumptions. The fact that this inference is clearly not valid indicates that a semantics for epistemic comparatives should not encode these properties or anything with the same effect.

### 3.4 Kratzer on graded modality

We have devoted much space to a discussion of Lewis' Comparative Possibility-based semantics for graded modals because a variant of this analysis has become the "standard" analysis of graded modality in much subsequent work in linguistic and philosophical semantics. This is due to a considerable extent, though not exclusively, to the influence of Kratzer (1981; 1991). Kratzer's theory is an impressive effort to unify the semantics of epistemic, deontic, and circumstantial modals, both graded and non-graded. However, as a theory of graded modality it inherits the limitations of Lewis' framework just discussed. The qualitative structures that Kratzer invokes are even weaker than Lewis' ordinal scales, with the result that even very weak quantitative statements remain uninterpretable. In addition, Kratzer's unification of the formal structures underlying graded deontic and epistemic modality has the undesirable consequence that they must pattern together with respect to inferences like (3.18) and (3.19). Kratzer's semantics, like Lewis', renders these inferences valid in both cases, including in the problematic epistemic case.

### 3.4.1 Premise sets and world-orderings

On face, Kratzer's proposal looks rather different from Lewis'. Lewis (at least in one version of his semantics) assumes as a primitive a function that provides each world  $w$  with a weak ordering  $\succsim_w$  on propositions. Kratzer proposes instead to derive a related ordering by rule from two contextual parameters, each of which determines a *premise set* for a given evaluation world  $w$ . The *modal base*  $\mathbf{f}$  fixes a set of propositions  $\mathbf{f}(w)$  representing the factual background relevant to the evaluation of a modal statement at a world  $w$ . The set of worlds where all of the facts in question are true,  $\cap \mathbf{f}(w)$ , plays a role analogous to the "accessible worlds"  $R(w)$  of classical modal logic. (Convention: when the modal base is empty— $\mathbf{f}(w) = \emptyset$ —we fix  $\cap \mathbf{f}(w) = W$ . This is to avoid the cumbersome locution " $\cap[\mathbf{f}(w) \cup \{W\}]$ ".)

The *ordering source*  $\mathbf{g}$  supplies an evaluation world  $w$  with a set of propositions  $\mathbf{g}(w)$ —intuitively: goals, norms, or expectations—that are used to induce an ordering over the possible worlds by the following rule.

$$(3.20) \quad u \succsim_{\mathbf{g}(w)} v \text{ iff } \{\phi \mid \phi \in \mathbf{g}(w) \wedge v \in \phi\} \subseteq \{\psi \mid \psi \in \mathbf{g}(w) \wedge u \in \psi\}$$

That is,  $u$  is at least as good/normal/etc. as  $v$  iff  $u$  satisfies every ordering source proposition (law, norm, etc.) that  $v$  does.  $u$  is strictly better than  $v$ — $u \succ_{\mathbf{g}(w)} v$ —iff  $u$  satisfies every ordering source proposition that  $v$  does, and some more besides. This relation is a **preorder**—reflexive and transitive—but it is not in general connected, because in many cases the norms that  $w'$  satisfies will not be a subset of those that  $w''$  satisfies, nor vice versa. (Note that when the ordering source is empty,  $\succsim_{\mathbf{g}(w)}$  is  $W \times W$ , the universal relation on worlds.)

For a concrete example, suppose that we are in world  $w$ , where there are three relevant norms, and that the following descriptions hold of the 10 worlds in  $W$ .

(3.21) NORMS:

N1. Children obey their parents. ( $= \{w_1, w_4, w_5, w_8, w_9\}$ )

N2. There is no trespassing. ( $= \{w_1, w_2, w_3, w_5, w_7\}$ )

N3. There is no murder. ( $= \{w_1, w_2, w_3, w_4, w_6\}$ )

(3.22)  $w_1$ : N1-N3 are obeyed.

$w_2, w_3$ : Only N1 violated.

$w_4$ : Only N2 violated.

$w_5$ : Only N3 violated.

$w_6$ : Only N1 and N2 violated.

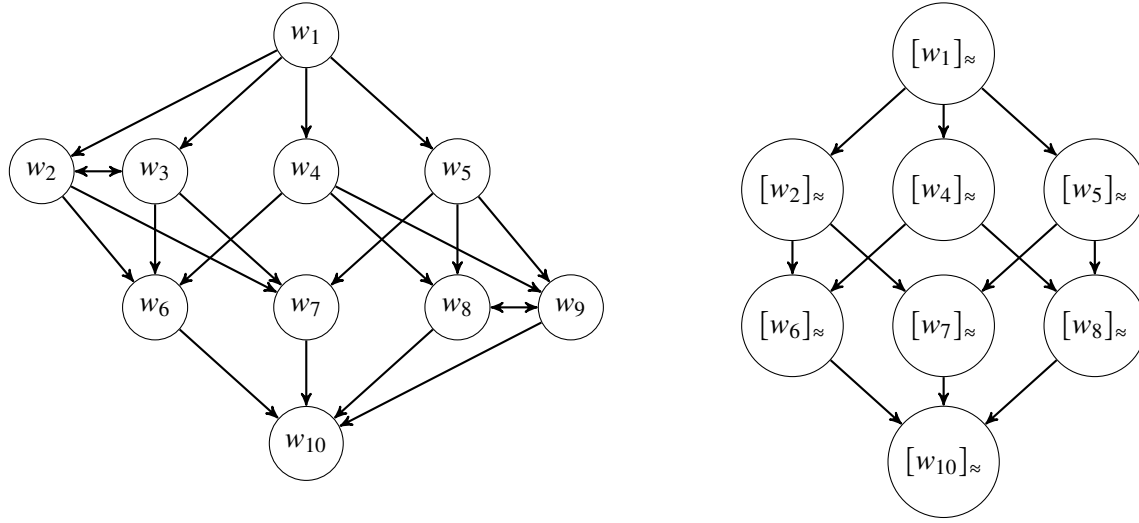
$w_7$ : Only N1 and N3 violated.

$w_8, w_9$ : Only N2 and N3 violated.

$w_{10}$ : N1, N2, and N3 all violated.

Then we are dealing with a structure  $\langle W, \succsim_{\mathbf{g}(w)} \rangle$  is depicted as in Figure 3.1.

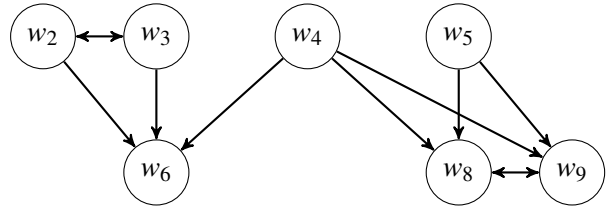
In cases like this in which the propositions in  $\mathbf{g}(w)$  are logically independent, the structure  $\langle W, \succsim_{\mathbf{g}(w)} \rangle$  is a **pre-Boolean algebra** (like a Boolean algebra, but allowing ties). Its reduction to equivalence classes (see §2.1) is a Boolean algebra, essentially because  $\succsim_{\mathbf{g}(w)}$  is defined in terms of the subset relation.



**Figure 3.1** Left: Kratzerian world-ordering corresponding to (3.21) and (3.22). Right: its reduction to equivalence classes. (Reflexive and transitive arrows are left implicit to avoid clutter.)

Kratzer takes care to construct the preorder  $\succsim_{\mathbf{g}(w)}$  so that it also defined when the ordering source propositions are inconsistent—intuitively, when there is some conflict between goals/norms/expectations/etc. For instance, imagine that the norms are the same as in the above example, but someone’s parent has instructed them to commit murder. In this case there is no possibility of violating no norms, and so there is no top-ranked world. If there is only one relevant way to obey parents and one way to commit murder, then it is impossible to fulfill either all or none of the norms, and we must make do with an impoverished world-ordering.

- (3.23)  $w_2, w_3$ : Only N1 violated.  
 $w_4$ : Only N2 violated.  
 $w_5$ : Only N3 violated.  
 $w_6$ : Only N1 and N2 violated.  
 $w_8, w_9$ : Only N2 and N3 violated.

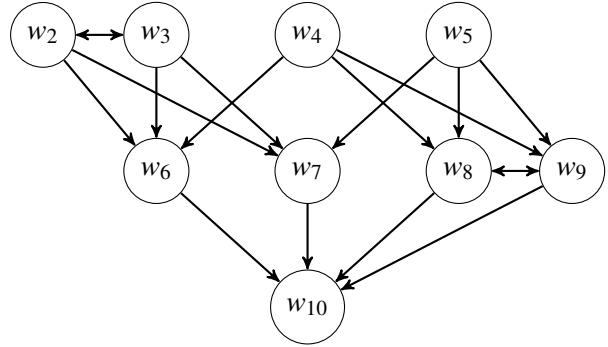


If the propositions in  $\mathbf{g}(w)$  are consistent, the “best” worlds are those where all of the ordering source propositions are simultaneously fulfilled— $\bigcap \mathbf{g}(w)$ . In the first example, where the norms were consistent, the set of best worlds was  $\{w_1\}$ . In the example at hand, where  $\bigcap \mathbf{g}(w)$  is empty, the best worlds are the worlds which are maximal in their respective branches:  $\{w_2, w_3, w_4, w_5\}$ .

If there are other ways of disobeying parents and committing murder which render these norms logically independent, however, then it is impossible to fulfill all of the norms, but it is possible to fulfill none of them. (You could obey your parent’s order to commit murder, but disobey some other order from a parent.) This means that the ordering is impoverished only in that there is no

top-ranked world.

- (3.24)  $w_2, w_3$ : N1 violated.  
 $w_4$ : N2 violated.  
 $w_5$ : N3 violated.  
 $w_6$ : N1 and N2 violated.  
 $w_7$ : N1 and N3 violated.  
 $w_8, w_9$ : N2 and N3 violated.  
 $w_{10}$ : N1, N2, and N3 violated.



Here, the best worlds are still  $\{w_2, w_3, w_4, w_5\}$ , but there is additional structure below which may be relevant to the truth-conditions of certain modal statements.

Kratzer's method of ordering worlds bears a close relationship to the version of Lewis' semantics for counterfactuals which was defined in terms of a world-ordering. In fact, Lewis (1981) proves the following correspondence. Suppose that we take a function from worlds to world-orderings as our contextual parameter, as Lewis does, but drop the assumption that orderings are centered and connected. Since we have dropped the connectedness assumption, Lewis'  $\succsim_w$  is no longer a weak order (reflexive, transitive, and connected), but a mere preorder (reflexive and transitive). It has the same formal structure as Kratzer's  $\succsim_{\mathbf{g}(w)}$ .

What Lewis proved is that Kratzer's premise semantics and his ordering semantics are expressively equivalent, in the sense that for every choice of  $\mathbf{g}(w)$  there is an equivalent world-ordering, and for every world-ordering there are many premise sets which generate that world-ordering by the rule in (3.20). This means, as Lewis puts it, that "[f]ormally, there is nothing to choose". Like the choice whether to use a semantics based on degrees or qualitative orders, the choice to begin with a world-ordering or to construct one using premise sets is primarily a matter of convenience.

(It is not hard to see how the proof goes: suppose we start with a Kratzerian premise set. The corresponding Lewisian world-ordering is simply Kratzer's  $\succsim_{\mathbf{g}(w)}$ . Conversely, suppose that we start with a preorder  $\succsim_w$ . An equivalent premise set is the set of propositions

$$\{\{w'' \mid w'' \succsim_w w'\} \mid w' \in W\},$$

which collects up, for each possible world  $w'$ , the proposition true of all and only worlds that are ranked at least as high in  $\succsim_w$  as  $w'$ . For example, in the first example of a Kratzerian  $\succsim_{\mathbf{g}(w)}$  above, this construction would give us a premise set like this (starting from the top):  $\{\{w_1\}, \{w_1, w_2, w_3\}, \{w_1, w_4\}, \{w_1, w_5\}, \{w_1, w_2, w_3, w_4, w_6\}, \dots, \{w_1, \dots, w_{10}\}\}$ . We could then re-induce the original ordering from this premise set using (3.20).)

Given this correspondence, we might expect that the problems that were discussed for Lewis' semantics for graded modals in the previous section should carry over to Kratzer's semantics, and they do. But before going through this, it is worth pausing to discuss Kratzer's significant additions to this semantic framework, involving (a) the lexical semantics of graded and non-graded modals and (b) the compositional interpretation of conditionals.



### 3.4.2 Modal comparisons

Kratzer's (1981; 1991) semantics represents the first sustained effort to situate graded modalities within a broader semantics for modals, and to articulate a lexical semantics for a wide variety of simple and complex modal expressions which captures logical relations between them. The key to the latter effort is to interpret all modals relative to a fixed set of related structures. The crucial components are the preorder on worlds  $\succsim_{\mathbf{g}(w)}$  defined above, and a corresponding preorder on propositions  $\succsim_{\mathbf{g}(w)}^s$  which is defined in terms of the world-ordering. Following Lewis, Kratzer labels the proposition-ordering "Comparative Possibility".

(3.25) COMPARATIVE POSSIBILITY:

$\phi$  is at least as good a possibility as  $\psi$  (in  $w$ , relative to  $\mathbf{f}$  and  $\mathbf{g}$ ) if and only if:

For all  $u \in \cap \mathbf{f}(w)$ : if  $u \in \psi$ , then there is a world  $v \in \cap \mathbf{f}(w)$  such that  $v \succsim_{\mathbf{g}(w)} u$  and  $v \in \phi$ .

In other words,  $\phi$  is at least as good/likely/desirable/etc. as  $\psi$  if and only if every  $\psi$ -world in the modal base is weakly dominated by some  $\phi$ -world in the modal base. An equivalent condition is the requirement that there cannot be any  $\psi$ -worlds in the modal base which either (a) outrank, or (b) are not comparable with, all  $\phi$ -worlds. I will use the abbreviation  $\succsim_{\mathbf{g}(w)}^s$ , with  $s$  for "set", to pick out a relation on propositions "at least as good a possibility as" which is derived from the relation  $\succsim_{\mathbf{g}(w)}$  on worlds as in (3.25).

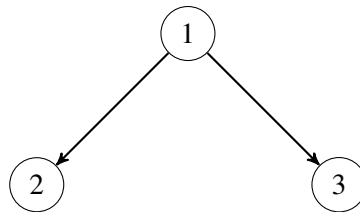
(3.26)  $\succsim_{\mathbf{g}(w)}^s =_{df} \{(\phi, \psi) \mid \forall u \in \psi \exists v \in \phi : v \succsim_{\mathbf{g}(w)} u\}$ , where  $u, v \in \cap \mathbf{f}(w)$ .

As usual, we can define "better possibility than" and "exactly as good a possibility as" from  $\succsim_{\mathbf{g}(w)}^s$ .

(3.27)  $\phi \succ_{\mathbf{g}(w)}^s \psi$  if and only if  $\phi \succsim_{\mathbf{g}(w)}^s \psi$  and  $\psi \not\succsim_{\mathbf{g}(w)}^s \phi$ . [“better possibility”]

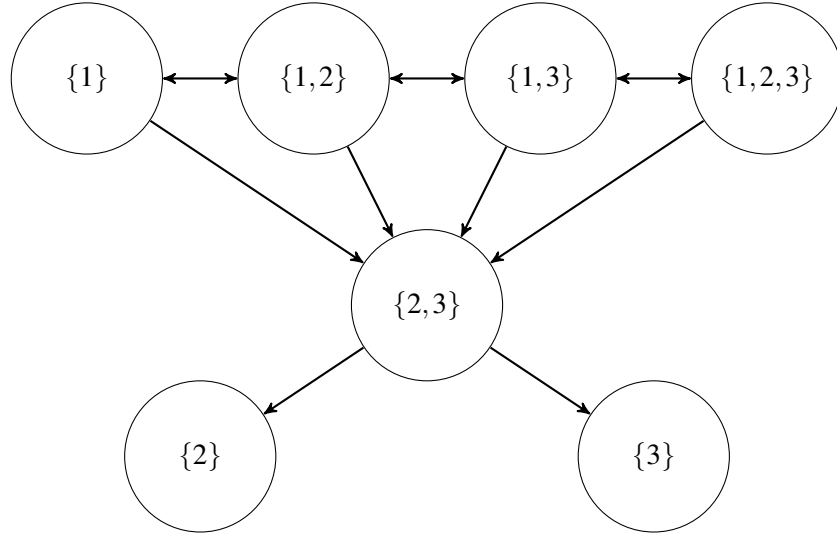
(3.28)  $\phi \approx_{\mathbf{g}(w)}^s \psi$  if and only if  $\phi \succsim_{\mathbf{g}(w)}^s \psi$  and  $\psi \succsim_{\mathbf{g}(w)}^s \phi$ . [“exactly as good a possibility”]

Kratzer's proposition-orderings inherit structural features from the world-ordering from which they are lifted: if  $\succsim_{\mathbf{g}(w)}$  is a mere preorder, then a Comparative Possibility relation defined from it will generally be a mere preorder as well. Figures 3.2 and 3.3 illustrate the lifting when domain of “accessible” worlds is  $\cap \mathbf{f}(w) = W = \{1, 2, 3\}$  and the ordering source is  $\mathbf{g}(w) = \{\{1\}, \{1, 2\}, \{1, 3\}\}$ . The corresponding lifting is an ordering on  $2^{|\cap \mathbf{f}(w)|} = 8$  propositions, with all propositions containing the “best” world (1) having the same position in the ordering. (Figure 3.3 omits the contradiction  $\emptyset$ .)



**Figure 3.2**

World-ordering with domain  $\{1, 2, 3\}$  and ordering source  $\{\{1\}, \{1, 2\}, \{1, 3\}\}$ .



**Figure 3.3** Comparative Possibility relation derived from world-ordering in Figure 3.2.

Note that in Figure 3.3 the unit set containing the “best” world,  $\{1\}$ , is exactly as good a possibility as the tautology  $W$ : that is,  $\{1\} \approx_{\mathbf{g}(w)}^s \{1,2,3\}$ . This is a reflection of the fact that Kratzer’s Comparative Possibility relation shares the **Maximality** and **Size-insensitivity** properties of Lewis’ relation of the same name. In general, whenever there are one or more “best” worlds in  $\cap \mathbf{f}(w)$ —i.e., ones that are at least as highly ranked in  $\geq_{\mathbf{g}(w)}^s$  as any other—the unit set containing any one of these worlds will be maximal in the Comparative Possibility relation, and so will be ranked as high as any arbitrary proposition, including the entire set of possible worlds. This will happen, for example, anytime the worlds in  $\mathbf{g}(w)$  are consistent or finite.

Also following Lewis (1973: §5), Kratzer proposes to use Comparative Possibility to provide interpretations for comparative modalities. The key difference is that, while Lewis restricts his proposal to deontic comparatives, Kratzer uses Comparative Possibility to interpret epistemic comparatives as well. A second, technical difference is that, rather than parametrizing the interpretation with a world- or proposition-ordering directly as Lewis does, Kratzer parametrizes it with a modal base  $\mathbf{f}$  and an ordering source  $\mathbf{g}$  which are used to induce these orderings.

- (3.29) a.  $\llbracket \phi \text{ is as likely/probable/good as } \psi \rrbracket_{\mathbf{f},\mathbf{g}}^{\mathcal{M},w} = 1 \text{ iff } \phi \geq_{\mathbf{g}(w)}^s \psi$   
 b.  $\llbracket \phi \text{ is more likely/probable/good than } \psi \rrbracket_{\mathbf{f},\mathbf{g}}^{\mathcal{M},w} = 1 \text{ iff } \phi >_{\mathbf{g}(w)}^s \psi$

As usual in Kratzer’s theory, epistemic and deontic interpretations have the same basic logic; differences are attributed to different conditions on the content of the modal base and ordering source, e.g., the requirement that the actual world be in  $\cap \mathbf{f}(w)$  for epistemic (as opposed to merely doxastic) interpretations.

The structure of the scale  $\langle \mathcal{P}(W), \geq_{\mathbf{g}(w)}^s \rangle$ —consisting only of a set and an ordering on the elements of the set—is reminiscent of the ordinal scales discussed in chapter 2. However, the scale induced by Kratzer’s Comparative Possibility relation is even weaker than an ordinal scale, since

$\succsim_{g(w)}^s$  is not in general connected. As a result, there will often be pairs of propositions which are modally incomparable. For example, the Comparative Possibility relation depicted in Figure 3.3 renders the propositions  $\{2\}$  and  $\{3\}$  incomparable: since the set of ordering source propositions satisfied by  $\{2\}$  is not a subset of those satisfied by  $\{3\}$ , nor vice versa. Given the proposed interpretation of comparatives in (3.29), then, both of the following will be false when this modal base and ordering source are operative.

- (3.30) a.  $\phi$  is as likely/good as  $\psi$ .  
 b.  $\psi$  is as likely/good as  $\phi$ .

Naturally, since  $\phi$  is *more likely than*  $\psi$  and  $\phi$  is *better than*  $\psi$  entail the corresponding equatives in (3.30), the strict comparisons will also be false in such models.

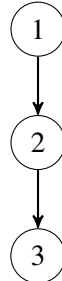
More generally, Kratzer's proposal predicts that, whenever there are rich, logically independent sets of norms—such as *No murder* and *Don't trespass*—there will be a considerable amount of modal incomparability. Some of the worlds in *Bill commits murder* violate exactly one norm, *Don't murder*. Some of the worlds in *Bill trespasses* violate exactly one norm, *Don't trespass*. These sets of worlds will not be comparable in  $\succsim_{g(w)}$ . As a result, neither *Bill commits murder* nor *Bill trespasses* is preferred by Comparative Possibility. Somewhat counter-intuitively, *It is better for Bill to trespass than to commit murder* comes out as false simply because the norms invoked in the ordering source are logically independent. This seems wrong: we want this comparison to come out true because murder is much worse than trespassing, but it is not obvious how to represent this information about relative priority of norms.

This might seem to be a problem for Kratzer's theory of modality, as I argued it was in Lassiter 2011a. But that conclusion was too hasty. We can ensure that *It is better for Bill to trespass than to commit murder* comes out true if we carefully select an ordering source in which the prohibitions against murder and trespassing are not logically independent. For example, instead of  $\{No\ murder, No\ trespassing\}$ , we could choose  $\{No\ murder\ and\ no\ trespassing, No\ murder\}$  (Silk 2012; Kratzer 2012).

One could object to this solution that it involves a heavy dose of reverse-engineering: rather than looking for constraints on premise sets and then checking whether they deliver intuitive results, we consult the intuitive results and work backward to the premise sets that are needed on the assumption that the theory is correct. Lewis (1981) showed that this strategy will always work for simple comparisons, but it seems to lack in motivation. Worse, it undermines the motivation for using premise sets to a considerable extent. If intuitions about comparisons are the primary datum, why not simply begin with information about such comparisons as Lewis does, rather than going through the detour of premise sets?

A resolution which has somewhat better conceptual motivation is to start with ordering sources in which certain norms are tagged as having priority over others—for example, by using a sequence of ordering sources rather than a single ordering source. Katz, Portner & Rubinstein (2012) suggest resolving the puzzle of ranked priorities by ranking propositions first on the basis of the status of their best worlds in relative to higher-priority norms (*No murder*), and then using lower-priority norms (*No trespassing*) only to break ties among worlds which agree on the truth-values of more important norms.

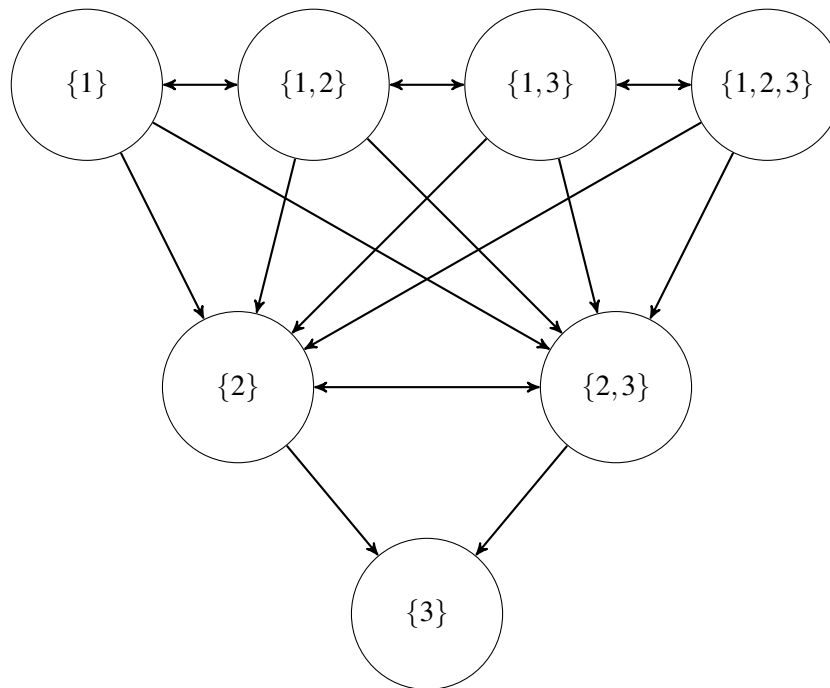
In any case, it is technically straightforward to construct a modal base and ordering source in which any world where a murder occurs is worse than a world with trespassing but no murder. For example, letting  $\cap \mathbf{f}(w) = W = \{1, 2, 3\}$  and  $\mathbf{g}(w) = \{\{1\}, \{1, 2\}\}$ , we obtain the world-ordering in Figure (3.3.4).



**Figure 3.4** World-ordering with domain  $\{1, 2, 3\}$  and ordering source  $\{\{1\}, \{1, 2\}\}$ .

---

As a result of the fact that the ordering source is nested—every proposition in it either entails or is entailed by every other—the Comparative Possibility relation induced by this modal base/ordering source combination is connected. Any two propositions are thus comparable, as in Figure 3.4.



**Figure 3.5** Comparative Possibility relation derived from world-ordering in Figure 3.4.

---

On an intuitive level, this ordering is what we would derive if the norms were not two logically

independent propositions—“No one commits murder” and “There is no trespassing”—but rather the logically nested propositions in (3.31).

- (3.31) a.  $\{1\}$  = No one commits murder and there is no trespassing.  
 b.  $\{1, 2\}$  = No one commits murder.

According to this ordering source, murder is worse than trespassing because the prohibition against murder occurs as a conjunct in all of the norms in which the prohibition against trespassing occurs, and some more besides. It has the desired feature that “Murder is worse than trespassing” comes out true.

### 3.4.3 Compositional interpretation using RTM

Kratzer’s proposals involving graded modality are unsatisfactory in one important respect: they are not compositional. Chunks of material, including non-constituents such as *is as likely as* and *is more probable than*, are interpreted as unanalyzed wholes, rather than being derived compositionally—as we would expect, given that we know how to do this for non-modal comparatives involving *happy*, *hot*, and the like.

However, as Portner (2009) suggests, it is not difficult to convert these proposals into a compositional form. The measurement-theoretic apparatus introduced in chapter 2 makes it straightforward to embed Kratzer’s interpretation of comparative modalities into a degree-based compositional semantics in the style of Kennedy (1997). Specifically, we can associate *likely*, *probable*, and *good* with measure functions taking propositions to degrees, and supervalue over degree scales that are admissible relative to the qualitative scale  $\langle \mathcal{P}(W), \succsim_{\mathbf{g}(w)}^s \rangle$  that is induced from a particular choice of modal base and ordering source. Here I sketch the derivation briefly, obscuring certain details that are syntactically important but less so semantically. (Notably, the preferred pronunciations of these comparisons include, for example, “It is more likely that  $\phi$  than it is that  $\psi$ ” and “VP-ing is more likely than VP’-ing”, the latter with implicit subjects for the VPs.)

- (3.32) a.  $\llbracket \text{likely} \rrbracket_{\mathbf{f}, \mathbf{g}}^{\mathcal{M}, w} = \lambda p_{\langle s, t \rangle} \cdot \mu_{\text{likely}}(p)$   
 b.  $\llbracket \text{probable} \rrbracket_{\mathbf{f}, \mathbf{g}}^{\mathcal{M}, w} = \lambda p_{\langle s, t \rangle} \cdot \mu_{\text{probable}}(p)$   
 c.  $\llbracket \text{good} \rrbracket_{\mathbf{f}, \mathbf{g}}^{\mathcal{M}, w} = \lambda p_{\langle s, t \rangle} \cdot \mu_{\text{good}}(p)$

We can then use the type-polymorphic denotation for *more* from chapter 1 (§1.3.1) to convert these into a function that compares the degree associated with the propositional argument of *likely/probable/good* to the degree denoted by the comparative clause. I illustrate using *likely*.

$$(3.33) \quad \llbracket \text{more/-er} \rrbracket_{\mathbf{f}, \mathbf{g}}^{\mathcal{M}, w} = \lambda K_{\langle \alpha, d \rangle} \lambda d_d \lambda k_{\alpha} . K(k) > d$$

$$(3.34) \quad \llbracket \text{more likely} \rrbracket_{\mathbf{f}, \mathbf{g}}^{\mathcal{M}, w} = \lambda d_d \lambda k_{\langle s, t \rangle} \cdot \mu_{\text{likely}}(k) > d$$

As usual, we assume that the comparative clause in (3.29b) contains an elided *is likely*, so that the material interpreted is *than  $\psi$  is likely*. Furthermore, the interpretation of the comparative clause involves some mechanism for abstracting out and maximizing over the set of degrees that make it true (see e.g. von Stechow 1984; Kennedy 1997; Heim 2001). The result is:

$$(3.35) \quad \llbracket \text{than } \psi \text{ (is likely)} \rrbracket_{\mathbf{f}, \mathbf{g}}^{\mathcal{M}, w} = \max(\lambda d'. \mu_{\text{likely}}(\psi) \geq d) = \mu_{\text{likely}}(\psi)$$

Combining this with (3.34) and then with the copula and subject, we have an interpretation for the full sentence.

$$(3.36) \quad \llbracket [\text{more likely}] [\text{than } \psi \text{ (is likely)}] \rrbracket_{\mathbf{f}, \mathbf{g}}^{\mathcal{M}, w} = \lambda k_{\langle s, t \rangle}. \mu_{\text{likely}}(k) > \mu_{\text{likely}}(\psi)$$

$$(3.37) \quad \llbracket \text{is} \rrbracket^{\mathcal{M}, w} = \lambda P_{\langle e, t \rangle} \lambda x_e. P(x)$$

$$(3.38) \quad \llbracket \text{is} [\text{more likely}] [\text{than } \psi \text{ (is likely)}] \rrbracket_{\mathbf{f}, \mathbf{g}}^{\mathcal{M}, w} = \llbracket [\text{more likely}] [\text{than } \psi \text{ (is likely)}] \rrbracket_{\mathbf{f}, \mathbf{g}}^{\mathcal{M}, w}$$

$$(3.39) \quad \llbracket \phi [\text{is} [\text{more likely}] [\text{than } \psi \text{ (is likely)}]] \rrbracket_{\mathbf{f}, \mathbf{g}}^{\mathcal{M}, w} = 1 \text{ iff } \mu_{\text{likely}}(\phi) > \mu_{\text{likely}}(\psi)$$

These truth-conditions make reference to measure functions, but the model does not provide a specific measure function associated with *likely*. Instead, it provides a qualitative scale  $\mathcal{S}_{\text{likely}} = \langle \mathcal{P}(W), \succ_{\mathbf{g}(w)}^s \rangle$  whose precise form is derived from the  $\mathbf{f}$  and  $\mathbf{g}$  parameters of the interpretation function as sketched above. The supervaluation procedure for interpreting measure functions given in chapter 2 gives us a final interpretation for this comparison.

$$(3.40) \quad \llbracket \phi \text{ is more likely than } \psi \rrbracket_{\mathbf{f}, \mathbf{g}}^{\mathcal{M}, w} \begin{cases} = 1 \text{ iff, for all } \mathcal{S}_{\text{likely}}\text{-admissible } \mu_{\text{likely}}, \mu_{\text{likely}}(\phi) > \mu_{\text{likely}}(\psi) \\ = 0 \text{ iff, for no } \mathcal{S}_{\text{likely}}\text{-admissible } \mu_{\text{likely}}, \mu_{\text{likely}}(\phi) > \mu_{\text{likely}}(\psi) \\ \text{is undefined otherwise} \end{cases}$$

As discussed in chapter 2,  $\mu_{\text{likely}}(\phi) > \mu_{\text{likely}}(\psi)$  will be true relative to all admissible  $\mu_{\text{likely}}$  if and only if  $\phi \succ_{\mathbf{g}(w)}^s \psi$ . Analogously, the interpretation of *phi is exactly as likely as psi* will be true if and only if they have the same degree according to all admissible  $\mu_{\text{likely}}$ , and this will hold if and only if  $\phi \approx_{\mathbf{g}(w)}^s \psi$ ; and so forth for the other comparative modalities. The result is an interpretation procedure which returns precisely Kratzer's proposals for the interpretation of comparative modalities, but does so in a strictly compositional manner.

This style of interpretation retains all of the qualitative information contained in the qualitative scale  $\mathcal{S}_{\text{likely}} = \langle \mathcal{P}(W), \succ_{\mathbf{g}(w)}^s \rangle$  while avoiding attributing to this scale any additional structure that might accidentally be present in some particular choice of  $\mu_{\text{likely}}$ .

### 3.4.4 Other graded modalities

Kratzer (1991) also suggests interpretations for a number of other graded modalities. For example, she proposes that *probably phi* is interpreted effectively as *phi is more likely than not phi*.

$$(3.41) \quad \llbracket \text{probably } \phi \rrbracket_{\mathbf{f}, \mathbf{g}}^{\mathcal{M}, w} = 1 \text{ iff } \phi \succ_{\mathbf{g}(w)}^s \neg \phi.$$

Another suggestion is to treat the intermediate modalities *slight possibility* and *good possibility* as follows.

$$(3.42) \quad \llbracket \text{There is a slight possibility that } \phi \rrbracket_{\mathbf{f}, \mathbf{g}}^{\mathcal{M}, w} = 1 \text{ iff}$$

- a.  $\phi \cap \cap \mathbf{f}(w) \neq \emptyset$  (i.e.,  $\phi$  is compatible with the modal base); and
- b.  $\neg \phi$  is probable in  $w$  (by the definition given above).

$$(3.43) \quad \llbracket \text{There is a good possibility that } \phi \rrbracket_{\mathbf{f}, \mathbf{g}}^{\mathcal{M}, w} = 1 \text{ iff } \llbracket \text{It is probable that } \neg \phi \rrbracket_{\mathbf{f}, \mathbf{g}}^{\mathcal{M}, w} = 0$$

Again, there is a lingering puzzle about how these interpretations can be derived compositionally. Ideally, the meaning of *slight possibility* should be derived using independently motivated compositional mechanisms from the meaning of *possibility* and *slight*, where the latter is interpreted in the same way as in *I have a slight headache*. Similarly, *a good possibility* should presumably be related to the intensifying use of *good* in modal and non-modal expressions such as (3.44).

- (3.44) a. There is a good chance that we will lose.  
 b. There were a good number of people at the game.

I will not pursue a compositional implementation of the specific constructions that Kratzer analyzes here. However, Kratzer's general point that comparison and degree modification constructions can be important sources of data for the semantics of modality will be crucial for us.

### 3.4.5 Lexical semantics of non-graded modals

Kratzer (1991) retains the core idea from classical modal logic that *must*, *necessarily*, etc. are universal quantifiers over worlds, and that *might*, *possibly*, etc. are existential quantifiers. However, rather than treating them as quantifiers over a set of worlds which is pragmatically determined once and for all, Kratzer analyzes these modals as quantifiers whose restriction is determined in a somewhat more complicated fashion by the modal base and ordering source. *Must* (and other strong modals, presumably) is defined as in (3.45).

$$(3.45) \quad \llbracket \text{must } \phi \rrbracket_{\mathbf{f}, \mathbf{g}}^{\mathcal{M}, w} = 1 \text{ iff } \forall u \exists v [v \succsim_{\mathbf{g}(w)} u \wedge \forall z : z \succsim_{\mathbf{g}(w)} v \rightarrow z \in \phi]. \quad (u, v, z \in \cap \mathbf{f}(w))$$

As Kratzer explains, the effect of (3.45) is that “a proposition is a necessity if and only if it is true in all accessible worlds which come closest to the ideal established by the ordering source”.

The definition is intended to be maximally general, but it is a bit obscure as stated. The complications are motivated by Kratzer's desire to avoid the “limit assumption”, i.e., the assumption that all branches of the ordering have maximal elements. However, finite ordering sources always satisfy the limit assumption, and many infinite ones do as well. If we simplify by making the limit assumption, (3.45) simplifies dramatically: now it is simply a universal quantifier over the set of “best” worlds, i.e., those which are undominated in  $\succsim_{\mathbf{g}(w)}$ .

$$(3.46) \quad \mathbf{BEST}(\mathbf{f}, \mathbf{g}, w) =_{df} \{v \mid v \in \cap \mathbf{f}(w) \wedge \neg \exists v' \in \cap \mathbf{f}(w) : v' \succ_{\mathbf{g}(w)} v\}$$

$$(3.47) \quad \llbracket \text{must } \phi \rrbracket_{\mathbf{f}, \mathbf{g}}^{\mathcal{M}, w} = 1 \text{ iff } \forall u : u \in \mathbf{BEST}(\mathbf{f}, \mathbf{g}, w) \rightarrow u \in \phi.$$

That is, *must*  $\phi$  is true iff  $\phi$  is true in all worlds which are maximal in their respective  $\succsim_{\mathbf{g}(w)}$ -branches. Viewed this way, Kratzer's definitions of *must* and other strong modals are quite close to those of classical modal logic.

Kratzer also follows classical modal logic in treating possibility as the dual of necessity: that is, a proposition is a possibility (etc.) if and only if its negation is not a necessity.

$$(3.48) \quad \llbracket \text{might } \phi \rrbracket_{\mathbf{f}, \mathbf{g}}^{\mathcal{M}, w} = 1 \text{ iff } \exists u : u \in \phi \wedge u \in \mathbf{BEST}(\mathbf{f}, \mathbf{g}, w).$$

### 3.4.6 Conditionals

Kratzer (1986, 1991) proposes an analysis of conditionals as restrictors of modals. This treatment of conditionals has become more or less standard in linguistics, and variants have seen recent popularity in philosophical work as well (e.g., Yalcin 2007; Kolodny & MacFarlane 2010; Egré & Cozic 2011; Rothschild 2011). The idea, in brief, is that the conditional is not a sentential connective but rather a device of domain restriction. That is, in a sentence *If  $\phi$  then  $\psi$* , the antecedent functions to restrict the domain of a modal expression contained in  $\psi$  to the worlds in which  $\phi$  holds.

$$(3.49) \quad \llbracket \text{If } \phi, \text{ then } \psi \rrbracket_{\mathbf{f}, \mathbf{g}}^{\mathcal{M}, w} = \llbracket \psi \rrbracket_{\mathbf{f}', \mathbf{g}}^{\mathcal{M}, w} \text{ where, for all } w, \mathbf{f}'(w) =_{df} \mathbf{f}(w) \cup \{w' \mid \llbracket \phi \rrbracket^{\mathcal{M}, w'} = 1\}.$$

Note that in order to make this theory work we must assume that there is always a modal in the consequent—otherwise, the antecedent would have no semantic effect. In bare conditionals, with no overt modal in the consequent, Kratzer assumes that there is a covert modal with a default epistemic interpretation.

In cases where the overt or covert modal is epistemic,  $\mathbf{f}(w)$  represents the set of propositions known at world  $w$ , and the analysis in (3.49) can be paraphrased along the following lines: to evaluate *If  $\phi$  then  $\psi$* , pretend that you know  $\phi$  and then evaluate  $\psi$  on the basis of that pretense. Since modal sentences' truth-conditions are sensitive only to facts about the position in  $\succsim_{\mathbf{g}(w)}$  of worlds which do satisfy all of the propositions in  $\mathbf{f}(w)$ , adding  $\phi$  to  $\mathbf{f}(w)$  will have the effect of rendering all non- $\phi$  worlds in  $\succsim_{\mathbf{g}(w)}$  irrelevant to the evaluation of the consequent.

This reasoning suggests a different way to state the restrictor theory of conditionals, which has the same semantic effect but does not rely on special features of Kratzer's theory of modality. Instead of interpreting a conditional by adding its antecedent temporarily to the modal base, we can treat the antecedent of the conditional as a restrictor of the binary order over worlds  $\succsim_{\mathbf{g}(w)}$  to worlds which satisfy the antecedent. First define the restriction operator  $\upharpoonright$  as:

$$(3.50) \quad \text{The **restriction** } \succsim \upharpoonright B \text{ of an order } \succsim \text{ to a set } B \text{ is defined as } \{(x, y) \mid x \succsim y \wedge x \in B \wedge y \in B\}.$$

Restricting a binary order to a set  $B$  means removing from the order any pair for which one or both members are not in  $B$ . We can now achieve the semantic effect of (3.49) in two steps. First, we relativize the interpretation of the sentence to a single parameter  $\mathbf{h}$  which, in the ordinary case, gives us an order equivalent to Kratzer's  $\succsim_{\mathbf{g}(w)}$  (though this time restricted to modal base-compatible worlds). Second, we define the conditional so that the antecedent restricts the order so derived for the purpose of evaluating the consequent.

$$(3.51) \quad \llbracket \text{If } \phi, \text{ then } \psi \rrbracket_{\mathbf{h}}^{\mathcal{M}, w} = \llbracket \psi \rrbracket_{\mathbf{h}'}^{\mathcal{M}, w}, \text{ where}$$

- a. For all  $w$ ,  $\mathbf{h}(w) =_{df} \succsim_{\mathbf{g}(w)} \upharpoonright \cap \mathbf{f}(w)$  as defined in (3.20) above.
- b. For all  $w$ ,  $\mathbf{h}'(w) =_{df} \mathbf{h}(w) \upharpoonright \{w' \mid \llbracket \phi \rrbracket_{\mathbf{h}}^{\mathcal{M}, w'} = 1\}.$

In (3.51) we allow the antecedent to restrict the order  $\succsim_{\mathbf{g}(w)}$  directly, rather than indirectly as in (3.49). In the context of Kratzer's theory of modality, these two approaches are equivalent modulo a few definitional tweaks.<sup>1</sup> However, (3.51) is more general because it can be applied to any theory of modality which makes use of a binary order over worlds, and not just to one which uses premise sets as an intermediary.

<sup>1</sup> In a handful of cases we have to change the definitions slightly in order to accommodate items for which Kratzer's



### 3.4.7 Logical relations among modal statements

Kratzer's work on modality has been very influential, and has had an major positive impact on the field by focusing attention on a number of issues that had been rather neglected. Kratzer stressed the importance of a unified semantics for modals, for example, by pointing out that there was strong cross-linguistic motivation for treating *must* and *may*, and their equivalents in other languages, as having unified meaning across different modal flavors. In Kratzer's theory, differences among epistemic, deontic, circumstantial, and other types of interpretations are induced by variation in the ordering source. The basic logical structure of all kinds of modals is the same modulo these parameters and the details of specific lexical entries.

Since there has been a tendency for theorists to focus on one kind of modality exclusively—or even on the interaction of one kind of modal with conditionals—Kratzer's work has had a salutary impact in its emphasis on constructing a general theory of modality in natural language, rather than taking a piecemeal approach. A second major impact of Kratzer's work has been to increase the amount of attention paid to graded and comparative modalities. While the literature continues to some extent to focus on a handful of items (the verbs *must*, *might*, *may*, *ought*, *should* and the adjectives *possible* and *necessary*), somewhat more attention is being paid to other modal items, to the logical relations between modal expressions, and to the compositional interaction of modals with conditionals.

Despite these undeniable attractions, I do not adopt Kratzer's theory of modality in this book, and I will not use it as a starting point for my own proposals (except as general inspiration). There are two primary reasons for this choice. First, the general prediction that all modalities have the same basic logic is falsified by the fact that epistemic and deontic comparatives differ in the inferences that they validate (as we saw already in §3.3.4). In fact, I will argue that the Kratzer's theory makes incorrect predictions about the inferences licensed by both epistemic and deontic comparatives—but for different reasons in each case. Second, the scalar structures that undergird Kratzer's semantics, when analyzed in light of the measurement-theoretic background of this book, do not provide sufficient structure to allow us to interpret quantitative modal language.

These problems are, of course, very closely related to the issues discussed above for Lewis' semantics for deontic comparatives. This is unsurprising given that Kratzer's theory is an adaptation of Lewis'. However, one of Kratzer's primary innovations—the proposal to extend Lewis' semantic proposal to include epistemic and other modalities—is also the theory's downfall. While the logical problems with Kratzer's and related semantic proposals are fairly subtle in the deontic case (see chapter 7), the problems involving epistemic modals are plain and very difficult to explain away. Furthermore, as we will discuss in detail in chapters 4-5, epistemic adjectives such as *likely* and *certain* support a much richer variety of quantitative language than deontic adjectives such as *good* do. There, we will see that measurement-theoretic considerations on this quantitative language point toward different scale types for the two types of modals.

---

official proposals do not make reference to  $\mathbf{g}(w)$ . For example, the first clause of the definition of *slight possibility* in (3.42) is  $\phi \cap \mathbf{f}(w) \neq \emptyset$ , the requirement that there are some  $\phi$ -worlds in the modal base. We can make this clause dependent on  $\mathbf{g}$  by changing it to  $\exists w \exists w' [w \in \phi \cap \mathbf{f}(w) \wedge w \succ_{\mathbf{g}(w)} w']$ , the requirement that some  $\phi$ -worlds in the modal base appear somewhere in the ordering; then, restricting the world-ordering as in (3.51) will have the same effect on the interpretation of *slight possibility* as restricting the modal base in (3.50).

The logical problems with Kratzer's theory as applied to epistemic modals have been discussed in detail in a number of recent papers, and I will sketch them only briefly here.<sup>2</sup> As we discussed in §3.4.2, Kratzer's theory inherits the **Maximality** and **Size-insensitivity** properties of Lewis' theory. As a result, if  $w' \succsim_{\mathbf{g}(w)} w''$  for all  $w'' \in \psi$ , then  $\{w'\} \succsim_{\mathbf{g}(w)}^s \psi$ —the unit set  $\{w'\}$  is at least as (likely/good/...) as  $\psi$ . In addition, all propositions containing  $\{w\}$  are at least as (likely/good/...) as  $\psi$ . The result is that both of the following (repeated from (3.18) and (3.19)) are predicted to be valid:

- |   |  |
|---|--|
| <p>(3.52) a. It's as good for Bill to perform <math>A</math> is as it is for him to perform <math>B</math>.</p> <p>b. It's as good for Bill to perform <math>A</math> is as it is for him to perform <math>C</math>.</p> <p>c. Therefore, it's as good for Bill to perform <math>A</math> is as it is for him to either perform <math>B</math> or perform <math>C</math>.</p> | <p>(3.53) a. It's as likely that Bill will perform <math>A</math> is as it is that he will perform <math>B</math>.</p> <p>b. It's as likely that Bill will perform <math>A</math> is as it is that he will perform <math>C</math>.</p> <p>c. Therefore, it's as likely that Bill will perform action <math>A</math> as it is that he will either perform <math>B</math> or perform <math>C</math>.</p> |
|---|--|

Kratzer's theory validates both inferences for the following reason. The truth of the (a)-premises require that for every  $B$ -world  $w_B$  there is some  $A$ -world  $w_A$  such that  $w_A \succsim_{\mathbf{g}(w)} w_B$ . Similarly, the (b)-premises require that for every  $C$ -world  $w_C$  there is some  $A$ -world  $w_A$  such that  $w_A \succsim_{\mathbf{g}(w)} w_C$ . Now consider  $B \cup C$ , the set of worlds where Bill either performs  $B$  or performs  $C$ . For each world  $w_X$  in this set, we can clearly find an  $A$ -world  $w_A$  such that  $w_A \succsim_{\mathbf{g}(w)} w_X$ , since  $w_X$  is either in  $B$  or in  $C$ . But then, by the definition of  $\succsim_{\mathbf{g}(w)}^s$ , we have  $A \succsim_{\mathbf{g}(w)}^s B \cup C$ . The conclusion thus follows.

It is quite plausible that the deontic variant (3.52) is valid, but it is very implausible that the epistemic version (3.53) is valid. In the example of rolling a fair die discussed above,  $A$  was rolling a 1,  $B$  was rolling a 2, and  $C$  was rolling a 3. Clearly,  $A$  could be as likely as  $B$  and  $C$  pairwise, but less likely than their disjunction. This simple contrast has an important implications: it is simply not the case that epistemic and deontic comparatives have the same basic logic. The two types of scales have different structural features, and as a result the goal of a completely unified semantics for modals of all flavors is not achievable.

Kratzer (2012) proposes a modified version of the Comparative Possibility relation. The proposal appears to be designed to fix a particularly striking instantiation of the inference in (3.53) which was pointed out by Yalcin (2010). If we set  $A = B$  and  $C = \text{not-}A$ , we can derive the following validity.

- (3.54) a.  $A$  is as likely as  $A$ .
- b.  $A$  is as likely as not- $A$ .
- c. Therefore,  $A$  is as likely as ( $A$  or not- $A$ ).

The first premise is trivial, so the only substantive assumption is (3.54b). But since  $A$  or not- $A$  is a tautology, the unwelcome conclusion is that, anytime a proposition is as likely as its own negation,

<sup>2</sup> Lassiter (2010) and Yalcin (2010), working independently, were the first to observe the problem involving epistemic comparatives and disjunction. Lassiter (2011a) pointed out that deontics and epistemics differ crucially in this respect, a point expanded on in Lassiter 2014. For further discussion see Holliday & Icard 2013; Lassiter 2015b.

it is as likely as a tautology. Note that this is a simple consequence of the fact—noted by Lewis himself and in the discussion of Kratzer’s theory above—that when maximal worlds exist, each singleton set containing a maximal worlds is ranked as high as the set of all possible worlds.

Kratzer’s (2012: 41) revised definition of the Comparative Possibility relation is given in (3.55).

$$(3.55) \quad \phi \succsim_{\mathbf{g}(w)}^s \psi \text{ iff } \neg \exists u \in (\psi - \phi) \forall v \in (\phi - \psi) [u >_{\mathbf{g}(w)} v] \quad (\text{where } u, v \in \bigcap \mathbf{f}(w))$$

This revision does have the effect of blocking the unwanted inference in (3.54), by virtue of ignoring worlds that are in both of the propositions being compared. However, as Lassiter (2011a, 2014) points out, it does not help in the general case. In the original, intuitive version of the puzzle, the alternative actions were stipulated to be mutually exclusive: performing *A* entails not performing *B* or *C*, and so forth. Lassiter proves that the undesirable inference in (3.53) remains valid according to the revised version of Comparative Possibility in (3.55), given disjointness of *A*, *B*, and *C*. Modifying the crucial relation in this way does not, therefore, help with the general semantic problem.

Another possible modification is suggested by Holliday & Icard (2013). They point out that, if we define Comparative Possibility instead as in (3.56), we can derive a proposition-ordering from a world-ordering, as Kratzer does, without validating (3.53).

$$(3.56) \quad \phi \succsim_{\mathbf{g}(w)}^s \psi \text{ iff there is an injection } h: \psi \rightarrow \phi \text{ such that, for all } w' \in \psi, h(w') \succsim_{\mathbf{g}(w)} w'.$$

The idea is that, if each  $\psi$ -world can be matched up with a *distinct*  $\phi$ -world which is at least as plausible as a candidate for being the actual world, then  $\phi$  is at least as likely as  $\psi$ . This solution is quite satisfactory as far as (3.53) is concerned. (In fact, as Holliday & Icard discuss, a scale for *likely* with this structure would be only slightly weaker than those that I will propose in the next chapter.)

We might be tempted to respond to the problem in (3.53) by simply replacing Kratzer’s version of Comparative Possibility with Holliday & Icard’s (3.56), and keeping the rest of the theory. We could then conclude that the former was on the right track after all, modulo a few technical details. This would be a mistake for several reasons. First, as we will discuss briefly in the next subsection and in some detail later in the book, *likely* and other epistemic adjectives admit a rich variety of quantitative modifiers. Giving a formal interpretation of quantitative epistemic language is not possible unless epistemic adjectives are associated with scales that have a richer structure than any of these variants of Comparative Possibility. Second, the deontic variant in (3.52) *is* valid, but it would not be if (3.56) provided the interpretation of Comparative Possibility for all modal flavors. On the other hand, if we modify Kratzer’s semantics by adopting the revised relation *solely for epistemic modals*, we have given up on the project of a unified semantics for all types of modals.

A third problem with an attempt to graft the relation in (3.56) into Kratzer’s semantics is that the result generates various incorrect predictions about logical relations between *likely* and the modal auxiliaries. This issue is discussed in some detail by Lassiter (2015b). For example, the modified theory under consideration would invalidate the intuitively obvious inference in (3.57) (Lassiter 2015b: 663-664).

- (3.57) a.  $\phi$  must be the case.  
b. Therefore,  $\phi$  is more likely than  $\neg\psi$ .

Here, the problem is created primarily by Kratzer's distinctive semantics for the auxiliaries (see §3.4.5), which Holliday & Icard (2013) do *not* endorse: they assume that *must* is a universal quantifier over epistemically accessible worlds, as usual in modal logic.

In any case, the example illustrates the interdependence of the semantics of adjectives and auxiliaries. We cannot theorize about these in a modular fashion: modifications to the semantics for adjectives, such as those that we will propose later, will generally require revisions to the semantics of auxiliaries as well. Once we have developed an alternative account of *likely*, we will return to this point in detail in chapter 6.

### 3.4.8 Quantitative modal language

Kratzer's account inherits the problems for Lewis' semantics involving quantitative modal language (§3.3.3). Lewis' assumption that Comparative Possibility is a mere ordinal scale (reflexive, transitive, and connected) predicts that even very weak quantitative modal language should not be interpretable. The same problem arises for Kratzer: formally, the only difference is that modals' scales are even weaker on her theory, since they may not be connected. Furthermore, since Kratzer extends the semantics to epistemic modals, both of the following are predicted to be uninterpretable.

- (3.58) a.  $\phi$  is much better than  $\psi$ .  
 b.  $\phi$  is much more likely than  $\psi$ .

In the case of epistemic modals, the problem is more extreme: overtly quantitative epistemic language is quite common, and all of this would be uninterpretable if epistemic scales have a structure as weak as that of a reflexive, transitive relation. Here are a few examples drawn from the web which illustrate the kinds of quantitative modification that we will discuss in detail in the following chapters.

- (3.59) a. I do think it is extremely good that people are aware and sensitive about these issues ...  
 b. "Mary Doe" of the companion Doe v. Bolton lawsuit, the mother of three whose real name is Sandra Cano, maintains that she never wanted or had an abortion and that she is "ninety-nine percent certain that [she] did not sign" the affidavit to initiate the suit.  
 c. NASA scientist Ellen Stofan ... considers it extremely possible that alien life will have been discovered in other planets of our galaxy within twenty years.  
 d. Children of the 90s are THREE times as likely to be obese as their parents and grandparents.

As this small selection already makes clear, both epistemic and deontic adjectives admit a rich variety of degree talk. In the case of epistemic modals, we need to be able to make sense not only of *much more likely than* but also of *n times as likely as* and *extremely possible*. Furthermore, the kinds of modifiers that occur with different modal adjectives vary in a way that is closely related to scalar boundedness and the minimum/maximum/relative distinction that were discussed in chapters 1-2. One of the most important goals of chapters 4-8 will be to make sense of the distribution of modifiers with various kinds of modals in light of what we know about modification and scales more generally.

### 3.5 Conclusion

This chapter discussed the analyses of graded and non-graded modal language provided by the influential theories of Lewis (1973) and Kratzer (1981, 1991). While innovative and inspiring in various ways, these accounts are not sufficient for a number of reasons. Both fail to explain how quantitative modal language can be interpreted. In light of the discussion of Measurement Theory in chapter 2, it seems that modal scales must be logically stronger than assumed by these prominent accounts. Lewis' semantics is meant only for deontics, and he expresses serious reservations about its correctness (a point that we will return to in chapters 7-8). Kratzer's generalization of Lewis' semantics inherits these problems, and makes demonstrably incorrect predictions about the meanings of epistemic comparatives. Finally, the fact that epistemic and deontic comparatives differ in their interaction with disjunction means that one of the main selling points of Kratzer's semantics—the hope of a completely unified semantics for modals of all flavors—may not be achievable.

However, if a key difference between epistemic and deontic comparatives resides in their interaction with disjunction, the typology of scales developed in chapter 2 may provide some hope for a moderately unified semantics. Disjunction is the realization of the join operation in Boolean domains, and a restricted version of this operation (concatenation) was one of the primary loci of variation among scale types. Could it be that *good* and *likely* are grammatically very similar—even identical—except that their scales differ along one or a small number of generally-motivated parameters? What would this imply for the interpretation of other epistemic and deontic modals? I will argue that fully closed, positive ratio scales play a critical role in the semantics of epistemic modals, and that fully open, intermediate interval scales are widely implicated in the interpretation of deontic modals. This distinction provides an explanation for the inferential difference between *good* and *likely* that was discussed in this chapter ((3.52) and (3.53)), along with a number of other differences. It is also supported by a variety of comparative data involving modal and non-modal adjectives.

## CHAPTER 4

### Epistemic Adjectives: *Likely* and *Probable*

We closed the last chapter by noting the puzzle that quantitative modal language provides for previous theories, which rely on binary orders with a relatively weak formal structure. This is the kind of puzzle that Measurement Theory was designed to help us resolve. Measurement Theory invites us to consider which kind of quantitative statements make sense, and which do not, and to use this information to infer which mathematical features the underlying qualitative structures must have, and which they may or should not. At the same time, we must also account for evidence about the (un)boundedness of scales provided by the linguistic tests surveyed in chapter 1 (§1.3.4).

The variety of data that we must account for is exemplified by the web examples in (4.1).

- (4.1)
- a. Occupy respondents were exactly twice as likely as Tea Party respondents to say that they favored the creation of a new political party.
  - b. Indonesia's search and rescue agency chief said on Tuesday he was 95 percent certain debris ... was part of the AirAsia jet ...
  - c. [I]t's extremely possible (maybe even likely) that Lawrence could be asked to take over the place-kicking duties next fall as a true freshmen ...
  - d. It appears extremely likely, nearly certain, that society matrons of the time would have been as meticulously groomed ... as any LA party girl is today, if not more so.

These examples illustrate two important features of the epistemic adjectives. First, they admit a variety of modifiers, including several quantitative modifiers which—as I argued in chapter 2—are acceptable only when the adjective's scale satisfies fairly strong structural conditions. Second, epistemic adjectives have complex internal relationships, as exemplified by *extremely possible* (*maybe even likely*) and *extremely likely, nearly certain* in (4.1c) and (4.1d). These examples suggest that *possible* is weaker than *likely*, and *likely* weaker than *certain*—much as *extremely warm, maybe hot* suggests that *warm* is weaker than *hot*.

Ideally, we would like to have a theory in which such examples are interpreted in the same way as the corresponding non-modal examples in (4.2).

- (4.2)
- a. [A]n established engineering firm ... estimated the structure to be exactly twice as expensive as the budget allocated to the project.
  - b. The seed lot averaged 95-percent full.
  - c. BASE jumping is extremely dangerous, but that's why BASE jumpers love it.
  - d. That water is very cheap, and nearly free ... means people think it's unlimited.

Based on the earlier discussion, we already have a fairly clear sense of what the examples in (4.2) mean. (Recall that  $\top_A$  and  $\perp_A$  pick out the inherent maximum and minimum elements of adjective  $A$ 's scale  $\mathcal{S}_A$ , when these bounds exist.)

- (4.3)
- a. (4.2a)  $\implies \mu_{\text{expensive}}(\text{structure}) = 2 \times \mu_{\text{expensive}}(\text{budget})$ .
  - b. (4.2b)  $\implies \mu_{\text{full}}(\text{seed lot}) = 0.95 \times \mu_{\text{full}}(\top_{\text{full}})$ .

c. (4.2c)  $\implies \mu_{dangerous}(\mathbf{BASE\ jumping})$  is much greater than  $\mu_{dangerous}(\perp_{dangerous})$ .

We should presumably analyze sentences like those in (4.1) along similar lines.

- (4.4) a. (4.1a)  $\implies \mu_{likely}(\phi) = 2 \times \mu_{likely}(\psi)$ .  
 b. (4.1b)  $\implies \mu_{certain}(\phi) = 0.95 \times \mu_{certain}(\top_{certain})$ .  
 c. (4.1c)  $\implies \mu_{possible}(\phi)$  is much greater than  $\mu_{possible}(\perp_{possible})$ .

As we discussed in detail in chapter 2, (4.3) places considerable constraints on the kinds of qualitative scales that we can assume for expense, fullness, and danger. The corresponding truth-conditions for (4.1) given in (4.4) license analogous inferences about *likely*, *certain*, and *possible*.

- In light of the interpretability of (4.1a) and (4.2a), we can infer that  $\mathcal{S}_{expensive}$  and  $\mathcal{S}_{likely}$  are ratio scales.
- From (4.1b) and (4.2b), we can infer that  $\mathcal{S}_{full}$  and  $\mathcal{S}_{certain}$  are scales with inherent maxima— $\top_{full}$  and  $\top_{certain}$  respectively—and that there is a stable fact about which elements have measure  $0.95 \times \mu_{full}(\top_{full})$  or  $0.95 \times \mu_{certain}(\top_{certain})$ . This means that these scales are either ratio, or fully-closed and interval.
- The acceptability of (4.1c) and (4.2c) suggests that  $\mathcal{S}_{danger}$  and  $\mathcal{S}_{possible}$  are lower-bounded, and that both contains at least minimal quantitative information; neither can be as weak as an ordinal scale, for example.

This chapter begins with a discussion of epistemic adjectives starting from this perspective. In the case of *likely* and *probable*, measurement-theoretic reasoning about the quantitative modifiers that these expressions accept can take us surprisingly far. The distribution of modifiers and the inferences licensed by comparative likelihood statements suggest that  $\mathcal{S}_{likely}$  and  $\mathcal{S}_{probable}$  are fully closed and ratio. The main point of uncertainty in this classification is connectedness—whether any two propositions are comparable in likelihood. If connectedness does hold, then this scale—like any fully closed ratio scale—is isomorphic to a FINITELY ADDITIVE PROBABILITY MEASURE.

After discussing these adjectives (especially *likely*) in detail, I turn to a consideration of how this classification fits in with the literature on scalar boundedness and the relative/absolute distinction. The theory of *likely* that I develop in §4.1 is not compatible with Kennedy’s (2007) effort to reduce the relative/absolute distinction classification to independently motivated parameters of scalar boundedness. This is because my account entails that *likely* is a relative adjective on a fully-closed scale, bounded above by the likelihood of a tautology and below by the likelihood of a contradiction. Kennedy’s theory predicts that this situation should be impossible, since it requires that adjectives whose scale has one or more endpoints must be absolute (minimum and/or maximum). I will argue that *likely* does not constitute a problem for the analysis given here. Instead, it is a counter-example to Kennedy’s theory, along with a number of other examples of relative adjectives that fall onto closed scales (such as *loud/quiet*, *fast/slow*, and *close/far*). If relative adjectives do occur on bounded scales, there is no reason to doubt the intuitive classification of the likelihood scale as fully closed.

In the next chapter I discuss the classification of the *certain* and epistemic *possible* within the typology of scalar adjectives. I argue that these adjectives’ scales are closely related—though

perhaps not identical—to the scales of *likely* and *probable*. Chapter 6 then draws out some implications for the semantics of the epistemic auxiliaries *must* and *might*. Given logical relations between the epistemic adjectives and auxiliaries, the conclusions of earlier sections involving the epistemic adjectives can be used to illuminate the better-studied auxiliaries and to rule out some otherwise plausible analyses.

#### 4.1 Scale structure and modifiers

*Likely* and *probable* seem to be nearly identical in meaning, but they are grammatically distinct: for example, *likely* is a raising adjective while *probable* is not.

(4.5) Bill is likely/\*probable to win the race.

I have been unable to identify any clear semantic difference between the two, and I treat them together here. Note, however, that *likely* may be slightly stronger: in a modal reasoning experiment reported in Lassiter & Goodman 2015b, participants were slightly but significantly more willing to judge a conclusion “probable” than “likely”, with the premises held fixed. If this result is replicated in further experiments, it would constitute an interesting new twist to the story given here, and one in need of additional semantic and/or pragmatic explanation.

*Likely* and *probable* both accept a variety of quantitative modifiers: intensifiers such as *very* and *extremely*, attenuators such as *quite*, ratio modifiers (*n times as Adj as*), and proportional modifiers such as *n%* (for *n* between 0 and 100). The following are all naturally-occurring examples drawn from the web.<sup>1</sup>

- (4.6) a. A source close to John Kasich said Sunday that the Ohio governor is “very likely” to run for president.  
 b. [U]nion leaders warned that a strike is all but inevitable. “I think it’s extremely likely,” said Josie Mooney, chief negotiator for Service Employees International Union Local 1021.  
 c. Quite likely one of the worst companies I’ve ever worked for.  
 d. A Russian diplomatic source told Reuters it was 70 percent likely that the the leaders would reach an agreement on the crisis during Wednesday’s talks.  
 e. Republicans are over three times as likely as Democrats to say there is no evidence the earth is warming.
- (4.7) a. General manager Neal Huntington admitted yesterday that it’s “very probable” starting catcher Chris Snyder will begin the season on the disabled list due to back soreness.

<sup>1</sup> These uses occur in texts of a variety of genres, spoken and written. Examples of *n times as probable* and *n% probable/likely* are, however, most frequent in academic and legal texts. This pattern might be attributed to the convergence of two independent factors. First, there is a difference in register between the two items, with *probable* (a French borrowing) retaining an association with formal and scientific usage; across the board, *likely* seems to be much more natural and informal. Second, it may be more common in academic and legal texts to focus on contexts in which sufficient information is available for precise estimation of likelihoods. This could explain the particularly skewed distribution of *n% probable* and *n times as probable* across genres. In any case, none of these uses are totally confined to such contexts.



- b. If you argue with a madman, it is extremely probable that you will get the worst of it; for in many ways his mind moves all the quicker for not being delayed by the things that go with good judgment. (G. K. Chesterton)
- c. [I]t's quite probable that Burger King violated some part of the Terms of Services ...
- d. [T]he threat posed by the 2013 TV135 is minor. Updated estimates show that it has a one in 14,000 chance of colliding with our planet, which is five times as probable as the initial estimate of 1 in 63,000, but still negligible.
- e. National climatologists are sticking to their prediction that an El Niño is about 80 percent probable in North America during fall and early winter.

In light of the conclusions in chapter 2, these examples support the following hypotheses.

- Since *likely* and *probable* accept ratio modifiers, both are associated with ratio scales. This also predicts the acceptability of weaker quantitative modifiers such as *extremely*, *very*, *quite*.
- Since both accept the proportional modifier *n%*, both are associated with a scale that has is at least asymptotically bounded above and below.

In addition to this linguistic evidence, there is a clear theoretical reason to suppose that these scales should be upper-bounded: their domain contains  $W$ , the set of all possible worlds. Plainly, nothing can be more likely than a proposition that is true always (a tautology). So, the inherent maximum  $\top_{likely}$  can be identified with  $W$ .

The same reasoning also leads to the conclusion that the scale should be lower-bounded, with  $\perp_{likely} = \emptyset$ : after all, nothing could be less likely than a contradiction, which is true never. (Here, we are setting ourselves up for a possible problem involving the linguistic features of positive-form adjectives on bounded scales: see §4.2 below.)

#### 4.1.1 The Disjunctive Inference revisited

In addition to rendering ratio and proportional modifiers interpretable, the hypothesis that  $\mathcal{S}_{likely}$  and  $\mathcal{S}_{probable}$  are ratio scales has the nice consequence of explaining why the “Disjunctive Inference” discussed in chapter 3 (sections 3.3.4 and 3.4.7) should come out as invalid. (This reasoning holds whether or not the scales are upper-bounded.)

- (4.8)    a.  $\phi$  is as likely as  $\psi$ .  
           b.  $\phi$  is as likely as  $\chi$ .  
           c. Therefore,  $\phi$  is as likely as  $(\psi \vee \chi)$ .

Bracketing the issue of boundedness, suppose that  $\mathcal{S}_{likely}$  is at least a ratio scale— $\langle \mathcal{Q}, \succsim_{likely}, \circ \rangle$ , where  $\mathcal{Q}$  is a set of propositions and  $\circ$  is interpreted as join (disjunction) restricted to disjoint elements of  $\mathcal{Q}$ . Since  $\mathcal{S}_{likely}$  is a ratio scale,  $\succsim_{likely}$  is **positive** with respect to  $\circ$ . (See chapter 2, §2.4.2.)

- (4.9)    **Positivity:** Suppose that  $\psi$  and  $\chi$  are disjoint, and that  $\chi$  is not exactly as likely as  $\perp_{likely}$ . Then  $\psi \vee \chi$  is more likely than  $\psi$ .

The logic behind this validity is exactly the same as in the case of ratio scales generally. The need to assume positivity is clear when we consider ratio scale adjectives such as *heavy* and *expensive*.

(4.10) Suppose that  $y$  and  $z$  are non-overlapping. Then their join  $y \sqcup z$  is heavier than  $y$ .

(4.11) Suppose that  $y$  and  $z$  are non-overlapping and  $z$  is not free. Then their join  $y \sqcup z$  is more expensive than  $y$ .

In addition to making intuitive sense, we saw in chapter 2 that this formal property is necessary in order to ensure that sentences making reference to ratios, such as  $x$  is (exactly)  $n$  times as heavy/expensive as  $y$ , are semantically interpretable. (See §4.1.3 below for more discussion and consideration of some possible alternative analyses.)

In the case of *heavy*, the analogue of the Disjunctive Inference (4.8) is (4.12).

- (4.12) a.  $x$  is as heavy as  $y$ .  
 b.  $x$  is as heavy as  $z$ .  
 c. Therefore,  $x$  is as heavy as  $(y \sqcup z)$ .

This inference is obviously invalid, and it is easy to see why: for example,  $x$ ,  $y$ , and  $z$  might all have the same weight. Since  $y \sqcup z$  must be strictly heavier than either  $y$  or  $z$ , it must also be heavier than  $x$  if these objects all have the same weight.

Similarly, the fact that Positivity holds if  $\mathcal{S}_{likely}$  is a ratio scale suggests that we should look for counterexamples to the Disjunctive Inference in cases where  $\psi$  and  $\chi$  are disjoint. And indeed, we saw already in chapter 3 that the most obvious counter-examples to (4.8) are cases in which  $\phi$ ,  $\psi$ , and  $\chi$  are equally likely. Consider, for example, drawing a single card at random from a well-shuffled pack. The inference in (4.13) would clearly be inappropriate.

- (4.13) a. The card is as likely to be a spade as it is to be a heart.  
 b. The card is as likely to be a spade as it is to be a diamond.  
 c. Therefore, the card is as likely to be a spade as it is to be a red card (i.e., either a heart or a diamond).

Indeed, the hypothesis that  $\mathcal{S}_{likely}$  is a ratio scale predicts that the inference in (4.14) should be valid.

- (4.14) a. It is not possible for the card to be any two of these three at the same time: a spade, a heart, a diamond.  
 b. The card is exactly as likely to be a spade as it is to be a heart.  
 c. The card is exactly as likely to be a spade as it is to be a diamond.  
 d. Therefore, the card is **less** likely to be a spade than it is to be a red card (i.e., either a heart or a diamond).

This prediction seems to correct. Its validity supports the claim that  $\succsim_{likely}$  is positive with respect to concatenations, consonant with the suggestion that  $\mathcal{S}_{likely}$  and  $\mathcal{S}_{probable}$  are ratio scales.

#### 4.1.2 What makes a ratio scale, again?

The hypothesis that  $\mathcal{S}_{likely}$  and  $\mathcal{S}_{probable}$  are ratio scales, when combined with the claim that both are double-bounded, is just a short step from an idea that has been suggested a number of times in

the recent formal semantics literature: that *likely* and *probable* are associated with a scale that is equivalent to ordinary numerical probability.

**Hypothesis:** The degree scales associated with *likely* and *probable* have the formal structure of FINITELY ADDITIVE PROBABILITY SPACES.<sup>2</sup>

The key component of a finitely additive probability space is a measure function that assigns values in the range  $[0, 1]$  to propositions, subject to the constraint of disjoint additivity.

(4.15) A FINITELY ADDITIVE PROBABILITY SPACE is a structure  $\langle W, \mathcal{Q}, P \rangle$ , where  $W$  is a set of possible worlds,  $\mathcal{Q} \subseteq \mathcal{P}(W)$ ,  $W \in \mathcal{Q}$ , and  $P$  is a total function from  $\mathcal{Q}$  to  $[0, 1]$  satisfying the following constraints:

- a. **Boolean:** Whenever  $A$  and  $B$  are in the domain of  $P$ , so are  $\neg A$  and  $A \cup B$ .
- b. **Additivity:**  $P(A \cup B) = P(A) + P(B)$  whenever  $A \cap B = \emptyset$ .
- c. **Normalization:**  $P(W) = 1$ .

In fact, these analyses are provably equivalent up to normalization, as long as we assume (as usual) that the domain of propositions  $D_{\langle s, t \rangle}$  is Boolean (Narens 2007: 32-35). In other words, the qualitative claim that  $\mathcal{S}_{likely}$  and  $\mathcal{S}_{probable}$  are fully closed ratio scales is, in the context of the broader framework outlined in this book, formally equivalent to the quantitative claim that *likely* and *probable* are degree adjectives on a scale of finitely additive probability. However, the qualitative characterization does have some advantages: its axiomatic characterization gives us the opportunity to break it down and examine the plausibility of the individual assumptions that must hold if it is correct. Therefore, after describing the class of measure functions that are admissible relative to fully closed ratio scales and explaining how this class relates to the finitely additive probability measures, I will devote the following sections to a consideration of the empirical validity of the assumptions that are implicit in the theory that  $\mathcal{S}_{likely}$  and  $\mathcal{S}_{probable}$  are fully closed ratio scales.

Recall from chapter 2 (§2.4.2) the key formal requirements that must be met in order for  $\mathcal{S}_{likely}$  to count as a ratio scale.

- $\succsim_{likely}$  is a weak order (reflexive, transitive, connected) on  $D_{\langle s, t \rangle}$ .
- **Monotonicity:** If  $\phi$  and  $\psi$  are both disjoint from  $\chi$ , then  $\phi \succsim_{likely} \psi$  iff  $\phi \vee \chi \succsim_{likely} \psi \vee \chi$ .
- **Positivity:** If  $\phi$  and  $\psi$  are disjoint and  $\phi \not\approx \perp_{likely}$ , then  $\phi \vee \psi \succ_{likely} \psi$ .
- **Solvability:** If  $\phi \succ_{likely} \psi$ , then there is some  $\chi \in D_{\langle s, t \rangle}$ , disjoint from  $\psi$ , such that  $\phi \succsim_{likely} \psi \vee \chi$ .
- **Archimedean:** for all  $\phi \in D_{\langle s, t \rangle}$ , there is no infinite sequence of disjoint  $\psi_1, \psi_2, \dots$  such that
  - for all  $j$ ,  $\psi_j \succ_{likely} \perp_{likely}$ ;
  - for all  $j$  and  $k$ ,  $\psi_j \approx_{likely} \psi_k$ ;

<sup>2</sup> See Swanson 2006, 2015; Yalcin 2007, 2010; Lassiter 2010, 2014, 2015b; Lassiter & Goodman 2015b; Moss 2015, among others.

– for all  $j$ ,  $\phi \succsim_{\text{likely}} \bigvee_{i=1}^j \psi_i$  (i.e.,  $\psi_1 \vee \psi_2 \vee \dots \vee \psi_j$ ).

In the next few subsections we will consider whether it is reasonable to suppose that  $\mathcal{S}_{\text{likely}}$  has all of these properties. If it does, then  $\mathcal{S}_{\text{likely}}$  is a ratio scale and all admissible  $\mu_{\text{likely}}$  are additive. Furthermore, all admissible  $\mu_{\text{likely}}, \mu'_{\text{likely}}$  are related to each other as in (4.16):

$$(4.16) \quad \exists r \in \mathbb{R}^+ : \forall \phi : \mu_{\text{likely}}(\phi) = r \times \mu'_{\text{likely}}(\phi)$$

That is, any two admissible measure functions are equivalent modulo rescaling. As a consequence, any two admissible  $\mu_{\text{likely}}, \mu'_{\text{likely}}$  are related as follows:

$$\frac{\mu_{\text{likely}}(\phi)}{\mu_{\text{likely}}(W)} = \frac{\mu'_{\text{likely}}(\phi)}{\mu'_{\text{likely}}(W)}$$

Any  $\mathcal{S}_{\text{likely}}$  with these properties has exactly one “normalized” admissible measure  $\mu_{\text{likely}}$  where  $\mu_{\text{likely}}(W) = 1$ . With the additional assumption that the domain  $D_{\langle s, t \rangle}$  is Boolean, this measure satisfies the conditions in (4.15) for counting as a probability measure. Indeed, we can convert any admissible  $\mu_{\text{likely}}$  into a finitely additive probability measure by dividing the measures of all propositions in the domain by  $\mu_{\text{likely}}(W)$ . In other words, for any admissible  $\mu_{\text{likely}}$ , the measure  $\mu'_{\text{likely}}$  defined so that  $\mu'_{\text{likely}}(\phi) = \mu_{\text{likely}}(\phi) / \mu_{\text{likely}}(W)$  is a finitely additive probability measure.

On these assumptions, then, the class of admissible measure functions is so restricted that we can without loss of generality choose any admissible  $\mu_{\text{likely}}$  to reason about for semantic purposes. In particular, we can choose the normalized measure with  $\mu_{\text{likely}}(W) = 1$ . Call this measure *prob*. We can reason equivalently about  $\mathcal{S}_{\text{likely}}$  and *prob*, without needing to worry about importing quantitative information that is not already contained in the qualitative source structure  $\mathcal{S}_{\text{likely}}$ , as long as we are careful to avoid statements that rely for their truth-values on normalization. Even this is not much of a restriction, since we can read any apparently uninterpretable statement of the form “ $\mu_{\text{likely}}(\phi) = r$ ” as short for the statement “ $\mu_{\text{likely}}(\phi) / \mu_{\text{likely}}(W) = r$ ”. The latter is interpretable, since its truth-value is stable across all admissible  $\mu_{\text{likely}}$ .

To the extent that the qualitative assumptions described above are reasonable for  $\mathcal{S}_{\text{likely}}$  and  $\mathcal{S}_{\text{probable}}$ , then, we have a purely qualitative derivation of the hypothesis that *likely* and *probable* are adjectives with probabilistic scales, with no quantitative assumptions.

This would be a convenient conclusion, since numerical probability is well-studied formally and broadly useful in philosophical, linguistic, and cognitive science applications. But is it reasonable to assume that  $\mathcal{S}_{\text{likely}}$  and  $\mathcal{S}_{\text{probable}}$  have all of the requisite properties— **upper- and lower-boundedness**, **reflexivity**, **transitivity**, **connectedness**, **monotonicity**, **positivity**, **solvability**, and **Archimedean**? Here is an quick overview, with pointers to later sections where more extensive discussion is required.

- **Upper-boundedness:** Apparent, as long as  $W$  is in the domain. (Nothing can be more likely than a tautology.)
- **Lower-boundedness:** Apparent, as long as  $\emptyset$  is in the domain. (Nothing can be less likely than a contradiction.)

- **Reflexivity:** Apparent from the meaning of “at least as ... as”.
- **Transitivity:** Fundamental to the concept of a scale.
- **Connectedness:** It is quite plausible that connectedness fails for  $\mathcal{S}_{likely}$  and  $\mathcal{S}_{probable}$ . See §4.1.5 below for discussion.
- **Monotonicity:** Apparent. (If drawing a heart is as likely as drawing a club, then drawing a heart *or a spade* is as likely as drawing a club *or a spade*.)
- **Positivity:** Apparent. (If drawing a club is possible, then drawing a heart *or a club* is more likely than drawing a heart.)
- **Solvability:** This assumption is quite restrictive, but there are decent theoretical arguments in its favor. See §4.1.4 below.
- **Archimedean:** Holds trivially if  $W$  is finite. With infinite  $W$ , this is a significant assumption, ruling out infinitesimal probabilities (those that are smaller than any real number). While I do not know of any airtight argument in favor of making this assumption across the board, I also do not know of any good arguments in favor of employing infinitesimals in a degree semantics for any area of natural language. For this reason, I will continue to assume that the Archimedean assumption holds quite generally for the scales that are lexicalized in natural languages.

### 4.1.3 The uses of ratio modifiers

**Barney:** Robin, think of the funniest thing that has ever happened.

**Robin:** Got it.

**Barney:** Now double that.

**Robin:** So, a chimpanzee wearing TWO tuxedos? ... I mean, what, did he—he forgot he put the first one on? Stupid monkey.

(*How I Met Your Mother*, season 4, episode 16)

A straightforward linguistic argument that we could make for the conjunction of the formal conditions just described is that this is what is needed to explain the acceptability of ratio modifiers with *likely* and *probable*. As noted above, (4.17) is an intuitively reasonable description of a situation in which one is about to throw a fair die.

(4.17) Throwing an odd number is three times as likely/probable as throwing a four.

$\mathcal{S}_{likely}$  and  $\mathcal{S}_{probable}$  must satisfy some fairly stringent formal conditions in order for this statement to be interpretable in the RTM sense (§2.2). A ratio scale classification is an obvious choice, especially given the empirical arguments for positivity that we saw above (§4.1.1).

Indeed, a number of authors have used the acceptability of ratio modifiers as a diagnostic for ratio scale status (Sassoon 2010b; van Rooij 2011; Lassiter 2011a, 2015b). For example, consider

the following contrast. ((4.18) and several further examples in this section are drawn from Lassiter 2015b.)

- (4.18) a. ✓ Mary is three times as tall/old/heavy as Bill is.  
 b. ?? Mary is three times as angry/hungry/lecherous as Bill is.

*Tall*, *old*, and *heavy* are good candidates for being ratio scale adjectives; *angry*, *hungry*, and *lecherous* are not. This is plausibly enough to explain the intuitive contrast in (4.18).

However, it is not generally the case that ratio modifiers can be used with an adjective *Adj* only if  $S_{Adj}$  is ratio. In naturally-occurring examples of this construction, several interesting classes of exceptions turn up. First, there are **hyperbolic** uses in which *n times as Adj as* is used to convey “much more *Adj* than”.

- (4.19) a. This is easily the best adaption of the Arthurian legends that I’ve ever read! It is an absolute tragedy that the book isn’t as well-known as, say, the Da Vinci Code, for it is a million times as good as that!  
 b. “We’re doing the Shamrock Fighting Championships at River City Casino,” Finney said. “It’s three times as nice as I thought it was going to be; stadium seating along with a stage and LED screen above the cage. It blew me away.”

Second, there are **coercive** uses in which an adjective that is not generally ratio is exceptionally associated with some ratio (or stronger) scale. Consider the examples in (4.20).

- (4.20) a. Pinterest is retaining and engaging users as much as 2-3 times as efficiently as Twitter was at a similar time in their history.  
 b. The Exploding Golf Ball Four Pack is exactly four times as funny as the regular Exploding Golf Ball, and that’s wicked funny!

*Efficient* is probably not in general a ratio scale adjective. However, in the context of a website detailing “Social Media Stats”, it is possible to interpret (4.20a) as talking about a scalar property which is ratio and is clearly tied to efficiency of this type—the proportion of new users that remain active after some specified period of time. Similarly, the humor of (4.20b) derives from the absurdity of what we must do in order to interpret it: coerce funniness into a ratio scale whose units are numbers of exploding golf balls.

Perhaps the most telling class of examples involve **metalinguistic** uses, in which a ratio modifier occurs under the scope of a negation or other entailment-canceling operator. In these examples, the message is precisely that the inferences that would follow from a ratio scale classification are not appropriate.

- (4.21) a. Nonetheless, inequality of happiness is usually less marked than inequality of income, at least in wealthy societies. A man earning \$500,000 a year is not usually 10 times as happy as a man earning \$50,000 a year.  
 b. She poured two gobletfuls from a bottle whose label I recognized—it cost around thirty dollars ... The wine really was nice, though hardly three times as nice as one that I would typically serve to company, nor fifteen times as nice as the Two-Buck Chuck that was my daily fare.

- c. Of course, intelligence is not a one-dimensional thing. We are not just ten times as smart as dinosaurs or a thousand times as smart as bombardier beetles or a million times as smart as bacteria. Our intelligence is qualitatively different from theirs.

In addition, many examples that turn up in corpora turn out to have an intentionally humorous effect, such as (4.22).

- (4.22) Rosen is the U.S. Attorney for D.C., and he is, in fact, a halfway decent person. This makes him exactly four times as decent as Olivia and eight times as decent as Fitz. (And 10 or 12 times as decent as Mellie, but far less awesome.)

Here, the humor is built around the phrasal idiom *halfway decent*. A literal interpretation of this idiom would be possible only if decency were an upper-bounded ratio scale, and so capable of supporting ratio modifiers: thus, *four times as decent*, etc.

Given that these interpretations are available and fairly well-attested, is the acceptability of ratio modifiers informative about whether an adjective's scale is ratio? This question is critical because the answer will determine whether we can use examples like (4.17) to fix the scale type of *likely* and *probable*. This example is not metalinguistic (because it does not occur under an entailment-canceling operator), or humorous; but could it be hyperbolic or coercive?

We can rule out hyperbolic uses, at least, by considering the effects of adding *exactly*. Our original example involving *likely*, (4.17), seems to be unaffected by the addition of *exactly*. Indeed, this was already the intuitive interpretation.

- (4.23) Throwing an odd number is exactly three times as likely/probable as throwing a four.

However, the addition of *exactly* renders hyperbolic uses exceedingly strange. Consider these modified versions of the examples from (4.19).

- (4.24) a. ?? This book is exactly a million times as good as the Da Vinci Code!  
b. ?? River City Casino is exactly three times as nice as I thought it was going to be.

This effect is, I suggest, a reasonable diagnostic for hyperbolic uses.

Another possibility is that examples like (4.23) involve coercion, so that *likely* and *probable* are not interpreted in their normal fashion. *Exactly* can occur in clear cases of coercive uses: for instance, it occurs in (4.20b), one of the examples that I used to motivate a coercion analysis above.

As far as I can see, it is not possible to conclusively rule out this alternative hypothesis. Coercion, in the sense that we are employing the term here, is an empirically necessary but exceedingly powerful mechanism, which essentially involves constructing *ad hoc* lexical interpretations. (Compare, for example, Clark & Clark's (1979) discussion of expressions such as *to Houdini one's way out of the closet*.) This kind of coercion contrasts with the much more constrained theoretical notion that is employed, for example, in Generative Lexicon theory (Pustejovsky 1995).

It is *always* possible in principle that a particular example involves *ad hoc* reinterpretation of an expression, but we should be hesitant to dismiss data points on this basis unless we have reason to believe that an expression is not being used in a normal way, and a plausible hypothesis about how and why an expression would be re-interpreted in a particular way in some context. In examples like (4.23), though, *likely* seems to be used in its basic sense. My belief is that we should not employ

such powerful theoretical mechanisms if the only reason to do so is to avoid the obvious theoretical implication of these examples: that  $\mathcal{S}_{likely}$  and  $\mathcal{S}_{probable}$  are ratio scales.

That said, since this argument is methodological and plausibilistic in nature, it seems worthwhile to go back and consider in detail the two ratio scale axioms that we expressed uncertainty about in §4.1.2: Solvability and Connectedness. If these could be substantiated, then the classification of  $\mathcal{S}_{likely}$  and  $\mathcal{S}_{probable}$  as ratio scales would be more secure.

#### 4.1.4 Solvability

Recall the **Solvability** assumption:

If  $\phi >_{likely} \psi$ , then there is some  $\chi \in D_{\langle s, t \rangle}$ , disjoint from  $\psi$ , such that  $\phi \succsim_{likely} \psi \vee \chi$ .

The force of the solvability condition is that it should be possible to identify portions of logical space with arbitrarily small likelihood. So, the domain of propositions must be fairly rich. For example, solvability may fail if there is a proposition  $\phi$  that has significant probability, and which is indivisible—e.g., a singleton, so that the domain of propositions does not contain any strict subsets of  $\phi$ .

It is not immediately obvious whether it is reasonable for a semanticist to assume that this condition is met. However, it is required to ensure that  $\mathcal{S}_{likely}$  only admits measure functions that are equivalent, up to normalization, to finitely additive probability measures. Without solvability, we would have a scale which shares some of the formal properties of a ratio scale, but admits measures that are not additive. (See Holliday & Icard 2013 for an insightful discussion of related scalar representations that are almost-probabilistic.)

What would we have to do in order to ensure that solvability holds? When discussing this condition for non-modal scales in chapter 2 (§2.4.2), we noted that solvability can be ensured if we add to the domain a set of “pseudo-objects”—in the case of the weight scale, a variety of weights. This assumption is not reasonable if we imagine ourselves trying to set up a weight scale for arbitrary sets of real-world objects. However, if we are modeling the cognitive representation of the weight scale, it does seem reasonable: after all, there is no difficulty in *imagining* adding a very small weight to any given object and weighing the result. Whether the actual world contains such an object is irrelevant.

In the case of probability, a similar effect could be achieved by allowing that the domain contains an unlimited variety of propositions that are logically independent of each other and of all others. This is not as dramatic as it sounds: all we need is for the universe to contain a fair coin which can be flipped an arbitrary number of times (cf. Savage 1954). Using this conceit, we can always create a proposition that is disjoint from  $\psi$  and arbitrarily fine-grained. For example, if  $\phi >_{likely} \psi$  and we are trying to check whether solvability is satisfied, we have to find a  $\chi$  disjoint from  $\psi$  such that  $\phi \succsim \psi \vee \chi$ . The following will do the trick: let  $\chi$  be  $\neg\psi$  and *heads on each of  $n$  consecutive flips of a fair coin*, for sufficiently large  $n$ . No matter how small the difference in likelihood between  $\phi$  and  $\psi$ , there is some  $n$  that will allow this proposition to satisfy the solvability assumption.

If we are interested in formalizing qualitative probability for arbitrary modeling purposes, this assumption is probably too restrictive. On the other hand, if we are doing lexical semantics with an



eye to cognitive science applications, it seems perfectly reasonable. We can conceive of and talk about fair coins, and the rich, compositional language that we speak clearly allows us to form the needed propositions and talk about their likelihoods: “The likelihood that it’s not raining and 17 coin flips all came up heads”. For natural language semantics, the solvability assumption seems to be a reasonable one.

#### 4.1.5 Connectedness

Another vexing question is whether we can reasonably assume that all comparisons involving *likely* and *probable* have determinate truth-values. Consider a statement like (4.25):

- (4.25) It’s more likely that the Dallas Cowboys will win the 2017 Super Bowl than it is that it will rain in Zanzibar on October 7, 2024.

If someone were to ask me whether (4.25) is true, I might answer “I don’t know”, or “That’s a very weird thing to ask”, or “I’ll get back to you once I’ve had a chance to learn more about the NFL and the weather patterns of coastal Tanzania”.

Perhaps, then, the fact that speakers may not be able to form an opinion about some comparisons involving *likely* should lead us to classify it with adjectives whose scales are plausibly not connected, like *beautiful* and *clever*.

- (4.26) a. The Sistine Chapel is as beautiful as the Aurora Borealis.  
b. Paul Simon is more clever than Sasha Baron-Cohen.

In these examples, forming a judgment is fairly impossible, and it feels perverse to try. There are so many different facets to beauty and cleverness that, unless we specify one of them as particularly relevant, there is no way that we can compare people and objects as diverse as these.

Obviously, though, we cannot conclude from the mere fact that people are unwilling or unable to pass judgment on a certain comparison’s truth that there is not a fact of the matter. For example, no one knows whether Lucy (a 3 million year old *Australopithecus* whose remains were unearthed in Ethiopia in 1974) was taller than her mother, and there are no practical or indeed imaginable means by which we could find out. Nevertheless, she was either taller than her mother, or she was not. Could it be that the likelihood/probability scale is connected after all, but that people have uncertainty about the relevant ordering, so that they are unwilling or unable to make judgments in certain cases?

An account along the latter lines would be most straightforward to implement using the interpretation procedure for expressions with a qualitative scalar semantics introduced in chapter 2. For obvious reasons, it was necessary to assume world-relative scales for heights, weights, and the like. If likelihood is similarly world-relative, uncertainty about the truth-value of a sentence like (4.25) could be explained as uncertainty about which world one resides in, even while the sentence has a determinate truth-value at each world. Of course, the semantics is equally compatible with world-relative, non-connected scales; the point is that the observation that people may demure on (4.25) may not show us anything about connectedness. Whether it does, I suggest, depends on our interpretation of the scales.

These questions are deeply related to difficult and much-discussed issues in the philosophy of probability, where the interpretation of probability scales has received much attention. While there are many divisions and controversies in this area, the most important issue for us involves the metaphysics of probability scales. One group of theories, including both the radical subjectivism of De Finetti (1989) and the logical interpretation of probability advocated by Keynes (1921), treat probability as a form of knowledge: “something which agents *possess*, or *attach* to particular propositions” (as Lawson (1988) summarizes).

The other group treats probability as a worldly fact—“something to be *discovered*, *learned* about, and *known*” (Lawson 1988: 33). This includes frequentists about probability, who identify probabilities with proportions of events in a reference class (e.g., von Mises 1957). It also includes propensity theorists, who treat probabilities as latent features of the world which are causally implicated in frequency distributions, but are not constituted by them (Popper 1959).

The key issue is whether probability *judgments* are involved in the metaphysics of probability in a constitutive way: do we learn about probabilities, or (under appropriate conditions) manufacture them? If (rational) judgment is constitutively involved in a proposition’s having a certain probability, then the observation that a (rational) agent may refuse to offer a judgment on some matter, say (4.25), could be construed as an indication that there is no fact. Keynes (1921: §3) argues at length for this conclusion.

Consider three sets of experiments, each directed towards establishing a generalisation. The first set is more numerous; in the second set the irrelevant conditions have been more carefully varied; in the third case the generalisation in view is wider in scope than in the others. Which of these generalisations is on such evidence the most probable? There is, surely, no answer; there is neither equality nor inequality between them. ... Is our expectation of rain, when we start out for a walk, always more likely than not, or less likely than not, or as likely as not? I am prepared to argue that on some occasions none of these alternatives hold, and that it will be an arbitrary matter to decide for or against the umbrella. ... (§3.8)

One possible response—perhaps attributable to de Finetti, who agreed with Keynes on the constitutive question but endorsed universal comparability—is to deny the data. Even though people may sometimes be squeamish about offering probability judgments, we know that they *are* able to form opinions when it matters because they are generally capable of forming a judgment about whether to take a bet.<sup>3</sup>

Another response is to endorse a different interpretation of probability. If we suppose that probabilities are frequencies or propensities, it is straightforward to conceive of a situation in which someone would refuse to offer a judgment about the relative probability of two events simply because they do not have sufficient information. For example, in Keynes’ rain example there might be a fact of the matter about how likely rain is, depending on non-deterministic causal relations among meteorological variables. On this interpretation, our expectations could incorporate significant uncertainty about the probability of rain, so much so that we are and should be unwilling

<sup>3</sup> See, however, Feduzi, Runde & Zappia 2011 for evidence that de Finetti’s position was subtler than this characterization suggests.

to render a judgment. As we noted above, this is the most straightforward interpretation of the world-relative probability scales that would result from a straightforward application of the theory of scalar predicates given in chapter 2.

For our purposes, it does not matter what the correct way to think about probabilities is in general (if this question is even well-posed). We are engaged in a cognitive modeling enterprise, trying to build a formal theory of epistemic meaning which helps to explain how speakers use words like *likely* and *probable* and how listeners understand them. The relevant question for us is whether the agents whose cognitive states we are modeling—ordinary speakers of English—*think of* probabilities as rational degrees of belief, as external objects of knowledge, or as something else.

Given this, what we ought to do is ask how people think and talk about the comparisons that are used to motivate incomparability. If people tend to think of probabilities as external objects of knowledge, then it should be natural to demur from judgment by saying things like “I don’t know”, or “I’ll try to find out”. On the other hand, if they think of probabilities as constitutively related to their own mental states, these responses should generally be strange; they would make sense only to the extent that a speaker’s introspective abilities are limited, or perhaps if the speaker has not been able to draw out the consequences of the available information (see Keynes 1921: §3.12).

For example, suppose that we are in a situation like the one that Keynes describes in the quote above, where we have been asked to judge the relative likelihood of rain and sunshine and are unable to offer an answer. The responses in (4.27) seem natural:

- (4.27) a. I don’t know whether rain is as likely as sunshine.  
b. Good question—let me check the weather report.

These responses both suggest that the speaker is treating the relative likelihood of rain and sunshine as an external object of knowledge: something that they could know or not (4.27a), and something that they could find out with further investigation (4.27b).

In contrast, if failure of judgment is most readily interpreted as indicating true incomparability, we might expect the responses in (4.28) to seem natural.

- (4.28) Is rain as likely as sunshine?  
a. It makes no sense to compare rain and sunshine in this way.  
b. Rain isn’t as likely as sunshine, nor is sunshine as likely as rain. They are too different in likelihood to be compared.

To my ear, at least, (4.28a) seems wrong, and (4.28b) sounds like an outright contradiction. This would be odd if the scalar concepts picked out by *likely* and *probable* were not connected: after all, it is not difficult to make sense of the responses in (4.29).

- (4.29) Is the Sistine Chapel as beautiful as the Aurora Borealis?  
a. It makes no sense to compare the Sistine Chapel and the Aurora Borealis in this way.  
b. The Sistine Chapel is not as beautiful as the Aurora Borealis, nor is the Aurora Borealis as beautiful as the Sistine Chapel. They are too different in beauty to be compared.

We might be tempted to infer from these observations that ordinary language users may conceive of the likelihood/probability scale as a mind-independent object of knowledge (relating, for example,

to the causal structure of a non-deterministic universe). With enough information, they are able to form reasonably confident estimates of probabilities, just as sufficient information enables them to form reasonably confident estimates of heights, weights, temperatures, and other worldly properties. This idea is consonant with much recent psychological work which suggests that our representations of the external world are structured like causal models, i.e., probabilistic models which encode non-deterministic causal relations between events (e.g., Sloman 2005; Danks 2014). (Note, by the way, that if it is true that people conceive of probabilities as external objects of knowledge, it does not necessarily follow that universal comparability *is* a property of  $\mathcal{S}_{likely}$ . All that this would show is that an influential argument based on intuitions about likelihood judgments is not relevant for semantic purposes.)

Nevertheless, this conclusion would probably be too quick in light of the following example (adapted from an example involving *possible* in DeRose 1991). Suppose that we have strong reason to suspect that Bill has a fairly common disease *X*: he displays several typical symptoms, and no other common disease leads to this cluster of symptoms. Given this, we are inclined to endorse (4.30).

(4.30) It is likely that Bill has disease *X*.

In fact, there is a fairly reliable test for disease *X*, and Bill has already been tested but the results have not come back yet. A positive result means that the patient probably has the disease, and a negative result means that the patient probably does not. While waiting for the results, we might also be inclined to endorse (4.31).

(4.31) We don't know yet whether it is likely that Bill has disease *X*. We'll find out when we get the test back.

On general theoretical grounds, we would expect these claims to be pragmatically inconsistent: one is not in general licensed to assert  $\phi$  when one does not know whether  $\phi$ . We could resolve the conflict if we suppose that *likely* is able to express, in a sense, both of the interpretations of probability that we have discussed. Somehow, likelihood judgments are sufficiently flexible in their subject matter than they can integrate elements of individual uncertainty with reasoning about the possible future evolution of one's epistemic state.

Ultimately, it is not clear to me whether the existence of examples like (4.25), where agents may refuse to make a probability judgment, is a threat to connectedness. It would also be useful to have a clearer picture about the empirical situation involving examples like (4.27)-(4.31)—for example, pinpointing the contextual factors that determine which judgments and responses are appropriate.

If connectedness does fail, then the scale of likelihood/probability will have a weaker structure than a fully closed ratio scale, with implications for the analysis of linguistic features that are related to these formal structures (among others, ratio and proportional modifiers). Non-connected qualitative scales are under certain conditions representable by *sets* of measures (Harrison-Trainor, Holliday & Icard 2015). Thus it may be possible to explain the interpretability of ratio and proportional modifiers in some cases, even without universal connectedness, if all of the measures in a set determined in this fashion agree on the relevant ratios.

From here on I will proceed, with some trepidation, to make use of the connectedness assumption. This is mostly for convenience: it is often easier to reason with numerical probabilities than

with qualitative scales or with sets of probability measures. Hopefully, future work will consider in more detail the implications of the use of a non-connected likelihood scale would affect the detailed semantics of epistemic expressions.

#### 4.1.6 Summary: The scales of *probable* and *likely*

Summarizing the results of this section, evidence from degree modification suggests that  $\mathcal{S}_{likely}$  and  $\mathcal{S}_{probable}$  are upper-bounded ratio scales. General theoretical considerations involving the likelihoods of tautologies and contradictions also indicate that they are both upper- and lower-bounded. This classification explains the failure of the Disjunctive Inference and the acceptability of ratio modifiers with *likely* and *probable*. While the latter data point is not proof positive, various alternative explanations for the occurrence of ratio modifiers, suggested by a consideration of corpus examples, do not seem to provide a convincing explanation of the acceptability of ratio modifiers.

If this classification is correct, then we have a qualitative derivation of the claim, made in a number of recent formal semantics papers, that *likely* and *probable* live on scales of finitely additive probability. Turning to a consideration of the individual axioms required to be sure of this conclusion—**upper- and lower-boundedness**, **reflexivity**, **transitivity**, **connectedness**, **monotonicity**, **positivity**, **solvability**, and **Archimedean**—I suggested that solvability and connectedness are the only ones that are potentially contentious, and that solvability is in fact quite reasonable when we restrict attention to semantic applications. Whether connectedness is also a reasonable assumption depends to a considerable extent on the appropriate interpretation the probability scale, and evidence is not yet clear on this front. Fortunately, there are generalizations of qualitative probability scales that do not assume connectedness and are representable by sets of probability measures. While further work will be needed to confirm that  $\mathcal{S}_{likely}$  and  $\mathcal{S}_{probable}$  have the requisite formal properties, these generalizations would at least have the potential to render ratio and proportional modifiers interpretable in the sense of chapter 2: comparability is not required, but merely that all of the admissible measures agree on the comparison in question.

## 4.2 Boundedness and the relative/absolute distinction

This section considers the previous section's conclusions from a different perspective: the relationship between scalar boundedness, degree modification, and the minimum/maximum/relative classification of adjectives. This has been an issue of considerable interest to formal semanticists in recent years, principally because of the seminal work of Kennedy & McNally (2005a) and Kennedy (2007). Situating the analysis of *likely/probable* just given within the literature on scalar adjectives in general is, of course, an important part of this book's overall mission to integrate theories of modal and scalar semantics. In addition, it raises a number of empirical issues that will be important for the analysis of *certain* and *possible* in chapter 5—especially in the initial discussion in sections 4.2.1–4.2.3.

We noted above that it is natural to suppose that  $\mathcal{S}_{likely}$  and  $\mathcal{S}_{probable}$  are upper-bounded by *W* (a tautology) and lower-bounded by  $\emptyset$  (a contradiction). In addition, these adjectives pattern on various tests with relative adjectives such as *tall* and *expensive*, and behave differently from absolute

adjectives—both minimum adjectives such as *dangerous*, *wet* and maximum adjectives such as *safe*, *dry*, *full*. However, these observations are in tension with each other if a prominent theory of the relative/absolute distinction due to Kennedy (2007) is correct: Kennedy’s principle of “Interpretive Economy” rules out the possibility that relative adjectives can ever occur on scales with maxima or minima. In this section I will explain why this is, discuss a proposed modification due to Klecha (2012, 2014), which involves assigning a non-obvious scale structure to *likely* and *probable*, and explain why I think the modification is both unmotivated and empirically problematic. I will argue that there are numerous examples in English of relative adjectives that are associated with bounded scales, and that *likely* and *probable* simply add two more to the list.

If this argument is correct, the tension between the relative classification of *likely/probable* and the theory of Kennedy (2007) does not give us a reason to assign a non-obvious structure to  $\mathcal{S}_{likely}$  and  $\mathcal{S}_{probable}$ . Instead, it simply provides another counter-example to the purported empirical generalization that relative adjectives cannot occur on bounded scales. The correlation between scalar boundedness and relative/absolute status that Interpretive Economy is designed to explain seems to be real. However, since it is not a perfect correlation, it should not be explained by this or any other general principle of grammar or processing. This conclusion leaves unexplained some empirical questions, first pointed out by Portner (2009), about the lack of a degree reading of modifiers such as *completely* and *slightly* with *likely/probable*. I will suggest that these data can be explained if the modifiers in question embed the positive form, rather than the adjectival root (compare Klein 1980; Kennedy & McNally 2005b; Burnett 2016). While this hypothesis is still speculative in the case of *completely*, it has substantial empirical support as applied to *slightly*, and may be more broadly (even generally) applicable.

What is at stake here? It varies, depending on whether you are primarily interested in modality or in scalar adjectives. For theorists interested in scalar adjectives, the discussion should be of considerable interest, since it calls into question some widely assumed empirical generalizations about, and a prominent theoretical model of, the relative/absolute distinction. For those mainly interested in modality, this section’s discussion will affect only the detailed analysis of the important items *likely* and *probable*. In particular, this book’s broader effort to situate modal semantics within scalar semantics would scarcely be affected if we were to adopt Kennedy’s theory, and with it Klecha’s endpoint-free analysis of *likely* and *probable*. Very little in the treatment of other modal items in subsequent chapters will hinge crucially on the conclusions of this section. The only important exception is that, based on the conclusions of this section, I will give serious consideration in chapter 5 to several hypotheses about *possible* and *certain* which would be trivially false if Kennedy’s (2007) theory were adopted wholesale. In any case, readers who are mainly interested in modality may, without loss of continuity, skip to chapter 5 after reading through section 4.2.3.

#### 4.2.1 Tests for Boundedness and Minimum/Maximum/Relative Classification

This section reviews briefly several linguistic diagnostics that have been used to infer boundedness properties and minimum/maximum/relative classification.

**1. Maximizing degree modifiers.** *Completely* is a polysemous modifier, but one common function is as a maximizer. An adjective modified by *completely* can have a “maximum degree”

interpretation only if it is a maximum-standard absolute adjective, like *full* or *straight*.<sup>4</sup> This generalization explains the contrast in (4.32), where *completely* induces a “maximum” interpretation with the maximum adjective *full* but not with the minimum adjective *bent* or relative adjective *tall*.

- (4.32) a. The room is completely full.  
 b. Bill is completely tall.  
 c. The rod is completely bent.

Similar observations hold of maximizers like *perfectly*, *totally*, *absolutely*, though there is a certain amount of variability among these modifiers in the availability and meaning of non-degree-modifying interpretations.

**2. *Almost*.** A related test involves the observation that, if an object is *almost Adj*, then it typically fails to be *Adj* by a small margin (Rotstein & Winter 2004). Among the adjective classes that we are interested in here, *almost* is generally acceptable only with maximum adjectives.

- (4.33) a. This area is almost safe.  $\leadsto$  This area is not completely safe.  
 b. The glass is almost full.  $\leadsto$  The glass is not completely full.  
 c. # Bill is almost tall.  
 d. # This rod is almost bent.

**3. “*Adj* but could be *Adj-er*”:** Minimum and relative adjectives are natural in the construction in (4.34), but maximum adjectives are not (Kennedy 2007).

- (4.34) a. The rod is bent, but it could be more bent.  
 b. Bill is tall, but he could be taller.  
 c. # The room is full, but it could be fuller.

This can be explained by the hypothesis that only maximum adjectives require in the positive form that the object have the maximum possible degree of the property, so that the continuation *but it could be Adj-er* is contradictory.

**4. *Slightly-modification*.** According to Rotstein & Winter (2004); Kennedy (2007), *x is slightly Adj* is true just in case *x* is *Adj* to a degree which differs by a small amount from the minimum degree in  $\mathcal{S}_{Adj}$ . Thus if an adjective can be modified by *slightly*, it is associated with a scale with a bottom element, and, in general, is a minimum-standard adjective.

- (4.35) a. The rod is slightly bent.  
 b. # Bill is slightly tall.  
 c. # The glass is slightly full.

Note that merely being on a scale that has a minimum element does not seem to be sufficient for *slightly-modification*. For example, *slightly full/empty* lack a “just greater than minimum” reading, despite the fact that both adjectives are scales that have a lower-bound. As Solt (2012) and Sassoon (2014) point out, the controlling factor appears to be the kind of meaning that the positive form

<sup>4</sup> Compare Kennedy & McNally 2005a; Kennedy 2007, who frame this generalization in somewhat different terms—see below.

has: *slightly*-modification is acceptable with minimum adjectives, among other constructions that provide sufficiently precise thresholds for comparison. (This point will be important later because it problematizes the theory of degree modification in Kennedy 2007, which assumes that modifiers attach directly to the root (denoting a measure function) rather than the full positive form (denoting a predicate of individuals).)

**5. Classification of antonym.** As Kennedy & McNally (2005a) point out, relative adjectives have relative antonyms. The antonyms of absolute adjectives are also absolute. (They may vary between either minimum or maximum if the scale is closed so that both options are formally possible.) The reader can verify these classifications by considering the antonyms in light of the other tests described here.

- (4.36) a. Bill is rich/poor. (relative/relative)  
 b. This is an expensive/cheap car. (relative/relative)  
 c. Skiing is safe/dangerous. (maximum/minimum)  
 d. Smoking and lung cancer are correlated/uncorrelated. (minimum/maximum)  
 e. The cafe is full/empty. (maximum/maximum)

**6. Negation entails antonym?** Minimum adjectives frequently have maximum antonyms. In such minimum/maximum pairs, the negation of the minimum member entails that the maximum antonym holds, and likewise the negation of the maximum entails that the minimum antonym holds. In contrast, relative adjectives have a “zone of indifference” (Sapir 1944; Kennedy 2007), so that it is possible to be neither  $A_{pos}$  nor  $A_{neg}$ .

- (4.37) a. The rod is not bent.  $\models$  The rod is straight.  
 b. The rod is not straight.  $\models$  The rod is bent.  
 c. Bill is not tall.  $\not\models$  Bill is short.  
 (4.38) a. # The rod is not bent, but it is not straight either.  
 b. # The rod is not straight, but it is not bent either  
 c. Bill is not tall, but he is not short either.

However, some maximum adjectives on fully-closed scales resemble relative adjectives in this respect, because their antonyms are also maximum adjectives. For example, both members of the maximum/maximum pair *full/empty* can be negated without contradiction.

- (4.39) a. The glass is not full.  $\not\models$  The glass is empty.  
 b. The glass is not full, but it is not empty either.

**7. Sensitivity to comparison classes.** Kennedy (2007) points out that relative adjectives are robustly sensitive to comparison classes, while absolute (minimum or maximum) adjectives are much less so. This feature emerges in two ways. First, contextually provided alternatives tend to produce much more variation in the interpretation of relative adjectives. For example, someone might be judged *tall* when surrounded by very short people, but not when surrounded by very tall people. (See Syrett, Kennedy & Lidz 2010 for experimental evidence corroborating this claim, and evidence of its (near-)absence in absolute adjectives.) Second, relative adjectives are more felicitous with explicit comparison classes.



- (4.40) a. Bill is tall for a fourteen-year-old.  
 b. ?? This room is full for a classroom.  
 c. ?? This cloth is wet for an cotton towel.

#### 4.2.2 The classification of *likely* and *probable*

Almost without exception, *probable* and *likely* behave like the relative adjectives *tall* and *heavy* (Lassiter 2010). For example, a degree-maximizing reading of *completely* is unavailable in both sentences in (4.41), though *completely* can have other interpretations (e.g., speaker confidence in the assertion's truth).

- (4.41) a. # Bill is completely tall.  
 b. # It is completely likely/probable that the Jets will win the Super Bowl.

This suggests that *probable* and *likely* are not maximum adjectives. This indication is supported by the observation that both allow the construction “*Adj* but could be *Adj*-er”—

- (4.42) a. Bill is tall, but he could be taller.  
 b. It is likely that the Jets will win, but it could be more likely.

—and that both are odd when modified by *almost*.

- (4.43) a. # Bill is almost tall.  
 b. # It is almost likely/probable that the Jets will win.

*Slightly*-modification is distinctly odd with both adjectives, indicating that they are not minimum adjectives.

- (4.44) a. # Bill is slightly tall.  
 b. # It is slightly likely/probable that the Jets will win.

Relative classification is directly supported by several observations. First, the antonyms of *likely* and *probable* are relative, as we would expect.

- (4.45) a. Bill is #completely/#slightly short.  
 b. It is #completely/#slightly unlikely/improbable that the Jets will win.

Second, both *likely* and *probable* show a zone of indifference, consistent with a relative classification.

- (4.46) a. Bill is not tall, but he is not short either.  
 b. It is not likely that the Jets will win, but it is not unlikely either.

Third, *likely* and *probable* resemble *tall* and *heavy* in being sensitive to contextual alternatives. Teigen (1988) and Windschitl & Wells (1998) provide extensive experimental evidence for this feature, using scenarios which vary in the distribution of probabilities among non-maximal alternatives. For example, one group of subjects might be asked to judge the description “A is likely to win” when A is described as having a 40% chance of winning and a 60% chance of not winning. Another group was asked to judge the same sentence when A was described as having a 40% chance of winning,

and six other individuals were described as having a 10% chance of winning. Participants were robustly more willing to endorse the description in the latter case. While Teigen and Windschitl & Wells interpret this result as evidence of reasoning errors, it is natural from a linguistic perspective to interpret it as the same type of alternative-sensitivity that relative adjectives generally display. See Yalcin 2010 and Lassiter 2011a: §4 for further discussion of this effect and its interpretation.

### 4.2.3 The curious case of proportional modifiers

A notable point of difference between *likely/probable* and the more familiar relative adjectives is that the former can sometimes be modified by certain proportional modifiers, as we saw in §4.1. Here are several more examples, drawn from news sources.

(4.47) [I]t's 80 percent likely that the iPhone will be coming to T-Mobile ...

(4.48) [T]he IPCC ... said it was “very likely” or more than 90 percent probable that human activities ... had caused most of the warming in the past half century.

Proportional modifiers are unacceptable with most relative adjectives, and they tend to be most natural with adjectives on fully closed scales, like *full/empty* and *open/closed* (cf. Kennedy & McNally 2005a: 352-3). This distribution cannot be attributed to the fact that these are maximum adjectives, since maximum adjectives on partially open scales—such as *safe* and *dry*—do not accept percentage modifiers.

- (4.49) a. # Bill is 70% tall/short.  
 b. The glass is 70% full/empty.  
 c. # The neighborhood is 70% dangerous/safe.

(4.49c) can of course be interpreted as quantifying over parts of the neighborhood: “70% of it is dangerous/safe”. Presumably, (4.49a) is totally unacceptable because quantification over parts does not make sense in this example.

Interestingly, though, there are certain proportional modifiers that seem to be equally unacceptable with all relative adjectives, including *half* and *mostly*.

- (4.50) a. # Bill is half/mostly tall.  
 b. # The Jets are half/mostly likely to win.

The difference between *n%* and *half/mostly* suggests that there may be two kinds of proportional modifiers. Recall the truth-conditions for simple sentences with *n%* and *half* suggested in ch.2, §2.4.4:

- (4.51) a.  $\llbracket x \text{ is half } P \rrbracket^{\mathcal{M},w} = 1 \text{ iff } \frac{\mu_P(x)}{\mu_P(\top_P)} = .5.$   
 b.  $\llbracket x \text{ is } n\% P \rrbracket^{\mathcal{M},w} = 1 \text{ iff } \frac{\mu_P(x)}{\mu_P(\top_P)} = n/100.$

From these proposals, we would expect *half* and *n%* to pattern together always, since both require a scale which (a) is upper-bounded; (b) is at least asymptotically bounded below, and (c) is able to support stable units—so, either interval or ratio. I argued above that *likely* and *probable* satisfy this

description: their scales are fully closed and ratio. But in this case, it is a mystery why *half likely* is not acceptable, with the same interpretation as *50% likely*, i.e. “exactly as likely as not”.<sup>5</sup>

My best guess at this point is that percentage modifiers are indeed interpreted as in (4.51), so that they place restrictions only on the structure of the scale underlying the adjective that they modify. *Half* and *mostly*, in contrast, have more stringent requirements. For clear semantic reasons, they can only modify adjectives on fully closed (interval or ratio) scales; but they also select somehow for maximum adjectives (and certain other constructions, such as equatives). According to this speculation, *half likely* and *mostly likely* are ruled out not because of *likely*’s scale structure, but because it is a relative adjective.<sup>6</sup>

Empirically, this account would predict that other relative and minimum adjectives on fully closed scales should accept percentage modifiers but reject modification by *mostly* and *half*. Such examples are difficult to find: for example, all of the examples of relative adjectives on closed scales discussed below (§4.2.7) involve scales with one bound only. However, chapter 5 will discuss one

<sup>5</sup> Note, however, that the first requirement could be weakened: as we noted in chapter 2 (§2.4.4), these modifiers could be given defined in terms of limits rather than bounds. If so, the requirement would be merely that the scale be asymptotically bounded above. As I mentioned there, I do not know of any way to test these alternative hypotheses directly. This alternative might seem congenial to Klecha’s (2012; 2014) account of *likely*, according to which the likelihood scale is  $(0, 1)$  with an asymptotic upper bound. However, in this case we would wrongly expect that *half likely* should be acceptable, in addition to *n% likely*. So, the divergence between *n%* and *half* presents a puzzle regardless of how we treat *likely/probable*.

Klecha’s explicit proposal is in fact different: he suggests that *half* is ruled out by the absence of an upper bound in  $(0, 1)$ , as we would expect on his proposal given (4.51). He explains examples like (4.47)–(4.48) by positing that the modifiers there are actually measure phrases which happen to be homophonous with percentage modifiers as used in e.g. *70% full*. A homophony analysis does not strike me as very attractive, and the single argument given in its favor—based on ambiguous cases like the following—is unconvincing.

(4.52) Clinton’s likelihood of winning increased 10% last week.

- a. ✓ The likelihood went from 60% to 70%.
- b. ✓ The likelihood went from 60% to 66%.

All that this example shows is that there is sometimes flexibility in which interval is relevant—either 0 to 1 (open or closed), or 0 to 0.6 (bounded above by Clinton’s previous likelihood of winning). The latter reading is also the relevant one in cases like *The item’s price increased 20%*, where the lack of a (true or asymptotic) upper bound renders only the context-sensitive interpretation available.

<sup>6</sup> Perhaps this issue is related to another set of divergences among *half*, *mostly*, and *n%*: the fact that *half* is acceptable as a modifier of equatives, but not of comparatives.

(4.53) a. Bill is 50%/half/#mostly as tall as Mary.  
b. Bill is 50%/#half/#mostly taller than Mary.

Comparatives behave in other respects like minimum adjectives—e.g., in accepting modification by *slightly*—so it is not too surprising on the present hypothesis that they would reject modification by *half* and *mostly*. Is there a sufficient connection between equatives and maximum adjectives like *full* to explain why they would both allow *half*-modification? Perhaps there is. Both allow *almost*-modification, unlike minimum adjectives and comparatives. Both are necessarily associated with scales that have fixed maxima—for example, *as tall as Mary* is associated with the height subinterval  $(0, \mu_{\text{height}}(\text{Mary})]$ . There are also some important disanalogies, e.g., *completely full* vs. ??*completely as tall as Mary*, and of course the nonexistence of comparatives built from the equative (*x is fuller than y* vs. the impossible *\*Bill is more as tall as Mary than Sue is*). In addition, it remains unclear why *mostly* is unacceptable as a modifier in (4.53b). It seems that more is needed to understand this class of modifiers fully.

such case—the minimum adjective *possible*. Regardless of how the main theoretical controversy discussed there is resolved (whether the scalar interpretation of this adjective is basic or derived), we will see considerable empirical evidence that the scalar *possible* is a minimum adjective and that it accepts percentage modifiers. The latter property is illustrated by the naturally-occurring examples in (4.54).

- (4.54) a. The DHS report says it is more than 50 percent possible that EMFs could cause a very small increased lifetime risk of childhood leukemia, adult brain cancer and Lou Gehrig's Disease.  
 b. So was it possible, even one thousandth of one percent possible, that I had married a murderer? No, of *course* not.

(See §5.2 below for more such examples.) However, *possible* appears to reject modification by *half* and *mostly*: if the (a)-examples below (modified from (4.54)) mean anything, it is certainly not what the (b)-examples describe.

- (4.55) a. The DHS report says it is half possible that EMFs could cause a very small increased lifetime risk of childhood leukemia ...  
 b. "It is 50% possible that ..."
- (4.56) a. Was it mostly possible that I had married a murderer?  
 b. "Was is more than 50% possible that I had married a murderer?"

This is as the hypothesis floated here about two kinds of proportional modifiers would lead us to expect: percentage modifiers are acceptable and interpretable, while *half* and *mostly* are not. So, the existence of this distinction does seem to enjoy some empirical support.

I do not have a detailed theoretical explanation of why some proportional modifiers would be sensitive to the meaning of the positive form while others are not, but I should note that this situation is predicted to be possible by the theory of degree modification proposed by Kennedy & McNally (2005b), which is presented as an effort to remedy some empirical defects of the theory in Kennedy & McNally 2005a. Kennedy & McNally (2005b) argue that some modifiers embed the root directly, while others embed the positive form and others still serve to transform the underlying scale. If *half* and *mostly* are of the type that embed the positive form, then syntactic and semantic characteristics of the positive form are in principle available to condition the meaning and acceptability of a complex expression containing the modifier. The contrast between *half full* and *# half likely/possible* may then be explicable in terms of the difference in positive-form meaning, attributable to the fact that *full* is a maximum adjective while *likely* and *possible* are not. This could hold even while their scales have the same formal structure.

If the suggestion is correct, it still leaves an interesting empirical issue to be explained: percentage modifiers with *likely* are not uniformly distributed in the 0-100 range. Several linguists have suggested to me that values of *n* in the 50-90% range are more natural than those in the 0-50% range, or those above 90%. I share this intuition, and my informal impression from examination of many corpus examples is that it is borne out by the statistical distribution of these modifiers. Interestingly, *n% likely* is most acceptable in the range in which unmodified *likely* would be natural. This gloss

even extends beyond *likely*: the examples of *n% possible* and *n% certain* that I have found seem to cluster in the probability ranges in which one might be inclined employ these adjectives in the positive form.

I have no theoretical explanation for this pattern. However, I suspect that it is not grammatical in origin, but rather falls into a broader category of non-categorical usage preferences—alongside, for example, ordering preferences in syntactic alternations (Wasow 2002; Bresnan 2007). While not very well-understood theoretically, there is no doubt that such preferences exist in a wide variety of strengths, and that they are not reducible to categorical rules/constraints of the type that we are focusing on here (see, for example, Bresnan, Cueni, Nikitina, Baayen et al. 2007; Bresnan & Ford 2010).

If so, we may expect in large corpora to find counter-examples to the observed statistical pattern, just as Wasow and Bresnan find many examples which go counter to the trend of the usage preferences that they document. And we do: it is easy to find naturally-occurring examples of *n% likely* outside of the 50-90 range.

- (4.57) a. If the condition develops untreated, intellectual impairment is 97 percent likely.  
 b. Foresters and preschool teachers are also less than 1 percent likely to have their job replaced by automation in the next 15 years.

#### 4.2.4 Can relative/absolute status be reduced to scalar boundedness?

The linguistic inferences derived from patterns of modification that were described in §4.2.1 were phrased in terms of a classification of adjectives due to Kennedy & McNally (2005a); Kennedy (2007). Two kinds of distinctions were discussed: the presence or absence of minimum and maximum elements on an adjective's scale, and whether the interpretation of the positive form is *relative* or *absolute*. The absolute adjectives subdivide into the *minimum* and *maximum* adjectives.

Relative/absolute status and scalar boundedness are clearly not unrelated: for example, it would not make sense for a minimum adjective to be on a scale with no minimum element, and similarly a maximum adjective cannot be on a scale with no maximum element. On the other hand, a relative adjective could in principle be on a scale with a maximum or minimum; and minimum/maximum/relative status places no *a priori* constraints on what kinds of adjectives could be located on fully closed scales. These facts suggest the following set of one-way implications, driven by purely conceptual considerations—by what is necessary before we can even talk sensibly about the positive-form interpretation involving a minimum or maximum degree.

- (4.58) Suggested typology from §4.2.1:  
 a. Minimum adjective  $\implies$  lower-closed **or** fully closed  
 b. Maximum adjective  $\implies$  upper-closed **or** fully closed  
 c. Relative adjective  $\implies$  fully open **or** lower-closed **or** upper-closed **or** fully closed

However, Kennedy (2007) argues that the relationship between boundedness and scale structure is closer: relative/absolute status is fully predictable from scalar boundedness. Conceptually, this means flipping the direction of the implication:

- (4.59) Typology from Kennedy 2007:
- a. Fully open scale  $\implies$  relative adjective
  - b. Lower-closed scale  $\implies$  minimum adjective
  - c. Upper-closed scale  $\implies$  maximum adjective
  - d. Fully closed scale  $\implies$  minimum or maximum adjective

Notably, since fully open scales are the only ones that can generate relative interpretations of adjectives, this implication is bidirectional: an adjective is relative if and only if its scale is fully open.

In contrast to the purely meaning-driven restrictions summarized in (4.58)—e.g., “no maximum adjectives on scales without maxima”—Kennedy’s alternative set of generalizations needs some further theoretical explanation. As he notes, there is no obvious reason why a relative adjective could not occur on a fully closed scale. Why doesn’t English have a relative adjective *schmull* meaning something like “fuller than most things in the reference class”?

Kennedy (2007) explains this apparent gap using two theoretical components. First, he proposes (Kennedy 2007: §2.3) that the positive form has a single, unified meaning:

- (4.60) A positive-form scalar adjective *Adj* is true of an individual *x* in a context *c* if and only if *x* “stands out” in *c* relative to the kind of measurement that *Adj* encodes.

While informal, this gloss seems to be more or less correct for relative adjectives (though see §4.2.8 below for some problems). However, (4.60) seems to predict incorrectly that minimum and maximum adjectives should have similarly context-sensitive meanings. To explain why they do not, Kennedy (2007: §4.3) proposes a general principle of interpretation/processing which he calls “Interpretive Economy” (IE):

- (4.61) Maximize the contribution of the conventional meanings of the elements of a sentence to the computation of its truth conditions.

The relevant conventional aspects of meaning, in this context, are assumed to be scalar boundedness properties. When there is a scalar endpoint, we can maximize its contribution to the interpretation if we resolve “stands out” as meaning “meets or exceeds the scalar endpoint”. This yields a minimum or maximum interpretation of the positive form. However, when no endpoint is available, ordinary context-sensitivity kicks in, and “stands out” is resolved by reference to comparison classes, expectations, and other aspects of context. This gives us a standard relative interpretation. The cumulative effect of (4.60) and (4.61), then, is that adjectives on open scales must be relative and context-sensitive, while adjectives on partially or fully closed scales must be absolute—maximum or minimum, as appropriate.

#### 4.2.5 Implications of Kennedy’s theory for *likely* and *probable*

Earlier in this chapter, I used the observation that nothing can be more likely than a tautology or less likely than a contradiction to motivate assigning *likely* and *probable* to fully closed scales. However, if the empirical generalizations that Kennedy (2007) is trying to capture hold without exception—or, perhaps more vividly, if (4.60) is correct and IE is an general constraint on the grammar and

processing of scalar adjectives—the conclusion that these adjectives’ scales are bounded is not compatible with the observation that they are relative adjectives. They must be on open scales, or they cannot be relative.

A second problem for the probabilistic semantics was first noted by Portner (2009: §3): these adjectives do not maximize with *completely*. For instance, (4.62a) cannot be understood as conveying (4.62b).

- (4.62) a. It’s completely likely/probable that it will snow.  
 b. “Snow has the maximum possible probability, i.e., probability 1.”

In Kennedy’s (2007) theory of degree modification, *completely* should combine freely with scalar adjectives whose scale has a maximum point, yielding a “maximum degree” interpretation. So, according to this theory of modification, (4.62) implies that  $\mathcal{S}_{likely}$  and  $\mathcal{S}_{probable}$  are not upper-bounded.

In addition, *likely* and *probable* do not get a “just above minimum” reading with *slightly*: (4.63a) is not equivalent to (4.63b).

- (4.63) a. It’s slightly likely/probable that it will snow.  
 b. “The probability of snow is just above the minimum possible degree of 0.”

Again, in Kennedy’s (2007) theory *slightly* should combine freely with scalar adjectives whose scale has a minimum point. So, on this account (4.63) shows that  $\mathcal{S}_{likely}$  and  $\mathcal{S}_{probable}$  are not lower-bounded.

Given the overwhelming evidence that *likely* and *probable* are relative adjectives, there seem to be two possible responses at this point. First, we could embrace Kennedy’s theory of modification and the positive form, and conclude that  $\mathcal{S}_{likely}$  and  $\mathcal{S}_{probable}$  are open scales after all. Second, we could reject Kennedy’s theory, as Lassiter (2010) does, opting for an account in which relative adjectives can fall onto closed scales, and where degree modifiers can be sensitive to minimum/maximum/relative status in addition to scalar boundedness properties. I will consider these options in turn.

#### 4.2.6 Klecha’s account

Klecha (2012, 2014), following an idea floated (but not endorsed) by Portner (2009: §3), proposes to resolve this tension by restricting  $\mathcal{S}_{likely}$  and  $\mathcal{S}_{probable}$  to the open interval  $(0, 1)$ . This scale is formally distinguished from the open interval  $(-\infty, \infty)$ —which forms the range of open-scale adjectives like *beautiful*—in that the excluded endpoints exist at the conceptual level, and it is still possible to state formal constraints that make reference to them. (This feature is crucial, for example, if you want the probability concept being invoked to relate to a broadly Bayesian conception of belief and information uptake, which would not work technically if the endpoints simply did not exist.) In effect, the suggestion is to assume the existence of a measure *prob* from which  $\mathcal{S}_{likely}$  and  $\mathcal{S}_{prob}$  are derived in a simple way.

$$\mu_{likely} = \mu_{probable} = \{(p, r) \in prob \mid r \notin \{0, 1\}\}$$

(This does not correspond precisely to Klecha’s proposal, but it is in the spirit of this proposal, and it avoids a formal problem.<sup>7</sup>) Here, the idea is that the ordinary probability is basic—it forms the “conceptual scale”, as I will call it—and the “linguistic scale” associated with the item *likely* is derived from it by removing the endpoints. More generally, Klecha (2014: §2) suggests that scalar endpoints that exist at the conceptual level can either be included in the range of a measure function, or excluded and approached only asymptotically. As he points out, this assumption expands the typology of boundedness-related scale types in Kennedy’s theory from four possibilities to nine.

On this account, *likely* and *probable* are additive, but they have open scales which exclude the endpoints of 0 and 1. This is an attractive move in several respects. It gives us an IE-compatible explanation of why *likely* and *probable* are relative adjectives. Since their scale is open, they must be relative. It also explains (4.62) and (4.63): if *completely* and *slightly* are scalar modifiers whose interpretation refers to scalar endpoints, we expect that they will not be able to modify adjectives whose scales lack these endpoints.

However, Klecha’s proposal also encounters some theoretical problems. The motivation for assigning a non-obvious scale structure to *likely/probable* is suspiciously theory-internal. The account is predicated on the assumption that Kennedy’s theory is correct in the essentials, and no independent evidence is given for the prediction that probability 1 events cannot be described as “likely” or “probable”, and that probability 0 events cannot be described as “unlikely” or “improbable”.

In fact, as Klecha acknowledges, there is evidence that such events **can** be so described. Consider the dialogue in (4.64):

- (4.64) a. **A:** It is likely to rain tomorrow?  
 b. **B:** Yes. In fact, it’s absolutely certain to rain tomorrow.

By responding “yes”, **B** affirms that the probability of rain falls somewhere above  $\theta_{likely}$  on the scale  $\mathcal{S}_{likely}$ . But, if Klecha’s account is correct, then the fact that the probability of rain falls anywhere on  $\mathcal{S}_{likely}$  implies that the probability of rain is not 1: that degree has been excluded from the scale. **B**’s response should thus be a contradiction, given that *absolutely certain* implies probability 1 (see chapter 5, §5.1).

Lest the discussion begin to rely too heavily on invented data, here are two examples from news reports that make the same point. Here, the author of (4.65) describes an event as *extremely likely* which they think may be certain, and the author of (4.66) considers the same event to be both *very unlikely* and *perhaps impossible*.

<sup>7</sup> Specifically, Klecha attempts to axiomatize the scale directly, assuming additivity but non-upper-boundedness. But absurd consequences follow if we do not restrict the additivity constraint to pairs of propositions whose disjunctions would not have probability 1 in a standard version of probability—e.g.,  $\phi$  and  $\neg\phi$  for arbitrary  $\phi$ . Otherwise, we would have to deny that it is ever sensible to talk about the likelihood of both  $\phi$  and  $\neg\phi$ , or else allow the tautology  $\phi \vee \neg\phi$  to have less-than-maximal likelihood. The alternative definition used in the main text avoids this problem, and also makes explicit the assumption that the endpoints really do exist underlyingly. Klecha also seems to make this assumption, in his appeal to “coercion” processes which can lead to the inclusion of probability 0 or 1 in the range of  $\mu_{likely}$  in examples like *It’s likely to rain—in fact, it’s certain to* (2014, p.63-64).

Note also that it not matter for our purposes whether we are dealing with unconditional probability, as I have been assuming for simplicity, or modified Popper functions as Klecha proposes.



- (4.65) The White House has been frantically briefing news organizations about the proposals President Obama is going to make tonight—and in any case it's extremely likely (perhaps certain?) that his entire agenda will be dead on arrival in the Republican-controlled Congress.
- (4.66) In multicultural Kashmir, mosques, temples and gurdwaras have offered shelter to flood-affected victims, irrespective of the religion to which they belonged. Would this be true of PoK? Not just very unlikely, but perhaps impossible for the simple reason that PoK, like the rest of Pakistan, is virtually an exclusively Islamic state in which other faiths have little, if any, presence.

In both cases, Klecha's theory of *likely* would lead us to expect that the author's speculations should be incoherent, since they are affirming a claim ("It is extremely likely/very unlikely") which is incompatible with another claim parenthetically indicated to be possible ("It is certain/impossible"). We should, I think, expect an effect similar to the bizarre move of denying one's own presupposition just after making it:

- (4.67) a. # It was Bill's sister (perhaps Bill doesn't have a sister) who I saw.  
 b. # Mary realizes that it's raining (perhaps it isn't raining).

Klecha (2014: 63-64) anticipates this objection and responds by suggesting that  $\mathcal{S}_{likely}$  can sometimes be modified to include the endpoints. As he puts it, "scales can be coerced to include end points when there is an intuitive way to include the end points" (p.63). An immediate problem with this suggestion is that the addition of an unconstrained coercion operation effectively neutralizes the predictive power of the theory. Absent some articulated theory of "coercion", the account seems worrying close to the vacuous claim that *likely*'s scale lacks endpoints except when it has them.

In addition, the invocation of scalar coercion in this context fails to distinguish between the two cases that Klecha crucially needs to explain: the acceptability of *likely*, *indeed certain* (4.64), and the unacceptability of degree-modifying readings of *completely/slightly likely* ((4.62) and (4.63)). If a coercion operation is available to modify the scale to include probability 1 in (4.64), why is this operation not also available when we are interpreting *completely likely*? There is an equally intuitive way to include the maximum point in such examples, and yet the predicted reading of *completely likely* as "absolutely certain" does not exist. The same goes for *slightly likely*, where a freely available coercion operation should be able to insert the minimum point as needed, yielding an interpretation paraphrasable by "slightly possible". Klecha's account of the surprising comparatives in (4.64) thus deprives us of an explanation of the oddity of (4.62) and (4.63).

Since Klecha's repair strategy is problematic, we have good reason to reconsider the assumption that relative *likely* and *probable* must be associated with open scales. Recall that this proposal relies heavily for motivation on Kennedy's (2007) theory of degree modification and positive form semantics: if Kennedy's account is correct, then surprising consequences follow for *likely* and *probable*. In what follows I will lay out a case that both planks of Kennedy's account are problematic. The elegant theory of modifiers like *slightly* and *completely* that it assumes is problematized by evidence that *slightly*, at least, is sensitive to the minimum/maximum/relative status of the positive form of the adjectives that they combine with. I will hypothesize that *completely* is as well. In addition, Interpretive Economy—the crucial assumption that rules out relative adjectives on closed

scales—is subject to numerous counter-examples which can at best be explained with awkward theoretical stipulations. In the end, I will suggest that there is no real tension between the relative status of *likely* and *probable* and the intuitive boundedness of their scales: relative adjectives do occur on closed scales, and certain degree modifiers fail to combine with them because their positive form has a relative interpretation.

#### 4.2.7 Other relative adjectives on closed scales

IE has a number of virtues, not least that it allows us to avoid stipulating minimum/maximum/relative classification of adjectives as an arbitrary lexical property. In the system of Kennedy & McNally (2005a), this classification is a lexical feature of adjectives, and the different meanings in the positive form are generated by combination with three different *pos* morphemes, which are sortally restricted to combine with adjectives of the right type. Kennedy's (2007) proposal is attractive in that it makes it possible to have a single *pos* morpheme and no item-by-item stipulations of an adjective's place within the minimum/maximum/relative typology.

However, IE has theoretical costs as well. As Potts (2008) points out, IE is “the only aspect of [Kennedy's] theory that is not grounded in functional denotations for morphemes or well-established forms of context dependency”. It is not clear what kind of principle IE is, or how it could be implemented concretely in the grammar or processing system. A second problem pointed out by Potts is that IE is “an optimization principle left unsupported by a theory of optimization”. While Kennedy claims to have derived specific predictions from the principle, the notion of “maximization” that these predictions derive from is not specified beyond the intuitive level. There is no indication of what metric the processor could use to determine whether a given interpretation satisfies IE relative to some set of alternative interpretations, or indeed of how the alternatives to be considered are selected in the first place.

In this section and the next I will discuss two more empirically-oriented objections to IE. First, the problematic involving *likely/probable* involved the supposed impossibility of relative adjectives on bounded scales; however, it turns out that English has a substantial number of relative adjectives that fall onto intuitively closed scales. These can be explained within Kennedy's theory by gerrymandering scale structure to exclude the endpoints, but this strategy has only theory-internal motivation, and becomes less plausible with repeated invocation. In addition, each such example encounters the same problem that we noted for Klecha's account above: it predicts that certain discourses should be ill-formed which are in fact natural and readily interpreted. Second, I will argue briefly that a crucial plank of Kennedy's theory—the “stands out” interpretation of the positive form in (4.60)—is problematized by the existence of multiple relative adjective adjectives falling on the same scale. Neither of these objections is absolutely killer: with the addition of further mechanisms, these data could probably be accounted for. But the cumulative effect of the stipulations that would be needed is to render the theory less compelling, and in the process its surprising implications for *likely* and *probable*.

In the significant literature on the relative/absolute distinction generated by Kennedy's (2007) seminal work, it has generally been taken for granted that the empirical generalization motivating IE is correct: relative adjectives occur only on open scales. However, Kennedy (2007) already noted

an apparent counter-example involving adjectives of cost. Cost is intuitively a lower-bounded scale, with zero cost serving as the lower bound. If so, IE predicts that *expensive*, since it is on a scale with a minimum point, should be a minimum adjective meaning “cost greater than zero”, i.e. “not free”. In addition, since its antonyms *inexpensive* and *cheap* are on an upper-bounded scale—the inverted cost scale—they should be maximum adjectives meaning “free”.

Kennedy argues that these examples do not constitute counter-examples to IE, but rather that they show that the linguistic scale associated with cost adjectives is not the same as the conceptual scale of cost. Instead, the linguistic scale is the conceptual scale with the zero point removed. (Note that the terminology of “linguistic” and “conceptual” scales is mine, not Kennedy’s, but it seems to capture the intent:  $\mathcal{S}_{expensive}$  is the subset of the scalar COST concept that is relevant to the item *expensive*.) This is essentially the same strategy that Klecha (2012, 2014) applies to *likely*, and it is subject to the same objections. Most importantly, this account of *free* predicts that it should be bizarre or nonsensical to talk about the cost of things that are free using *(in)expensive* or *cheap*. The naturalistic examples in (4.68) indicate that this prediction is incorrect.

- (4.68) a. But fret not, because not only is this ringtone very cheap, it is in fact FREE!  
 b. [T]he old trusted spreadsheet has really had its day when it comes to your accounting system. Why? Whilst it may seem like a very cheap, or in fact free option, using a spreadsheet can be very limiting for your business.  
 c. The drinks are very cheap here today. In fact some are even free.

For another contrast, consider the quite reasonable examples of *expensive/cheap* in (4.69) alongside the bizarre examples with *tall/short* in (4.70).

- (4.69) a. Yachts and caviar are much more expensive than clean water, which is free.  
 b. In contrast to yachts and caviar, clean water is very cheap. In fact, it is free.  
 (4.70) a. # NBA players are much taller than dignity, which has no height.  
 b. # In contrast to NBA players, dignity is very short. In fact, it has no height.

If  $\mathcal{S}_{expensive}$  and  $\mathcal{S}_{cheap}$  exclude the point of zero cost, the examples with *expensive/cheap* should pattern with those involving *tall/short*—but they do not. While it is true that it often feels odd to talk about whether and to what extent a free item is cheap, I suspect that the explanation is fairly mundane: given that we have an unambiguous expression for zero cost (*free*), by choosing to talk about how cheap or expensive an item a speaker may implicate that she is not in a position to describe it as “free”.

Now, we could try to employ Klecha’s coercion strategy here: *expensive* and *cheap* are indeed associated with open scales, but it is possible to coerce these scales to include their intuitive bounds when necessary. Absent some articulated theory of coercion which distinguishes the two cases, it is not obvious why the cost adjectives in (4.68) and (4.69) should differ from the height adjectives in (4.70). Perhaps there is some way to distinguish them (as Klecha (2014: 63-64) claims, though I do not understand how his suggestion is meant to work). Even so, the objection from degree modification that I raised above would apply. If  $\mathcal{S}_{cheap}$  can be coerced to include zero cost in the examples in (4.68), it is a mystery why the same coercion operation is not available in (4.71a), generating an interpretation paraphrasable as (4.71b).

- (4.71) a. This ringtone is completely cheap.  
b. “This ringtone has zero cost.”

There are a number of additional examples of relative adjectives that seem to fall on bounded scales. Consider, for example, adjectives of speed, loudness, wealth, distance, and weight.

- Since zero speed is a conceptual possibility that is widely attested (by anything that is motionless), the speed scale seems to be lower-closed. Nevertheless, as Sassoon (2014) points out, *fast* is not a minimum adjective meaning “non-zero speed”, as IE would lead us to expect. Instead, it is a relative adjective indicating high speed. Similarly, *slow* is not a maximum adjective meaning “zero speed”, but a relative adjective indicating low speed.
- There is a point of maximal quietness, silence, and yet *quiet* does not mean “silent”, and *loud* does not mean “more than zero noise”.
- Since it is possible to have no money, IE would lead us to expect that *rich* should mean “having non-zero money”, and *poor* should mean “having no money”.
- There is a minimum possible distance between two objects. Absent additional stipulations, IE thus wrongly predicts that *close* should be a maximum adjective meaning “in the same place”, and *far* should be a minimum adjective meaning “not in the same place”.
- Weight adjectives constitute a slightly more recondite counter-example. In contexts in which the possibility of having zero weight is highly salient—such as space travel—IE would lead us to expect that *heavy* should be interpreted as “having non-zero weight”, and *light* should indicate zero weight. Similarly, *slightly heavy* and *completely light* should be acceptable with the respective degree-modifying interpretations “not weightless” and “weightless”. However, the salience of zero weight as a live possibility does not have the expected effect of allowing these modifiers to have a degree interpretation.

All of these adjectives are relative, rather than absolute as IE would lead us—at least, if we take the obvious tack of identifying the linguistically relevant scales with the scalar concepts that they appear to pick out.

Of course, it is possible to relax this assumption, making room for *ad hoc* adjustments—here, stipulating that the conceptual lower bound does not count in each of these cases for the purpose of applying IE. However, the plausibility of this maneuver declines rather sharply when it must be used repeatedly in order to evade apparent counter-examples. Worse, for each problematic adjective pair we can find naturalistic examples analogous to (4.69), where semantic exclusion of the zero point from the relative adjective’s scale would render the discourse contradictory, or some portion of it trivially false.

- (4.72) a. [C]ombat in The Banner Saga seems about as slow as the motionless sun that hangs in the sky.  
b. The first guy was incredibly quiet. In fact, he was completely and utterly silent.  
c. [I]n case you are wondering about the price of food, it is usually very cheap. the very poor simply have no money at all.

d. Yeah, we're close on the map, too—in fact, our green dots are in exactly the same place.

If objects whose degree is at the relevant scalar endpoint were not “on the scale” of the relative adjectives *slow*, *quiet*, *poor*, and *close*, we would not be able to make sense of these examples. And, again, invoking an unexplained “coercion” mechanism in each of these cases would create problems involving degree modification. If this mechanism is available, we no longer have a motivated explanation for the absence of degree-maximizing readings of *completely slow/quiet/poor/close*, and of degree-minimizing readings of *slightly fast/loud/rich/far*.

The clearest way out of this theoretical muddle, I suggest, is not to add further epicycles but to abandon ship. Relative adjectives can occur on bounded scales, and so scalar boundedness properties do not directly determine minimum/maximum/relative classification directly, as IE would have it. I conclude that IE is not a generally applicable constraint of grammar or processing. At best, it serves as a summary of an interesting but imperfect empirical correlation that is in need of deeper explanation.

#### 4.2.8 Multiple non-synonymous adjectives on the same scale

A second kind of problem for Kennedy's theory of positive form semantics involves adjectives on the same scale, such as the pairs *warm/hot* and *cool/cold*. *Warm* and *hot* both make reference to comparisons of temperature, both have the same polar orientation, and are both relative adjectives. However, whenever there are two adjectives on a scale, the unified account of the positive form in (4.60) appears to predict that they should have exactly the same meaning. Relative to a fixed context *c*, *warm* and *hot* should both indicate that the object they are predicated of “stands out” in *c* relative to temperature. Similarly, *cool* and *cold* should both indicate that an object “stands out” in context relative to inverted temperature.

The prediction that any two scalar adjectives that are on the same scale must be synonymous flies in the face of a commonplace observation: we often have multiple adjectives of varying strength that are used to talk about important scalar properties (see, for example, Hatzivassiloglou & McKeown 1993; Sheinman, Fellbaum, Julien, Schulam & Tokunaga 2013). For example, *peckish*, *hungry*, and *famished* are adjectives which pick out varying grades of hunger, and are ordered by strength: in any context,  $\theta_{peckish} < \theta_{hungry} < \theta_{famished}$ . Similarly, we surely would not want to rule out the possibility that any two of our rich inventory of beauty adjectives could be on the same scale without being synonymous. Further examples of plausibly co-scalar adjectives include  $\langle ajar, open \rangle$ ,  $\langle good, great, wonderful \rangle$ ,  $\langle bad, terrible \rangle$ ,  $\langle tasty, delicious \rangle$ ,  $\langle funny, hilarious \rangle$ ,  $\langle smart, brilliant \rangle$ ,  $\langle dumb, idiotic \rangle$ , and probably many more.

Concentrating for the moment on temperature adjectives, perhaps we could try to make room for non-synonymy by supposing that *warm/hot* are somehow not on the same scales, and likewise for the pair *cool/cold*. This is not a very promising line, though. First, the comparatives *x is warmer than y* and *x is hotter than y* are truth-conditionally equivalent. This would be hard to explain if the measure functions associated with *warm* and *hot* had different domains. Indeed the latter hypothesis would predict that **B**'s bizarre response in the following dialogue should be felicitous in some circumstances:

(4.73) a. **A**: Today is hotter than yesterday was.

- b. **B**: [No, you're wrong/No, that doesn't make sense.] But it's true that today is warmer than yesterday was.

If today, yesterday, or both are outside the domain of *hot*, then **A**'s claim should be infelicitous. Further, this situation would be compatible with both today and yesterday being in the domain of *warm*, rendering **B**'s correction "Today is warmer than yesterday" accurate. But it is difficult to imagine any circumstance in which **B**'s response could fail to be self-defeating.

A second kind of problem applies to the claim that the domains of  $\mu_{warm}$  and  $\mu_{hot}$  vary in a specific way: that one measure function is defined on any object that can have temperature, while the other is defined on a strict subset of these items. The most plausible version of this hypothesis would be that  $\mu_{hot}$  is defined, relative to some context *c*, only in the range of temperatures in which the positive form *hot* could be used appropriately in *c*. This means treating *hot* essentially as Morzycki (2012) analyzes extreme adjectives such as *brilliant* and *scorching*. On this account, it should be infelicitous to assert both *x is hotter than y* alongside *y is not hot*, or any sentence which entails this. This is because, on the theory under consideration, *x is hotter than y* is undefined if either *x* or *y* could not be described as "hot" in context.

As Morzycki discusses, this prediction seems to be correct for extreme adjectives: for example, (4.74) is odd.

- (4.74) ?? Tim is not brilliant, but he is more brilliant than Harry.

But the prediction is not correct for *hot*. Here are two naturalistic counter-examples.

- (4.75) a. On the way home from our first trip after installation, I noticed the tire temperature ... was about 40° hotter than the rest. Still not hot, only about 140°F but hotter than the other three tires.
- b. No aircraft would be doing Mach 8 on final approach. By the time it closed to its destination, descended, approached the airport, landed and taxied to the terminal, it would be no hotter than a lukewarm coffee mug.

In these examples, something that is described as *not hot* or as *lukewarm* (and so not hot) serves as one of the relata of *hotter*. These examples thus indicate that *hotter* can be used to relate things that do not count as *hot*. Computing the truth-conditions of *x is hotter than y* requires checking that  $\mu_{hot}(x)$  is greater than  $\mu_{hot}(y)$ . So,  $\mu_{hot}$  cannot be restricted to things that count as *hot* in the examples in (4.75). If it were, then these examples would be odd, perhaps (on Morzycki's account) for the same reason that (4.74) is.<sup>8</sup>

A third problem is that temperature adjectives are context-sensitive in the way that we expect for relative adjectives, and they seem to move in lockstep: *hot* suggests warmer temperature than *warm*, and *cool* warmer temperature than *cold*, whether we are talking about stars or spring days

<sup>8</sup> A parallel suggestion about restricting the *warm* scale would encounter the same problem. For example,

- (4.76) a. The underwater photographer breaks through arctic sea ice dropping into a cold -2C water—although still warmer than the -30C world up above.
- b. Snow crystals trap small amounts of air, and polar bears will burrow into the snow to create an insulating blanker which, while cold, is still warmer than the outside world.

in northern Europe. In particular, none of these adjectives is restricted to any particular range of temperatures. For example, a 2014 report in the Milwaukee *Journal-Sentinel* quotes the astronomer David Kaplan describing as “extremely, extremely cold” a white dwarf star whose temperature is around 3,000 Celsius, or 5,432 Fahrenheit. With the right point of comparison, it seems that just about any temperature could count as “cold”, “cool”, “warm”, or “hot”. If the scales associated with *warm* and *hot* were different, the formal structure of the scales themselves would have to shift in context in a way that mimics contextual shifts in meaning of the positive form items *warm* and *hot*. This would constitute a totally new type of context-sensitivity, and one with little or no independent support.

I do not see any well-motivated alternative to the following conclusion: multiple non-synonymous positive-form adjectives can occur on the same scale, and their non-synonymy is enforced by a lexical constraint on the relative strength of their interpretations. Specifically, there must be a lexical constraint requiring that, in any context *c*, the positive form of *hot* requires a greater temperature than the positive form of *warm*, and similarly with *cold* and *cool*, *peckish* and *hungry*, and so on. If this is right, then the meaning of the positive form cannot be reduced to a single, general schema that applies to all scalar adjectives, such as “standing out” on a given scale (see (4.60)). At a minimum, information about the ordering of co-scalar expressions is also needed.

While this conclusion may appear at first glance to be related only indirectly to our immediate concerns in this chapter, it is important for several reasons. First, Kennedy’s “Interpretive Economy” constraint was one of the primary troublemakers with regard to the boundedness of  $S_{likely}$  and  $S_{probable}$ , and the motivation for positing this constraint depends crucially on the assumption that Kennedy’s unified interpretation of the positive form is correct. Second, in the next chapter we will consider the suggestion that the absolute adjectives *possible* and *certain* are on the same scale as *likely/probable*. While the evidence for these hypotheses is mixed, there are decent arguments in their favor, and they are certainly not trivially wrong as they would have to be if Kennedy’s (2007) theory of the positive form were correct. If it were, then the co-scalar hypothesis would have the obviously false consequence that *possible*, *likely*, and *certain* should all mean “stands out in context relative to probability”. As a result, problematizing the “stands out” interpretation of the positive form is crucial to establishing the plausibility of several leading hypotheses about the meaning of *possible* and *certain*.

The third—and least important—reason for mentioning the issue of co-scalar adjectives is that I am invested in a different theory of the positive form, due to Lassiter & Goodman (2013, 2015a), which treats it the threshold simply as a free variable subject to pragmatic inference (spelled out in terms of a probabilistic computational model of conversational reasoning). On this account there is no problem with having multiple adjectives on a scale, whether both are relative or one is relative and one is absolute. While promoting this alternative account is not my primary concern here, in the context of a rejection of Kennedy’s theory of the positive form it may be useful to note that there is a worked-out alternative available. I will say a few more words about this alternative in a moment (§4.2.10).

#### 4.2.9 Degree modifiers can be sensitive to positive form meaning

If we wish to hold onto the intuitive claim that  $S_{likely}$  and  $S_{probable}$  are closed on both ends, we must account for the behavior of degree modifiers. This is relevant in two ways. First, as noted above *completely likely/probable* does not have a maximizing reading (“as likely as a tautology”), and *slightly likely* does not have a greater-than-minimum reading (“a bit more likely than a contradiction”). But these are the readings that Kennedy’s (2007) theory of degree modification predicts on the assumption that the scales in question are fully closed. On that theory, degree-modifying readings of *completely Adj* are available if and only if  $S_{Adj}$  is upper-bounded. Likewise, *slightly Adj* has a degree-modifying interpretation if and only if  $S_{Adj}$  is lower-bounded. This makes sense if we suppose that these constructions have the syntax  $[MOD\ Adj]$ , where *Adj* is the root; the root already contains the information necessary to determine boundedness properties.

However, a number of theorists have suggested analyzing (some or all) degree modifiers as embedding the full positive form—for example, Klein (1980); Rotstein & Winter (2004); Burnett (2016), and indeed Kennedy & McNally (2005b) in a different incarnation. On such an account, at least some degree modification structures have the form  $[MOD\ [pos\ Adj]]$  (or something that is close enough for current purposes). A modifier that embeds the full positive form has access to any syntactic and semantic information that is contained in the latter, even if it is not part of the meaning of the root—including the information that the positive form has a minimum, maximum, or relative interpretation.

There is evidence that this is the right way to think about *slightly*-modification, and that *slightly*-modification of relative and maximum adjectives is ruled out in a principled way that does not make direct reference to scale structure. Solt (2012) and Sassoon (2014) make a convincing case that it is the detailed interpretation of the positive form, rather than mere lower-boundedness, that controls the acceptability of *slightly*-modification. For example, Kennedy’s (2007) theory would predict that the sentences in (4.77) and (4.78) should have the (b) readings, since both are associated with lower-bounded scales (cf. Kennedy 2007: §4.2). The experimental and corpus work reported by Sassoon and Solt indicates that this prediction is not correct.

- (4.77) a. The glass is slightly full.  
b. “The glass is almost but not completely empty”

- (4.78) a. The glass is slightly empty.  
b. “The glass is almost but not completely full”

Note that Kennedy’s (2007) account makes the right predictions for closed-scale adjective pairs such as *open/closed* and *opaque/transparent*, all members of which are acceptable with both *slightly* and *completely*. This is to be expected if *slightly* prefers to combine with the positive form of minimum adjectives. As Kennedy (2007: §4.3) notes, these four adjectives are—unlike *full/empty*—genuinely ambiguous between minimum and maximum interpretations in the positive form. The contrast with *full/empty* confirms that there is a correlation between positive form interpretation and the acceptability of *slightly*-modification that does not depend on scalar boundedness. This means, in turn, that the lower-boundedness of an adjective’s scale is not the only factor controlling the acceptability of *slightly*.



Thus, as Solt (2012: 8) concludes, “incompatibility with *slightly* cannot be used as a diagnostic for the absence of a scalar minimum point”. The absence of a degree-modifying reading of *slightly likely* does not show that  $\mathcal{S}_{likely}$  is not lower-bounded, just as the failure of *slightly full* to mean “almost but not quite empty” does not show that  $\mathcal{S}_{full}$  is not lower-bounded.

Solt and Sassoon give different accounts of *slightly*, both designed to explain the fact that this item also occurs in a variety of additional constructions, such as:

- (4.79) a. Bill is slightly older than Bob.  
 b. Bill is slightly too old to play.  
 c. Bill is slightly old for a linebacker.

While the details of their accounts differ, the relevant upshot is that the distribution of *slightly* depends on the meaning of the item modified, construed as a predicate of individuals. In the case of *slightly bent* vs. *slightly old*, Solt and Sassoon trace the contrast to the meaning of the positive form of the adjectives *bent* and *old*—and specifically the question of whether the positive form exhibits vagueness. Showing little or no vagueness is a feature shared by a variety of constructions which readily admit *slightly*-modification: minimum adjectives, comparatives, excessive constructions, and relative adjectives with sufficiently precise contextual standards, among others. On such a theory, the oddity of *??slightly likely/probable* is explained, without further ado, by the fact that these adjectives have relative interpretations in the positive form, and are typically vague. There is no need to make *ad hoc* adjustments to the structure of their scales.

In addition, suppose that it is true as I argued above that *cheap*, *inexpensive*, *slow*, *quiet*, *close*, *light*, *likely* and *probable* are relative adjectives that are associated with upper-closed scales. If so, we must allow that degree-modifying *completely* selectively combines with the positive form of maximum adjectives (or some broader category containing these adjectives) if we are to explain why none of these adjectives has a degree-maximizing reading when modified by *completely*. This is admittedly a promissory note, but—given the failure of even the quite complex open-scale-plus-coercion analysis of these adjectives that was discussed above—it is perhaps not too bold to endorse this generalization and hope that future work will uncover a principled rationale for it.

Finally, the contrast between *n% likely* and *#half/#mostly likely* discussed in §4.2.3 suggests that proportional modifiers vary in whether they can modify non-maximum adjectives. If this is correct, we have another example of modifiers which are sensitive to minimum/maximum/relative status in addition to scalar structure. This gives us another example of a *prima facie* problem for the closed-scale analysis of *likely* and *probable* which is best explained, not by modifying our analysis of the adjectives, but by attending to further complexities in the theory of degree modification.

#### 4.2.10 Summary: *Likely* and *probable* in the context of scalar adjective typology

What does all of this imply for the adjectives we are interested in here—*likely* and *probable*? Intuitively, these adjectives are on closed scales, with the bounds provided by the likelihood of a tautology and that of a contradiction. This intuition comes into theoretical conflict with a prominent theory of adjective semantics, degree modification, and scalar structure due to Kennedy (2007). We can resolve this tension either by removing the intuitively present bounds (but re-inserting them via

coercion when the data demand), or by rejecting the theory that created the problems and retaining the intuitively present scalar endpoints.

I argued that there are multiple reasons to take the latter route. If we remove the endpoints, we encounter a dilemma involving the posited coercion operation. Without it the theory undergenerates, ruling out examples such as “likely, indeed certain” and “cheap, in fact free”. With it the theory overgenerates, predicting that *completely likely/cheap* should have degree-maximizing readings equivalent to “certain” and “free” respectively. In addition, there are several more general problems with the broader theoretical framework that served as the primary motivation for assigning a counter-intuitive structure to  $S_{likely}$ , especially involving the prediction that relative adjectives cannot occur on bounded scales. In §4.2.7 I noted a number of apparent counter-examples to this claim and argued that an effort to fix the problem by appeal to a coercion mechanism fails. In this light, it is natural to interpret the observation that *likely* and *probable* are relative as simply adding another two problematic cases to the list. In addition, I argued that the problematic cases involving degree modification may be explicable in terms of a more elaborate account of these modifiers.

It would doubtless be possible to patch up these problems by adding further mechanisms to the theory. However, their cumulative effect is to suggest that the elegant theory of degree modification and the positive form proposed by Kennedy (2007) is incorrect in several relevant respects. If so, there is no reason to associate *likely* and *probable* with a counter-intuitive scalar structure: that move was entirely motivated by the assumption that the predictions about adjective typology and degree modification associated with Kennedy’s theory are accurate.

The empirical picture that emerges is that

- a) relative adjectives can occur on either bounded or unbounded scales;
- b) multiple adjectives can occur on the same scale, without synonymy, as long as they are lexically ordered;
- c) degree modifiers (at least sometimes) take the full positive form of an adjective as their argument, giving them access to information about this interpretation including its minimum/maximum/relative status.

Does this mean that we must revert to Kennedy & McNally’s (2005a) assumption that minimum/maximum/relative status is associated with scalar adjectives on an arbitrary item-by-item basis? I hope not. Here is a possible alternative. Properties (a) and (b) can already be accounted for within an alternative theory of the relative/absolute distinction proposed by (2013; see also Lassiter & Goodman 2015a). In brief, this theory treats adjective interpretation as an active process involving pragmatic reasoning about alternative ways to express a scalar property, subject to linguistic and discourse constraints. This is all given a precise computational implementation within a Bayesian model of interpretation that builds closely on game-theoretic pragmatics. The theory accounts for the relative/absolute distinction in terms of statistical differences between different kinds of reference classes which are correlated with, but not fully determined by, scalar boundedness properties. It also has no difficulty making room for the presence of multiple non-synonymous adjectives on a scale. In the case of *warm* and *hot*, for example, we can simply add constrain the pragmatic inference by requiring that the estimated value of  $\theta_{warm}$  is less than that of  $\theta_{hot}$ . Simulations which extend

Lassiter & Goodman’s model to the pragmatic competition between *warm* and *hot* indicate that this constraint is sufficient to generate plausible, context-dependent interpretations for both items, with a clear division of semantic labor.

The conclusions of this section, if correct, have significant implications for the theory of scalar adjectives. I don’t doubt that there is much more to be said here. Indeed, it would be reasonable at this point for theorists who are attracted to Kennedy’s (2007) theory to look for other ways to construe the theory and maintain its core insights. However, I hope I have at least made a case in this brief space that theories of modal semantics should not be overly constrained by the details of this one theoretical approach, influential though it may be.

As far as the theory of modality is concerned, the importance of this section’s discussion is mostly limited to the detailed characteristics of the important items *likely* and *probable*. As the lightly modified version of Klecha’s account discussed here demonstrates, we can have additivity without boundedness if we want it; but, I argued, there is little reason to want it, and empirical and theoretical problems arise in the attempt. The only theoretical issue discussed in this section that is critical for the rest of the book is the claim that multiple adjectives can occur on the same scale without synonymy, which will be presupposed in chapter 5’s discussion of *certain* and *possible*.

### 4.3 Chapter summary

*Likely* and *probable* are the scalar adjectives that seem to provide the most natural way to express epistemic comparisons such as *Mary is as likely to win as Bill is*. The first main part of this chapter considered these adjectives from the perspective of measurement theory, as presented in chapter 2. The acceptability of ratio modifiers with *likely* and *probable*—for example, in *Rain is exactly twice as likely as snow*—suggests that they are associated with ratio scales, similarly to *heavy* and *tall*. This classification explains the failure of the Disjunctive Inference, the crucial empirical problem with Kratzer’s (1991) account of epistemic comparison. The key insight is that  $S_{likely}$  and  $S_{probable}$  have the formal property of **additivity**, which follows from a ratio scale classification.

In addition,  $S_{likely}$  and  $S_{probable}$  are upper- and lower-bounded, as evidenced by the observation that nothing can be more likely than a tautology, or less likely than a disjunction. (As we will see in chapter 7, this property is crucially not shared by their deontic counterpart, *good*.) These scales are thus fully-closed ratio scales, making them the qualitative counterpart of a finitely additive probability measure (Narens 2007). If we are working with a theory of scalar language that makes use of degrees, this implies that the scales of *likely* and *probable* are simply ordinary probability scales.

These conclusions were corroborated to a considerable extent by a detailed examination of the ratio scale axioms presented in chapter 2. Ratio scales are by definition upper- and lower-bounded, reflexive, transitive, connectedness, monotone, positive, solvable, and Archimedean. Of these properties, the only one that plausibly fails for  $S_{likely}$  and  $S_{probable}$  is connectedness. The rest of the book tentatively assumes connectedness, but the validity of this assumption should be given further scrutiny, as well as its precise meaning in the context of a theory of epistemic adjectives.

Turning to the interpretation of the unmodified (“positive”) form of *likely* and *probable*, we saw that they behave as relative adjectives, much like *heavy*, *tall*, and *happy*. However, they differ in

allowing the proportional modifier *n%*, as in # *Bill is 80% heavy* vs. the naturally-occurring *It's 80 percent likely that the iPhone will be coming to T-Mobile*. This difference makes sense if *heavy*'s scale lacks a maximum point—as it intuitively does—while *likely*'s scale has one.

The classification of *likely* and *probable* as relative adjectives on a fully closed scale is in conflict with an influential theory of the semantics of the positive form and of degree modification due to Kennedy 2007. On Kennedy's account, adjectives on closed scales must be either minimum (like *bent*) or maximum (like *straight* or *full*). As Klecha (2012, 2014) points out, we could resolve this tension by proposing that  $\mathcal{S}_{likely}$  and  $\mathcal{S}_{probable}$  are formally similar to probability scales, but lack the maximum and minimum points. While this move is technically feasible, I argued that it has empirical problems which persist even if we help ourselves to a “coercion” operation to explain anomalous data. Further, I argued that Kennedy's (2007) theories of the positive form and of degree modification admit of various counter-examples among non-modal adjectives, including the existence in English of a variety of additional relative adjectives that are associated with closed scales. Since Klecha's account was motivated by the assumption that Kennedy's theory is correct to rule out this possibility, these observations do much to undermine the motivation for assigning a counter-intuitive scale structure to *likely* and *probable*.

With due caveats about connectedness, then, I suggest that *likely* and *probable* are relative adjectives, falling onto an upper- and lower-bounded ratio scale. Against the theoretical background of Measurement Theory, this conclusion turns out to be equivalent to the conclusion that their scales are probabilistic, as Swanson (2006); Yalcin (2007, 2010); Lassiter (2010, 2011a) propose. Crucially, this conclusion is stronger than the one reached by Holliday & Icard (2013), who are able to give empirical motivation only for a “qualitative additivity” property which is weaker than the finite additivity that a ratio scale classification implies. The difference in conclusions is fundamentally due to two factors. First, the language that we are investigating is larger than the one that Holliday & Icard consider, and crucially includes ratio modifiers (cf. Lassiter 2015b: §2.2). In addition, we have helped ourselves to a number of theory-internal assumptions that are associated with the broader projects of intensional and degree semantics for natural language, which were crucial in reaching the conclusion that  $\mathcal{S}_{likely}$  is finitely additive. In any case, our study has not led to any findings that are in conflict with those of Holliday & Icard (2013). However, the integration of a measurement-theoretic perspective into an intensional degree semantics in the tradition of Montague (1973) allows us to reach certain stronger conclusions—ones which are also, perhaps, more exciting because they draw direct connections between the semantics of epistemic modals and the philosophically and psychologically important concept of probability.

## CHAPTER 5

### Certainty and possibility

#### 5.1 *Certain and sure*

*Certain* and *sure* are gradable epistemic adjectives that are closely related to *likely*. (They are also, as far as I can see, interchangeable modulo register. I will mostly focus on *certain*.) Certainty entails likelihood: it would be incoherent for me to claim that it is certain to rain, but not likely; or that it is sure to snow, but that snow is not probable. This section works to clarify the scale structure of *certain* and its relationship to other modal adjectives.

To set the stage, I can't resist quoting at length from an article that illustrates a number of interesting linguistic features of *certain* (Merchant 2013).

#### Things Scientists Are Less Sure of Than Climate Change

If you are one of the few humans who has not yet been persuaded by the overwhelming scientific evidence that our activities are heating up the planet, or are under the impression that scientists are still uncertain as to whether dumping carbon dioxide into the atmosphere is causing global temperatures to rise, then consider this brief guide for your benefit. ... [T]he world's climatologists are now gearing up to officially proclaim that they are 95 percent certain that humans are to blame for global warming.

That 5 percent gap may seem large. It is not. In science, nothing is 100 percent sure—not even the law of gravity. ... [H]ere are a few things that scientists are just as or less certain of than climate change:

- that cigarettes kill
- the age of the universe
- that vitamins make you healthy
- that dioxin in Superfund sites is dangerous

... Science is an arduous, time-intensive process about which certainties of the magnitude found in the field of climatology are few and far between. The fact that a stunning 97 percent of scientists in a given field agree on a dominant theory is all but unheard of—and that's how many agree that human activity is driving the rise in global temperatures.

Merchant uses *certain* and *sure* interchangeably. Here are some further features that he demonstrates:

- (5.1) Equatives and comparatives:
  - a. *things scientists are less sure of than climate change*

- b. *things that scientists are just as or less certain of than climate change*
- (5.2) Percentage modifiers:
- a. *they are 95 percent certain that humans are to blame for global warming*
  - b. *In science, nothing is 100 percent sure*

Since *certain* and *sure* occur in the comparative and equative, they are gradable adjectives (5.1). In addition, the examples in (5.2) indicate that they can, at least sometimes, be used with percentage modifiers.

Another linguistically interesting point is that Merchant specifically decries the inference from *not 100% certain* to *uncertain*: “If you ... are under the impression that scientists are still uncertain as to whether dumping carbon dioxide into the atmosphere is causing global temperatures to rise ...”, strongly implicating that they are not. In light of evidence that *certain* is a maximum adjective (see §5.1.1 below), the tests reviewed in §4.2.1 above would lead us to expect *not 100% certain* to entail *at least a little bit uncertain*. Merchant’s article is specifically designed to combat the suggestion that this little bit of uncertainty is large enough to be meaningful. While this inference may be technically correct, the message is that it is disingenuous to focus on it, as opponents of action on climate change have.

A third interesting feature that emerges in this passage involves the alternation between *It is certain/sure that  $\phi$* , with a dummy subject *it*, and *x is certain/sure that  $\phi$* , for some attitude-holder *x*. This alternation distinguishes *certain/sure* from *likely*, *probable*, and *possible*, none of which admit a subject that represents the bearer of an attitude. A sentence like *\*Bill is likely/possible that it’s snowing* is simply ungrammatical. This syntactic alternation is also closely related to a distinction between “epistemic” and “subjective certainty”. This distinction is common in the philosophical literature on knowledge and skepticism (e.g., Stanley 2008), and is also supported by experimental evidence in the psychological literature on uncertainty expressions (Løhre & Teigen 2015). Roughly, *Bill is certain that it will snow* attributes a maximal *subjective* degree of belief in snow to Bill. This sentence does not comment on whether Bill’s beliefs are reasonable. In contrast, the truth of *Snow is certain*, *It is certain to snow*, or *It is certain that it will snow* depends not on anyone’s psychological state, but on the degree of belief in snow that it would be rational to have given a certain body of evidence.

Interestingly, in the global warming discussion this distinction is not maintained clearly: Merchant speaks in one breath of subjective certainty—*scientists’* degree of certainty in climate change—and in the next of epistemic certainty—*nothing is 100 percent sure*. Presumably, this choice has to do with the presumption that the scientists in question are rational and have taken into account all of the available information: if *they* are 95% certain of global warming, then *it* is 95% certain that global warming is occurring. We will see a similar alternation in the examples in (5.19) below, where reporters discussing a rescue operation alternate, in a fairly haphazard way, between discussing the (epistemic) certainty of an eventuality and the (subjective) certainty of a certain well-informed person. This choice makes sense, since the person whose subjective state is under discussion is the individual who happens to have the best information about the eventuality in question. The reporters thus have every reason to think that his opinion has been formed on the basis of rational consideration of all of the available evidence.

The distinction between two kinds of certainty is an important one, and its syntactic and

argument-structural properties deserve extensive investigation. In addition, divergences in the acceptability and entailments of numerical statements of subjective and epistemic certainty, as documented by Løhre & Teigen (2015), deserves further investigation. However, I will mostly gloss over this distinction in the following discussion, since—as far as I have been able to discern—the two kinds of certainty do not differ in structural properties, and this is the issue that we are mainly interested in.

### 5.1.1 Classification of *certain* and *sure*

On many of the tests from §4.2.1 above, *certain* behaves like *full*, a maximum adjective on a fully-closed scale (Lassiter 2010; Beddor 2015). For example, both admit degree-modifying *completely*.

- (5.3) a. The glass is completely full.
- b. It is completely certain/sure that the Jets will win the Super Bowl.

*Slightly*-modification is distinctly odd; both fail the “*Adj* but could be *Adj*-er” test; and both can be modified by *almost*.

- (5.4) a. # The glass is slightly full.
- b. # It is slightly certain/sure that the Jets will win.
- (5.5) a. # The glass is full, but it could be fuller.
- b. # It is certain that the Jets will win, but it could be more certain.
- (5.6) a. The glass is almost full.
- b. It is almost certain/sure that the Jets will win.

### 5.1.2 Proportional modifiers

*Certain* and *sure* also resemble *full*—and differ from *safe*—in admitting percentage modifiers. Compare (5.7a), from the passage quoted above, with the Twitter post quoted in (5.7b).

- (5.7) a. [T]hey are 95 percent certain that humans are to blame for global warming.
- b. Michigan stadium more than 95 percent full 20 mins before match time.

As we have seen, *safe* is on a scale that is closed above but not below. As we might expect, *n% safe* is generally quite odd (except in the case where *n* = 100, which is plausibly a synonym for degree-maximizing *completely*: see Kennedy & McNally 2005a). However, *n% safe* does sometimes occur in contexts where safety can be directly equated with the probability that some undesirable event is prevented. In this case safety can be measured on a fully closed, additive scale.

- (5.8) When used properly, condoms are 97 percent safe.

This use is apparently another example of the ill-understood phenomenon of scalar coercion (§4.1.3).

A further question involves the status of the proportional modifiers *mostly* and *half*. I suggested in §4.2.3 that these modifiers differ from *n%* in that percentage modifiers care only whether an adjective has the right kind of scale structure, while *mostly* and *half* are also restricted to maximum adjectives. Since *certain* is maximum, both *mostly certain* and *half certain* should be acceptable. Naturalistic examples seem to corroborate this expectation:

- (5.9) a. I'm mostly certain that the thing jutting out of the dead kangaroo's nether regions is its tail. Mostly.  
b. The response options include (1) completely or almost completely uncertain; (2) mostly uncertain; (3) slightly more uncertain than certain; (4) slightly more certain than uncertain; (5) mostly certain; and (6) completely or almost completely certain. (Rubin, Rubin, Graham, Perse & Seibold 2009: 309)
- (5.10) a. I'm half certain that nobody really knows what a hipster is...  
b. I'm half-certain that Christopher Lee didn't die; Death just realised he was the second-best person to play his role and stood aside.

### 5.1.3 A puzzle: Type of antonym

There are at least two interesting points of difference between *certain/sure* and *full*, both relating to the linguistic properties of their antonyms. *Full*'s antonym *empty* is a maximum adjective indicating maximal emptiness (= minimum fullness; cf. also *completely*/*#slightly empty*, etc.). In contrast, the antonyms of *sure* and *certain* are minimum adjectives. They do not generally indicate maximum possible uncertainty, but rather some amount of deviation from total certainty in the negative direction.

- (5.11) [I]t is uncertain that I can fulfill my duty not to do my neighbour harm if I speak.  
a. "It is not certain that I can fulfill my duty"  
b. # "It is maximally uncertain that I can fulfill my duty"

Modification tests provide some additional evidence for this classification.<sup>1</sup>

- (5.12) a. Michael Kors has been crowned the "King of jet-set American style," though it is slightly uncertain whether he received this title by popular consensus in the fashion world or simply from his colleagues on Bravo's hit show Project Runway.  
b. I suspect the future of the GIMP's Scheme interpreter is slightly uncertain.  
c. Haydon is slightly uncertain about the name of the latter.

<sup>1</sup> However, very few of the clearly acceptable corpus examples that I have located involve the constructions that we are focusing on here, of the form *It is slightly uncertain that  $\phi$*  or  *$x$  is slightly uncertain that  $\phi$* . Most involve either *wh*-complements or subjects that are readily associated with events or propositions. This pattern may be attributable to whatever is responsible for the overall low frequency of *It is uncertain that* as opposed to *It is uncertain whether* and constructions of the type *event  $e$  is uncertain*. It is also possible that it is indicative some subtle difference in meaning that depends on the syntactic or semantic properties of the subject and/or complement, and is reflected in scalar structure or minimum/maximum status.



- d. The exact time of the boundary between the Phanerozoic and the Proterozoic is slightly uncertain.

Since their antonyms are minimum adjectives, *certain* and *sure* may resemble more closely *safe*, which is a maximum adjective on a scale with an upper bound but no lower bound.

(5.13) # The neighborhood is slightly safe.

(5.14) The neighborhood is slightly unsafe/dangerous.

A second point of difference between *certain* and *full* is that there is not a robust extension gap between *certain* and its antonym. In this, it again resembles *safe*:

(5.15) a. The glass is not full, but it is not empty either.

b. ?? It is not certain that the Jets will win, but it is not uncertain either.

c. ?? The neighborhood is not safe, but it is not unsafe either.

One point of unclarity in this classification involves modification by *completely*. While *full*'s antonym *empty* readily accepts modification by *completely* with a degree-maximizing interpretation, *safe*'s antonyms *unsafe/dangerous* do not.

(5.16) The glass is completely empty. "It could not possibly be emptier."

(5.17) The neighborhood is completely unsafe/dangerous.

a. "It's dangerous any way you look at it."

b. "Every part of it is dangerous."

c. # "It could not possibly be more dangerous."

It is not entirely clear to me whether *uncertain* and *unsure* admit an interpretation along these lines with *completely*. Consider the following naturally-occurring examples:

(5.18) a. She is completely uncertain that she can start a career at this point in her life.

b. If it is not even certain whether Paul entered the primary Nabatean territory in the first place, then it is completely uncertain that he missioned among the Nabateans.

In these examples, *completely uncertain* does not indicate that the appropriate degree of belief in the embedded proposition could not possibly be lower—i.e., that it is impossible. Instead, both examples seem to indicate something rather different—that the question of *whether* this proposition is true is maximally uncertain. Here, the paraphrase "could not be more uncertain" is not unreasonable.

It may be that this reading can be reduced to one of the alternative interpretations of *completely*—for example, "uncertain any way you look at it" is not an implausible gloss. However, (5.18) might also be interpreted as evidence for an alternative question-based semantics for *certain* discussed in §5.1.7 below, which is inspired by observations of Klecha (2012) and fleshed out in information-theoretic terms.

#### 5.1.4 Certainty and likelihood

Contrary to the conclusions of Lassiter (2010, 2011a), it appears that *certain* and *sure* do not fit neatly into any of the categories of adjectives that we have studied previously. However, it is clear

that their scales must be closely related to the likelihood scale. Here are several arguments in support of this close connection.

First, there is an entailment relationship between the positive forms: certainty asymmetrically entails likelihood. It would be inconsistent to claim that  $\phi$  is certain, but deny that  $\phi$  is likely. Similarly, if  $\phi$  is not likely then it is not certain either.

Second, the certainty and likelihood scales are related monotonically, in the following sense: comparative certainty entails comparative likelihood. Intuitively, if  $\phi$  is more certain than  $\psi$ , then  $\phi$  is also more likely. (See Lassiter & Goodman 2015b for experimental evidence for this monotonicity relationship.)

Third, both scales are upper-bounded by the probability/certainty of a tautology. Just as nothing can be more likely than a tautology, nothing can be more certain.

Fourth, authors sometimes use *n% likely* and *n% certain* interchangeably. The most dramatic naturalistic example of this equation that I have located is in the transcript of a 12/30/14 CNN broadcast on the 2014 AirAsia crash shows multiple speakers discussing a probability estimate of special interest. They alternate repeatedly between *95 percent likely* (5 times) and *95 percent certain* (3 times). Here are all instances of *percent likely/certain* in the report:

- (5.19) a. (Headline) “95 Percent Likely” Debris is from AirAsia Plane  
 b. (Berman) Investigators say they are 95 percent certain it is part of the vanished jetliner.  
 c. (Romans) [T]he head of Indonesia’s search and rescue operation now says it is 95 percent likely the debris spotted floating on the surface in the Java Sea, that debris is from missing AirAsia Flight 8501.  
 d. (Stevens) The head of the search and rescue saying the only reason he is not saying this is 100 percent certain is he has not had his own eyes on that debris field.  
 e. (Romans) Officials say it is 95 percent likely to be from the AirAsia flights.  
 f. (Berman) Indonesian officials say it is 95 percent likely ...  
 g. (Berman) The debris found, investigators 95 percent certain, it is connected to the vanished jetliner.  
 h. (Berman) An Indonesian rescue official says it is 95 percent likely ...  
 i. (Berman) Investigators say they are 95 percent certain ...

Relatedly, Løhre & Teigen (2015; experiment 4) found that Norwegian university students produced statistically indistinguishable numerical estimates for a variety of soccer-related forecasts in the Norwegian equivalents of the frames “It is \_\_% certain” and “There is a \_\_% probability”.<sup>2</sup>

Maximum adjective status shows that  $S_{certain}$  has a true upper bound. The data involving proportional modifiers indicates that this scale is (a) also at least asymptotically lower-bounded, and (b) rich enough to support percentage modification—so, either interval or ratio (§§2.2-2.3). The key formal property that remains unspecified is positivity, and it is hard to imagine that  $S_{certain}$  could

<sup>2</sup> Norwegian *Det er \_\_% sikker* and *Det er \_\_% sannsyilig*, respectively. Interestingly, the mean numerical values produced in the frame “I am \_\_ certain” (*Jeg er \_\_% sikker*) were almost perfectly correlated with the other two frames, but consistently around 10% higher. See Løhre & Teigen 2015 for discussion and an interpretation.

fail to have this property. A denial of positivity would have the bizarre consequence that there could be disjoint propositions  $\phi$  and  $\psi$  such that each has some degree of certainty, but  $\phi$  is exactly as certain as  $\phi \vee \psi$ .<sup>3</sup>

If  $\mathcal{S}_{\text{certain}}$  has all of these properties, then it is difficult to see how it could be distinguished formally from  $\mathcal{S}_{\text{likely}}$ . The next sections will discuss three alternative analyses:

- $\mathcal{S}_{\text{certain}}$  is structurally identical to  $\mathcal{S}_{\text{likely}}$ .
- $\mathcal{S}_{\text{certain}}$  differs from  $\mathcal{S}_{\text{likely}}$  in that the former is not lower-bounded (Klecha 2012, 2014).
- $\mathcal{S}_{\text{certain}}$  orders objects of a different semantic type: sets of propositions, rather than propositions (cf. Uegaki 2015).

In the end, I am unsure which analysis is best: each has some things going for it. Let’s consider them in turn.

### 5.1.5 Analysis 1: Identical scales

Lassiter (2010, 2011a) argues that *likely* and *certain* are on the same scale—a fully closed ratio scale, as proposed in §4.1 for  $\mathcal{S}_{\text{likely}}$ . On this analysis, the two adjectives differ only in that *likely* is a relative adjective and *certain* is a maximum adjective, and other linguistic differences between them are attributable to this difference. This hypothesis relies on the assumption that multiple adjectives can be associated with the same scale and have different interpretations. As we saw in chapter 4, pairs like *warm* and *hot* give us good reason to think that this is indeed the case.

The co-scalar hypothesis easily accounts for the fact that  $\phi$  *is certain* entails  $\phi$  *is likely*. These adjectives are on the same scale, but the positive-form meaning of *certain* is associated with the maximum point of the scale, while the positive-form meaning of *likely* is context-sensitive and generally lower. Likewise, the fact that both are upper-bounded is predicted trivially, as is the monotonicity relationship:  $\phi$  *is more certain than*  $\psi$  entails  $\phi$  *is more likely than*  $\psi$  because they have the same truth-conditions. (They are not, however, identical in their typical use:  $\phi$  *is more certain than*  $\psi$  strongly suggests that one or both is highly likely. This difference probably has an independent explanation given that similar asymmetries are observed with the comparatives of co-scalar adjective pairs like *warm/hot*, *cool/cold*, and *big/enormous*. See §5.2.2 below for further discussion.)

---

<sup>3</sup> Note, however, that *certain* and *likely* may differ in their tendency to describe “epistemic” vs. “aleatory” uncertainty (Kahneman & Tversky 1982; Fox & Ülkümen 2011; Ülkümen, Fox & Malle 2015). As Ülkümen et al. (2015) show using corpus and experimental studies, *likely* is more strongly associated with events that are essentially stochastic (aleatory uncertainty). In contrast, *certain* and related expressions are more frequently used to describe propositions whose truth-values are simply unknown (epistemic uncertainty), such as the probabilities of single events. However, it does not appear that this tendency is categorical, considering (for example) the alternation between *likely* and *certain* in (5.19). All of these examples are instances of epistemic uncertainty, where a non-repeatable past event is under discussion—whether certain debris came from a particular plane. In addition, epistemic and aleatory uncertainty appear to differ in the interpretation of the scale, but not in the structural properties that are our primary focus here. This is analogous to the formal similarity between the two very different interpretations of *long* (spatial extent vs. time). As a result, while the epistemic/aleatory distinction is important and deserving of attention from linguists, I will not discuss it further here.

Third, this analysis explains why both *likely* and *certain* accept percentage modifiers, but only *certain* can be modified by *mostly* and *half*. As we saw already in §4.2.3, both kinds of modifiers select for closed scales, but *mostly* and *half* seem to require in addition that the adjective modified has a maximum standard.

The co-scalar hypothesis does not offer an explanation of the following closely related features of *certain/sure*:

- Why *certain* differs from *full*, and resembles *safe*, in having a minimum-standard antonym.
- Why *completely uncertain*, when (if) it has a degree-modifying interpretation, does not mean “impossible” (cf. (5.18) above).

These features are related in the sense that we would not generally expect *completely* to have a degree-modifying interpretation when combining with a minimum-standard adjective. However, if we somehow allow that (5.18) are (exceptionally) degree-modifying interpretations of *completely uncertain*, then the co-scalar hypothesis would predict that it should mean “probability 0”.

The failure of this analysis to account for these features is not necessarily devastating. Perhaps some other difference between these adjectives could be invoked to explain the puzzle around antonym type, or perhaps there is a certain amount of lexical arbitrariness in the type of antonyms when the scale structure makes available multiple options.

As for the interpretation of *completely uncertain*: it is not at all obvious that these examples are really degree modification structures, and the natural move for an advocate of the co-scalar hypothesis would be to deny that they are. However, even if these data points can be dismissed, a related observation of Klecha (2012, 2014) probably cannot. Klecha points out that for  $\phi$  to be as uncertain as it could possibly be is not the same thing as it is for  $\phi$  to be as unlikely as it could possibly be. The former suggests uncertainty about the question *whether*  $\phi$ , not maximal certainty in the truth of  $\neg\phi$ . If this is correct, the co-scalar hypothesis may be in trouble.

In any case, it would seem worthwhile to consider the prospects of some alternative scalar theories of certainty.

### 5.1.6 Analysis 2: Distinct boundedness properties

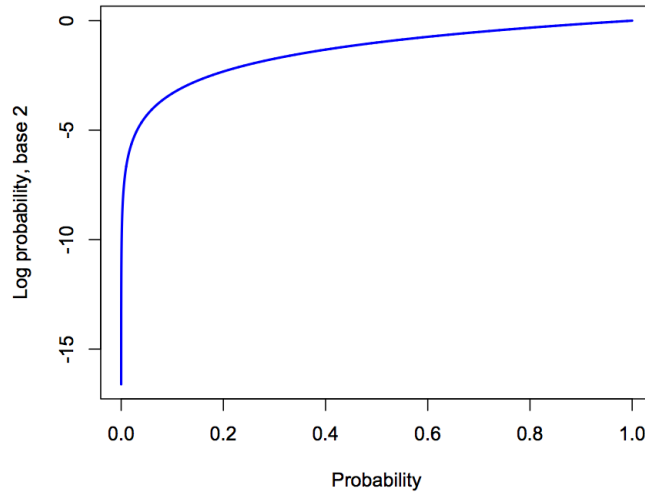
A second possibility is to suppose, following Klecha 2014, that  $\mathcal{S}_{certain}$  is not lower-bounded. Klecha argues that *certain* expresses confidence, but says little about the formal properties of the confidence scale. However, it is perhaps in the spirit of his analysis to suppose that  $\mathcal{S}_{certain}$  and  $\mathcal{S}_{likely}$  are related in either of the following two ways.

First,  $\mathcal{S}_{certain}$  might be equivalent to  $\mathcal{S}_{likely}$  except that  $\mathcal{S}_{certain}$  lacks an intrinsic minimum corresponding to  $\perp_{likely}$ . In the language of degree semantics, this corresponds to the following claim about their relationship. (*prob* is a probability measure, either equal to or deriving  $\mu_{likely}$ : see §4.2.5.)

$$\mu_{certain} = \{(p, r) \in prob \mid r \neq 0\}$$

In effect, this hypothesis says that the certainty scale is the probability scale restricted to the range  $(0, 1]$ .

Second, the certainty scale might correspond to some monotonic transformation of the probability scale. For example, logarithmic transformation of the probability scale is common in information theory and psychology (Shannon 1948; Luce 1959). This transformation maps the  $[0, 1]$  interval to the upper-closed interval  $(-\infty, 0]$ , with no lower bound because  $\log(0)$  is undefined. Figure 5.1 illustrates this relationship with base 2. (The relevant formal properties of the logarithmic transformation do not depend on the base.) As the figure makes clear, this transformation is monotonic:  $\mu_{\text{likely}}(\phi) \geq \mu_{\text{likely}}(\psi)$  iff  $\log(\mu_{\text{likely}}(\phi)) \geq \log(\mu_{\text{likely}}(\psi))$  as long as the latter is defined (i.e., neither likelihood is zero).



**Figure 5.1** Log probability as a function of probability.

These two versions of the “no-minimum” theory make generally similar predictions. Both predict the entailment relationship between *certain* and *likely*, the close relationship between the comparative forms, and the fact that  $\mathcal{S}_{\text{certain}}$  (like  $\mathcal{S}_{\text{likely}}$ ) is upper-bounded by the certainty of a tautology.

In addition, both versions of this hypothesis explain a fact that was puzzling for Analysis 1. On this account, the certainty scale has the same boundedness properties as the safety scale. *Uncertain/unsure* could not be maximum adjectives because, unlike the inverted fullness scale, the inverted certainty scale lacks a maximum element. It is not surprising, then, that the pair *certain/uncertain* resembles the max/min pairs *safe/dangerous* and *dry/wet*, rather than the max/max pair *full/empty*.

There are two major kinds of problems for Analysis 2. The first applies only to the the log probability variant, which is no even asymptotically bounded below (cf. Figure 5.1). Since proportional modification requires minimally that both scalar endpoints exist asymptotically, the log-probability version fails to explain the acceptability of proportional modifiers (*n%*, *mostly*, *half*) in examples like (5.9), (5.10), and (5.19).<sup>4</sup> One way out would be to analyze all of these examples as cases of scalar coercion: certainty is sometimes temporarily equated with probability,

<sup>4</sup> Klecha (2014: §2.3.1) claims that *mostly/half certain* are unacceptable, but the quite natural-sounding corpus examples that we have seen cause me to doubt the reliability of intuitions on this point. Note, however, that it is not clear that the

even though they are not the same thing in general. As we have seen in various points already, the *Deus ex machina* mechanism of scalar coercion does seem to be necessary in some cases. However, one should be hesitant about invoking it, and a theory which is able to avoid such maneuvers should generally be preferred to one which uses them, all else being equal. As a result, I am skeptical that certainty can be equated with log probability.

Second, both versions of Analysis 2 share a problem with Analysis 1: they do not account for Klecha's observation that there is a sharp conceptual contrast between maximal uncertainty and maximal unlikelihood. If anything, Analysis 2 amplifies the problem since it entails that there *is* no such thing as maximal uncertainty.

### 5.1.7 Analysis 3: Covert type-theoretic distinction

According to the third analysis that I will consider here,  $\mathcal{S}_{certain}$  is upper- and lower-bounded like  $\mathcal{S}_{likely}$ , but they have a different internal structure and in fact order objects of different types entirely. This analysis is mainly inspired by Klecha's observation, which I glossed above along the following lines: being maximally uncertain that  $\phi$  is not the same thing as judging  $\phi$  to be maximally unlikely. Rather, it suggests maximal uncertainty as to *whether*  $\phi$ , or judging  $\phi$  and  $\neg\phi$  to be about equally likely.

Klecha (2014: 58) illustrates this point with the examples similar to those in (5.20).

- (5.20) a. It couldn't be less certain that Obama will be reelected.  
b. It couldn't be less likely that Obama will be reelected.

It seems clear that these examples are not identical in meaning. (5.20a) sounds like the hand-wringing of a nervous but dedicated supporter, while (5.20b) indicates that we should abandon hope. In addition, (5.20a) is naturally paraphrased with a *wh*-complement as in (5.21a). The (b) examples are of course not equivalent, since (5.21b) is ungrammatical.

- (5.21) a. It couldn't be less certain whether Obama will be reelected.  
b. \* It couldn't be less likely whether Obama will be reelected.

As I will develop this idea—with inspiration from the compositional analysis of question-embedding predicates in Uegaki 2015—*certain* really orders sets of propositions, rather than individual propositions. To model (5.20), we need a semantics which assigns a minimal degree of certainty to sets of propositions in which probability is spread evenly among the propositions. This indicates maximal uncertainty about which of the propositions obtains. A natural choice here is NEGATIVE ENTROPY, the inverse of the information-theoretic concept of entropy (see Shannon 1948; Cover & Thomas 1991; MacKay 2003 and, for semantic/pragmatic applications, van Rooij 2003, 2004). On this account,  $\mu_{certain}$  would be defined as follows when  $\mathcal{Q}$  is a partition of  $W$ :

$$(5.22) \quad \mu_{certain}(\mathcal{Q}) = \sum_{q \in \mathcal{Q}} \text{prob}(q) \times \log(\text{prob}(q))$$

---

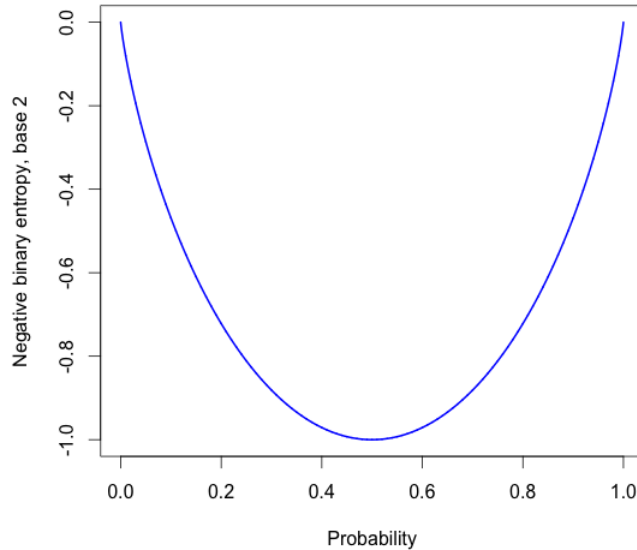
data from proportional modification are a problem for Klecha's analysis of *certain*, if we spell out it out formally as the claim that the relevant interval is the asymptotically bounded  $(0, 1]$ . As noted in chapter 2, §2.4.4, we could state the definitions of proportional modifiers in terms of asymptotic bounds if desired.

If we do not wish to assume that  $\mathcal{Q}$  exhaustively covers  $W$  (cf. below), the following more general definition, which assumes only that the elements of  $\mathcal{Q}$  be mutually exclusive, may provide a plausible modification.

$$(5.23) \quad \mu_{\text{certain}}(\mathcal{Q}) = \sum_{q \in \mathcal{Q}} \text{prob}(q) / \text{prob}(\cup \mathcal{Q}) \times \log(\text{prob}(q))$$

By this measure, a set of propositions  $\mathcal{Q}$  has certainty 0, the maximum possible value, when there is a single proposition that has all of the probability in  $\mathcal{Q}$ .  $\mathcal{Q}$  is as uncertain as it could possibly be when the total probability in its elements is divided evenly among them: for each  $q \in \mathcal{Q}$ ,  $\text{prob}(q) = \text{prob}(\cup \mathcal{Q}) / |\mathcal{Q}|$ .

If  $\mathcal{Q}$  happens to be a partition of  $W$ , then this point occurs when each cell has probability  $1/|\mathcal{Q}|$ . For instance, when  $\mathcal{Q}$  is  $\{\phi, \neg\phi\}$ —the question whether  $\phi$ — $\mu_{\text{certain}}$  is the NEGATIVE BINARY ENTROPY function. As Figure 5.2 illustrates, this function reaches its maximum when  $\text{prob}(\phi)$  is 0 or 1, and is minimized when  $\text{prob}(\phi) = 1/2$ .



**Figure 5.2** Negative entropy of  $\{\phi, \neg\phi\}$  as a function of  $\text{prob}(\phi)$ .

This analysis has the nice feature of giving insight into the entailment from (5.24a) to (5.24b):

- (5.24) a. It is certain who will win the race.  
b. There is some  $x$  such that it is certain that  $x$  will win the race.

Suppose that *who will win the race* denotes a partition of  $W$ , with the form

$$\{\{w \mid \llbracket x \text{ wins the race} \rrbracket^{\mathcal{M}, w} = 1\} \mid x \in D_e\}$$

(Groenendijk & Stokhof 1984). Then we can feed this set of propositions as an argument to *certain*, deriving the following interpretation:

$$(5.25) \quad \llbracket (5.24a) \rrbracket^{\mathcal{M},w} = 1 \text{ iff } \mu_{\text{certain}}(\llbracket \text{who will win the race} \rrbracket^{\mathcal{M},w}) = \mathbf{max}(\mu_{\text{certain}}) = 0$$

This is true iff there is some cell in  $\llbracket \text{who will win the race} \rrbracket^{\mathcal{M},w}$  that has probability 1, which in turn is true iff there is some  $x$  such that  $x$  wins with probability 1.

This does not quite explain the entailment from (5.24a) to (5.24b), though. To do that, the semantics under consideration would have analyze *It is certain that  $x$  will win* as “ $x$  wins with probability 1”. But this semantics does not assign this sentence any interpretation at all: *certain* is looking for a set of propositions, but *that  $x$  will win* just denotes a proposition. So, mutatis mutandis, for nearly all of the examples that we considered earlier in this section.

There are two obvious solutions. One is to follow Uegaki (2015) in allowing  $\phi$  to type-shift into the singleton set  $\{\phi\}$  when the composition demands it. Uegaki suggests that a variety of embedding predicates that can embed both *that*-clauses and questions—*know* and *discover*, for example—are really specialized to take sets of propositions as complements. He argues that their interpretation with *that*-complements are special cases of the question-embedding semantics when the predicate’s argument happens to be a singleton set of propositions.

In this special case, we have to use the definition in (5.23), which does not assume that the set of propositions covers  $W$ . Then the negative entropy analysis reduces to the log probability account sketched in §5.1.6: the maximum possible certainty in  $\{\phi\}$  is when  $\text{prob}(\phi) = 1$ , in which case

$$\begin{aligned} \mu_{\text{certain}}(\{\phi\}) &= \sum_{q \in \{\phi\}} \text{prob}(q) / \text{prob}(\bigcup \{\phi\}) \times \log(\text{prob}(q)) \\ &= \text{prob}(\phi) / \text{prob}(\phi) \times \log(\text{prob}(\phi)) \\ &= 1 \times \log(1) \\ &= 0. \end{aligned}$$

More generally, as the second line shows, the certainty of the singleton  $\{\phi\}$  will always be the same as the log probability of  $\phi$ . The analysis under consideration thus derives the log-probability variant of Analysis 2 as a special case when the complement of *certain* is a *that*-clause.

This analysis has some very nice properties, then: a single semantics can account for *that*-complements—sharing the positive features of the log-probability analysis—and also for the interpretation of *certain* with *wh*-complements, including the validity of the inference in (5.24) and the maximum-entropy interpretation of *completely uncertain whether  $\phi$* .

However, this analysis shares the major empirical drawback of the log-probability analysis: it does not account for proportional modification. In addition, it fails to explain the fact that *completely uncertain that  $\phi$*  can sometimes be interpreted as meaning the same as *completely uncertain whether  $\phi$* . Similarly, *it couldn’t be less certain that  $\phi$* , *it is maximally uncertain that  $\phi$* , etc. are predicted to be generally infelicitous, since there is no lower bound to the certainty of  $\{\phi\}$  according to this definition.

The second possible solution is designed to deal with this puzzle. Instead of shifting the interpretation of the complement clause in *completely uncertain that  $\phi$*  into  $\{\phi\}$ , we shift it into  $\{\phi, \neg\phi\}$  (cf. Roelofsen & Farkas 2015). The negative entropy of this set is lower-bounded, with maximum uncertainty at  $\text{prob}(\phi) = \text{prob}(\neg\phi) = .5$ . This analysis would allow us to make sense of the intuitive equivalence of the examples in (5.26).



- (5.26) a. She is completely uncertain that she can start a career at this point in her life.  
 b. She is completely uncertain whether she can start a career at this point in her life.

The second solution has a serious drawback, though. If this type-shift is generally available, we predict nonexistent interpretations of *that*-clauses with question-embedding predicates. For example, *Bill is certain that  $\phi$*  would have an interpretation which ascribes to Bill maximal certainty about  $\{\phi, \neg\phi\}$ . These truth-conditions are satisfied whenever Bill assigns probability 0 to  $\phi$ , and probability 1 to  $\neg\phi$ —but this is obviously not a situation in which *Bill is certain that  $\phi$*  could be true on any interpretation. Similar objections would hold for *know* and *discover*: for example, we would predict that *Bill knows that it's raining* could mean that Bill knows *whether* it raining. The latter could, of course, be true if Bill knows that it's not raining, as long as it is indeed not raining.

### 5.1.8 Summary: The scale of certainty

*Certain* and its near-synonym *sure* are maximum adjectives; their scales must therefore be upper-bounded. Several arguments including the acceptability of proportional modifiers (*n%*, *half*, *mostly*) suggests that the certainty scale is either lower closed, or asymptotically lower-bounded.

We considered three concrete analyses of certainty, all of which had pros and cons. Certainty and likelihood are closely connected, and it is tempting to simply equate them, attributing differences to the fact that *likely* is a relative adjective and *certain* is maximum. However, this analysis does not predict that *certain*'s antonym should be a minimum adjective, rather than being a maximum adjective like that of *full*. In addition, this analysis fails to account for Klecha's (2012; 2014) observation that minimum certainty is intuitively different from minimum likelihood. An interesting alternative is to treat *certain* as embedding sets of propositions, measuring certainty in terms of the distribution of probability among elements of a set. This analysis provides a unified analysis of question- and proposition-embedding *certain*, but an effort to analyze maximal uncertainty in these terms generates incorrect predictions about the basic meaning of *It is certain that  $\phi$* .

While none of the analyses canvassed here is fully satisfying empirically, the discussion should hopefully provide a starting point for future elaboration and a set of interesting empirical and theoretical puzzles that are in need of resolution.

## 5.2 Epistemic Possible

### 5.2.1 Possibility as a scalar concept

Traditionally possibility has been analyzed as something that you either have or you don't. The classical semantic analysis assumes this: *possible* denotes an existential quantifier over a set of accessible worlds, with variation in meaning (epistemic, metaphysical, ...) determined by the identity of the selected set. The existential quantifier is not graded. Either some worlds in the relevant set have the property in question, or none do.

Many commentators on English usage seem to agree with the classical semantic analysis on this point. A video lesson on "Modal Verbs" from *Oxford Online English* (no connection to Oxford University) asserts that there are no grades of possibility, or indeed certainty.

Possibility and certainty are both ‘yes/no’ qualities. Either something is possible, or it isn’t. There’s nothing in between. Certainty is the same: you can’t be half-certain about something. Either you’re certain that something is true, or you aren’t.

Probability isn’t like this. Probability can have different levels. It doesn’t make sense to say that one thing is more possible than something else, but it is clear if you say that one thing is more probable.

Similarly, a commentator from the WordReference.com Language Forums informs an English learner that *more possible* is not acceptable in English.

**Lulu0wns:** I would like to know if both of their comparative forms—**more possible** and **more probable**—mean the same? also their superlative forms ...

**heypresto:** I don’t think you can say ‘more possible’ or ‘the most possible’ - something is either possible or it’s not. You can, however, say ‘more probable’. Something that has say, an 80% likelihood is more probable than something with a 55% likelihood. ...

Others have difference advice for learners. One book advises teachers of English to introduce students to the concept of “degrees of possibility”, employing the phrases *more possible* and *less possible* as if they were obviously meaningful.

Possibility and doubt are closely related ideas and they are best treated as one facet of a combined idea. Something is possible, but there are degrees of possibility—something can be more possible or less possible i.e. more doubtful. (Buckmaster 2014: 109)

As the latter example suggests, it is—despite the admonitions of certain commentators—not difficult to find uses of *possible* and *possibility* that would make sense only if the underlying concept were graded. For instance, the common phrase *increase the possibility that/of* would be difficult to interpret if possibility were truly an all-or-nothing concept. Here are a few naturalistic examples.

- (5.27)
- A weaker heliosphere increases the possibility that the earth will be exposed to harm from intergalactic cosmic materials.
  - [T]he presence of a gun increases the possibility that a normal family fight or drinking binge will become deadly.
  - Warmer weather increases the possibility of fire.

Interestingly, *possibility* seems to be interchangeable in these examples with the nouns *likelihood* and *chance*, which clearly express graded epistemic concepts.

- (5.28)
- A weaker heliosphere increases the likelihood that the earth will be exposed to harm from intergalactic cosmic materials.
  - The presence of a gun increases the chance that a normal family fight or drinking binge will become deadly.
  - Warmer weather increases the likelihood of fire.

This equivalence points again to the very close connection between epistemic possibility and scalar likelihood.

More generally, I suggest that this construction can be used to diagnose whether an abstract nominal concept refers to a concept which is underlyingly scalar. Abstract nouns that pick out scalar concepts—for example price, speed, length, strength, size, anger, skill, rate of growth, frequency, temperature, success, and wealth—are felicitous as the subject of intransitive *increase* and the direct object of its transitive counterpart.

- (5.29) a. His wealth increased through the years.  
b. She increased her speed by stepping on the gas.  
c. Bill is so mad, surely his anger cannot increase any further.

(So, of course, are nouns referring to scalar modal concepts like likelihood, certainty, chance, ability, goodness, and obligation.) Abstract nouns referring to non-scalar concepts—like *thought*, *truth*, and *nationality*—generally are not.

- (5.30) a. ?? Bill's thought about Mary increased through the years.  
b. ?? She increased the deposition's truth by saying more.  
c. ?? Bill is so patriotic, surely his American nationality cannot increase any further.

The fact that *possibility* patterns here with *price*, *size*, and other nouns expressing scalar concepts strongly suggests that the concept that it expresses is fundamentally a scalar one. Since the existential quantifier is all-or-nothing, rather than scalar, it follows that the concept of possibility is not adequately captured by an analysis in terms of existential quantification.

Data involving epistemic *possible* indicate that this adjective not only expresses a scalar concept, but is in addition grammatically gradable. For example, it is easy to locate examples of this item being used in comparative and equative constructions—a core diagnostic for gradability.

- (5.31) In 1923 the black trade union leader Asa Philip Randolph received a severed hand through the post with the warning that he himself may suffer such mutilation if he did not cease his opposition to Garvey. Ostensibly sent by the Ku Klux Klan, it was no less possible that the grisly package came from UNIA members or supporters. (Verney 2013: 119)
- (5.32) Cutler In a Bears Uniform Is More Possible Than Ever ... The Broncos have agreed to trade Cutler. ... This is our chance to get a franchise QB for the first time since 1950. Here's to getting our hopes up!

In addition, we find corpus examples of epistemic *possible* with various degree modifiers. Here are several with *too* and *very*, perhaps the most unambiguous modifiers we could use to diagnose gradability.

- (5.33) Examples of *too possible*
- a. My friend that's in has a step mom that called [his parole officer] on him. That resulted in a 5 year sentence then he did something inside that got him another 10. You don't want to be part of that. It's too possible that you'd regret it.
- b. I almost went for the triggering of MP's recorded parts, but it's too possible that he'll attend a show, see them doing that, then there will be an ugly lawsuit.

- c. I also wouldn't seek to do something that doesn't interest both people, as it's too possible that it would turn awkward. Quickly, at that.
- (5.34) Examples of *very possible*
- a. It's very possible that landline phones will be gone by 2020.
  - b. Siegel, who has projected the Dow Jones industrial average could hit 20,000 this year, said it is still "very, very possible" that could happen.
  - c. It is very possible that our sighting on March 9 was actually a wandering local dog who had gotten out of its fence.

### 5.2.2 But is *possible* really a gradable adjective?

Two analyses seem to be available. The first is that the traditional semantic analysis is wrong, along with the various commentators who have intuited that possibility is not a scalar concept and *possible* is not a gradable adjective. Lassiter (2010, 2011a) argues along these lines that *possible* is a minimum-standard gradable adjective. For instance, naturalistic examples such as (5.35) suggest a minimum adjective classification for *possible*, *open*, and *dangerous* alike, since *slightly* tends to occur with minimum adjectives (cf. Rotstein & Winter 2004; Kennedy 2007 and discussion in ch.4).

- (5.35)
- a. It's actually not uncommon for people to sleep with their eyes slightly open.
  - b. The wire is also very springy, so it can be difficult and slightly dangerous.
  - c. In a tense situation, it's slightly possible that an asteroid entering our atmosphere could trigger a nuclear war.

Pointing to naturally-occurring examples of *n% possible* (cf. (5.38)-(5.40) below), Lassiter suggests that  $\mathcal{S}_{possible}$  is formally identical to  $\mathcal{S}_{likely}$ . This account explains the various examples of *more/very/too possible* that we have seen, but does not explain why commentators often intuit that these examples are not acceptable. It also does not offer an explanation of Klecha's (2012) observation that degree modification of *possible*, while attested, is less frequent than with certain gradable adjectives.

The second possibility is that the traditional analysis is correct, and that the attested examples of gradable *possible* involve coercion of a basic meaning to a gradable use. This is proposed by Klecha (2012, 2014) (see also Herburger & Rubinstein 2014, where it is extended to the German item *möglich*). On this analysis, the basic meaning of *possible* is the non-graded one—existential quantification over a set of possibilities—and all gradable uses involve coercion of *possible* into some related scalar concept.<sup>5</sup> The examples of gradable *possible* that we have seen above are thus of the same kind as Miracle Max's use of *mostly dead*:

<sup>5</sup> Klecha's main arguments (his chapter 2) involve (i) intuitions about the acceptability of degree modifiers with *possible*, (ii) a study of patterns in the frequency of degree modifiers in COCA (Davies 2008-), and (iii) an experimental study testing for the acceptability of degree modifiers. I disagree with many of the intuitions that Klecha reports around (i), and naturalistic examples of the constructions that he considers infelicitous are easy to find in several cases; some examples are reported in this section. The corpus study (point ii) is inconclusive, and its interpretation relies in any case on the questionable assumption that adjectives that share grammatical and semantic characteristics should pattern together in a transparent way in terms of the frequency of use of certain modifiers. Many factors affect frequency of use other than grammatical and semantic features, including—crucially in the case of epistemic adjectives—the availability

**Miracle Max:** He probably owes you money huh? I'll ask him.

**Inigo:** He's dead. He can't talk.

**Miracle Max:** Woo-hoo-hoo, look who knows so much. It just so happens that your friend here is only MOSTLY dead. There's a big difference between "mostly dead" and "all dead". Mostly dead is slightly alive. With all dead, well, with all dead there's usually only one thing you can do.

**Inigo:** What's that?

**Miracle Max:** Go through his clothes and look for loose change.

(*The Princess Bride*, 1987)

(Assuming, that is, that *dead* really is a non-gradable adjective. This is usual, but—as Rotstein & Winter (2004: 265) and Kennedy & McNally (2005a: 359) point out—expressions such as *almost/half dead* suggest that *dead* is a gradable adjective on a fully closed scale. This is what Miracle Max's use of *mostly dead* also hints at. Rotstein & Winter suggest that *dead* is unusual mainly in being unnatural in the comparative.)

Unfortunately, the dialectic about whether *possible* is "really" gradable is set up in a way that makes it very difficult to resolve. In fact, the theories seem to be indistinguishable in terms of what they predict to be possible or impossible. Consider first the positive form meaning. According to the classical semantics *It is possible that  $\phi$*  is true whenever there is an epistemically accessible world where  $\phi$  holds. On the scalar alternative, this sentence is true whenever  $\phi$  has possibility greater than  $\theta_{possible}$ . If  $\theta_{possible} = 0$  and  $S_{possible} = S_{probable}$ , as Lassiter (2010) proposes, then the analysis is simply that  $\phi$  is possible whenever  $prob(\phi) > 0$ . If we constrain *prob* so that all of the probability mass resides in the accessible worlds, it is very difficult to distinguish these proposals empirically. Their predictions diverge only in cases which involve propositions that are epistemically possible but have zero probability—for example, the proposition that a certain car, at a certain instant, is moving at exactly  $r$  miles per hour for some real number  $r$ . Typically there will be uncountably many possible values of  $r$ , each of which has zero probability of being the true value.

We might conclude from this description that possibility cannot be equated with non-zero probability, as e.g. Hájek (2003); Yalcin (2007) do. However, the argument is semantically relevant only if we suppose that natural language sentences involving real values are interpreted with zero granularity—so that *The car is moving at 35 miles per hour* would be false if it were in fact moving

---

of competing forms to express an intended message. (I suspect that this point plays a role in the only coherent cluster that emerges, involving extreme adjectives—all of which are in pragmatic competition with relative adjectives.) The same point applies to Klecha's experimental study (iii), given that acceptability is sensitive to a variety of pragmatic factors, not to mention language users' knowledge of frequencies. But more troubling is that Klecha's experiment compares *possible* to *likely*, rather than to other purported minimum adjectives. It is, I think, no surprise to anyone that *possible* and *likely* behave differently with respect to degree modification. So, Klecha's study does not establish that degree modification is less natural with *possible* than with uncontroversially gradable minimum adjectives. If it is somewhat less natural, though—as I suspect it is—this would be the usual case when an intended message can be expressed by multiple adjectival constructions. In general, the construction involving the relative adjective is less marked. (For example, compare the intuitive contrast between *very likely* vs. *?very possible* to that between *very big* vs. *?very enormous*. All of these constructions are possible and attested, but they vary in their intuitive naturalness.) The analogy to extreme adjectives will be discussed further in the main text in a moment.

35.000000000001 miles per hour. As I have argued in some detail elsewhere (Lassiter 2015b,c), the existence of granularity in the interpretation of numerical expressions in natural language (e.g., Lewis 1979; Krifka 2007) renders this difference moot. If *35 miles per hour* is interpreted as  $35 \pm g$ , where  $g$  can be *any* real value greater than 0, then the event picked out by *The car is moving at 35 miles per hour* will not generally have zero probability. The argument involving possible events with zero probability is only problematic if we assume that numerical expressions in natural language denote with the precision of pure mathematics, and this assumption is not especially plausible. But if this is the only avenue for distinguishing empirically between the existential-quantification analysis and the non-zero-probability analysis by looking exclusively at the interpretation of the positive form, we are back where we started.

Grammatical gradability seems initially more promising as a way to distinguish empirically between these analyses. Unfortunately, even the numerous naturalistic examples given above as evidence for the gradability of *possible* are not strictly probative. We would, I suspect, not hesitate to interpret these examples as showing that *possible* is gradable if it were a word with less theoretical history. Still, since an advocate of the classical semantics can make use of the coercion mechanism to explain away any individual example which suggests gradability, the existence of these examples does not constitute a knock-down argument for the scalar analysis. An extreme example may illustrate the point here: consider a theory according to which there are no gradable adjectives in English, and the many apparent examples that we see are all analyzed as coercion of a basic, non-gradable meaning into a gradable use. No amount of empirical evidence could disprove this theory. Still, it would hardly be worth taking seriously.

Ultimately, methodological and plausibilistic considerations like these may be all that we can appeal to. As I have suggested at various points, it is methodologically desirable to limit appeals to “coercion” because there are no clear constraints on the theoretical power of this mechanism. We should not use it as a blunt instrument to dismiss recalcitrant data.

How, then, should we weight the fact that many English speakers—including some semanticists—intuit that *more/too/very possible* are unacceptable? If there is no alternative explanation of these gut feelings, perhaps they can be taken to index the “basic” nature of the possibility concept, with increasing discomfort indexing the inherent difficulty associated with the coercion operation (whatever this is). However, this analysis assumes that linguistic intuitions give us a direct pipeline into our linguistic competence. I believe that such a theory of intuitions is implausible: intuitions are another kind of linguistic performance, and their relation to the competence grammar is complex and indirect. In the case of grammaticality judgments, for example, it is well-known that linguists’ intuitions are unstable, influenced by theoretical commitments, and highly sensitive to considerations of pragmatic acceptability (Schütze 1996; Wasow & Arnold 2005). In addition, experimental evidence suggests that grammaticality intuitions can be influenced by frequency (Glass & Lau 2003). Presumably intuitions of acceptability are equally susceptible to these effects; if so, perhaps we can appeal to the infrequency of the comparative and modified forms of *possible* to explain the unstable intuition that they are unacceptable.

This only pushes the question back one step, though. How can a scalar analysis explain the fact that *more/less/very/too possible* are relatively infrequent? It is hard to be sure, since many factors influence frequency that are related to the the grammar only marginally or not at all (pragmatic as

well as narrowly linguistic; on the latter see, e.g., Bresnan & Ford 2010). However, a scalar analysis may be able to appeal to the claim that *possible* is co-scalar with the relative adjective *likely* as a partial explanation of the frequency facts. Here is one way that such an analysis might go.

When multiple adjectives are available for expressing a single comparison, one of them is often “unmarked”, i.e., more frequent and strongly preferred intuitively. For example, “extreme” adjectives like *enormous* and *scrumptious* have sometimes been claimed to be impossible in the comparative. In reality, they are possible and attested, but there is usually a strong preference for using the less marked relative counterparts *big* and *tasty* (cf. Morzycki 2012).

- (5.36) a. Dogs are  $\left\{ \begin{array}{l} \text{bigger} \\ ?? \text{ more enormous} \end{array} \right\}$  than mice.  
 b. Everything is  $\left\{ \begin{array}{l} \text{tastier} \\ ?? \text{ more scrumptious} \end{array} \right\}$  than natto. (Morzycki 2012, (10c))

Of course, the intuition that the comparative is odd with extreme adjectives like *enormous*, *tremendous*, *marvelous*, *scrumptious* does not show us that these examples are unacceptable. Indeed, naturalistic counterparts to the examples in (5.36) are easy to find.

- (5.37) a. Astrophysicists have found an out-of-place supermassive black hole—12 billion times more enormous than the sun—that mysteriously formed when the cosmos was less than 900 million years old.  
 b. As for the (ludicrous) plan itself: It’s as though Trump read a copy of the Jeb Bush plan, thought about it for a moment, and then tossed it at an underling, yelling, “We should do this, but make it more tremendous, more marvelous!”  
 c. We navigate through a maze of unfamiliar dishes, each more scrumptious than the last—albacore tataki, masago unagi, ebi fry, ikura this and hotategai that.

In general, the relative adjectives associated with these properties provide the default form for the comparative and modified forms, with the extreme adjectives being reserved for cases in which a speaker wishes to emphasize the (positive-form) extremeness of some degree. My hunch is that the situation with *possible* and *likely* is related: relative *likely* provides the default comparative form, and *more possible than* is generally reserved for cases in which a speaker wishes to emphasize the relative improbability of both things being compared. This is a mirror image of the situation with extreme adjectives, whose comparative emphasizes that both comparata have a high degree of the property in question.

There is a final empirical consideration that seems to favor the scalar analysis. As we saw in (5.27) above, the phrase *increase the possibility of/that* is readily acceptable, as much so as non-modal cases such as *increase the size of*. The modal and non-modal examples cited there are, I take it, unobjectionable for all, and a coercion analysis does not seem applicable: the interpretation does not require grammatical gradability *per se*, but rather that possibility, like size, be the sort of thing that could *increase*. Existential quantification does not seem to have this character, but the scalar concepts of size and possibility do.

While I favor the scalar analysis for these reasons, I realize that the argument is not airtight. In the rest of this section I will develop the scalar analysis of *possible* a bit more. Even for readers

who remain convinced of the classical analysis of *possible*, this should be an interesting activity: for them, it represents an analysis of the secondary, coerced scalar interpretation of *possible* that their semantic theory must also admit.

### 5.2.3 Relationship between gradable *possible* and *likely*

(Scalar) possibility is clearly closely related to likelihood. We saw this already in examples like (5.27), where substituting *likelihood* for *possibility* seemed to result in no change of meaning. Many of the diagnostics that we used to relate certainty to likelihood are applicable here as well.

- Likelihood entails possibility: It would be incoherent to claim that  $\phi$  is likely, but deny that it is (epistemically) possible.
- Comparative possibility entails comparative likelihood: It would be incoherent to claim that  $\phi$  is more (epistemically) possible than  $\psi$ , but deny that  $\phi$  is more likely than  $\psi$ .
- Both scales are lower-bounded by the likelihood of a contradiction. Nothing can be less possible, or less likely, than something that is impossible.

Many examples of *n% possible* seem to indicate that the relevant scale is one of probability.

- (5.38) If you have a mother and a father that are allergic, then it's most likely their children will be allergic. If only one of them is, then it may be 25 to 30-percent possible that their children will be allergic.
- (5.39) I felt that if it was 80-90 percent possible that [the cancer] hadn't spread, I didn't want the hysterectomy.

In addition to providing another example in which *n% possible* is intuitively interpreted as naming a probability, example (5.40) explicitly draws the connection with likelihood: the same outcome is described both in quantitative terms as “2% to 10% possible”, and in qualitative terms as “very unlikely ... but not impossible”.

- (5.40) It is very unlikely—2% to 10% possible—but not impossible, that residential or occupational EMFs [electromagnetic fields] could be responsible for even a small fraction of birth defects, low birth weight, neonatal deaths, or cancer generally.

### 5.2.4 Degree modification with *possible* and *likely*

Further examples indicating the close relationship between these concepts are in (5.41).

- (5.41) a. They shook that loss off, so it is very possible, and perhaps even probable, that they will do the same here.
- b. [I]t's extremely possible (maybe even likely) that Lawrence could be asked to take over the place-kicking duties next fall as a true freshmen ...



Interestingly, the examples in (5.41) seem to equate a very high degree of possibility with a moderate degree of likelihood/probability. At first glance, this might seem to be an argument against analyzing  $\mathcal{S}_{possible}$  and  $\mathcal{S}_{likely}$  as being identical. After all, in the best-known compositional analysis of degree modification—due to Kennedy & McNally (2005a); Kennedy (2007)—degree modifiers are in morphological competition with the *pos* morpheme. The interpretation of modified variants of these adjectives should thus be sensitive only to information carried in the adjectival root—notably, information about scalar structure—and not to special properties of the positive form.

However, as we saw already in chapter 4, there is no deep commitment in the Kennedy/McNally theory to the claim that this is the *only* kind of degree modification. In §4.2.3 and §4.2.9, I pointed to evidence that certain degree modifiers, such as *slightly*, *half*, and *mostly*, embed the positive form of the adjective and adjust its interpretation. (So, they presumably have semantic type  $\langle\langle e, t \rangle, \langle e, t \rangle\rangle$ .) In addition, Kennedy & McNally (2005b) argue in other work on modification that intensifiers such as *very* (and presumably *extremely*) operate in this way. If so, we have no reason to expect *very/extremely possible* and *very/extremely likely* to pattern together, even if they are on the same scale, since their positive form meanings are plainly different.

Further evidence for an analysis along these lines comes from the observation that co-scalar adjectives like *warm* and *hot* have different meanings when combined with *very* and *extremely*. For instance, the examples in (5.42) strongly imply that something can be *extremely warm* without clearly being hot.

(5.42) Only other major difference is that it gets extremely warm, borderline hot - not anything to be worried about, but the MINIX stays fairly cool under load ...

(5.43) The recent extended period of extremely warm (almost hot) weather has the majority of us gardeners racing here and yonder while trying to catch up with Mother Nature ...

These examples are, of course, very similar to the examples with *possible* and *likely* in (5.41). *Warm* and *hot* measure temperature but differ in the temperature that the positive form requires, relative to context. *Very* and *extremely* take this positive-form meaning and adjust it. Similarly, *possible* and *likely* both measure probability, but differ in the requirements of their positive form. The positive form meaning provides the starting point for interpreting *very/extremely possible* and *very/extremely likely*. Contrary to initial appearances, then, the co-scalar analysis actually does a good job of accounting for cases like *extremely possible*, *perhaps even likely*, and the observation that something can be very possible without being very likely.

Further evidence for the co-scalar analysis comes from the behavior of what Kennedy & McNally (2005b) call “true degree modifiers”. These are modifiers which saturate the degree argument of the root (type  $\langle\langle d, e \rangle, \langle e, t \rangle\rangle$ , if we assume Kennedy’s (2007) measure function analysis). Kennedy & McNally (2005b) suggest that *that*, *how*, and measure phrases like *three feet* fit this description. If so, the addition of such a modifier should neutralize any distinction between co-scalar adjectives, even if they differ in the positive form and when modified by *slightly*, *very*, *extremely*, etc. The following examples test this prediction for *possible* and *likely*:

(5.44) a. How possible is it that Mary has tuberculosis?

b. How likely is it that Mary has tuberculosis?

(5.45) [Mary says, while pointing to Bill’s test results which indicate a 10% chance of cancer:]

- a. If it were only *that* possible that I have cancer, I wouldn't want chemotherapy.
- b. If it were only *that* likely that I have cancer, I wouldn't want chemotherapy.

I detect no difference in meaning between the (a) and (b) examples here. (The examples with *likely* are generally more natural, but this seems to be the default when a relative competitor is available, as we noted above for comparatives.) This is as the co-scalar hypothesis would lead us to expect, assuming Kennedy & McNally's (2005b) treatment of *how* and *that*. In contrast, a theory which assigned different scales to *likely* and *possible* would have to take on the burden of explaining why the difference in meaning between them would be neutralized by the "true" degree modifiers.

The behavior of degree modifiers with *possible* and *likely* is complex, and is rather puzzling if we assume both the co-scalar hypothesis and the theory of modification in Kennedy & McNally 2005a. However, when we consider Kennedy & McNally's (2005b) more elaborate theory of degree modification we find reason to expect the observed differences, along with unexpected support for the co-scalar hypothesis in the behavior of Kennedy & McNally's (2005b) class of "true degree modifiers".

### 5.2.5 Summing up

The balance of evidence suggests that *possible* is a minimum-standard gradable adjective, rather than an existential quantifier as in the classical analysis. Entailments with *likely* and various aspects of modification are consistent with the hypothesis that  $\mathcal{S}_{possible}$  and  $\mathcal{S}_{likely}$  are identical, with differences in behavior attributable to the fact that *possible* is minimum while *likely* is relative. Future developments in our understanding of degree modification will hopefully clarify the situation further.

## CHAPTER 6

### Implications for the epistemic auxiliaries

The auxiliaries *must* and *might*, and the quasi-auxiliary *have to*, are not obviously gradable. However, there are significant semantic relationships between them and the adjectives that we discussed above—*likely*, *probable*, *certain*, *sure*, and *possible*. As a result, conclusions about the semantics of the epistemic adjectives have the potential to illuminate the semantics of the epistemic auxiliaries, ruling out otherwise plausible candidates for their interpretation.

A roadmap for this chapter: first we will consider a number of empirical issues around the relationship between the epistemic adjectives and auxiliaries. Then, in §§6.3–6.6 we will discuss a variety of theories of *must* and *might*—quantificational, scalar, and mixed. Next, I provide intuitive support for the claim that *must* and *might* are context-sensitive, and then present the results of an experiment which sheds light on the details of *might*'s context-sensitivity. The experiment also reveals that *possible* is weaker than *might*, falsifying the usual assumption that these items are equivalent in their epistemic interpretations. Finally, §6.7 considers the purported epistemic readings of *ought* and *should*, discussing two recent proposals as to their meaning, how they relate to the evidential meaning of *must*, the concept of normality, and the scalar semantics of *normal*.

#### 6.1 *Must*, *certain*, and (in)directness of evidence

In many contexts, *must*  $\phi$  seems to be roughly equivalent to  $\phi$  *is certain*. The first question to address, then, is how these two expressions are related. (The epistemic interpretation of *has to*  $\phi$  behaves similarly to *must*  $\phi$  in most relevant respects. I will assume that its interpretation is the same as *must*'s.)

As Palmer (1979) and von Stechow & Gillies (2010) point out, *must* has an indirect evidential meaning: it is strange to say that something *must* be the case if one has direct observational evidence for it. For instance, if one is looking directly at a reliable timepiece which reads 3PM, it is very strange to utter “It must be 3 o'clock”. This utterance is, however, acceptable if the conclusion that it is 3 o'clock has been formed by reasoning from indirect evidence: say, I just observed Sam having tea, and I know that Sam generally has tea at 3 o'clock.

*Certain*, in contrast, does not seem to be associated with an evidential meaning. If so,  $\phi$  *is certain* should be usable in some situations in which *must*  $\phi$  is not, when the speaker is maximally confident about  $\phi$  whether or not she has direct evidence for the truth of  $\phi$ .

Nevertheless, in cases where *must*'s requirement of indirect evidence is satisfied,  $\phi$  *is certain* does seem to entail *must*  $\phi$ . For example, it is intuitively contradictory to assert that  $\phi$  is certain while denying that it must (or has to) be the case. Suppose that I see someone off in the distance and notice that the person is wearing a unique hat. My friend Bill is the only person I have only ever seen wearing such a hat. I might utter any of the following:

- (6.1) a. I'm certain that that is Bill approaching (I recognize his funny hat).  
b. It's certain that that is Bill approaching.  
c. That must be Bill approaching.

- d. That has to be Bill approaching.

It would be very strange, though, to assert (6.1a) or (6.1b) while denying (6.1c) or (6.1d):

- (6.2) a. ?? I'm certain that that is Bill approaching, though it doesn't have to be Bill.  
b. ?? It's certain that that is Bill approaching, though it doesn't have to be Bill.

(We must use *have to* in these examples because *must* resists scoping directly under negation.)

The key question, then, is whether the converse—an assertion of *must* or *have to* coupled with a denial of certainty—is acceptable.

- (6.3) a. That must be Bill approaching, though I'm not certain that it is.  
b. That must be Bill approaching, though it's not certain that it is.

Here, there is disagreement in the literature. The proposal of von Fintel & Gillies (2010), on the most obvious interpretation, would predict that examples like those in (6.3) are totally unacceptable. If the prediction is correct, it would provide evidence for their conclusion that *must*  $\phi$  expresses maximal confidence in the truth of  $\phi$ —like  $\phi$  is *certain*, but with the additional entailments of truth and indirect evidence. In contrast, I find the examples in (6.3) to be quite natural. For example, I might utter either of these if recognition of Bill's unusual hat has made me highly confident that the person in the distance is him, but I also wish to hold out the slight possibility that he has lent his hat to someone else. (Perhaps I know that he is sometimes generous in this way.)

A related empirical issue involves the behavior of *must* and *certain* in the *A, in fact B* construction. Typically, this construction is unremarkable if *A* is strictly weaker than (asymmetrically entailed by) *B*, but distinctly odd if *A* entails *B*. This pattern can be used as a diagnostic of differences in strength, as in (6.4) and (6.5).

- (6.4) a. It's warm today—in fact, it's hot.  
b. ?? It's hot today—in fact, it's warm.
- (6.5) a. Mary ate most of my cookies—in fact, she ate all of them.  
b. ?? Mary ate all of my cookies—in fact, she ate most of them.

Injecting *must* and *certain* into this frame, my intuitions pattern similarly: there is an asymmetry between (6.6a) and (6.6b) which suggests a difference in strength.

- (6.6) a. That must be Bill approaching—in fact, I'm certain that it is.  
b. ?? I'm certain that that is Bill approaching—in fact, it must be.

If so, then we have another argument that *must* is associated with a lower confidence/probability requirement than *certain* is. (The oddity of (6.6b) is particularly striking since von Fintel & Gillies's (2010) theory seems to predict that *must* is strictly stronger than *certain*. If so, it should be (6.6a) that is odd, rather than (6.6b))

Presumably von Fintel & Gillies would have the opposite intuition about (6.6) as well, though. Intuition clashes being hard to resolve, we should look for guidance to examples produced by speakers with no theoretical stake in the issue. In fact it is possible to find naturally-occurring examples in which speakers use *must* to make a proposition while explicitly indicating less than

full confidence in its truth. Lassiter (2016) gives a number of such examples, two of which are reproduced here.

- (6.7) There's one missing pepper on the ground a few feet away. A closer look reveals it has been chewed by something. I wouldn't put it past that pesky blue jay to have teeth but then I think it is unlikely he does. It must be a squirrel. What else can get onto a second floor balcony?
- (6.8) I refuse to believe that this one game, Lost Planet 2 DX11, which was previously 100% stable remember, is crashing because my overclock is unstable .... It's not impossible, granted, but IMO it is highly unlikely. There must be some other cause.

In both of these examples, the author explicitly indicates that there are other possible explanations for the evidence observed, and nevertheless chooses to mark the most probable explanation by *must*. While these examples fall short of an explicit conjunction of the form *must  $\phi$  and  $\phi$  is not certain*, it is clear that the authors would not wish to commit to the certainty of the proposition marked by *must*. In (6.7), *It is certain that it is a squirrel* would contradict the earlier implication that the blue jay *might* be the culprit. In (6.8), *It is certain that there is some other cause* would contradict the author's explicit assertion that "It is not impossible" that an unstable overclock is the cause of the problem.

Here are some additional arguments in favor of the supposition that *must  $\phi$*  does not entail  *$\phi$  is certain*. First, authors sometimes gloss *must* as *almost certain*.

- (6.9) That face... was it not the same face as the shepherd he had met on the way? The one who gave him the straw? It must be, he was almost certain.
- (6.10) There must be some sort of limitation. I'm almost sure in that I cannot redeem this free night voucher towards my \$3k / night stay.

This example is of course not formally incompatible with the thesis that *must* entails maximal confidence. However, it would be rather odd for an author to use *almost certain* here if the briefer, logically stronger *certain* would have accurately depicted their epistemic state.

Second, authors sometimes explicitly conjoin an expression of the form *must  $\phi$*  with a statement which entails non-maximal subjective certainty in the truth of  $\phi$ . The authors of (6.11) and (6.12) attribute do so explicitly, and (6.13) attributes a set of beliefs with this content to a famous 19<sup>th</sup>-century geologist.

- (6.11) I have an injected TB42 turbo and dont like the current setup. There is an extra injected located in the piping from the throttle body.. Must be an old DTS diesel setup but Im not certain. Why would they have added this extra injector?
- (6.12) Now this mortician had this great idea. He must have been Irish but I don't know for sure.
- (6.13) More than a century ago Sorby (1879) found that the specific gravity of both perforate and non-perforate coral skeletons was about 2.75. He concluded that scleractinian coral exoskeletons must be wholly aragonite, although he was not certain that calcite was entirely absent. (Constantz & Meike 1989: 201)

A third point is that *must* and *might* are usually assumed to be duals, as are *certain* and *possible*. In §6.5 I will show experimentally that *might* is stronger than *possible*. It follows, if the usual duality assumptions are correct, that the assertive component of *must* is weaker than that of *certain*—and indeed, Lassiter (2016) shows experimentally that *certain* has stricter confidence requirements than *must*, once we control for the indirectness requirement of the latter.

In light of these considerations, I conclude that  $\phi$  *is certain* has stronger probability-related entailments than *must*  $\phi$ . (This is orthogonal to the issue of indirectness, and logically independent of the question of whether these sentences entail the truth of  $\phi$ . See Lassiter 2016 for a detailed discussion of the relationship among these meaning components.) In other words, epistemic *must* is “weak”—not in the sense of being extremely permissive, but in the sense of requiring less than maximal commitment to the truth of its complement. This point was first made by Karttunen (1972) and has been repeated by Kratzer (1991) and many other linguists. *Must* is **not** “strong” in the sense of entailing maximal certainty, as is usually assumed in modal logic. Recently, von Stechow & Gillies (2010) have given a number of philosophical and linguistic arguments in favor of a strong semantics. See Lassiter (2016) for a detailed response, including several additional empirical and theoretical arguments in favor of a “weak” semantics.

Our first empirical constraints on a theory of the epistemic auxiliaries, then, involve indirectness and the relationship with certainty. *Must*  $\phi$  is inappropriate when  $\phi$  is known through direct experience; when this condition is satisfied, it indicates a level of confidence which is high, but still lower than the level indicated by *certain*. As we will now see, though, the requirement cannot be *much* lower, or we would fail to encode the right semantic relationship between *must* and *likely*.

## 6.2 Entailments among epistemic auxiliaries, *likely*, and *possible*

This section discusses some further inferential connections among the adjectives and auxiliaries which are, I hope, uncontroversial. First:

(6.14) *must*  $\phi$  entails  $\phi$  *is likely*.

For example, it would be very strange to assert that  $\phi$  *must* be the case while denying that  $\phi$  is likely, or asserting something that entails that  $\phi$  is not likely.

- (6.15) a. # It must be raining, but it’s not likely that it is.  
b. # It must be raining, though it’s probably not raining.

Second, and relatedly:

(6.16) *must*  $\phi$  entails  $\phi$  *is much more likely than*  $\neg\phi$ .

- (6.17) a. # It must be snowing, but more likely it isn’t snowing.  
b. # It must be raining, though it’s only a tiny bit less likely that it’s snowing.

Given these connections, the confidence in  $\phi$  required by *must*  $\phi$  must reside somewhere in between that required by *much more likely than not*, and that required by *certain*. Suppose for a moment that *must* has a probabilistic component as part of its meaning: if (but not only if) *must*  $\phi$  is

true, then  $\text{prob}(\phi) > \theta_{\text{must}}$ . We will consider momentarily whether this is a reasonable assumption. If it is, we can summarize these constraints like this:

$$\theta_{\text{likely}} < \theta_{\text{must}} < \theta_{\text{certain}}.$$

Turning to *might*, there are some obvious connections with likelihood:

(6.18)  $\phi$  is likely entails *might*  $\phi$ .

(6.19)  $\phi$  is more likely than  $\neg\phi$  entails *might*  $\phi$ .

In addition, epistemic *possible* and *might* are very close in meaning. A standard assumption is that (6.20) holds generally.

(6.20) In their epistemic interpretations,  $\phi$  is possible and *might*  $\phi$  are logically equivalent.

Again assuming temporarily a probabilistic interpretation—*might*  $\phi$  entails that  $\text{prob}(\phi) > \theta_{\text{might}}$ —this would give us the refined ordering:

$$\theta_{\text{possible}} = \theta_{\text{might}} < \theta_{\text{likely}} < \theta_{\text{must}} < \theta_{\text{certain}}.$$

However, this assumption is problematized by experimental evidence presented below (§6.5). The results presented there suggest that the correct ordering may be

$$\theta_{\text{possible}} < \theta_{\text{might}} < \theta_{\text{likely}} < \theta_{\text{must}} < \theta_{\text{certain}}.$$

Regardless of whether a scalar semantics turns out to provide the right scaffolding for *must* and *might*, this constraint set can serve as a succinct summary of the probability-related entailment relations that we must account for.

A final point that will be important in the following is the relationship between Comparative Possibility and *might*. Suppose that, in a given year, Yale has a much better basketball team than Harvard does. Nevertheless, the odds against either team winning the NCAA March Madness tournament are astronomical. Against this background, it seems reasonable to assert (6.21a) while also denying (6.21b) (Lassiter 2015b: §4).

(6.21) a. Yale is more likely to win the NCAA basketball tournament than Harvard is.

b. Yale might win the NCAA basketball tournament.

The use of *might* in (6.21b) not only expresses that a Yale victory is *not impossible*, but also conveys wrongly that it is *a serious possibility*.

In agreement with Kratzer (1991); Willer (2013), then, I think that we do **not** want to make the traditional assumption that *might* simply conveys compatibility with one's epistemic state. When thinking in a probabilistic frame, the latter assumption would correspond to the stipulation that

$$\theta_{\text{possible}} = \theta_{\text{might}} = 0.$$

Instead, we want to allow that  $\theta_{\text{might}}$  may be greater than 0. In fact, as I will suggest later (§6.3.3), the degree of commitment associated with the use of *possible* and *might* exhibits a certain amount of context-dependence.

### 6.3 Theoretical accounts of *must* and *might*

In addition to the division of theories into “strong” and “weak” in terms of the relative strength of *must* and *certain*, we can also classify theories in terms of the semantic scaffolding that they employ: “Quantificational”, “Scalar”, or “Mixed”. These classifications are formally orthogonal to each other, and at least five of the six possibilities have been advocated explicitly in existing literature.

	Quantificational	Scalar	Mixed
“Strong”	von Fintel & Gillies 2008	Yalcin 2005	—
“Weak”	Kratzer 1991	Swanson 2006; Lassiter 2011a, 2016	Swanson 2015

**Table 6.1**

A 2x3 typology of theories of *must*, with examples of papers advocating some of them.

The dialectic, as I will present it, is this: “Strong” theories of any stripe are empirically inadequate for reasons discussed above, and the best-known “Weak” quantificational theory—Kratzer’s—fails to enforce the entailments between *must* and *likely* that we discussed in §6.2. Absent some improved version of the “Weak” quantificational semantics (with explicit connections to a plausible theory of the gradable epistemic modals), we seem to be drawn either to a “Weak” scalar semantics, or some kind of mixture of the weak scalar and quantificational semantics.

#### 6.3.1 Strong quantificational theories

The traditional analysis of *might* and *must* from modal logic treats them as, respectively, existential and universal quantifiers over a set of worlds  $\mathcal{E}_w$  that are “epistemically accessible” from the evaluation world  $w$ .

$$(6.22) \quad \begin{aligned} \text{a. } \llbracket \text{must} \rrbracket^{\mathcal{M},w} &= \lambda p_{\langle s,t \rangle}. \forall w' \in \mathcal{E}_w : p(w') = 1 \\ \text{b. } \llbracket \text{might} \rrbracket^{\mathcal{M},w} &= \lambda p_{\langle s,t \rangle}. \exists w' \in \mathcal{E}_w : p(w') = 1 \end{aligned}$$

This is a “strong quantificational” semantics. The theory of von Fintel & Gillies (2008) is an elaboration of the traditional account, containing a number of additional components designed to account for the evidential requirements of *must* and *might*.

As I argued earlier in chapters 4-5, the epistemic adjectives *likely*, *probable*, *certain*, *sure*, and *possible* have meanings which are either explicitly probabilistic or are defined in terms of some function of probabilities. Given the inferential connections between the epistemic auxiliaries and adjectives, a theory of the epistemic auxiliaries must have some way of encoding an explicit connection between epistemic accessibility and probability. An obvious way to do this, from the perspective of a strong quantificational semantics, is to require that the set of epistemically accessible worlds contains all of the probability mass— $\text{prob}(\mathcal{E}_w) = 1$ , or (qualitatively)  $\mathcal{E}_w \approx_{\text{likely}} W$ . (Formally, this could be done in a variety of ways whose details are not crucial here. See, for example, Yalcin 2010; Holliday & Icard 2013.) Taken together with (6.22), we immediately derive the implication



that *must*  $\phi$  entails  $\text{prob}(\phi) = 1$ , since *must*  $\phi$  is true if and only if  $\phi$  is true throughout (a superset of)  $\mathcal{E}_w$ .

A theory of this form will therefore validate the intuitive connections between *must* and *likely* given in §6.2. Since *must*  $\phi$  entails that  $\phi$  has probability 1, clearly it also entails both  $\phi$  is *likely* and  $\phi$  is *more likely than not*. It also follows correctly that  $\phi$  is *likely* (*more likely than not*) entails *might*  $\phi$ , since the latter is true whenever  $\phi$  overlaps  $\mathcal{E}_w$ . (Proof: If *might*  $\phi$  were false,  $\phi$  would lie entirely outside of  $\mathcal{E}$ . Since  $\text{prob}(\mathcal{E}_w) = 1$ , we also know that  $\text{prob}(W - \mathcal{E}_w) = 0$ .  $\text{prob}(\phi)$  must therefore be zero; but then  $\phi$  is *likely* and  $\phi$  is *more likely than not* are both false. By contraposition, *might*  $\phi$  is true whenever either  $\phi$  is *likely* and  $\phi$  is *more likely than not* is according to the theory under consideration.)

However, a strong quantificational theory also predicts wrongly that *must*  $\phi$  should entail  $\phi$  is *certain*, regardless of what the threshold  $\theta_{\text{certain}}$  is. (This conclusion does not depend on the choice among the three alternative analyses of *certain* that we considered in §5.1.) In addition, this style of theory predicts that  $\phi$  is *more likely than*  $\psi$  entails *might*  $\phi$ . While this inference may seem reasonable in the abstract, we saw an intuitively reasonable counter-example above, where *Yale is more likely to win than Harvard is* did not intuitively license the inference that *Yale might win* (example (6.21)). This counter-example provides further support to the contention that the simple, elegant account suggested by the  $\Box$  and  $\Diamond$  of classical modal logic is not the correct theory of English *must* and *might*.

Since there is not (to my knowledge) any alternative proposal on the market which would encode the inferential connections between the epistemic adjectives and the strong quantificational interpretation of the auxiliaries, we have good reason to ask whether an alternative semantics for the auxiliaries is able to do better.

### 6.3.2 Weak quantificational theories

The best-known weak quantificational theory is Kratzer 1991, which was discussed in chapter 3. I argued there that, as a theory of graded and comparative epistemic modals, this theory has a number of fatal flaws involving the logic of epistemic comparatives and the interpretation of quantitative modal language. However, this conclusion does not rule out the possibility that Kratzer's theory could be used to interpret the epistemic auxiliaries.

Recall from chapter 3 (§3.4.5) the basic form of Kratzer's semantics for the auxiliaries. (For simplicity, we continue to make the limit assumption.) *Must/might*  $\phi$  indicates that all/some of the worlds that are “best” according to the ordering  $\succsim_{\mathbf{g}(w)}$  are worlds in which  $\phi$  is true.

$$(6.23) \quad \mathbf{BEST}(\mathbf{f}, \mathbf{g}, w) =_{df} \{v \mid v \in \cap \mathbf{f}(w) \wedge \neg \exists v' \in \cap \mathbf{f}(w) : v' \succ_{\mathbf{g}(w)} v\}$$

$$(6.24) \quad \llbracket \text{must } \phi \rrbracket_{\mathbf{f}, \mathbf{g}}^{\mathcal{M}, w} = 1 \text{ iff } \forall u : u \in \mathbf{BEST}(\mathbf{f}, \mathbf{g}, w) \rightarrow u \in \phi.$$

$$(6.25) \quad \llbracket \text{might } \phi \rrbracket_{\mathbf{f}, \mathbf{g}}^{\mathcal{M}, w} = 1 \text{ iff } \exists u : u \in \mathbf{BEST}(\mathbf{f}, \mathbf{g}, w) \wedge u \in \phi.$$

Can we continue to make use of this proposal to account for *must* and *might*, while also using some form of probabilistic semantics, as sketched in chapters 4-5, to account for the epistemic adjectives? The basic problem is to account for the logical relations between epistemic adjectives and auxiliaries—most obviously, the fact that *must*  $\phi$  cannot be true when  $\phi$  is *less* likely than  $\neg\phi$ ,

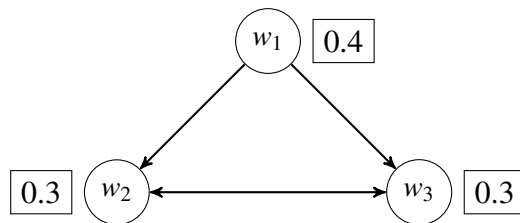
or exactly as likely, or only a tiny bit more likely. As Lassiter (2015b: §3) shows in detail, it is difficult to achieve this goal while maintaining Kratzer’s (1991) semantics for the auxiliaries.

In its most basic form, the hybrid theory under consideration fails because it would allow *must*  $\phi$  to be true when  $\phi$  has any probability at all—even zero. As stated, nothing in the semantics would constrain the proposition-ordering  $\succsim_{\mathbf{g}(w)}^s$  to have any particular probability, since there is no formal connection between this ordering and the likelihood scale  $\mathcal{S}_{\text{likely}}$ .

An obvious strengthening would be to require that  $\succsim_{\text{likely}}$  and  $\succsim_{\mathbf{g}(w)}$  agree on the ordering of singleton propositions:

$$(6.26) \quad \text{If } w_1 \succsim_{\mathbf{g}(w)} w_2, \text{ then } \{w_1\} \succsim_{\text{likely}} \{w_2\}.$$

However, examination of even very small models reveals that this constraint will not have the desired effect. For example, consider a model in which  $W = \{w_1, w_2, w_3\}$ , where  $\mathbf{f}$  and  $\mathbf{g}$  are such that  $\succsim_{\mathbf{g}(w_2)}$  has the structure in Figure 6.1.



**Figure 6.1**

Kratzerian world-ordering for the model described in the main text. Boxes next to nodes describe a probability distribution on singletons respecting the constraint in (6.26).

Here,  $w_1$  is ranked above both  $w_2$  and  $w_3$ , which are tied. If we set  $\text{prob}(\{w_1\}) = 0.4$ ,  $\text{prob}(\{w_2\}) = 0.3$ , and  $\text{prob}(\{w_3\}) = 0.3$ , the constraint in (6.26) is respected, and we predict that the following will both be true:

$$(6.27) \quad \llbracket \text{must } \{w_1\} \rrbracket_{\mathbf{f}, \mathbf{g}}^{\mathcal{M}, w_2} = 1.$$

$$(6.28) \quad \llbracket \{w_2, w_3\} \text{ is more likely than } \{w_1\} \rrbracket_{\mathbf{f}, \mathbf{g}}^{\mathcal{M}, w_2} = 1.$$

Since this is absurd, (6.26) cannot be the right bridge between probability and the orderings used in Kratzer’s semantics.

In Lassiter (2015b: §3) I explore this and several other apparently plausible ways to connect Kratzer’s theory of *must/might* with a probabilistic semantics for *likely*. Possibilities considered there include several ways to connect scalar likelihood with the lifted Comparative Possibility relation of Kratzer (1991), the revision proposed in Kratzer 2012, and the variant suggested by Holliday & Icard (2013). All of the proposals under consideration admit models in which *must*  $\phi$  is true even though  $\phi$  is just barely more probable than  $\neg\phi$  (e.g.,  $\text{prob}(\phi) = .500000001$ ), or even less probable in some cases. This is clearly not adequate. However, unless some better bridging principle can be articulated and shown to make reasonable predictions, Kratzer’s semantics is in trouble.

It is, of course, always possible that future work will manage to articulate some unforeseen principle which does a better job of encoding the relationship between comparative likelihood and Kratzer’s *must*. However, the prospects do not look especially good at present. The only reconciliation that I can envision is the brute-force method of adding a meaning postulate to the effect that the set of “best” worlds must have high probability. This is essentially the same as giving *must* a conjunctive meaning—true in all of the “best” worlds, and also high probability—and it is the mixed proposal of Swanson (2015) that we will consider in §6.7 once we have a scalar alternative on the table.

### 6.3.3 Scalar theories

Swanson (2006), and following him Lassiter (2011a), proposed that *must* and *might* are scalar expressions with interpretations along the following lines.

- (6.29) a.  $\llbracket \text{must} \rrbracket^{\mathcal{M},w} = \lambda p_{\langle s,t \rangle} \cdot \text{prob}(p) \geq \theta_{\text{must}}$   
b.  $\llbracket \text{might} \rrbracket^{\mathcal{M},w} = \lambda p_{\langle s,t \rangle} \cdot \text{prob}(p) > \theta_{\text{might}}$

If we assume as usual that *must* and *might* are duals, then we can add an additional constraint to render *must* equivalent to *not might not*.

$$\theta_{\text{might}} = 1 - \theta_{\text{must}}.$$

This skeletal proposal could be spelled out in either “weak” or “strong” versions. The “strong” version results when  $\theta_{\text{must}}$  is required to be 1, and  $\theta_{\text{might}}$  to be 0 (e.g., Yalcin 2005). Any semantics which allows greater flexibility could be considered a “weak” semantics. In light of the discussion in §6.2, I will assume that  $\theta_{\text{must}}$  is not required to equal 1, and  $\theta_{\text{might}}$  is not required to equal 0. I will also assume that *must* and *might* are duals, though this assumption is not critical.

Recall the discussion in chapter 4 (§4.2.7) of lexical ordering constraints like  $\theta_{\text{warm}} < \theta_{\text{hot}}$ , which constrains the interpretation of the co-scalar adjective pair *warm/hot*. As I noted there, some constraint along these lines seems to be necessary in order to explain how two expressions can live on the same scale and yet have different meanings. Along similar lines, we could encode the semantic relations discussed in §6.2 in a simple way by adopting this ordering directly:  $\theta_{\text{possible}} = \theta_{\text{might}} < \theta_{\text{likely}} < \theta_{\text{must}} < \theta_{\text{certain}}$ . These lexical relationships would then constrain the behavior of the pragmatic inference machinery which assigns contextual interpretations to particular items (for example, using the theory of Lassiter & Goodman 2013, 2015a; Goodman & Lassiter 2015).

This proposal is not quite sufficient, though, because it does not encode the requirement of indirect evidence noted by von Fintel & Gillies (2010): *must*  $\phi$  is interpreted essentially as “It is almost certain that  $\phi$ ”. However, Lassiter (2016) suggests a way of capturing this meaning component within an enriched probabilistic semantics. Lassiter adapts von Fintel & Gillies’s (2010) proposal, according to which *must* presupposes that neither the propositional argument, nor its negation, is entailed by a relevant set of propositions which are known by direct experience. The probabilistic variant is inspired by graphical models (Pearl 1988, 2000; Koller & Friedman 2009), which make crucial use of a partition of the epistemic space by means of random variables (essentially, question meanings). This additional structure provides the opportunity to mark certain

questions as ones whose answer has been observed directly. The proposal of von Fintel & Gillies (2010) can then be translated into a requirement that *must*  $\phi$  is inappropriate if  $\phi$  is an answer to a question whose true answer has been directly observed. This can then be added as an additional presupposition to the “weak” version of the scalar semantics in (6.29). See Lassiter 2016 for details.

This proposal is surely not the only way that the evidential meaning component (presupposition?) of *must* can be captured. However, it serves to illustrate that the existence of this component does not constitute an argument against a scalar semantics in the style of (6.29). (See §6.7 below for discussion of another way to capture this meaning component.)

#### 6.4 Context-sensitivity of *might* and *must*

Given that no one has a complete story about the context-sensitive interpretation of scalar expressions generally, a scalar theory of *must* and *might* should not be expected to provide a complete account of how  $\theta_{must}$  and  $\theta_{might}$  are contextually valued. (This is comparable to the fact that quantificational theories generally say little or nothing about how these items’ domains of quantification are determined, and/or about the content of the modal base and ordering source.) However, it would obviously make the story more convincing if some general constraints could be identified, especially if these constraints were related in a clear way to ones known to be operative in the interpretation of scalar expressions more generally.

At first glance, it may seem that the interpretation of *must* is not hugely context-sensitive in the way that the interpretations of, say, *warm* and *hot* are. For example, there is surely no context in which *must*  $\phi$  can be true when  $\phi$  has probability .3. Yet, as we discussed in §4.2.7 above, something can be “hot” when its temperature is very low, as long as the items that it is explicitly or implicitly compared to have even lower temperature.

This difference is related to at least two separate sets of theoretical considerations. First, *must* is required to have a stronger meaning than *likely*—probably as part of a set of conventional (lexical) ordering constraints. This constraint feeds into pragmatic reasoning about its interpretation so as to prevent *must* from ever having a very weak interpretation. Second, the epistemic scale is closed on both ends, and closed-scale scalar expressions like *full* and *visible* tend not to have highly context-sensitive meanings (Kennedy 2007). However, the pragmatic story does predict that there should be some context-sensitivity in the interpretation of *must* and *might*.

One reason to suppose that *must* and *might* really do have context-sensitive meanings is that their interpretations are goal-sensitive in the same way as absolute adjectives. Consider, for example, the following scenarios involving two speakers, Reasonable and Annoying.

- (6.30) a. **Annoying:** Want a top-up?  
       b. **Reasonable:** No thanks, my glass is full.  
       c. **Annoying:** Well, no, your glass is not quite full—I could still pour a tiny bit more water in it, see?<sup>1</sup>

- (6.31) a. **Annoying:** Let’s get this sofa inside.

<sup>1</sup> “... OK, technically it’s still not full now—it would be possible to add several more H<sub>2</sub>O molecules on the top—but it’s closer ...”

- b. **Reasonable:** Wait, the door isn't open.
- c. **Annoying:** Yes, it is. But it isn't open far enough to get the sofa through! There, now it's open.

Reasonable's statements in these dialogues make use of an absolute adjective with the intention of indicating sufficient proximity to, or lack of distance from, a scalar endpoint. The obvious intention is that "sufficient" be resolved relative to a practical goal that is highly salient in the conversation. (Compare, in particular, Fara 2000.) Annoying responds by pointing out that there is a way of interpreting the adjective that would make Reasonable's statement false, if the obvious conversational goal is ignored.

Conversely, when Reasonable has made a relatively weak statement involving a scalar expression with a context-sensitive meaning, Annoying might choose to point out that there is a way of interpreting Reasonable's statement that makes it trivially true, and thus uninformative. Again, this requires ignoring the obvious conversational goal:

- (6.32)
- a. **Annoying:** Hand me that towel—I need to dry off.
  - b. **Reasonable:** The towel is wet.
  - c. **Annoying:** Of course it is—all macroscopic objects on Earth have some amount of water on them. The question is, is it dry enough that I can towel off?

Both kinds of interaction have counterparts in the interpretation of the epistemic auxiliaries. An example of the first involving *must* and *might* is:

- (6.33)
- a. **Reasonable:** That must be Bill—I recognize his funny hat.
  - b. **Annoying:** Well, no, it might be someone else who just happens to have a hat identical to Bill's (however improbable that may be). Or, it might be someone who isn't wearing a hat at all, and you and I are having identical hallucinations. Or ...

Roughly, what Reasonable is trying to convey here is "Given the evidence available, our goals, and the particular stakes of the conversation at hand, it is reasonable to be highly confident in the conclusion 'That is Bill'". That is, Reasonable intended "The probability that that is Bill is greater than  $\theta_{must}$ ", where  $\theta_{must}$  is high but lower than 1. Annoying's response is annoying because it ignores the features of the conversation that render it reasonable to intend  $\theta_{must} < 1$ —the presumably low stakes, among other factors (such as the lack of readily available means for verifying the individual's identity).

The flipside of this interaction—a case in which a reasonable and informative instance of *might* is judged trivial by an annoying interlocutor—will likely be familiar to some readers.

- (6.34)
- a. **Annoying:** We need to start making party plans. Who's going to be there?
  - b. **Reasonable:** Bill might come.
  - c. **Annoying:** Well, sure, anyone who is alive *might* come—Barack Obama, or Arnold Schwarzenegger, or even Bill.

If Reasonable can contain her annoyance long enough to formulate a polite response, she might respond along the following lines: her intention was obviously **not** to communicate that there is no

conclusive evidence that Bill will not come to the party. Rather, it was to communicate that Bill's attendance is probable enough that they ought take it under serious consideration when making party plans. That is, the meaning is not "Bill will attend with probability greater than 0"; it is "Bill will attend with probability greater than  $\theta_{\text{might}}$ ", where  $\theta_{\text{might}}$  is some value significantly greater than 0. What counts as "significant" will, of course, depend in complex ways on factors such as the importance of being highly confident before acting.

Annoying's brand of pragmatic fail in these interactions has been analyzed in two very different ways in the literature. According to Unger (1975); Lasersohn (1999); Kennedy (2007); Klecha (2014), Annoying is technically correct from the get-go: absolute scalar expressions have perfectly strict maximal standards, so that water glasses in the real world are never actually full, roads are never flat, lines never straight, people never satisfied, and so forth. Nevertheless, in many contexts real-world objects can be close enough to the relevant Platonic ideals that we can communicate, approximately, about real-world objects using sentences about flat roads and full glasses—even though these sentences are guaranteed to be false.

As Unger (1971, 1975) points out, this analysis of absolute adjectives (notably *certain*) leads straight to global skepticism about the possibility of certainty and knowledge. After all, mere mortals such as us can almost never be *absolutely certain* about non-logical truths, in the sense that it would be impossible for an annoying person to raise some bizarre objection that could in principle be cause for doubt. If the true meaning of *certain* is rigidly fixed to the maximum possible degree of certainty, though, it seems to follow that almost all certainty ascriptions are literally false. It is a short step from here to the conclusion (which Unger endorses) that all non-trivial knowledge ascriptions are false. Klecha (2014) points out along similar lines that the arguments for treating minimum and maximum adjectives as having context-insensitive, endpoint-oriented meanings also extend directly to counterfactuals. This line of reasoning seems to lead inexorably to Hájek's (2007) conclusion, which Klecha endorses: "while we use them nonchalantly in daily conversation, and while they are staples of numerous philosophical analyses, most counterfactuals are false".

While Unger and Klecha bite their respective bullets with relish, I find both of them rather unappetizing. Is there an alternative? There is: Lewis (1979) gives a contextualist theory of absolute adjectives that is meant to apply to knowledge ascriptions as well (cf. Lewis 1996). On this approach, the meanings of these expressions are somewhat context-sensitive, and Reasonable's statements are literally true relative to her intended resolution of the context-sensitive meanings of the key expressions. (See Lassiter & Goodman 2013 for a formal quantitative theory of absolute adjectives that is broadly in the spirit of Lewis' analysis. This style of analysis also extends to the counterfactual case: see von Fintel 2001; Gillies 2007. Ichikawa (2011) draws out the analogy between counterfactuals and Lewis' theory of *knows* explicitly.) On Lewis' analysis, the pragmatic oddity of the interactions discussed above is due to the fact that the two parties are not really communicating: what the latter says is technically correct only under a *different* interpretation of the key expressions than the one that Reasonable intends. Annoying's response is annoying because it is uncooperative, in the sense of Grice (1975).<sup>2</sup>

<sup>2</sup> Klecha (2014: §4) offers several positive arguments against the general line pursued here, and in favor of an account in terms of pragmatic slack. The main arguments involve (i) that speakers sometimes retract their statements when presented with an Annoying response, and (ii) the related fact that the kind of context-sensitivity in question behaves

I find Lewis' strategy much more plausible on general methodological grounds. It relieves us of the need to assume that much of the ordinary discourse of cooperative and well-informed speakers is false, or to explain what it would mean to have a community of speakers who have coordinated on a communicative system in which they exchange mostly true information in an effective manner, but most of what they say is false. (For instance, such a theory entails massive violation of Davidson's (1973) Principle of Charity in ordinary language learning.) Lewis' approach also relieves us of the burden of defending a series of highly counter-intuitive claims: that knowledge and certainty are inaccessible for mere mortals, that all ordinary counterfactual sentences are trivially false, and that no real-world object is or ever has been flat, straight, or full.

In addition to these philosophical considerations, there is a direct empirical argument involving epistemic modals against the pragmatic slack theory as applied in this domain. Klecha treats non-endpoint-oriented uses of *might*, *possible*, *must*, *have to*, and *certain* uniformly in terms of pragmatic slack. This theory faces difficulty in accounting for two sets of empirical challenges that have been discussed in this chapter. First, there is the relative weakness of *must* vis-a-vis *certain* (Lassiter 2016 and §6.1 above). If *must* and *certain* were both oriented toward maximum confidence, as Klecha's analysis implies, then it is difficult to see how they could come apart in this way. Second, in the next section I will demonstrate experimentally that *might* is stronger than *possible*. This divergence is even more difficult to explain for a theory in which the only relevant kind of context-sensitivity is pragmatic slack. *Might* and *possible*, like *must* and *certain*, should pattern together pragmatically—absent some desperate maneuver like adding item-by-item stipulations about the amount of slack permitted.<sup>3</sup> I do not see any obvious alternative short of abandoning the entire line of analysis, at least as far as the epistemic modals are concerned.

I will thus treat the context-sensitivity of epistemic modals—and, where relevant, absolute adjectives—as truth-conditional in nature. The analogy between the examples with absolute adjectives and those with epistemic auxiliaries suggests they invoke related forms of context-sensitivity. This could take many forms. (For example, see Willer 2013 for a well-developed story about epistemic modals in the context of a dynamic, quantificational semantics.) However, given other observations that led us above to select a probabilistic interpretation, it is natural to interpret this context-sensitivity as a reflex of the simple scalar semantics for *must* and *might* given in (6.29)

---

differently from anaphora in various respects. I do not find these arguments fully convincing. First, retroactive revision of this type is one of the main phenomena treated in the large literature on contextualism/relativism/expressivism about epistemic modals, deontic modals, predicates of personal taste, future contingent sentences, etc., which Klecha does not engage with. (See, for example, von Stechow & Gillies 2008; Yalcin 2011; MacFarlane 2014.) There are numerous analyses available in this literature that could be employed to explain the phenomena that Klecha is discussing without invoking "loose talk". Second, not all semantic context-dependence is anaphoric in nature. So, we cannot conclude that if some context-dependent phenomenon does not behave like anaphora, it is not truth-conditionally relevant.

See also Lewis 2016 for a detailed defense of contextualism about counterfactuals in the face of skeptical arguments, which serves well as a response to Klecha's arguments against the semantic relevance of context-shifting in the counterfactual case. Parallel arguments in the case of modals and absolute adjectives should be available.

<sup>3</sup> Actually the situation is somewhat worse than this description suggests: since pragmatic slack is a device for weakening meanings, not strengthening them, it cannot account for the fact that the understood interpretations are stronger than the literal meanings on the theory under consideration. A pragmatic account in terms of relevant granularity (Rudin 2015) fares better in this respect, but still provides no insight into why two expressions with the same literal meaning—*might* and *possible*—would show different pragmatic behavior.

above. On this account, *must* and *might* have context-sensitive meanings for the same reason that many other scalar expressions do. Our linguistic knowledge imposes constraints on admissible valuations of their threshold arguments, and pragmatic inference is required to narrow them down further.

## 6.5 Alternative-sensitivity of *might* and *must*

A second potential source of context-sensitivity involves sensitivity of  $\theta_{must}$  and  $\theta_{might}$  to the distribution of alternative outcomes. (Compare Teigen 1988; Windschitl & Wells 1998; Yalcin 2010; Lassiter 2011a: §4 on related issues involving *likely*.) To see why this issue is important for a scalar theory of the epistemic auxiliaries, consider an objection from lotteries (Kyburg 1961), which was lodged against the proposal under consideration here by Yalcin (2005). Let  $\theta_{might}$  be any value greater than 0—say, .005. Suppose that I have a single ticket in a raffle with, say, 1000 tickets. Then my chance of winning is just .001, which is less than  $\theta_{might}$ . So, “I might win the raffle” comes out false, even though my ticket has the same chance of winning as any other. This may seem intuitively incorrect. Further, the result can be reproduced for any setting of  $\theta_{might} < 1$  by inventing a sufficiently large raffle.

Two responses are in order. First, it is not obvious that the prediction *is* incorrect: perhaps some significant proportion of speakers *would* judge “I might win the raffle” to be false in such scenarios. (As I will show in a moment, this is in fact the case.) Second, even if it is true that many or most speakers would judge the statement true, this would not constitute a conclusive refutation unless  $\theta_{might}$  were fixed in a way that is insensitive to the number of tickets. More precisely, we might suppose that one of the factors that goes into fixing  $\theta_{might}$  is the way that probabilities are distributed over the various alternative outcomes—here, whether the other tickets are held by many people or only a few. My hunch is that distributions that concentrate most of the probability in one or a few answers will favor higher values of  $\theta_{might}$ , while those that spread the probability among many alternative answers, each with lower probability, will favor lower settings of  $\theta_{might}$ . For example, consider the three raffles described in Table 6.2.

Even		Somewhat skewed		Highly skewed	
Individual	Tickets	Individual	Tickets	Individual	Tickets
$x_1$	1	$x_1$	1	$x_1$	1
$x_2$	1	$x_2$	1	$x_2$	1
$x_3$	1	...	...	$x_3$	1
...	...	$x_{100}$	1	$x_4$	994
$x_{998}$	1	$x_{101}$	10	$x_5$	1
$x_{999}$	1	...	...	$x_6$	1
$x_{1000}$	1	$x_{190}$	10	$x_7$	1

**Table 6.2** Three raffles with 1000 tickets each: one relatively even, one somewhat skewed, one highly skewed.



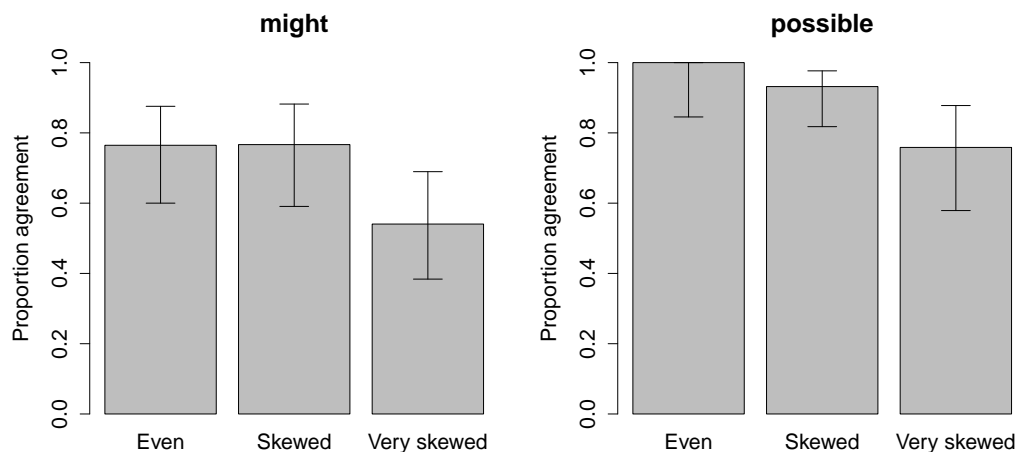
In the **Even Raffle**, one thousand individuals have one ticket each. In the **Somewhat skewed Raffle**, ninety individuals have ten tickets each, and one hundred individuals have one ticket each. In the **Highly skewed Raffle**, six individuals have one ticket each, while one has 994. The key question is whether and how the choice of scenario affects the truth of the sentences in (6.35).

- (6.35) a.  $x_1$  might win the raffle.  
b. It is possible that  $x_1$  will win the raffle.

Crucially, in all three scenarios it is epistemically possible that  $x_1$  will win, and the probability that  $x_1$  will win is held fixed at 1/1000.

If this hunch is correct, we should be able to influence the probability that people will judge (6.35) true by the choice of one of these three raffles. To test this prediction, I conducted a one-shot experiment using Amazon’s Mechanical Turk platform. 200 participants were paid \$0.15 each for answering a single question. Each participant saw a text description of one of the three raffles described in Table 6.2, selected at random, with the individual names “ $x_n$ ” replaced by ordinary English names (“Bill”, “Mary”, “Jane”, etc.). Each was presented with one of the prompts in (6.35), selected at random, with “ $x_1$ ” replaced by “Bill”. Bill was always described first, as an individuals who had only one ticket. They were asked to select “Agree” or “Disagree” in response to the prompt.

I discarded the responses of five participants who reported a native language other than English. The remaining 195 responses are plotted in Figure 6.2. Error bars indicate 95% binomial confidence intervals. The numerical data is presented in Table 6.3. Because of the random selection of conditions, Ns per condition ranged from 21 (*possible*, even) to 44 (*possible*, skewed).



**Figure 6.2**

Even in this small experiment, it is clear that there was an effect of the choice of raffle, as predicted. While responses in the **Even** and **Somewhat skewed** raffles were mostly similar, participants were less likely to endorse each of the two prompts in the **Highly skewed** condition. A nested model comparison using generalized linear models written in R (R Core Team 2015)

Might			
	Even	Somewhat skewed	Highly skewed
Agree	26	23	20
Disagree	8	7	17
Total	34	30	37
Proportion	.76	.77	.54

Possible			
	Even	Somewhat skewed	Highly skewed
Agree	21	41	22
Disagree	0	3	7
Total	21	44	29
Proportion	1.00	.93	.76

**Table 6.3** Numerical results from the experiment described in the main text.

confirmed a significant main effect of raffle choice ( $G^2 = 11.81$ ,  $df = 2$ ,  $p < .005$ ). When there was a single individual holding a large majority of the tickets, participants were more likely to disagree with “It is possible that Bill will win the raffle”, or with “Bill might win the raffle”, than they were to reject the same prompts when the tickets were spread more evenly among those participating. This was true even though the probability of Bill winning remained the same throughout all conditions.

A second, unexpected finding was that, across all three raffles, participants were more likely to endorse the *possible* description than the *might* description. This result suggests that the conditions for applicability of *possible* are less stringent than those of *might*, which is generally set to higher values. A nested model comparison confirms this impression: there was a highly significant main effect of the choice of modal ( $G^2 = 12.72$ ,  $df = 2$ ,  $p < .001$ ). There was not, however, strong evidence of an interaction between raffle condition and modal ( $G^2 = 3.34$ ,  $df = 2$ ,  $p \approx .19$ ). This suggests that the contextual manipulation influenced both modals’ thresholds in a qualitatively similar way, but that the threshold  $\theta_{possible}$  is in general lower than that of  $\theta_{might}$ .

The results of this experiment have a number of interesting implications. First, the results caution against a habit, common in epistemology, of taking intuitions about lottery cases too seriously (e.g., Williamson 2000; Hawthorne 2004). Ordinary language users appear to be less equivocal than philosophers in their endorsement of *Bill might win* when Bill has probability 0.001 of winning.

Second, the results demonstrate that *might* has a context-sensitive meaning, and in particular that it is sensitive to the distribution of probabilities among contextual alternatives. This does not yet show that the same holds of *must*, but this would be the obvious inference on standard assumptions: *must*  $\phi$  is usually thought to be equivalent to *not might not*  $\phi$ . Or, in probabilistic terms,  $\theta_{must} = 1 - \theta_{might}$  relative to any context. If so, manipulations of  $\theta_{might}$  should have a symmetric effect on  $\theta_{must}$ .

Third, the results problematize the usual assumption that epistemic *might* and *possible* are

equivalent, motivating instead the ordering

$$\theta_{possible} < \theta_{might}.$$

This conclusion is consonant with the findings of Turri (2015: Experiment 1), whose participants gave higher Likert-scale ratings to *possibly not* statements than to *might not* statements across a variety of scenarios.<sup>4</sup>

Fourth, on the standard assumption that *must* and *might* are duals, and that *certain* and *possible* are duals, the conclusion that  $\theta_{might} > \theta_{must}$  also implies that  $\theta_{certain} > \theta_{must}$ . This lends further support to the conclusion of §6.1 above, where I argued on the basis of intuition data and corpus examples that *must* is weaker than *certain* once the indirectness implication of *must* is controlled for.

The context-sensitivity that we have observed in the meaning of *must* and *might* is exactly what the scalar theory of their meaning would lead us to expect. It also supports the following refined ordering of epistemic modals that we have discussed in terms of their strength:

$$\theta_{possible} < \theta_{might} < \theta_{likely} < \theta_{must} < \theta_{certain}.$$

## 6.6 Are *must* and *might* gradable?

The scalar semantics in (6.29) was noncommittal about one important detail: whether *must* and *might* are grammatically gradable. In principle, we could generate the same interpretation of bare *must* sentences with either of the lexical entries in (6.36). Abstracting away from the evidential component of the meaning, this would give us something like the following two options for each:

$$(6.36) \quad \begin{array}{ll} \text{a. } \llbracket \text{must} \rrbracket^{\mathcal{M},w} = \lambda p_{\langle s,t \rangle} \cdot \text{prob}(p) \geq \theta_{must} \\ \text{b. } \llbracket \text{must} \rrbracket^{\mathcal{M},w} = \lambda p_{\langle s,t \rangle} \cdot \text{prob}(p) \end{array}$$

$$(6.37) \quad \begin{array}{ll} \text{a. } \llbracket \text{might} \rrbracket^{\mathcal{M},w} = \lambda p_{\langle s,t \rangle} \cdot \text{prob}(p) > \theta_{might} \\ \text{b. } \llbracket \text{might} \rrbracket^{\mathcal{M},w} = \lambda p_{\langle s,t \rangle} \cdot \text{prob}(p) \end{array}$$

In the first case, *must* would have a threshold value which is potentially context-sensitive, but cannot be bound by other operators in the sentence. If the second entry is right, *must* simply denotes a measure function on propositions. It could in principle be gradable, and its meaning should also be vague and context-sensitive in the unmodified form. The latter would have to behave compositionally like gradable adjectives in Kennedy's (2007) theory, with a type-shifting operation or a silent *pos* morpheme that converts the root into a simple predicate of propositions. This meaning would then refer to a contextually-supplied value  $\theta_{tall}/\theta_{must}/\theta_{might}$ .

$$(6.38) \quad \text{a. } \llbracket \text{tall} \rrbracket^{\mathcal{M},w} = \lambda x. \mu_{tall}(x)$$

<sup>4</sup> This held in Turri's experiment only in the condition where participants were provided with information specific to the case at hand. When they were asked to reason from general statistical information about a class, Turri's participants were at ceiling for both *might* and *possible*. As Turri points out, this pattern is readily explained by the well-known tendency to value conclusions derived from case-specific information more highly than those derived by purely statistical reasoning (Lagnado & Sloman 2004).

- b.  $\llbracket pos\ tall \rrbracket^{\mathcal{M},w} = \lambda x. \mu_{tall}(x) \geq \theta_{tall}$
- (6.39) a.  $\llbracket must \rrbracket^{\mathcal{M},w} = \lambda p. prob(p)$   
 b.  $\llbracket pos\ must \rrbracket^{\mathcal{M},w} = \lambda p. prob(p) \geq \theta_{must}$
- (6.40) a.  $\llbracket might \rrbracket^{\mathcal{M},w} = \lambda p. prob(p)$   
 b.  $\llbracket pos\ might \rrbracket^{\mathcal{M},w} = \lambda p. prob(p) > \theta_{might}$

Note that gradable analyses of *must* and *might* give rise to “positive-form” meanings that are identical to the basic entries of the non-gradable analyses. The only difference that we can hope to observe is in whether the expressions are grammatically gradable.

Now, until recently it has been almost universally assumed that *must* and *might* are not grammatically gradable. This is explicitly argued, for example, by Barker (2009); Lassiter (2011a); Klecha (2014), based on intuition data like that reported in (6.41).

- (6.41) a. ?? That must very much be Bill.  
 b. ?? That very much must be Bill.  
 c. ?? That must be Bill more than it must be Harry.
- (6.42) a. ?? That might very much be Bill.  
 b. ?? That very much might be Bill.  
 c. ?? That might be Bill more than it might be Harry.

For Lassiter (2011a), who argues for a scalar interpretation of these items, this indicates that (6.36a) and (6.37a)—the “graded-but-not-gradable” interpretations—are preferable to the measure function treatments in (6.36b) and (6.37b).

Importantly, the oddness of the examples in (6.41) and (6.42) is not a general feature of auxiliary modals: *ought* and *should* are gradable at least in their deontic interpretation, and possibly also in their epistemic interpretation (Lassiter 2011a: §5-6, and §8.13 of this book). (6.41) and (6.42) thus suggests either that the gradable and non-gradable auxiliaries differ in their lexical semantics, or that they are morphosyntactically different in some way that is reflected in their ability to enter into comparative constructions.

Portner & Rubinstein (2015) propose a rather different interpretation of the contrast between *must* and *ought/should*. According to them, *must* is an “extreme” expression, analogous to extreme adjectives like *gorgeous* and *terrible*. *Ought* and *should*, on the other hand, are comparable in many respects to the relative adjectives. The crucial observation of Portner & Rubinstein (2015) is that these items differ along a number of empirical dimensions in a way that mirrors empirical diagnostics for the relative/extreme distinction in adjectives identified in Morzycki 2012.

- (6.43) Unacceptability of ordinary intensifiers
- a. Bill is very  $\left\{ \begin{array}{l} \text{big} \\ ?? \text{enormous} \end{array} \right\}$ .
- b. Bill  $\left\{ \begin{array}{l} \text{should} \\ ?? \text{must} \end{array} \right\}$  very much be home by now. (epistemic)

- c. Bill  $\left\{ \begin{array}{c} \text{should} \\ ?? \text{ must} \end{array} \right\}$  very much leave. (deontic)
- (6.44) Acceptability of *downright*, etc.
- a. Bill is  $\left\{ \begin{array}{c} \text{downright} \\ \text{positively} \end{array} \right\} \left\{ \begin{array}{c} ?? \text{ big} \\ \text{enormous} \end{array} \right\}$ .
- b. Bill  $\left\{ \begin{array}{c} \text{downright} \\ \text{positively} \end{array} \right\} \left\{ \begin{array}{c} ?? \text{ should} \\ \text{must} \end{array} \right\}$  be home by now. (epistemic)
- c. Bill  $\left\{ \begin{array}{c} \text{downright} \\ \text{positively} \end{array} \right\} \left\{ \begin{array}{c} ?? \text{ should} \\ \text{must} \end{array} \right\}$  leave. (deontic)
- (6.45) Oddness of comparative
- a. Bill is  $\left\{ \begin{array}{c} \text{bigger} \\ ?? \text{ more enormous} \end{array} \right\}$  than Sam.
- b. Bill  $\left\{ \begin{array}{c} (?) \text{ should} \\ ?? \text{ must} \end{array} \right\}$  be home more than Mary  $\left\{ \begin{array}{c} \text{should} \\ \text{must} \end{array} \right\}$ . (epistemic)
- c. Bill  $\left\{ \begin{array}{c} \text{should} \\ ?? \text{ must} \end{array} \right\}$  leave more than Mary  $\left\{ \begin{array}{c} \text{should} \\ \text{must} \end{array} \right\}$ . (deontic)
- (6.46) Acceptability of equative
- a. Bill is  $\left\{ \begin{array}{c} \text{as big} \\ \text{as enormous} \end{array} \right\}$  as Sam.
- b. Bill  $\left\{ \begin{array}{c} \text{should} \\ \text{must} \end{array} \right\}$  be home as much as Mary  $\left\{ \begin{array}{c} \text{should} \\ \text{must} \end{array} \right\}$ . (epistemic)
- c. Bill  $\left\{ \begin{array}{c} \text{should} \\ \text{must} \end{array} \right\}$  leave as much as Mary  $\left\{ \begin{array}{c} \text{should} \\ \text{must} \end{array} \right\}$ . (deontic)

These similarities suggest, tantalizingly, that the limited gradability of *must* may be attributable to whatever is responsible for the limited gradability of extreme adjectives. If so, it would appear that *must* picks out a measure function as in (6.36b) rather than a property of propositions as in (6.36a).

However, I am hesitant to place too much weight on these observations in choosing a theory of *must*. At present, the restricted gradability of extreme adjectives is not very well-understood, either empirically or theoretically. (For example, the intuitive contrast between equatives and comparatives with extreme adjectives is baffling, and available theories such as Morzycki 2012 do little to explain it.) In addition, the single diagnostic that seems to pick out extreme adjectives uniquely—the acceptability of modifiers like *positively* and *downright*—actually diagnoses something else: the correlated, but distinct, property of surprisingness. Being an extreme adjective is neither necessary nor sufficient for the acceptability of these modifiers, as (6.47) shows.

- (6.47) This  $\left\{ \begin{array}{c} \text{shed} \\ ?? \text{ skyscraper} \end{array} \right\}$  is  $\left\{ \begin{array}{c} \text{positively} \\ \text{downright} \end{array} \right\}$  tall!

We generally expect sheds to be fairly short. In contrast, once you know that something is a skyscraper, the fact that it is (merely) tall will raise few eyebrows. The felicity of modification by *positively/downright* in (6.47) seems to be sensitive to these facts. As Morzycki (2015: §3.7.4) hints, the driving force behind the distribution of *downright* and *positively* may be the surprisingness of an object having a certain property, rather than its status as an extreme or non-extreme adjective. This hypothesis would also account for the fact that extreme adjectives generally are modifiable by *positively/downright*: the property of being enormous/scrumptious/gorgeous is highly correlated with the property of having a surprising degree of size/tastiness/beauty.

Corroborating this analysis of the constructed example in (6.47), here are a few naturalistic examples of *positively* and *downright* with relative, non-extreme adjectives.

- (6.48) a. Some cheesecakes are actually very light, and some are not just substantial, they're downright heavy and meant to be just that.  
 b. [T]his is one of the best designs as of late. However, this car is positively big, as compared to most two seaters like Ferrari, Lamborghini or Aston Martin.

The force of (6.48a) is that we *expect* cheesecakes to be light, and it is surprising when they are heavy. This comparison is even more explicit in (6.48b), where the car in question (a new model of Maserati) is explicitly compared to two other brands of luxury cars that are typically very small. Given that the expectation is that a car in this class should be petite, simply being “big” relative to the broader class of cars is surprising enough to warrant the modifier *positively*.

Along similar lines, it could be that the acceptability of *downright/positively must*  $\phi$  is restricted to contexts where the prior expectation was that there would be more uncertainty about the truth of  $\phi$ . This would explain away at least one of the empirical similarities between *must* and the extreme adjectives. It may be that the other similarities can be explained along similar lines. In the end, Portner & Rubinstein's (2015) observations are intriguing, but more work needs to be done before we can conclude confidently that *must* is gradable.

In any case, I do not know of any reason to think that *might* is ever gradable. As far as its basic semantics is concerned, it seems to lack a degree argument.

## 6.7 A mixed theory, and the status of “epistemic ought”

Swanson (2015) proposes an interesting blend of Kratzer's (1991) premise semantics and a scalar semantics along the lines sketched above. The core of the proposal for *must* and *might* is, in effect, that *might* and *must* express the conjunction of the scalar semantics just described, and the premise-semantic interpretation that we discussed and rejected in §6.3.2. In other words (simplifying somewhat), *must*  $\phi$  and *might*  $\phi$  hold if and only if  $\text{prob}(\phi)$  exceeds the relevant threshold, and  $\phi$  is true in some/all worlds that are maximal with respect to the ordering induced by the modal base and ordering source. As long as we maintain the constraints  $\theta_{\text{certain}} > \theta_{\text{must}} > \theta_{\text{likely}}$ , and allow  $\theta_{\text{might}}$  to take on a value greater than 0, this proposal will satisfy the main desiderata listed above for a theory of the auxiliaries, for the same reasons that the scalar theory does.

One of the attractions of this mixed theory is that it is able to combine probabilistic constraints with a plausible formalization of the evidential component of *must*. Specifically, suppose that we interpret the ordering source  $\mathbf{g}$  as a function supplying a world  $w$  with a set of premises, which

together determine a set of arguments that are relevant at  $w$ . As Swanson suggests, if we require that  $\phi$  be both highly probable and entailed by the relevant premises (relative to the set of epistemically accessible worlds), this may well be enough to explain the source of *must*'s evidential meaning. The only addition that seems necessary is to require that  $\phi$  itself not be among the premises (or, perhaps, that no individual premise entails  $\phi$ ).

However, Swanson's (2015) proposal has the counter-intuitive consequence that *must*  $\phi$  and *might*  $\neg\phi$  can both be false—that is, that *must* and *might* are not duals. Specifically, Swanson interprets *might*  $\phi$  as the conjunction of Kratzer's (1991) premise semantics (" $\phi$  is true in at least one of the best worlds") with the scalar account described above (" $prob(\phi) > \theta_{might}$ "). Suppose that  $prob(\phi) > \theta_{might}$  ( $=1 - \theta_{must}$ ), but  $\phi$  is not true in any of the best worlds. Then *might*  $\phi$  is false. But while  $\neg\phi$  is true in all of the best worlds, the fact that  $prob(\phi) > 1 - \theta_{must}$  entails  $prob(\neg\phi) < \theta_{must}$ . So, *must*  $\neg\phi$  is also false. This seems problematic, given that examples like (6.49) look a lot like contradictions.

(6.49) It doesn't have to be raining, but it's not true that it might be.

We could try to improve the intuitive situation by taking *must* to be basic and defining *might* as its dual. However, this will not work since it would give *might* a disjunctive meaning "Either probable enough, or true in some of the best worlds". A theory along these lines would then fail to enforce the condition that we argued for above, that *might* is stronger than *possible*: on any plausible interpretation of *possible*, this *might* would be either weaker or logically independent.

Bracketing this objection for the moment, one argument in favor of Swanson's mixed theory is that it elucidates the connection between *must* and *might* and the "epistemic" interpretation of *ought* and *should*. Many authors have suggested that sentences like (6.50a) have an epistemic interpretation that can be paraphrased in terms of likelihood (e.g., Horn 1972; Finlay 2010; Lassiter 2011a: §3). If so, then (6.50a) ought to be roughly equivalent to (6.50b).

- (6.50) a. Bill should/ought to be home by now.  
b. It is likely that Bill is home by now.

However, this paraphrase is not quite right: there is a marked difference in acceptability between the continuations in (6.51) (Copley 2004).

- (6.51) a. Bill should/ought to be home by now, but he isn't.  
b. # It is likely that Bill is home by now, but he isn't.

As Swanson points out, this observation is striking because the other epistemic modals—even the weakest ones, *might* and *possible*—are not acceptable in this frame.

- (6.52) a. # Bill must/has to be home by now, but he isn't.  
b. # Bill might be home by now, but he isn't.  
c. # It is possible that Bill is home by now, but he isn't.

That is, it is not correct to maintain a simple linear ordering in terms of strength—*must* stronger than *ought/should* stronger than *might*. On some relevant dimension, *might* is in fact stronger than *ought* and *should*.

Swanson suggests that *ought* and *should* differ qualitatively from the other epistemic items: they are pure inferential modals, whose semantics is essentially what Kratzer (1991) proposed for *must*. They are strictly weaker than *must* because they place no restrictions at all on the probability of their propositional arguments, but only require that they follow from some contextually relevant premises. As a result, it makes sense to assert *ought/should*  $\phi$  while denying  $\phi$ : the net effect is that the true state of affairs is not what the relevant argument(s) would lead us to expect. Since epistemic *ought* is epistemic *must* without the probabilistic meaning component, the latter is equivalent to *ought* + *almost certain*.

Yalcin (2015) discusses this interpretation of *ought* and *should* in more detail, considering their relationship to epistemic as well as deontic items. His formal account is broadly similar to Swanson's (unsurprisingly, since both build on the Lewis/Kratzer tradition surveyed in chapter 3). However, Yalcin suggests that we have simply miscategorized this interpretation of *should* and *ought*: it is not epistemic, but "pseudo-epistemic" or (as he eventually concludes) a "normality modal". Since normality is merely correlated with likelihood, it is not too surprising that maximally normal propositions can nevertheless fail to be true. This suggests a difference in orientation from Swanson's, but it is not immediately obvious that the difference between them is more than terminological.

While Yalcin and Swanson are noncommittal about the relationship between *should/ought* and gradable expressions of normality, it seems reasonable to ask at this point how the latter behave semantically. In particular, consider an inference pattern analogous to the "Disjunctive Inference" considered in chapters 3-4. Is it valid?

- (6.53) Given a well-shuffled pack of cards,
- a. ... it is as normal for the top card to be a spade as it is for it to be a heart.
  - b. ... it is as normal for the top card to be a spade as it is for it to be a diamond.
  - c. Therefore, it is as normal for the top card to be a spade as it is for it to be a red card.

It seems to me that the answer is "no": the premises are clearly true, but the conclusion false. If you have doubts on this point, consider a modified version, where we list all of the red cards individually, comparing the normalcy of their being the top card to the normalcy of the event of a certain individual card—the 8 of spades—being on top.

- (6.54) Given a well-shuffled pack of cards,
- a. ... it is as normal for the top card to be the 8 of spades as it is for it to be the 2 of hearts.
  - b. ... it is as normal for the top card to be the 8 of spades as it is for it to be the 3 of hearts.
  - c. ...
  - d. ... it is as normal for the top card to be the 8 of spades as it is for it to be the ace of hearts.
  - e. ... it is as normal for the top card to be the 8 of spades as it is for it to be the 2 of diamonds.
  - f. ... it is as normal for the top card to be the 8 of spades as it is for it to be the 3 of diamonds.
  - g. ...



- h. ... it is as normal for the top card to be the 8 of spades as it is for it to be the ace of diamonds.
- i. Therefore, it is as normal for the top card to be the 8 of spades as it is for it to be a red card.

The conclusion of (6.54) seems inappropriate: surely it is more normal for a well-shuffled deck to have a red card on top than it is for it to have the 8 of spades on top.

This indicates that degrees of normality do not validate the Disjunctive Inference:

- (6.55)
- a.  $\phi$  is as normal as  $\psi$ .
  - b.  $\phi$  is as normal as  $\chi$ .
  - c.  $\phi$  is as normal as  $(\psi \vee \chi)$ .

But then comparative structures derived from the Lewis/Kratzer tradition, such as those invoked by Swanson and Yalcin, do not provide the correct formalization of degrees of normalcy. If they did, (6.55) would be valid.

A detailed investigation of the lexical semantics of *normal* will have to await another occasion. For the moment, I will content myself with noting that *normal* appears to be a maximum-standard adjective whose antonym is minimum-standard. Consider, for example, these naturalistic examples of a maximizing degree modifier with *normal*, and of *slightly* with *abnormal*.

- (6.56)
- a. There are people in the country who think it is perfectly normal to sprinkle sugar onto their toast.
  - b. Oh sure, it is slightly abnormal to use a casting rod for a golf club, but that's the way it's done.

While the maximum classification indicates the existence of an upper bound, it seems that this scale is not closed below—that is, there may be no conceptual lower bound to normality. While *completely abnormal* seems to be fairly common, it is generally used in the distributive sense that we have seen many times before: “ $\phi$  is completely abnormal” means something like “ $\phi$  is abnormal no matter how you look at it”.

Getting back to (pseudo-)epistemic *ought* and *should*, there seem to be two possibilities. The first is that the meanings of these expressions are not, in fact, as closely related to normality as Yalcin indicates. Perhaps they are, as Swanson proposes, simply about the entailments of contextually relevant arguments, with no special connection to the item *normal*.

The second possibility is that (pseudo-)epistemic *ought* and *should* are in fact closely related to the concept of normality, but that we need to identify a more elaborated concept of normality in order to pin down the semantic behavior of these three items. Hopefully future work will refine the picture of the scalar structure of *normal* developed here, clarify the connection between normality, *should*, and *ought*, and elucidate the connection of all three items and the fundamental epistemic concept of probability.

As a first step in this direction, we can note that—as we saw in the last subsection—*ought* and *should* are gradable, accepting certain degree modifiers and forming comparatives.

- (6.57)
- a. Bill ought very much to be home by now.

- b. Mary ought to be home now more than Bill ought to be.

This is closely related to the gradability of deontic *ought* and *should*, which we will discuss further in chapter 8.

The Kratzerian semantics that Yalcin and Swanson adopt, since it analyzes both items as restricted universal quantifiers over worlds, does not seem to have the resources to deal with gradability. However, an explicitly scalar variant of these theories could treat unmodified (pseudo-)epistemic *ought* and *should* as the vague, context-dependent positive forms of the gradable modals that feed into the interpretation of modified forms.

The question, of course, is what kind of scale is implicated. At first glance, it seems unlikely that unmodified *ought* or *should* can be equated with the positive form of *normal*, or even an expression explicitly picking out the maximal degree of normality. Both of these can be construed as being much weaker.

- (6.58) a. # The top card of a well-shuffled deck ought to be the eight of spades.  
b. ✓ It is (totally) normal for the top card of a well-shuffled deck to be the 8 of spades.

However, the intuitive contrast in (6.58b) conceals a confound involving the focus-sensitivity of *normal*. To my ear, there is a sharp difference in acceptability between the following ways of pronouncing this sentence.

- (6.59) a. It is [NORMAL]<sub>F</sub> for the top card of a well-shuffled deck to be the eight of spades.  
(“This is not an unusual or surprising event”)  
b. # It is normal for the top card of a well-shuffled deck to be [THE EIGHT OF SPADES]<sub>F</sub>.  
(“Typically, the card at the top is the eight of spades”)

The second example in (6.59) seems to be odd in the same way that (6.58a) is: both indicate, wrongly, that it is the usual state of a well-shuffled deck to have the eight of spades on top. Perhaps, then, we can rescue the normality analysis of (pseudo-)epistemic *ought* and *should* by associating them with a proper subset of the available interpretations of (*totally*) *normal*. We could, perhaps, then treat the modified and comparative forms in (6.57) as the verbal counterparts of the relevant interpretations of *very normal* and *more normal than*.

The question of whether this analysis can be made to work in the details will, however, have to await another occasion. If it can, then it may be that the last vestiges of quantificational premise semantics can be eliminated in favor of a thoroughly scalar analysis of epistemic (and pseudo-epistemic) modals. Crucially, though, a variety of scales will have to be involved in the modal meanings traditionally labeled “epistemic”—normality, probability, and possibly certain functions of these such as log probability or negative entropy.

## 6.8 Conclusion

This concludes our discussion of epistemic modals. Let’s take stock.

Chapter 4 discussed *likely* and *probable* from two different perspectives: measurement-theoretic and degree-theoretic. We first considered whether  $\mathcal{S}_{likely}$  and  $\mathcal{S}_{probable}$  are fully-closed ratio scales, as suggested by the failure of the Disjunctive Inference discussed in ch.3. Several lines of evidence

support this classification, including the acceptability of ratio modifiers and the fact that these scales are intuitively bounded above and below (by a tautology and a contradiction, respectively). In addition, axiom-by-axiom examination of a measurement-theoretic formalization of qualitative probability suggests either that this classification is correct, or that these scales are at most slightly weaker if connectedness fails.

In §4.2, we discussed some objections to these conclusions articulated by Klecha (2012, 2014). The fact that *likely* and *probable* are relative adjectives is incompatible with their being on closed scales, according to Kennedy's (2007) prominent theory of the interpretation of scalar adjectives. I argued that the relevant aspects of Kennedy's theory are falsified by independent evidence involving non-modal adjectives, and that *likely* and *probable* simply add to the growing list of problems. While it remains very much up in the air what the best theory of degree modification and the positive form of gradable adjectives will look like, there is every reason to expect that the best theory will make room for the treatment of *possible* and *likely* that was motivated on structural and inferential grounds in the first part of chapter 4.

Chapter 5 turned to some other key epistemic items, starting with the maximum-standard adjectives *certain* and *sure*. I argued on the basis of linguistic and inferential connections between these adjectives and *likely* that their scales are either formally equivalent to  $S_{likely}$ , or involve some monotonic transformation. We also considered some puzzles around the interpretation of *uncertain*; the question of whether the certainty scale is lower-bounded; and the possibility of using a new semantics for embedded questions due to Uegaki (2015) to give a unified, compositional semantics for proposition- and question-embedding *certain* built around information-theoretic concepts. While this section did not settle on a single proposal for the meaning of *certain*, it will hopefully contribute to a better understanding of the numerous puzzles that such a theory must resolve, and the pros and cons of the various accounts discussed.

§5.2 began with the question of whether *possible* is gradable. While there is a widely-held belief among commentators (professional linguists and others) that the answer is negative, a variety of empirical and theoretical considerations indicate that possibility is a scalar concept, and that *possible* is a gradable adjective. This conclusion is problematic for the usual analysis of *possible* as an existential quantifier over epistemically accessible worlds, but it fits neatly with the hypothesis that *possible* is a minimum-standard adjective with a scale closely related (perhaps identical) to  $S_{likely}$ . *Possible* indicates deviation from a minimum point, and so—in effect—non-zero probability.

Finally, chapter 6 discussed the relation between the modal adjectives and auxiliaries, and the implications of the previous sections' conclusions for the auxiliaries. After considering a variety of semantic connections among the two groups of modals—notably, that *certain* is stronger than *must*, and *might* is stronger than *possible*, modulo the evidential component of the auxiliaries—we discussed several theoretical accounts of the latter. The relationship with modal adjectives creates serious difficulties for pure quantificational theories of the auxiliaries. While a strong theory such as von Stechow & Gillies 2010 makes incorrect predictions about the relationship between *must* and *certain*, a weak theory such as Kratzer 1991 fails to encode the semantic connection between *must* and *likely*. We also considered the mixed theory of Swanson (2015), which avoids both of these problems but appears to make puzzling predictions about the relationship between *must* and *might*.

The last theory standing is the pure scalar account, which treats *must* and *might* as stating

constraints on the probability of their propositional arguments—together with an evidential inference which was not discussed in detail here. (See von Fintel & Gillies 2010; Lassiter 2016; Goodhue 2016; Mandelkern 2016 for discussion of the evidential component of the meaning of the epistemic auxiliaries.) There are several ways to spell out such a theory, and it is supported by experimental evidence involving the relationship between *might* and *possible* as well as considerations showing that these items' meanings are context-dependent in roughly the same way that scalar adjectives are.

Scalar thinking is not merely useful in reasoning about modals—it is, I would suggest, indispensable. The next two chapters will continue this line of thought, applying the tools of scalar semantics to derive some surprising conclusions about the deontic adjective *good* and a number of semantically related modal verbs.

## CHAPTER 7

### Scalar Goodness

Most discussions of deontic language start with the modal verbs *may*, *must*, *should*, and especially *ought*. I will start instead with the adjectives *good* and *bad*—specifically, with the use of these adjectives that are associated in one way or another with propositional arguments, such as those in (7.1).

- (7.1) a. It is good/bad that Bill is here.  
b. It would be good/bad for Bill to leave.  
c. It is good/bad to give money to that charity.  
d. Giving money to charity is good/bad.

There are two reasons for this choice. First, it follows our general strategy of building theories in the first place around grammatically flexible expressions of a concept when these are available. Such items are likely to be more revealing about the basic logic of a concept than items whose interactions with other operators are more limited. Second, the next chapter will evaluate the prospects for using the scalar concept of goodness to provide an analysis of the important deontic items *ought* and *should*, as many theorists have proposed (e.g., Sloman 1970; Lewis 1973). The grammatical and inferential properties of the adjective *good* are thus of special interest, even for those who are mostly interested in *ought/should*, since they are likely to reveal in a fairly direct way the formal properties of a scale which strongly constrains the meaning of the latter.

*Good* and *bad* have a second, complex life as adnominal modifiers, and a third life as adverbs.

- (7.2) a. Bill is a good/bad dancer.  
b. Bill dances well/badly.

These uses are complex, and sufficiently different that I will not attempt to treat them together with proposition-embedding *good* here.

Two quick notes on terminology and assumptions. Throughout this chapter and the next, I use the word “deontic” in a very permissive way, including not only the strictly moral uses of *good*, *should*, *ought*, *must*, etc. but also their various other normative and deliberative uses. This probably constitutes a slight abuse of terminology. However, I am in the business of arguing that these uses share certain core semantic features, with variation in interpretation (normative, teleological, bouletic, etc.) located in the interpretation of the value function, and so outside of the semantics proper. As a result, I will not generally worry too much about the precise kind of normativity that is relevant in a particular context. However, it should be emphasized that discussions of the validity or invalidity of some argument are, unless otherwise, framed against the background assumption that the source of norms/values/goals/etc. is being held constant throughout the argument.

#### 7.1 Minimum/maximum/relative classification

In chapter 4 we introduced a number of tests that were used to classify adjectives within the minimum/maximum/relative typology (§4.2.1). These tests indicate that *good* is a relative adjective,

similar to *tall* or *heavy*.

- (7.3) The truck is completely heavy.  
a. “It’s clearly heavy; it’s heavy any way you look at it.”  
b. # “Nothing could be heavier.”
- (7.4) It is completely good to talk to him.  
a. “It’s clearly good; it’s good any way you look at it.”  
b. # “Nothing could be better.”

*Slightly*- and *almost*-modification are generally odd—

- (7.5) a. The truck is ??slightly/??almost heavy.  
b. It is ??slightly/??almost good to talk to him.

— and both pass the “*Adj* but could be *Adj-er*” test.

- (7.6) a. The truck is heavy, but it could be heavier.  
b. It is good to talk to him, but it could be better.

Both adjectives have antonyms that are also relative adjectives, for example:

- (7.7) a. The truck is ??slightly/??almost light.  
b. It is ??slightly/??almost bad to talk to him.

- (7.8) a. The truck is light, but it could be lighter.  
b. It is bad to talk to him, but it could be worse.

And both show a zone of indifference:

- (7.9) a. The truck is not heavy, but it is not light either.  
b. It is not good to talk to Bill, but it is not bad either.

Finally, *good* resembles relative adjectives (including *likely/probable*, see §4.2.2) in being sensitive to contextual alternatives. In the case of ordinary relative adjectives, this is most apparent through the use of comparison classes.

- (7.10) This truck is heavy for a flatbed.  
“Compared to an average flatbed truck, this one is heavy”

The alternative-sensitivity of the positive form of *good* is somewhat subtler. Explicit comparison classes, for example, do not seem to be especially natural.

- (7.11) Giving money to charity is good for a gesture to impress one’s girlfriend.  
?? “Compared to the usual things one might do to impress a girlfriend, giving money to charity is good”

However, the alternative-sensitivity of *good* reveals itself when we consider the effects of focus. See §7.9 below for discussion. It could turn out that this property is unrelated to the status of *good* as a relative adjective. However, given that relative adjectives quite generally have context-sensitive positive-form meanings that vary depending on the properties of a set of alternatives, it seems plausible that the alternative-sensitivity of *good* reflects the same forces that trigger the more general phenomenon.

## 7.2 Aspects of scale structure

Since we are restricting attention to *good/bad* taking propositional arguments, the scale that we are interested in is an ordering on propositions. Like  $\mathcal{S}_{\text{likely}}$ ,  $\mathcal{S}_{\text{good}}$  is built around a binary relation  $\succsim_{\text{good}} \subseteq \mathcal{P}(W)^2$ —that is, an ordering on (some subset of)  $\mathcal{P}(W)$ , the set of all propositions.

The next set of questions to ask involve the formal structure of  $\mathcal{S}_{\text{good}}$ : where (if anywhere) it fall into the typology of scales developed in chapter 2. Is it ordinal, ratio, interval, or something else? Upper- and/or lower-bounded? Connected? Positive, maximal, intermediate, or something else?

On Kennedy’s (2007) assumptions, we could infer from the fact that *good* is a relative adjective that its scale is unbounded on both ends. In chapter 4 (§4.2.4) I argued that this inference is not generally valid. Numerous relative adjectives live on scales that are bounded on one or both ends, including *likely* and *probable*. However, in the case of *good* I suspect that we can affirm the conclusion. Intuitively,  $\mathcal{S}_{\text{good}}$  does not have an intrinsic maximum or minimum: like kindness and beauty, there is no conceptual limitation on how good (or bad) a proposition could be. If so,  $\mathcal{S}_{\text{good}}$  contrasts intuitively with  $\mathcal{S}_{\text{likely}}$ , where it seems apparent that tautologies and contradictions provide likelihood with principled upper and lower bounds (respectively). However, nothing in what follows will depend crucially on the assumption that  $\mathcal{S}_{\text{good}}$  is unbounded.

Another basic question about  $\mathcal{S}_{\text{good}}$  is whether it is ordinal, interval, ratio, or something else entirely. A ratio scale classification is not viable in light of the conclusions of the next section (ratio scales are by definition positive, and  $\mathcal{S}_{\text{good}}$  is not). Nor is an ordinal scale classification promising, for reasons that were discussed in some detail in chapter 3, §3.3.3. There I pointed out that an ordinal scale classification would render even all quantitative comparisons uninterpretable, even very weak ones such as  $\phi$  is *much* better than  $\psi$ . Since comparisons of this type are unremarkable,  $\mathcal{S}_{\text{good}}$  must have additional structure. At a minimum, the relative sizes of gaps—the ratio of  $\mu_{\text{good}}(\phi) - \mu_{\text{good}}(\psi)$  to  $\mu_{\text{good}}(\phi') - \mu_{\text{good}}(\psi')$ —must be interpretable quantities.

This suggests that  $\mathcal{S}_{\text{good}}$  may be an INTERVAL SCALE. Interval scales have the property that all admissible measure functions are related by some positive affine transformation:

If  $\mathcal{S}_A$  is interval, then, for any two admissible measure functions  $\mu_A, \mu'_A$ , for all  $x$  in its domain  $X$ ,

$$\mu_A(x) = \alpha \times \mu'_A(x) + \beta$$

for some  $\alpha \in \mathbb{R}^+$  and some  $\beta \in \mathbb{R}$ .

Temperature, danger, and clock time are examples of interval scales that were discussed in chapter 2: see §2.2.3 for further details. Similarly, if  $\mathcal{S}_{\text{good}}$  is an interval scale, then this qualitative scale can be faithfully represented in terms of the properties of a set of admissible measure functions. Later in

this chapter I will consider a specific hypothesis about the form of these functions; however, for the moment we will focus primarily on the question of which qualitative constraints are appropriate.

Potentially problematic for this classification is that interval scales are required to be connected, and it is not at all clear that this assumption is reasonable in the case of  $\mathcal{S}_{good}$ . Are all pairs of propositions comparable in goodness? The theoretical situation is, I think much the same as in the extensive discussion of the connectedness of  $\mathcal{S}_{likely}$  in chapter 4 (§4.1.5). If human judgments of goodness are constitutively involved in the scale, then it is very tempting to conclude that the scale is not connected on the grounds that it is sometimes hard to form a judgment of relative goodness. On the other hand, if goodness is something that people learn about rather than creating—or, more precisely, if the agents that we are modeling believe that it is—then it is possible that  $\mathcal{S}_{good}$  is connected even while people may sometimes be unable to form judgments about comparative goodness.

In any case, it may be desirable to treat  $\mathcal{S}_{good}$  as having much of the structure of an interval scale, but without making the connectedness assumption. I do not know what a qualitative scale with these characteristics would look like formally. However, it is fairly straightforward to construct one indirectly from a collection of interval scales, using a supervaluation trick. The basic idea is the same that is the one discussed in a related context by Lassiter 2011a, and is broadly similar to the “ordering supervaluationism” of Swanson (2011, 2014). I will sketch this idea briefly in chapter 8, §8.11 with specific reference to *ought/should*, and discuss its prospects as an account of conflicting *oughts*.

### 7.3 Is goodness positive, maximal, or intermediate?

Recall from earlier chapters the epistemic DISJUNCTIVE INFERENCE, involving arguments of the form in (7.12).

- (7.12)    a.  $\phi \succsim_{likely} \psi$   
           b.  $\phi \succsim_{likely} \chi$   
           c.  $\therefore \phi \succsim_{likely} (\psi \vee \chi)$

A semantics that renders argument of this form valid is incorrect. For example, suppose that I have just drawn a card from the top of a well-shuffled-pack. Premises (7.13a) and (7.13b) are both true here, but the conclusion in (7.13c) is clearly false. In reality, it’s more likely that the card will be a red card than it is that the card will be a spade—exactly twice as likely, in fact.

- (7.13)    a. It’s as likely that the card is a spade as it is that it’s a heart.  
           b. It’s as likely that the card is a spade as it is that it’s a diamond.  
           c. Therefore, it’s as likely that the card is a spade as it is that it’s a red card (i.e., either a heart or a diamond).

In chapters 3–4 I used the intuitive failure of this inference to argue that the likelihood scale is not MAXIMAL but rather POSITIVE. These properties correspond to assuming the validity of the following inferences, for disjoint  $\phi$  and  $\psi$ :

- (7.14)    **Maximality of  $\mathcal{S}_{Adj}$ :** If  $\phi$  is at least as *Adj* as  $\psi$ , then  $\phi$  is exactly as *Adj* as  $\phi \vee \psi$ .



(7.15) **Positivity of  $\mathcal{S}_{Adj}$ :**  $\phi \vee \psi$  is more *Adj* than  $\phi$ , unless  $\mathcal{S}_{Adj}$  is lower-bounded and  $\psi$  is exactly as *Adj* as the lower bound.

Turning to the scale  $\mathcal{S}_{good}$ , there is a notable difference from  $\mathcal{S}_{likely}$ : the inference corresponding to (7.13) seems to be valid.

- (7.16) a.  $\phi \geq_{good} \psi$   
 b.  $\phi \geq_{good} \chi$   
 c.  $\therefore \phi \geq_{good} (\psi \vee \chi)$

Consider, for example, a scenario in which there are various monetary prizes attached to the event of drawing particular cards. You are planning to put any winnings from your draw toward charitable causes. The following inference pattern seems to be reasonable.

- (7.17) a. It's as good for the card to be a spade as it is for it to be a heart.  
 b. It's as good for the card to be a spade as it is for it to be a diamond.  
 c. Therefore, it's as good for the card to be a spade as it is for it to be a red card (i.e., either a heart or a diamond).

(7.17a) indicates that drawing a spade will generate (at least) as much revenue for your charitable cause as drawing a heart. Similarly, (7.17b) suggests that drawing a spade will generate (at least) as much revenue for this purpose as drawing a diamond. If these premises are both true, then clearly drawing a spade will generate (at least) as much revenue for this purpose as drawing a red card, since any red card is either a spade or a heart. So, drawing a spade is at least as good as drawing a red card.

The fact that this inference is intuitively reasonable in the example in (7.17) does not, of course, demonstrate that (7.16) is unrestrictedly valid. I believe that it is, though. (At least, I have failed to generate any plausible counter-examples after several years of trying.) However, we don't need to establish this validity to show that  $\mathcal{S}_{good}$  is not positive: the fact that the sentences in (7.17) are jointly satisfiable already proves that it cannot be. In the typology of scales developed in chapter 2,  $\mathcal{S}_{good}$  must differ from  $\mathcal{S}_{likely}$  along this crucial parameter.

This does not yet settle whether  $\mathcal{S}_{good}$  is maximal, though. **Intermediacy** is an alternative possible relationship between scalar properties and disjunction (join). In chapter 2 (§2.5.2) I argued that at least two scalar properties associated with adjectival scales in English—temperature and danger—have this relationship to the join operation.

(7.18) **Intermediacy of  $\mathcal{S}_{Adj}$ :**

- If  $\phi \geq_{Adj} \psi$ , then  $\phi \geq_{Adj} \phi \vee \psi \geq_{Adj} \psi$ ; and
- If  $\phi >_{Adj} \psi$ , then  $\phi >_{Adj} \phi \vee \psi >_{Adj} \psi$ —unless either  $\phi$  or  $\psi$  receives zero weight in the mixture.

The validity of the inference pattern in (7.16) is predicted equally by a semantics in which  $\mathcal{S}_{good}$  is either maximal or intermediate.

The definition in (7.18) is refined slightly relative to its original version in chapter 2, since—when talking about temperature and danger—we did not encounter any reason to worry about joins

involving objects that receive zero weight in determining the degree of the result. If we had, a similar “zero weight in the mixture” caveat would have been needed to account for trivial joins in which, for example, one object has zero volume and therefore does not influence the temperature of the result. In the modal domain there is considerable reason to be careful about such trivial cases, since the weighting function that I will propose to employ below—conditional probability—may sometimes assign zero weight to a proposition  $\psi$  when combining it with some  $\phi$ . When this happens, the goodness of  $\phi \vee \psi$  will be equal to, not less than, that of  $\phi$ . In many cases, though, it does little harm to rely on a simplified definition of intermediacy:

(7.19) **Intermediacy of  $\mathcal{S}_{Adj}$**  (over-simplified):

$$\text{If } \phi >_{Adj} \psi, \text{ then } \phi >_{Adj} \phi \vee \psi >_{Adj} \psi.$$

Of the two options that correctly validate the inference in (7.17), which one provides the right account of scalar goodness—maximality or intermediacy? It turns out that this question is essentially the same as the question of the **monotonicity** properties of goodness and related concepts. If it turns out that *ought* and *should* should be constrained, or even defined, in terms of the goodness scale, the monotonicity properties of this scale will also influence the logical properties of these better-studied items. We will return to this issue in some detail in chapter 8.

## 7.4 Problems with maximality

In chapter 3 we studied the qualitative scalar semantics of Lewis (1973: §5), which is built around a “Comparative Possibility” relation. The scalar aspect of this theory forms a crucial part of the *de facto* standard semantics for deontic modals, assumed, elaborated, or used as a starting point for investigation by many subsequent theorists. As we saw, Lewis expresses serious reservations about the maximality property of the scalar analysis of goodness that he has just introduced.

This is not instrumental or intrinsic betterness of any familiar form, but rather *maximax betterness*. Roughly, we are comparing  $\phi$ -at-its-best with  $\psi$ -at-its-best, and ignoring the non-best ways for  $\phi$  and  $\psi$  to hold. (Lewis 1973: 101)

Why does Lewis flag this as a concern? Presumably, it is because he thinks that the maximality property does not hold of the “familiar form” of betterness. One way to gloss the maximality property is the following:

**Maximality (equivalent definition, assuming connectedness):** If  $\Psi = \{\psi_1, \dots, \psi_n\}$  is a finite partition of  $\phi$ , where  $\bigcup_i \psi_i = \phi$ , then

$$\mu_{good}(\phi) = \max_i \mu_{good}(\psi_i),$$

for any admissible  $\mu_{good}$ .

In other words,  $\phi$  is exactly as good as the best cell in any finite partition of  $\phi$ .<sup>1</sup>

If  $\mathcal{S}_{good}$  is maximal and connected, this property holds for arbitrary partitions, including the maximally fine-grained partition with each cell containing exactly one world. As a result, maximality entails that any proposition  $\phi$  is exactly as good as its highest-ranked singleton subset—that is, whatever the best  $\{w\}$  is for any  $w \in \phi$ ,  $\phi$  is exactly that good. All information about the distribution of non-maximal possibilities—e.g., whether the other worlds in  $\phi$  are mostly very good, or mostly very bad—is ignored.

(Without the connectedness assumption, the statement of the property is less simple. However, this gloss continues to hold true of connected sub-parts of the ordering  $\succsim_{good}$ . None of the puzzles discussed below relies crucially on connectedness in general, but only on comparability among the relevant options.)

In addition, the maximality property applies recursively to the cells  $\psi_i$  of any partition of  $\phi$ : each is itself exactly as good as the best proposition in an arbitrary finite partition of  $\psi_i$ .

Adversarial games serve to illustrate why the maximality property leads to absurd consequences. Suppose that we are playing chess, and I am choosing among the  $k$  legal moves  $\{m_1, \dots, m_k\}$  that are available to me at a given point in time. Given the structure of chess,  $k$  will always be finite. For each of my possible moves  $m_i$ , there are  $k'$  legal moves that you could take in response, and so on. If I take my  $i$ th legal move, you take your  $j$ th legal response, and then I take my  $n$ th legal response to your response, ..., we write this sequence as  $\langle m_i, m_{ij}, m_{ijn}, \dots \rangle$ .

What is my best move? (Let's simplify, inconsequentially, by assuming that there is a unique best move.) Since we are assuming that  $\mathcal{S}_{good}$  is maximal, this means that the goodness of the proposition *I take move  $m_i$*  is equal to the goodness of the best cell in an arbitrary cell of any partition of this proposition. In particular, we can consider the partition which divides *I take move  $m_i$*  into cells of the form *I take move  $m_i$  and you take move  $m_{ij}$* , one for each move  $m_{ij}$  that you could take in response to my move  $m_i$ . By maximality, the goodness of *I take move  $m_i$* , from my perspective, is equal to the goodness of the *best* sequence of the form  $\langle m_i, m_{ij} \rangle$ , for any possible move  $m_{ij}$  that you could make. But since chess is a purely adversarial game, the best sequence from my perspective is the worst from yours.

Thus, the goodness of my taking move  $m_i$  is equal to the goodness, from my perspective, of the *worst possible move* that you could make in response to  $m_i$ . As a result, the *best* move out of all of my options will be the one that for which your worst possible response is worst. By applying this procedure similarly to your response, we find that my best move is the move for which my best possible response to your worst possible response is best. This, in turn, is required by maximality to be equivalent to the best possible response to your worst possible response to my best possible

<sup>1</sup> **Proof:** Let  $\{\psi_1, \psi_2, \dots\}$  be a partition of  $\phi$ . Given connectedness, either  $\psi_1 \succsim_{good} \psi_2$  or  $\psi_2 \succsim_{good} \psi_1$ . In the former case,  $(\psi_1 \cup \psi_2) \approx_{good} \psi_1$  by maximality. In the latter case,  $(\psi_1 \cup \psi_2) \approx_{good} \psi_2$ .

Similarly, if  $\psi_3 \succsim_{good} (\psi_1 \cup \psi_2)$  then  $(\psi_1 \cup \psi_2 \cup \psi_3) \approx_{good} \psi_3$  by maximality, and (by transitivity and reflexivity)  $\psi_3 \succsim_{good} \psi_i$  for all  $i \leq 3$ . Otherwise,  $(\psi_1 \cup \psi_2 \cup \psi_3) \succsim_{good} \psi_3$ , in which case either  $(\psi_1 \cup \psi_2 \cup \psi_3) \approx_{good} \psi_1$  or  $(\psi_1 \cup \psi_2 \cup \psi_3) \approx_{good} \psi_2$ , depending on whether  $\psi_1 \succsim_{good} \psi_2$ .

Continuing this procedure, we find that, for any  $j$ ,  $\psi_j \approx_{good} \bigcup_{i=1}^j \psi_i$  if and only if  $\psi_j \succsim_{good} \psi_i$  for all  $i \leq j$ ; and, if not  $\psi_j \approx_{good} \bigcup_{i=1}^j \psi_i$ , then there is some  $k < j$  such that  $\psi_k \approx_{good} \bigcup_{i=1}^j \psi_i$ . Given the definition of admissibility, it clearly follows that  $\mu_{good}(\phi) = \mu_{good}(\psi_k)$  for some cell  $\psi_k$  with the property that  $\psi_k \succsim_{good} \psi_i$  for all  $\psi_i$ .  $\square$

response to your worst possible response to ...

But what if I do not think that you are a terrible chess player? The intuitive concept of goodness seems to make room for reasoning like this:

If I take move  $m_i$ , he might take  $m_{ij}$ —which would be very good for me —or  $m_{ij'}$ , which would be very bad for me. However, if I take  $m_i$ , he will probably anticipate that  $m_{ij}$  would be good for me and bad for him, and if he does he will take  $m_{ij'}$ . So, move  $m_i$  is a bad move for me, *even though its best possible outcome is very good*.

Maximax goodness does not leave room for this kind of reasoning, though. If  $m_{ij}$  and  $m_{ij'}$  are relevantly possible response moves, then the maximality property forces us to identify the goodness of  $m_i$  with the goodness of the best possible response  $m_{ij}$ —from your perspective, the worst possible response that you could take.

The maximality property would oblige me to reason about the comparative goodness of my choices in an adversarial game *as if* I were certain that my opponent is a fool. This is unreasonable. The only way to get around this consequence without abandoning maximality is to exclude the better options from the ordering—in the above, to declare  $m_{ij}$  not to be a relevant possibility. While this helps in some cases, it is not sufficiently general. If I am playing an opponent and I know nothing about his chess abilities, it may be possible that he will respond to  $m_i$  with either the very bad  $m_{ij}$  or the very good  $m_{ij'}$ . (He might be a grandmaster or a novice.) At the same time, I would probably consider  $m_i$  to be a bad move because I think it is *probable* that my opponent will not take his worst possible move—even though it is a relevant possibility that he might.

For another problematic case, consider the following story.<sup>2</sup>

**Juliet** is considering whether to feign death by taking the drug that Friar Laurence has offered her. If she does, it will put her in a coma, and she will die unless Friar Laurence administers the antidote exactly 10 hours later. If she takes it and the Friar does administer the antidote, she will succeed in convincing her family of her death and she will be able to live happily ever after with Romeo. If she does not take the drug, she will live a long life without Romeo and will be less happy; this is much better than being dead, though. Unfortunately, the Friar is known for being cruel and capricious, and it is extremely likely (though not totally certain) that he will ‘forget’ to administer the antidote if she takes the drug.

Which of the following represents the best outcome?

- (A) Juliet takes the drug and the Friar administers the antidote. (drug^antidote)
- (B) Juliet takes the drug and the Friar does not administer the antidote. (drug^¬antidote)
- (C) Juliet does not take the drug. (¬drug)

---

<sup>2</sup> The story is taken from Lassiter 2014, and was inspired by Cariani’s (2013) chancy variant of Jackson & Pargetter’s (1986) famous “Professor Procrastinate” story.

To my mind, the scenario suggests answer (A). (C) is clearly the next best outcome, followed by the worst outcome, (B).

Now, please consult your intuitions about the following sentence:

(7.20) It is better for Juliet to take the drug than it is for her not to take the drug.

This sentence seems to be clearly false in the scenario at hand. If Juliet takes the drug, the Friar will probably fail to administer the antidote, and so the worst possible outcome—death—will probably result. So, it is better for her not to take the drug.

If comparative goodness has the maximality property, these intuitions are formally inconsistent. Since  $\phi$  is logically equivalent to  $(\phi \wedge \psi) \vee (\phi \wedge \neg\psi)$ , *Juliet takes the drug* is equivalent to *Either Juliet takes the drug and the Friar administers the antidote, or she takes the drug and he does not*.  $\{\mathbf{drug} \wedge \mathbf{antidote}, \mathbf{drug} \wedge \neg\mathbf{antidote}\}$  thus constitutes a partition of **drug**. By maximality, then, the goodness of **drug** is equivalent either to the goodness of **drug** $\wedge$ **antidote** or the goodness of **drug** $\wedge$  $\neg$ **antidote**, whichever is better. Since **drug** $\wedge$ **antidote** is clearly better, the consequence is that *Juliet takes the drug* is exactly as good as *Juliet takes the drug and the Friar administers the antidote*—even though we know that taking the drug will probably lead to death.

This reasoning is not just intuitively incorrect; it leads to a formal contradiction, since we have already established that **drug** $\wedge$ **antidote** is intuitively better than  $\neg$ **drug**. Given maximality and transitivity of  $\succsim_{good}$ , this is inconsistent with the denial of (7.20), which would entail that **drug**—a proposition with the same degree of goodness as **drug** $\wedge$ **antidote**—is not better than  $\neg$ **drug**.

The diagnosis that emerges is the following: the problem with maximality is that it writes an unreasonable optimism into the logic of goodness. The problem was summed up nicely in the *Port-Royal Logic* of 1662:

Many people fall into an illusion which is more deceptive the more reasonable it appears to them. They only consider the greatness of the consequences of the advantage that they wish for, or of the inconvenience that they fear, without considering in any way the probability that that advantage or inconvenience will occur or not.<sup>3</sup>

Later in the chapter I will lay out a possible-worlds semantics of scalar goodness that takes probability into account in the way that Arnauld & Nicole recommend, using the probabilities of the cells of a partition of  $\phi$  to weight their input into the goodness of the proposition. This is, I will suggest, a decent (if probably still imperfect) formalization of how people reason intuitively about scalar goodness. When we think about the relative goodness of Juliet's options, or of the various moves that I might make at a certain point in a chess game, we do not focus exclusively on the best—or worst—possible outcomes that might result. Instead, we strive to consider both how good the relevant outcomes would be, and how likely they are to occur:

To judge what one must do in order to obtain a good or avoid an evil, one must consider not only the good and the evil in itself, but also the probability that it will

---

<sup>3</sup> Antoine Arnauld & Pierre Nicole, *La Logique ou l'Art de Penser*, 1662, Fourth Part, Chapter XVI. This and the following quote are my translation from p.331 of the 1992 Gallimard edition, edited by Charles Jourdain.

occur or not, and to view geometrically the proportion that all these things have together.<sup>4</sup>

## 7.5 The logic of scalar goodness

The preceding discussion makes it possible to say a good deal about the logic of scalar goodness without yet settling on a specific semantic theory. If goodness is neither Positive nor maximal, what *is* the relationship between goodness and disjunction/join? There are many logically possible analyses, some of which were described in chapter 2, §2.5.3. However, only one of those alternatives is both logically plausible and known to be instantiated in the scalar expressions of English: intermediacy, as defined in §7.3 and repeated here.

(7.21) **Intermediacy of  $\mathcal{S}_{Adj}$ :**

- If  $\phi \succsim_{Adj} \psi$ , then  $\phi \succsim_{Adj} \phi \vee \psi \succsim_{Adj} \psi$ ; and
- If  $\phi \succ_{Adj} \psi$ , then  $\phi \succ_{Adj} \phi \vee \psi \succ_{Adj} \psi$ —unless either  $\phi$  or  $\psi$  receives zero weight in the mixture.

The idea that scalar goodness has the intermediacy property, or something closely related to it, has appeared occasionally in the philosophical literature. B. Hansson (1968) mentions it, Chisholm (1975) endorses it, and S. O. Hansson (2004) calls it “quite plausible”. However, its application to specific puzzles has perhaps not been explored sufficiently in previous semantic work.

Goodness, on this proposal, is the modal analogue of temperature—just as likelihood is the modal analogue of the additive properties of volume and weight. When we combine the contents of two bowls of water  $b_1$  and  $b_2$ ,

- the volume of the result is the sum of the volumes of  $b_1$  and  $b_2$ ;
- the temperature of the result is intermediate between the temperatures of  $b_1$  and  $b_2$  (with a precise value that depends on the volumes of  $b_1$  and  $b_2$ ).

Similarly, I suggest, when we combine two disjoint propositions  $\psi_1$  and  $\psi_2$  using the disjunction operation,

- the likelihood of the result is the sum of the likelihoods of  $\psi_1$  and  $\psi_2$ ;
- the goodness of the result is intermediate between the goodness of  $\psi_1$  and  $\psi_2$  (with a precise value that depends on the likelihoods of  $\psi_1$  and  $\psi_2$ ).

This proposal explains why the deontic version of the disjunctive inference should be valid. If  $\phi \succsim_{good} \psi$  and  $\phi \succsim_{good} \chi$ , then  $\phi \succsim_{good} (\psi \vee \chi)$  because  $\psi \vee \chi$  cannot be better than  $\psi$ , or better than  $\chi$ . It also makes room for the intuitively correct results in the scenarios that I argued were problematic for maximality. Specifically, if we assume connectedness, we can gloss intermediacy as follows:

<sup>4</sup> *Ibid.*. This quote is also the epigram of chapter 1 of Jeffrey (1965).

**Intermediacy (equivalent definition, assuming connectedness):** If  $\psi_1, \dots, \psi_n$  is a finite partition of  $\phi$ , then the goodness of  $\phi$  is somewhere in between the goodness of the best and worst cells in the partition: for any admissible  $\mu_{good}$ ,

$$\max_i \mu_{good}(\psi_i) \geq \mu_{good}(\phi) \geq \min_i \mu_{good}(\psi_i),$$

with both inequalities strict if all  $\psi_i$ 's have positive likelihood.

In the **Juliet** scenario from §7.4, the problem was that the goodness of *Juliet takes the drug* was intuitively **less than** the goodness of *Juliet takes the drug and the Friar administers the antidote*. In other words, a proposition of the form  $\phi \wedge \psi$  was intuitively **better than** a proposition  $\phi$ . This is impossible if  $\mathcal{S}_{good}$  is maximal, but it is predicted to be possible if  $\mathcal{S}_{good}$  is intermediate. The latter allows  $\phi \wedge \psi$  to be better than  $\phi$ , as long as  $\phi \wedge \neg\psi$  (Juliet takes the drug and the Friar does not administer the antidote) is worse.

In the **Chess** example, the problem was that the goodness of taking a move did not depend only on the best possible outcome (i.e., the opponent's worst possible response). Rather, the full range of possible responses appears to be relevant in determining how good it is to make a certain move, along with the probability that each response would be taken. The intermediacy constraint does not require that probabilities be taken into account (though, as we will see shortly in §7.6, it is compatible with one plausible way of doing so). It does, however, have the virtue of constraining the goodness of making a move to be somewhere in between the goodness of the best and worst possible outcomes. The absurd consequences of maximality do not arise for an intermediate semantics.

This does not yet constitute a proposal about the semantics of *good*. Rather, the claim is that any empirically adequate semantic theory of scalar goodness must have certain formal properties and lack others. The argument in favor of these particular constraints is simply that a semantics which obeys them is in principle capable of getting the basic data right: the deontic version of the Disjunctive Inference is valid, and no obviously incorrect predictions about **Juliet** and **Chess** are generated.

In the next section, I will propose a specific possible-worlds semantics that has the necessary features. I will then use this proposal to develop new solutions to some puzzles involving the information-sensitivity of *good*, close counterparts of which have much been discussed in recent literature.

## 7.6 A possible-worlds semantics for scalar goodness

We are looking for a semantics according to which  $\mathcal{S}_{good}$  is an interval scale that is intermediate with respect to disjunction. One clear candidate is the decision-theoretic concept of **expected value**. There are many different formalizations of this concept. The one that I will explore here follows most closely Jeffrey (1965), as applied to normative language by Goble (1996) (with considerable inspiration from Jackson (1985, 1991)).

We start with a **value function**  $V$  which takes possible worlds to real numbers:  $V : W \rightarrow \mathbb{R}$ . That is, if we can specify a state of affairs in all relevant detail—with no remnant uncertainty—this function will tell us exactly how good it would be for the world to be like that. Importantly, the

interpretation of the value function is not specified by the model: it could be moral goodness, instrumental goodness relative to a particular goal, desirability for a given individual, consequentialist value, deontological value, or something else entirely. Since there are many “flavors” of goodness, there will be correspondingly many value functions that are potentially relevant to the interpretation of this expression on a given occasion. Which one is active in a particular instance will be a context-dependent matter.

The standard assumption (going back to von Neumann & Morgenstern 1944) is that the value function is unique up to positive affine transformation—that is, the only “meaningful” information in the value function is what would be preserved in the transformation  $V'(w) = \alpha \times V(w) + \beta$  ( $\alpha \in \mathbb{R}^+, \beta \in \mathbb{R}$ ). Furthermore, all such transformations lead to equivalent empirical predictions. As we saw in detail in chapter 2, whenever there is a set of functions that meets these conditions, there is in the background a qualitative scale which obeys the interval scale axioms given in §2.5.1. The functions in question are exactly the admissible measure functions on this interval scale. So, the usual assumptions about the meaningful information in the value function  $V$  are equivalent to assuming that  $V$  is an admissible measure function on some qualitative interval scale  $\mathcal{S}_{\text{world-value}}$ .

As noted above, the things that we predicate goodness and badness of are not possible worlds but propositions. One could lift a scale representing the values of worlds to a scale representing the values of propositions in myriad ways. In decision theory, a standard way to do this is **expected value**: a weighted average of the values of the worlds in the proposition, representing our best guess about how good things will be if the proposition obtains. The weights are given by the conditional probabilities of the various worlds, assuming that the proposition obtains.

(7.22) The **expected value** of a proposition  $\phi$ , relative to a domain  $D$ , is a weighted average of the actual values of worlds in  $\phi \cap D$ .

$$\mathbb{E}_V(\phi) = \sum_{w \in \phi \cap D} V(w) \times \text{prob}(\{w\} \mid \phi \cap D).$$

(This is for the finite case; the infinite case is slightly more complex, but the technicalities do not affect the substance of our point here. Note also that we assume that  $\mathbb{E}_V(\emptyset)$  is undefined.)

In many cases of interest, the domain  $D$  can be equated with the epistemically possible worlds. If these happen also to be the worlds with positive probability, then all and only possible worlds that are not in  $D$  receive zero weight. So, the formula simplifies to

$$\mathbb{E}_V(\phi) = \sum_{w \in \phi} V(w) \times \text{prob}(\{w\} \mid \phi).$$

This is the form that I will generally use, except when the specification of the domain is of particular importance. More generally, I will assume  $\text{prob}(D) = 1$  except when otherwise noted.

Given the assumption that world-value is an interval scale, we can prove that expected value is also an interval scale. This is a result of the way that expected value is defined from value. Given any expected value function  $\mathbb{E}_V$  derived from a value function  $V$  using the definition in (7.22), we can show that any positive affine transformation of  $\mathbb{E}_V$  is an expected value function  $\mathbb{E}_{V'}$ , and that  $\mathbb{E}_{V'}$  is derived by (7.22) from a value function  $V'$  which is a positive affine transformation  $V$ , with



the same *prob* function and coefficients  $\alpha$  and  $\beta$ . To see this, note that if we apply a positive affine transformation to the values of all worlds— $V'(w) = \alpha \times V(w) + \beta$ —then we have

$$\begin{aligned}\mathbb{E}_{V'} &= \sum_{w \in \phi} [\alpha \times V(w) + \beta] \times \text{prob}(\{w\} \mid \phi) \\ &= \sum_{w \in \phi} (\alpha \times V(w) \times \text{prob}(\{w\} \mid \phi)) + (\beta \times \text{prob}(\{w\} \mid \phi)) \\ &= \sum_{w \in \phi} (\alpha \times V(w) \times \text{prob}(\{w\} \mid \phi)) + \sum_{w \in \phi} (\beta \times \text{prob}(\{w\} \mid \phi)) \\ &= \alpha \sum_{w \in \phi} V(w) \times \text{prob}(\{w\} \mid \phi) + \beta \sum_{w \in \phi} \text{prob}(\{w\} \mid \phi)\end{aligned}$$

Since  $\sum_{w \in \phi} \text{prob}(\{w\} \mid \phi) = 1$ , this reduces to

$$\begin{aligned}\mathbb{E}_{V'} &= \alpha \times \left( \sum_{w \in \phi} V(w) \times \text{prob}(\{w\} \mid \phi) \right) + \beta \\ &= \alpha \times \mathbb{E}_V(\phi) + \beta.\end{aligned}$$

Any time a scale is reflected by a set of measure functions that includes all and only the positive affine transformations of any element of that set, it is an interval scale (Krantz et al. 1971, cf. ch.2, §2.5.1). So, for any  $V$  reflecting an interval scale of world-comparisons  $\mathcal{S}_{\text{world-value}}$ ,  $\mathbb{E}_V(\cdot)$  reflects an interval scale of proposition-comparisons  $\mathcal{S}_{EV}$ . We can write the qualitative comparison underlying it as  $\succsim_{EV}$ , with  $\phi \succsim_{EV} \psi$  if and only if  $\mathbb{E}_V(\phi) \geq \mathbb{E}_V(\psi)$  for every  $\mathcal{S}_{EV}$ -admissible measure function  $\mathbb{E}_V(\cdot)$ .

Expected value is an intermediate scale. In general, whenever  $\phi$  and  $\psi$  are disjoint, the expected value of their disjunction can be calculated as a probability-weighted average of their individual expected values (Jeffrey 1965: §5).

$$\mathbb{E}_V(\phi \vee \psi) = \frac{\mathbb{E}_V(\phi) \times \text{prob}(\phi) + \mathbb{E}_V(\psi) \times \text{prob}(\psi)}{\text{prob}(\phi) + \text{prob}(\psi)}$$

Equivalently, we can find the weighted sum of the values, where the weights are the conditional probabilities *on the assumption that one or the other is true*.

$$\mathbb{E}_V(\phi \vee \psi) = \mathbb{E}_V(\phi) \times \text{prob}(\phi \mid \phi \vee \psi) + \mathbb{E}_V(\psi) \times \text{prob}(\psi \mid \phi \vee \psi)$$

Suppose that  $\mathbb{E}_V(\phi) > \mathbb{E}_V(\psi)$ . Then clearly  $\mathbb{E}_V(\phi \vee \psi) = \mathbb{E}_V(\phi)$  if  $\text{prob}(\phi \mid \phi \vee \psi) = 1$ , so that  $\text{prob}(\psi \mid \phi \vee \psi) = 0$ . On the other hand, if  $\text{prob}(\psi \mid \phi \vee \psi) = 1$  then  $\mathbb{E}_V(\phi \vee \psi) = \mathbb{E}_V(\psi)$ . In any other case,  $\mathbb{E}_V(\phi \vee \psi)$  will be a mixture of the two: greater or lesser depending on the relative probabilities, but never greater than  $\mathbb{E}_V(\phi)$  or less than  $\mathbb{E}_V(\psi)$ . As a result, the statement

$$\mathbb{E}_V(\phi) \geq \mathbb{E}_V(\phi \vee \psi) \geq \mathbb{E}_V(\psi)$$

holds whenever  $\mathbb{E}_V(\phi) > \mathbb{E}_V(\psi)$ . This means that  $\mathcal{S}_{EV}$  is an intermediate scale.

Expected value thus has the crucial properties that were identified above. It is an interval scale, and it is intermediate rather than maximal. It remains to be seen whether it does a good job of formalizing the concept of scalar goodness. In the remainder of this section I will suggest that it does: at least to a first approximation, the equation

$$\mathcal{S}_{good} = \mathcal{S}_{EV}$$

is reasonable.

There are many other formalisms for reasoning about the interaction of information and values, and there may well be more complex representations that do a better job of handling certain phenomena. For example, it would be interesting to consider whether Prospect Theory (Kahneman & Tversky 1979) provides a superior account in some respects. Another possibility is that the account given here needs to be refined in order to account for situations in which some version of causal decision theory makes better predictions than Jeffrey's epistemic decision theory (cf. Charlow 2015). It does seem likely that it will ultimately be necessary to adopt a semantics whose predictions agree with those of a causal decision theory. However, I will not pursue this refinement here for three reasons: (i) the difference does not matter for most of our key points; (ii) the appropriate formalization of causal decision theory is still under debate; (iii) it is possible to render a causal decision theory without modifying the definitions given here, as long as we ensure that the background probabilistic model is structured appropriately (Meek & Glymour 1994, cf. Lassiter 2015c).

In any case, I certainly do not wish to commit here to the claim that the semantics used here is the final word on scalar goodness. What is really important is that any decent account of scalar goodness will have to have certain key formal features:

- $\mathcal{S}_{good}$  is rich enough to support **quantitative comparisons**;
- $\mathcal{S}_{good}$  is **intermediate**;
- $\mathcal{S}_{good}$  is **information-sensitive**.

These constraints are not specific enough to force the conclusion that  $\mathcal{S}_{good} = \mathcal{S}_{EV}$ , but they do rule out most of the proposals that have been offered in the formal semantics literature. In any case, the main interest of the expected value theory here is that it has the needed properties, and that it is spelled out in sufficient formal detail to allow us to derive and evaluate interesting predictions.

## 7.7 Applications

### 7.7.1 The Juliet scenario

In §7.4 I used the **Juliet** scenario to argue against the common assumption that comparative goodness is maximal. In this case, there was a strong intuition that (A) below was better than (C), even while (C) was better than its own negation. Assuming maximality, a formal contradiction follows immediately, since (C) is exactly as good as (A) or (B), whichever is better.

- (A) Juliet takes the drug and the Friar administers the antidote. (drug $\wedge$ antidote)

(B) Juliet takes the drug and the Friar does not administer the antidote. (**drug**  $\wedge$   $\neg$ **antidote**)

(C) Juliet does not take the drug. ( $\neg$ **drug**)

The only way to avoid the contradiction while maintaining maximality would be to somehow suppose that the domain shifts while we are evaluating these propositions: perhaps, when we evaluate *Juliet does not take the drug*, we temporarily ignore the worlds in which the best possible outcome happens (she takes the drug and the antidote is administered). Then,  $\neg$ **drug** will be temporarily treated as being exactly as good as the terrible **drug**  $\wedge$   $\neg$ **antidote**, rather than the wonderful **drug**  $\wedge$  **antidote**. But this would be a desperate maneuver, especially in light of the fact that the story specifically emphasizes that the Friar *might* administer the antidote if she takes the drug: the problem is not that we *know* he won't, but rather than we *can't safely assume* that he will.

The puzzle is resolved if we suppose that scalar goodness has the formal structure of expected value. For example, suppose that  $\mathbb{E}_V(\mathbf{drug} \wedge \mathbf{antidote})$  is 100, and  $\mathbb{E}_V(\mathbf{drug} \wedge \neg \mathbf{antidote})$  is  $-100$ .  $\mathbb{E}_V(\neg \mathbf{drug})$ , on the other hand, has the middling value of 0. Furthermore, fix the probability that the Friar will administer the antidote *if* Juliet takes the drug to some lowish value, say .1. This information allows us to calculate  $\mathbb{E}_V(\mathbf{drug})$ , the expected value of (C).

$$\begin{aligned}\mathbb{E}_V(\mathbf{drug}) &= \mathbb{E}_V((\mathbf{drug} \wedge \mathbf{antidote}) \vee (\mathbf{drug} \wedge \neg \mathbf{antidote})) \\ &= \mathbb{E}_V(\mathbf{drug} \wedge \mathbf{antidote}) \times \text{prob}(\mathbf{antidote} \mid \mathbf{drug}) \\ &\quad + \mathbb{E}_V(\mathbf{drug} \wedge \neg \mathbf{antidote}) \times \text{prob}(\neg \mathbf{antidote} \mid \mathbf{drug}) \\ &= 100 \times .1 + -100 \times .9 \\ &= -80.\end{aligned}$$

In this model of the **Juliet** scenario **drug**  $\wedge$  **antidote** is better than **no drug**, with expected values 100 and 0 respectively. This is true even while  $\neg$ **drug** is much better than **drug**—0 vs.  $-80$ . This is the intuitively correct result.

This conjunction of judgments does not lead to a contradiction for the expected-value semantics, because the calculation of expected value does not only take into account information about which outcomes are best (or worst). Instead, it weighs the goodness or badness of the various outcomes against their probability of occurring.

Implicit in this discussion is the importance of probabilistic information. In the **Juliet** scenario, the goodness of **drug** *could* be as high as that of **drug**  $\wedge$  **antidote**, as low as **drug**  $\wedge$   $\neg$ **antidote**, or anywhere in between. What controls this outcome is the conditional probability of **antidote** given **drug**.

Intuitively, this seems right. If we knew for certain that the Friar would give the antidote, then (given the other assumptions in the scenario) it would clearly be better for Juliet to take the drug than not. On this theory at hand, this is because **drug**  $\wedge$   $\neg$ **antidote** would receive zero weight in the calculation of  $\mathbb{E}_V(\mathbf{drug}) = \mathbb{E}_V((\mathbf{drug} \wedge \mathbf{antidote}) \vee (\mathbf{drug} \wedge \neg \mathbf{antidote}))$ . The judgment that taking the drug is a better choice would probably remain true if there were a 99% or 98% chance that the Friar would give the antidote conditional on this choice. But the judgment that taking the drug is better becomes increasingly uncertain as this probability declines, until—at the opposite extreme—taking the drug is clearly the worse option when the chance of getting an antidote is 20%, or 5%, or zero.

### 7.7.2 Information-sensitivity: Modified Procrastinate

A number of puzzles involving deontic concepts in the literature yield to a similar analysis. Most of these have been discussed with reference to *ought* or *should*, but they have exact parallels in the case of the semantically simpler scalar item *good*. One such case is the famous **Professor Procrastinate** example of Jackson & Pargetter (1986). The Procrastinate example has usually been discussed with reference to the item *ought*, but the key points extend also to scalar goodness, and the intuitions are perhaps cleaner.

Here is a slightly embellished “chancy” version of the story due to Cariani (2013).

Prof. Procrastinate is invited to review a book on which he is the only fully qualified specialist on the planet. Procrastinate’s notable character flaw, however, is his inability to bring projects to completion. In particular, if Procrastinate accepts the request to review the book, he *might* write it, but it is very likely (though not quite certain) that he will forget and not end up writing the review. In the eyes of the editor, and of the whole scientific community, this is the worst possible outcome. If Procrastinate declines, someone else will write the review—someone less qualified than him, but more reliable.

(7.23) strikes me as clearly correct here:

(7.23) It is better for Professor Procrastinate to accept and write the review than for him not to accept, or to accept but not write.

At the same time, since he will probably forget if he accepts, I also find (7.24) compelling:

(7.24) It is better for Professor Procrastinate to decline the invitation than to accept.

Semantically, this presents a challenge that is identical to that of the **Juliet** scenario: the only difference is that we are considering the choices at two different time slices of a single, forgetful agent, instead of those of two agents acting independently.

The pattern of judgments in (7.23) and (7.24) would be incoherent if goodness were maximal, but it is readily intelligible if goodness is an intermediate scale. This first lesson of the **Professor Procrastinate** scenario does not rely on the reader sharing the judgments I described for (7.23) and (7.24): all that is required is that this pattern of judgments is *not incoherent*. Any intermediate semantics would have this effect, not just the expected-value semantics.

I happen to think that these judgments are not only sane but eminently reasonable. From this perspective, the shift to the expected value semantics has benefits that go beyond merely providing an intermediate interval scale that permits these scenarios to be logically consistent. When combined with a reasonable choice of probability measure, the expected value semantics allows us to explain *why* these judgments are reasonable, as opposed to merely not inconsistent. Given that the forgetful Professor must make a choice about whether to accept now, with uncertainty about his own future behavior, he must weigh the relative goodness of his possible actions against the probabilities of the various outcomes, good and bad. His decision problem is exactly the same as the one that Juliet faces, except that the unreliable person whose behavior he is trying to predict is his own future self.

### 7.7.3 Modified Miners' Puzzle and conditional expected value

Some further scenarios in the literature involving information-sensitivity (of *ought*, *should*, *want*) that are readily modified to involve scalar goodness include:

- The Medicine scenario of Jackson (1991); Goble (1996): a doctor must choose to prescribe either a safe but mediocre medicine, or a risky medicine which could kill the patient or cure her completely.
- The Chicken scenario of Jackson (1985): two agents must make independent decisions, where the choice of each could be very good or very bad depending on the choice of the other.
- The Insurance scenario of Levinson (2003): an agent must decide whether to buy insurance against some catastrophic event. The best-case outcome is that the agent saves money by not buying insurance, and the event insured against does not occur. However, the gap in value between the best-case scenario and the intermediate outcome (insurance but no catastrophe) is small compared to the large gap between the intermediate outcome and the worst (catastrophe and no insurance).
- The Miners' Puzzle discussed by Kolodny & MacFarlane (2010) (who attribute it to Regan (1980)). We will discuss this scenario shortly.

Since we have already seen the general formula for dealing with these cases, I will not survey each case here. (For some of the gory details—though focusing on *ought* and *want*—see Lassiter 2011a: §6 and §7.9 below.) All of these scenarios have two general features. First, the choices involve fine-grained manipulations of probabilities and values, and intuitions are sensitive to both in ways that make good sense from the perspective of the expected-value semantics. Second, they are highly problematic for any semantics which focuses on extreme (best or worst) outcomes, or otherwise fails to incorporate probability weighting.

In this section I want to focus on the last case—the Miners' Puzzle—which is especially interesting due to the inclusion of conditional reasoning. Quoting from Kolodny & MacFarlane (2010) (example numbers have been changed):

Ten miners are trapped either in shaft *A* or in shaft *B*, but we do not know which. Flood waters threaten to flood the shafts. We have enough sandbags to block one shaft, but not both. If we block one shaft, all the water will go into the other shaft, killing any miners inside it. If we block neither shaft, both shafts will fill halfway with water, and just one miner, the lowest in the shaft, will be killed.

We take it as obvious that the outcome of our deliberation should be

(7.25) We ought to block neither shaft.

Still, in deliberating about what to do, it seems natural to accept:

(7.26) If the miners are in shaft *A*, we ought to block shaft *A*.

(7.27) If the miners are in shaft *B*, we ought to block shaft *B*.

We also accept:

(7.28) Either the miners are in shaft *A* or they are in shaft *B*.

But (7.26), (7.27), and (7.28) seem to entail

(7.29) Either we ought to block shaft *A* or we ought to block shaft *B*.

And this is incompatible with (7.25). So we have a paradox.

We'll consider in the next chapter what we can say about the original, *ought*-involving puzzle. However, most interesting here is that there the paradox can be replicated without invoking *ought*, using just sentences that are analyzed directly in terms of scalar goodness. This is useful because we now know a good bit about scalar goodness that can be used to constrain our solution.

Specifically, in this scenario (7.30) seems compelling:

(7.30) Out of the available options—block *A*, block *B*, or block neither—it is best to block neither.

These premises also seem to be correct:

(7.31) If the miners are in shaft *A*, it is best to block *A*.

(7.32) If the miners are in shaft *B*, it is best to block *B*.

(7.33) of course remains true:

(7.33) Either the miners are in shaft *A* or they are in shaft *B*.

But on face these premises would seem to entail (7.34), which is incompatible with (7.30).

(7.34) Either it is best to block *A*, or it is best to block *B*.

As in the original version, we have here an apparent contradiction. The inference to (7.34) is valid on standard assumptions (notably, unrestricted validity of modus ponens). However, this inference is in conflict with the truth of (7.30): if *A* is the best of some options  $\{A, B, C, \dots\}$ , then *A* is better than each other option. Obviously, then, neither of the disjuncts of (7.34) can be true. (Note that by “it is best” and similar locutions I always intend “it is uniquely best”. This is meant to rule out the possibility that the option in question is merely tied for best. When the latter is intended I use a plural noun, e.g., “is among the best options”.)

In their discussion, Kolodny & MacFarlane (2010) target Modus Ponens as the source of the problem: this assumption is required in the transition to (7.29) (and (7.34)), and the derivation can be blocked if we deny that Modus Ponens is valid in conditionals with deontic consequents. In addition, Kolodny & MacFarlane introduce a notion of “serious information-dependence”, which essentially means treating the deontic ordering as a two-place function of worlds and information states, where information gain can re-order worlds in arbitrary ways.

I agree that deontic conditionals reveal that modus ponens is not an unrestrictedly valid inference rule (more in a moment). Still, simply denying modus ponens is a fully satisfying solution. On

the philosophical side, Charlow (2013) points out a number of theoretical problems with the crucial notion of “serious information-dependence”. In addition, Charlow points out that the proposal succeeds in blocking the derivation of a formal contradiction, but does little to explain why the scenario in question so strongly prefers this pattern of judgments instead over some other. Most importantly, Kolodny & MacFarlane’s (2010) solution fails to illuminate one of the most semantically interesting aspects of the Miners’ Puzzle: the fact that the *best* option is one that is *guaranteed sub-optimal*, and known to be so (cf. Lassiter 2011a; Cariani, Kaufmann & Kaufmann 2013). Even though the best-case outcomes of the other choices are preferable, the option the we intuitively judge “best”—blocking neither shaft—is one that will guarantee that at least some miners die. The reason that this option is best is not that its best-case outcome is better than the best-case outcome of the other options. Rather, it is that the uncertainty inherent in the situation renders it safer than the others options.

Glossed this way, the puzzle sounds a good deal like the **Juliet** and **Professor Procrastinate** scenarios that were discussed above: all of them serve to illustrate that scalar goodness is intermediate and information-sensitive. Consider, for concreteness, the following model. Each world’s value is determined by the number of miners saved: so, if we block *A* and guess right, this has value 10, but if we guess wrong it has value 0. In the story we have no idea which shaft the Miners

---

	block <i>A</i>	block <i>B</i>	block neither
miners in <i>A</i>	10	0	9
miners in <i>B</i>	0	10	9

---

**Table 7.1**

A model for the original Miners’ Puzzle, with values (numbers of lives saved) for worlds in each possible outcome.

---

are in; this suggests  $\text{prob}(\text{miners in } A) = \text{prob}(\text{miners in } B) = .5$ . In this case, we can calculate the expected values of the various options.

- $\mathbb{E}_V(\text{block } A) = 10 \times .5 + 0 \times .5 = 5$
- $\mathbb{E}_V(\text{block } B) = 0 \times .5 + 10 \times .5 = 5$
- $\mathbb{E}_V(\text{block neither}) = 9 \times .5 + 9 \times .5 = 9$

By this reckoning the best option is indeed to block neither shaft, even though this option is guaranteed sub-optimal. This is because blocking neither optimizes the expected number of lives saved. It is neither too risky nor too risk-averse.

The next task is to explain the status of the conditional sentences (7.31) and (7.32). Like Kolodny & MacFarlane (2010), I will adopt a variant of Kratzer’s (1986) semantics for conditionals, according to which conditionals function to temporarily restrict the domain of a modal in their consequent. In the variant I have in mind, a conditional accomplishes this restriction by restricting the domain *D* of worlds which is used to calculate expected value:

$$(7.35) \quad \llbracket \text{If } \phi, \psi \rrbracket^{\mathcal{M}, w, D} = \llbracket \psi \rrbracket^{\mathcal{M}, w, D'}, \text{ where } D' = D \cap \llbracket \phi \rrbracket^{\mathcal{M}, w, D}.$$

This has the same effect on meaning that we would get by conditionalizing the relevant probability measure on the truth of the antecedent. (See chapter 3, §3.4.6 for discussion of a qualitative version of the same operation with reference to Kratzer’s original proposal.) In effect, the interpretation of conditionals conforms to the following informal rule:

*If  $\phi$ ,  $\psi$* —interpreted relative to  $V$  and  $prob$ —is equivalent to  $\psi$  interpreted relative to  $V$  and  $prob(\cdot | \phi)$ .

If the consequent  $\psi$  happens to contain an expression sensitive to expected values, then the information in the antecedent will influence the calculation of expected values for the purpose of interpreting the consequent. All worlds where the antecedent  $\phi$  is false will receive zero weight, and all others will receive weight proportional to their original probability. So, for example,

(7.36) If  $\phi$ ,  $\psi$  is better than  $\chi$

interpreted relative to  $V$  and  $prob$  is the same as

(7.37)  $\psi$  is better than  $\chi$

interpreted relative to  $V$  and  $prob(\cdot | \phi)$ , which corresponds to a claim about the relative magnitude of two reweighted averages:

$$\left( \sum_{w \in \psi} V(w) \times \begin{cases} 0 & \text{if } w \notin \phi \\ prob(\{w\} | \phi \wedge \psi) & \text{otherwise} \end{cases} \right) > \left( \sum_{w \in \chi} V(w) \times \begin{cases} 0 & \text{if } w \notin \phi \\ prob(\{w\} | \phi \wedge \chi) & \text{otherwise} \end{cases} \right)$$

This is the just a long-winded way of comparing the **conditional expected values** of the two propositions given the antecedent.

$$\mathbb{E}_V(A | B) =_{df} \sum_{w \in A} V(w) \times prob(\{w\} | A \cap B)$$

So:

(7.38) (7.36) is true iff  $\mathbb{E}_V(\psi | \phi) > \mathbb{E}_V(\chi | \phi)$ .

The effect, then, is that a deontic conditional is evaluated by temporarily updating expected values to conditional expected values given the antecedent. (7.36) means essentially: “On the supposition that  $\phi$  is true,  $\psi$  would be better than  $\chi$ ”.

On this assumption, the model just described predicts the truth of (7.31) and (7.32). If the miners are in  $A$ , the worlds in which the miners are in  $B$  are temporarily ignored, and we derive conditional expected values that are equal to the values in the first row of Table 7.1.

- $\mathbb{E}_V(\text{block A} | \text{miners in A}) = 10 \times 1 = 10$
- $\mathbb{E}_V(\text{block B} | \text{miners in A}) = 0 \times 1 = 0$
- $\mathbb{E}_V(\text{block neither} | \text{miners in A}) = 9 \times 1 = 9$



Thus, we predict that *If the miners are in A, it is best to block A* comes out true. The right prediction follows by identical reasoning for *If the miners are in B, we ought to block B*.

Importantly, though, (7.34) does not follow on this semantics: we can consistently endorse both of the conditionals and *We ought to block neither shaft*, while denying *Either we ought to block A, or we ought to block B*. This serves to illustrate an important point: Modus Ponens is already invalid in the widely used conditional semantics of Kratzer (1986) (cf. Kolodny & MacFarlane 2010; Cariani et al. 2013).<sup>5</sup> The restrictor semantics does not always permit us to conclude from  $\phi$  and *If  $\phi$ ,  $\psi$*  that  $\psi$ . The Miners' Puzzle suggests that this is a good property of the semantics.

Of course, simply denying Modus Ponens—even if just for the English item *if*, and even if just for some kinds of consequents—seems a bit crazy. Consider a modus ponens argument with a deontic expression in the consequent, such as:

- (7.39) a. If  $\phi$ ,  $\psi$  is better than  $\chi$ .  
       b.  $\phi$ .  
       c. So,  $\psi$  is better than  $\chi$ .

This is an inference pattern that just seems *obviously* right. As Kolodny & MacFarlane (2010: §4.5) note, though, we can explain why it is so compelling—while denying that it is a logically valid principle—by noting that it is “pragmatically valid” in a certain sense. In the semantics proposed here, argument (7.39) has the feature that

in every context in which the premises could appropriately be asserted or supposed,  
 it is impossible for anyone to accept the premisses without committing himself to  
 the conclusion ...  
(Stalnaker 1975: 271)

If the argument in (7.39) is pragmatically valid, then any agent who is in a position to endorse both (7.39a) and (7.39b) will also be in a position to endorse (7.39c). When we consider whether the inference is reasonable, we generally do so by imagining *having the information* that each premise is true. This procedure is reasonable, but it cannot distinguish true logical consequence from pragmatic validity.<sup>6</sup>

<sup>5</sup> Yalcin (2012b: §3) questions whether the labels “modus ponens” and “modus tollens” should be used to describe the relevant argument forms if we adopt a restrictor semantics. After all, this theory denies that conditionals have the logical form of two-place sentential connectives. In the current context, I am simply interested in whether English sentences with the surface syntactic form *If  $\phi$  then  $\psi$*  validate modus ponens arguments in the usual, classical sense of “validate”—i.e., if the truth of  $\phi$  guarantees the truth of  $\psi$  on the assumption that the conditional holds. The terminological issue of whether there is some reason to use the term “modus ponens” differently is beside the point. Note in addition that it does not matter here whether there is some other notion of consequence that one could employ in evaluating modus ponens arguments—e.g., whether substituting Yalcin’s “informational consequence” for classical consequence would allow us to validate modus ponens (cf. Yalcin 2012a, fn. 14; Bledin 2015). As we will see in a moment, this move would indeed have the intended effect—but this means validating not modus ponens as traditionally conceived, but a different principle which results from using a different notion of consequence that amounts to interpreting the premisses as if they were modalized (cf. Schulz 2010). I see no reason to think that there is a single “correct” notion of consequence that should be employed in such arguments. There are many interesting concepts of consequence and inference that are likely to be useful for different purposes. The interpretation that we are currently considering merely happens to have a certain amount of historical priority, which renders it the default reading of “consequence” in such discussions.

<sup>6</sup> My “pragmatic validity” is inspired by what Stalnaker calls “reasonable inference”. However, I will avoid this

The semantics given here renders (7.39) pragmatically valid, in some suitably fleshed-out sense. There are many ways to render this idea; here is one.<sup>7</sup> Suppose that we are evaluating information- and value-sensitive sentences relative to probability measure *prob* and value function *V*. Then, as noted above, (7.39a) is true if and only if  $\mathbb{E}_V(\psi | \phi) > \mathbb{E}_V(\chi | \phi)$ . Now, suppose that  $\phi$  is not just *actually* true, but believed to be true with maximal confidence—i.e.,  $\text{prob}(\phi) = 1$  in the relevant *prob*. In this case, the conditional expected value *given*  $\phi$  is the same as the expected value simpliciter—and so the truth-condition of (7.39a),  $\mathbb{E}_V(\psi | \phi) > \mathbb{E}_V(\chi | \phi)$ , is equivalent to the condition that  $\mathbb{E}_V(\psi) > \mathbb{E}_V(\chi)$ . But this is just the interpretation of (7.39c), the conclusion. More generally, anyone whose state of information and value is well-represented by (7.39a), and who assigns probability 1 to (7.39b), is also in a state of information and value that is well-described by (7.39c).

The argument also goes through if we read it in a dynamic frame of mind. Suppose that (7.39a) is known and then that (7.39b) is *learned*. On the usual Bayesian assumption that learning  $\phi$  leads a rational agent to update her prior distribution  $\text{prob}(\cdot)$  to a posterior distribution  $\text{prob}(\cdot | \phi)$ , the result will be a posterior where  $\phi$  has probability 1. Then, the fact that the pair  $\langle \text{prob}, V \rangle$  makes  $\mathbb{E}_V(\psi | \phi) > \mathbb{E}_V(\chi | \phi)$  true implies that  $\langle \text{prob}(\cdot | \phi), V \rangle$  makes  $\mathbb{E}_V(\psi) > \mathbb{E}_V(\chi)$  true. But then (7.39c) is true relative to  $\langle \text{prob}(\cdot | \phi), V \rangle$ .

Pragmatic validity does not, however, imply logical validity. Counter-examples with deontic expressions in particular—including forms of *good*—are easily generated in cases in which the antecedent is *actually* true, but no one knows it to be true. For a fanciful case that nevertheless makes the point clearly, consider conditional reasoning about events that we could not, in principle, have any information about.

- (7.40)    a. If Alpha Centauri A went supernova 4 years ago—so that Earth will be destroyed in a few months—it is better for Bill to spend all his money now than to save for retirement.  
           b. Alpha Centauri A went supernova 4 years ago.  
           c. So, it is better for Bill to spend all his money now than to save for retirement.

This seems like valid reasoning, and it is—in the sense that, if an agent knows that the premises are true, her information will also support the conclusion. However, since Alpha Centauri is 4.37 light-years away, the information that it went supernova 4 years ago is not in our light cone: according to current physics, at least, no one could actually possess this information. A more fleshed-out version of the argument should thus look like this:

- (7.41)    a. If Alpha Centauri A went supernova 4 years ago—so that Earth will be destroyed in a few months—it is better for Bill to spend all his money now than to save for retirement.  
           b. In actuality, Alpha Centauri A went supernova 4 years ago (but no one on earth knows about this event yet).  
           c. So, it is better for Bill to spend all his money now than to save for retirement.

---

terminology because the non-technical meaning of the term is much weaker than Stalnaker's intended sense.

<sup>7</sup> This informal description is, I think, as precise as it is necessary to be here. For more general and formal versions, see Yalcin's (2007) "informational consequence" and Kolodny & MacFarlane's (2010) "quasi-validity". The dynamically-minded version described below corresponds roughly to what Charlow (2013: §3.2) calls "*ersatz* modus ponens", or to the notion of "dynamic validity" discussed by Malte Willer in various places, e.g., Willer 2010.

Only a godlike being could know that Bill's life will end violently in 0.37 years. Even though this may determine what Bill should do relative to the information of such a being, it does not decide what he should do given **his** information. The conclusion is valid only from a god's-eye perspective.

The semantic package that I have argued for predicts—correctly, I believe—that modus ponens is not **unrestrictedly** valid. When there is a modal in the conclusion of a conditional, the other premises and the conclusion must have a certain modal status as well. Facts about what the world is **actually** like are not always relevant to whether the conclusion of a modus ponens argument follows: what we need to know is whether the information is available in the reasoning or deliberation of a deliberating agent.

This complexity serves to illustrate the importance of holding fixed a state of information (e.g., a probability function) when reasoning about *good*, *ought*, and the like. If we allow it to shift, we will derive apparent paradoxes.

One such apparent paradox appears already in the Miners' Puzzle. As Cariani et al. (2013) point out, it is also possible to understand the scenario from a god's-eye perspective. Consider the following line of reasoning:

In the original story, the Miners are either in A or in B. Given this, it really **is** either best to block A, or best to block B—the best one to block is whichever shaft they are in. Unfortunately, we don't know which one this is!

The god's-eye perspective favors the truth of the conclusion that Kolodny & MacFarlane (2010) reject, (7.29) (and the corresponding *good*-premise (7.34)).

The only way to make this judgment coherent with our original judgment, that it is best to block neither, is to suppose that *best* is being interpreted differently. On this reading, it is (7.30)—*It is best to block neither*—that has to go: blocking neither simply is not the best option, and it is not what we ought to do. This is not a semantic paradox, but a reflection of the fact that the goodness ordering is calculated with reference to a state of information—specifically, in the account under consideration, to a probability measure. The choice of measure is sometimes fully determined by compositional semantics or contextual information, but in some contexts it must be inferred from contextual clues whose information is under discussion.

It is not immediately obvious how to capture the omniscient, god's-eye interpretation compositionally. One possibility, suggested by Wedgwood (2015), is to use the trivial probability measure that assigns probability 1 to all true propositions and 0 to all false propositions. This can be made to work compositionally if we recall that the identify of this function must depend on which world is actual: the god's eye-perspective is a world-relative probability measure.

A sentence  $\phi$  is interpreted from the **god's-eye perspective** just in case its interpretation is relativized to a probability measure  $P_w$  which covaries with the world of evaluation  $w$ , and where, for any  $A \subseteq \mathcal{P}(W)$ ,

$$P_w(A) = \begin{cases} 1 & \text{if } w \in A \\ 0 & \text{otherwise.} \end{cases}$$

(This is unproblematic from the perspective of our broader semantics, of course: as we discussed

in chapter 4, scales are generally world-relative, and there is no clear reason to expect probability scales to differ in this respect.)<sup>8</sup>

In the Miners scenario, the worlds are partitioned into two sets: those where the miners are in *A*, and those where the miners are in *B*. Since we have adopted the god's-eye perspective, any world where the miners are actually in *A* will be one in which *The miners are in A* has probability 1. As a result, the expected value of blocking *A* is 10, and the expected value of blocking *B* is 0. In a world *w* where the miners are in *A*, the best thing to do from this perspective is to block *A*. Similarly, in

---

	block <i>A</i>	block <i>B</i>	block neither
miners in <i>A</i>	10	0	9
miners in <i>B</i>	0	10	9

---

**Table 7.2** The Miners' Puzzle from the god's-eye perspective: the expected value of each action at a world is simply its value given the actual location of the miners at that world.

---

any world where the miners are in *B*, the best thing to do from the god's-eye perspective is to block *B*.

The effect, then is that an assertion of *Either it is best to block A, or it is best to block B* will be true, relative to the specified value function, just in case the actual world is either one in which the miners are in *A*, or one in which the miners are in *B*. The presentation of the scenario stipulated that this holds, and so we predict that (7.34) is true from the god's-eye perspective.

The analysis is more complex when we consider the conditionals. Following the reasoning above, an assertion of *If the miners are in A, it is best to block A* should be true from the god's-eye perspective if and only if *It is best to block A* is true when we restrict the domain—i.e., when we conditionalize the probability measure on the truth of the antecedent.

*If in A, it is best to block A* is true relative to *V* and *prob*

iff

*It is best to block A* is true relative to *V* and  $\text{prob}(\cdot \mid \text{in } A)$ .

This will of course be true in the worlds where the miners are in fact in *A*. But what about the worlds where they are in *B*? Since *prob* assigns zero probability to all sets that don't contain the actual world, attempting to conditionalize on *The miners are in A* when they are actually in *B* means conditionalizing on a proposition with zero probability. On standard assumptions, this means that conditionalization on *The miners are in A* is undefined when they are in *B*.

What we would like is for *If the miners are in A, it is best to block A* to come out true from the god's-eye perspective in all possible worlds, and not just those where the miners are in fact in *A*.

---

<sup>8</sup> A slightly weaker option would be to consider the probability measure on a set of facts about the actual world. This would have the effect that some, but not necessarily all, propositions that are true of the actual world receive probability 1. This might be useful if there are intermediate cases where we want the reasoning to include consideration of some, but not all, unknown facts (Cariani et al. 2013). However, it would encounter the same technical issue noted for Wedgwood's suggestion below.

While I am unsure what the best possible solution is, there are several ways to massage the issue that will achieve the right result for the present problem. For example, we could modify the evaluation rule for conditionals so that, when an undefined conditional probability statement is encountered, the evaluation function will consider the most similar worlds at which the statement is defined.

Whatever the best resolution is to the technical problem just mentioned, a nice property of the expected value semantics for *good* is that it offers a way to unify the ordinary (“deliberative”) and god’s-eye (“objective”) interpretations of *good*. All that varies between them, on this account, is the way that the relevant probability measure is determined.

## 7.8 Varying probability and values in the Miners’ Puzzle

One point that favors the expected-value semantics over salient competitors is that only the former makes specific and correct predictions about the way that shifting probabilities and outcomes influences our judgments about these scenarios. While the original scenario was maximally uninformative about the miners’ location, we can also consider variants in which we can make a better-than-chance guess.

Ten miners are trapped either in shaft *A* or in shaft *B*. We do not know which, but we do know that they spend 99% of their time in shaft *B*. Flood waters threaten to flood the shafts ...

In this case, there is a strong temptation to conclude that

(7.42) It is best to block shaft *B*.

After all, it’s much more likely that the miners are in *B*, and so if we block *B* we have a very good chance of saving them all.

If comparative goodness were maximal, this manipulation should make no difference. The question of which outcome is the *best possible* is not sensitive to non-categorical shifts in probability, but only to factors which exclude some worlds from consideration entirely. A mirror image of this same problem arises for a semantics which ties the goodness of a proposition to the goodness of the **lowest**-ranked worlds in the proposition. (This idea is inspired by the proposal in Cariani et al. 2013, though that paper focuses on *ought* and does not explicitly endorse a maximin semantics for comparative goodness.) An account along these lines would make the right predictions for the uninformed case, but would not be sufficiently sensitive to shifts in probability: moving from 50% to 99% in favor of *B* is not enough to influence such pessimistic deliberations. A maximin semantics would continue to focus on the small but non-zero chance that the miners are in *A* (with terrible consequences if we block *B*), concluding wrongly that blocking *A* and blocking *B* are equally bad because both lead to the worst possible outcome with non-zero probability.

The expected-value semantics, on the other hand, gracefully incorporates information about shifts in probability to derive reasonable predictions. Specifically, in the revised scenario we derive the following expected values:

- $\mathbb{E}_V(\text{block } A) = 10 \times .01 + 0 \times .99 = .1$

- $\mathbb{E}_V(\text{block } B) = 0 \times .95 + 10 \times .05 = 9.9$
- $\mathbb{E}_V(\text{block neither}) = 9 \times .05 + 9 \times .95 = 9$

In this case, we correctly derive the prediction that it is best to block *B*.

Simultaneous manipulations of probabilities and the relative value of outcomes are also handled elegantly by the expected-value theory. Consider for instance the following elaboration:

Ten miners are trapped either in shaft *A* or in shaft *B*. We do not know which, but we do know that they spend a majority (75%) of their time in shaft *B*.

However, shaft *A* is much more dangerous than *B* in a flood. If we choose to block *A* and they're really in *B*, two miners will die. However, if we choose to block *B* and they're really in *A*, all of the miners will die. Doing nothing is also more dangerous if they are in *A*, with 7 dying instead of 1.

Here is a summary of the new, more complicated story:

	probability	value: block <i>A</i>	value: block <i>B</i>	value: block neither
miners in <i>A</i>	.25	10	0	3
miners in <i>B</i>	.75	8	10	9

**Table 7.3** A model for the updated Miners' Puzzle, with unequal probabilities and values.

This is a difficult dilemma. I would personally be drawn to the option of blocking *A*, even though it's less likely that they are there—because there is a non-negligible chance that they are, and sandbagging *A* has a much greater effect in the case that they are in there. Indeed, the expected-value semantics predicts the following.

- $\mathbb{E}_V(\text{block } A) = .25 \times 10 + .75 \times 8 = 8.5$
- $\mathbb{E}_V(\text{block } B) = .25 \times 0 + .75 \times 10 = 7.5$
- $\mathbb{E}_V(\text{block neither}) = .25 \times 3 + .75 \times 9 = 7.5$

Here, the best option seems to be to block *A*.

(Of course, the prediction that blocking *A* should be the clear winner in Table 7.3 follows only on the assumption that the value function is sensitive to the number of lives saved, and nothing else. This is not part of the semantics, and other possibilities abound: for example, we might place some positive or negative value on certain kinds of choices, as people commonly do in Trolley problems probing intuitions around active vs. passive harm. See Lassiter 2015c: §4.1 for further discussion of this point.)

Consider, on the other hand, the predictions of an optimistic semantics in the Lewis/Kratzer style, and of a pessimistic semantics in the mold of Cariani et al. The most optimistic scenario is

one in which we pick a shaft and happen to guess right. So, the best options are to block *A* and to block *B*, and these are equally ranked.

The pessimistic semantics would get the right outcome for the variant in Table 7.3: maximizing the worst-case outcome of our choice means blocking *A*, which is reasonable. However, this judgment is fragile in a way that the pessimistic semantics is not. Suppose we change the scenario so that blocking neither has a much better expected outcome—7 saved instead of 3 when they are in *A*. Intuitively, this renders doing nothing a much more appealing option. However, in the pessimistic semantics it would have no effect on the comparison, since the worst-case outcome of blocking neither is still worse than the worst-case outcome of blocking *A*.

	probability	value: block <i>A</i>	value: block <i>B</i>	value: block neither
miners in <i>A</i>	.25	10	0	7
miners in <i>B</i>	.75	8	10	9

**Table 7.4** Another iteration, with a more optimistic projection for the do-nothing option when the miners are in *A*.

In contrast, this manipulation has an important effect in the expected-value semantics: blocking neither shaft is now tied with blocking *A*. (This may account for the ambivalence that one would likely feel about this choice.)

- $\mathbb{E}_V(\text{block } \mathbf{A}) = .25 \times 10 + .75 \times 8 = 8.5$
- $\mathbb{E}_V(\text{block } \mathbf{B}) = .25 \times 0 + .75 \times 10 = 7.5$
- $\mathbb{E}_V(\text{block neither}) = .25 \times 7 + .75 \times 9 = 8.5$

We could continue playing this game, but I trust that the point is clear: a hyper-optimistic semantics and a hyper-pessimistic semantics will both generate fairly coarse predictions, and are not sufficiently sensitive to subtle manipulations in the space of probabilities and of outcomes. The expected-value semantics, in contrast, makes smart use of all the information we have about the relative values and probabilities of outcomes. This leads to predictions that match much more closely the subtle ways that our intuitions of the relative goodness of various actions are influenced by these manipulations.

These considerations are of course not a conclusive demonstration of the correctness of the expected-value semantics for *good*. However, they do show that the correct semantics must integrate information about values and probabilities in some way, and cannot simply throw away information about outcomes that are neither maximal nor minimal. It could easily be that the correct theory of scalar goodness will be still more complex, integrating information about further factors—for example, in the style of Prospect Theory (Kahneman & Tversky 1979). However, I am quite confident that the correct theory not turn out to be *less* so: the expected-value semantics provides a lower bound on what a theory of goodness must be able to do.

## 7.9 Positive form *good*

The discussion up to now has been framed almost entirely in terms of comparative goodness and the superlative *best*, which is readily definable from it. So, the examples we have considered have all had the abstract form in (7.43).

- (7.43) a.  $\phi$  is as good as  $\psi$ .  
b.  $\phi$  is better than  $\psi$ .  
c. Out of some set of alternatives,  $\phi$  is the best.

In this section I will consider some ways of relating positive-form goodness to scalar goodness. Here are some salient options:

- (7.44) a.  $\phi$  is *good* is true iff  $\phi$  is (significantly) better than  $\neg\phi$ .  
b.  $\phi$  is *good* is true iff  $\phi$  is (significantly) better than every other member of some set of contextual alternatives.  
c.  $\phi$  is *good* is true iff  $\phi$  is (significantly) better than the union of some set of contextual alternatives.

I will assume that the “(significantly)” caveat applies throughout, as it does generally in the interpretation of the class of adjectives that *good* falls into. That is, as e.g. Fara (2000); Kennedy (2007) discuss, relative adjectives are vague, resisting such crisp interpretations, and someone who is 1 mm taller than average for some reference class will not generally be judged “tall”. It would be very surprising if *good* failed to be vague in a similar way: if  $\phi$  just barely meets the relevant criterion, it will not clearly count as “good”. Requiring a “significant” difference is not meant to be an explanation of anything, but only a summary of the fact that the vagueness of these expressions places limits on the precision with which we can state their truth-conditions.

Many authors have discussed the possibility that *good* means “better than not” (Brogan 1919; Von Wright 1972; Hansson 1990 among others). While this definition is attractively simple, it has various problems. One common objection to this definition is that it does not leave enough room for indifferent propositions (Chisholm & Sosa 1966). We could allow non-trivial space for indifferent propositions by adding the “significantly” proviso discussed above. Even supposing that this is sufficient, though, the “better than not” definition encounters an insuperable problem: it does not leave room for the alternative-sensitivity of *good*, including its sensitivity to prosodic focus.

We can see the effects of focus on the interpretation of positive-form *good* by considering the following scenario, which is modified from Dretske 1972.

Clyde’s dying father wishes desperately for him to get married, but does not care to whom. All else being equal, it would be a great moral good to grant his father’s wish so that he may die happy. Ambivalent about marriage but eager to fulfill this filial duty, Clyde asks his friend Bertha to marry him, and she agrees.

Consider in this context the examples in (7.45), where capitalization indicates prosodic focus.

- (7.45) a. It is good that Clyde MARRIED Bertha.



b. It is good that Clyde married BERTHA.

(7.45a) seems to be true here, but (7.45b) is false: any suitable marriage partner would do. Bertha has no special status in this regard.

If  $\phi$  is *good* means “ $\phi$  is better than  $\neg\phi$ ”, it is difficult to see how (7.45a) and (7.45b) could differ in truth-value. Both would mean simply “Marrying Bertha was better than not marrying Bertha”. However, a theory in which the meaning of *good* is sensitive to contextual alternatives—as in (7.44b) or (7.44c)—is able to distinguish them appropriately, in the following way.

Focus is generally thought to invoke sets of alternative propositions, corresponding to the different possible answers to a contextual Question Under Discussion (Dretske 1972; Rooth 1992; Roberts 1996; Beaver & Clark 2008). Thus, (7.45a) and (7.45b) differ because they invoke different alternative sets as a consequence of the placement of focus.

(7.46) It is good that Clyde MARRIED Bertha. (Alternatives: {Clyde *marries* Bertha, Clyde *dates* Bertha, Clyde *ignores* Bertha, ... })

(7.47) It is good that Clyde married BERTHA. (Alternatives: {Clyde *marries* **Bertha**, Clyde *marries* **David**, Clyde *marries* **Susan**, ... })

Both alternative-sensitive interpretations under consideration make reasonable predictions for this example. If *good* means *better than each alternative*, then—assuming the expected-value semantics discussed above—(7.46) comes out as (7.48), which is presumably true in this scenario.

(7.48) For all relevant alternatives  $R$  to marriage, it is better that Clyde married Bertha than it would have been if Clyde had stood in relation  $R$  to Bertha.

(7.47), on the other hand, is interpreted as the clearly false (7.49).

(7.49) For all relevant alternatives  $x$  to Bertha, it is better that Clyde married Bertha than it would have been if Clyde had married  $x$ .

The union interpretation (7.44c) makes slightly different predictions, which are also reasonable here. The union of a set of alternatives of the form {Clyde *marries* Bertha, Clyde *dates* Bertha, Clyde *ignores* Bertha, ... } comes down to *Clyde stands in some relevant relation to Bertha*. So, the predicted interpretation is

(7.50) It is better that Clyde married Bertha than it would have been if Clyde had stood in any old relation to Bertha.

An intermediate theory of scalar goodness such as the expected-value semantics predicts a sensible interpretation here. Marrying Bertha is better than standing in any old relation to her, because of the considerable chance that the relation chosen would be worse than marriage (e.g., ignoring her).

In contrast, if goodness were maximal this interpretation reduces to the superlative semantics. This is because maximality leads immediately to the prediction

$$\phi >_{good} \bigcup \Psi \Leftrightarrow \forall \psi \in \Psi : \phi >_{good} \psi,$$

via the fact that  $\bigcup \Psi$  is exactly as good as the best  $\psi \in \Psi$ . (Perhaps this fact accounts for the failure to consider the union interpretation in previous literature, where maximality is often assumed. It is only when we reject maximality that the predictions of (7.44b) and (7.44c) can diverge.)

Similarly, the union of a set of alternatives of the form {*Clyde marries Bertha*, *Clyde marries David*, *Clyde marries Susan*, ... } comes down to *Clyde marries some relevant person*. The union semantics thus predicts the interpretation

- (7.51) It is better that Clyde married Bertha than it would have been if Clyde had just married any old person (e.g., by random choice).

In the story at hand, this is false, since the key to his father's happiness is simply that Clyde be married.

While both accounts make acceptable predictions in this case, the union semantics strikes me as preferable on both theoretical and empirical grounds. There are two relevant theoretical points. First, it seems odd that English would have both expressions—*good* and *best*—if they were synonymous. More generally, it would be surprising if these expressions were equivalent given that this is not generally the case among adjectives: *angry* does not mean *angriest*, *sandy* does not mean *sandiest*, and so forth.

A second, related point is that it is theoretically desirable to unify the alternative-sensitivity of *good* with the sensitivity of other relative adjectives—*tall*, *long*, *happy*, and the like—to reference classes. But in the latter case, it is not at all plausible that the correct interpretation could require exceeding *all* members of the reference class. In the case of *tall*, for example, a semantics analogous to the all-alternatives interpretation of *good* would interpret *tall for a basketball player* as meaning *taller than all other basketball players*. But this is clearly too strong: the right interpretation is something like *significantly taller than average for basketball players*, which is akin to the union semantics for *good* in (7.44c).

Empirically, we can look to cases where maximality may be too strong—that is, where the alternative in question is clearly “above average” in the relevant sense, but not the absolute best. That this is a conceptual possibility is suggested by naturalistic examples such as the following.

- (7.52) With much national pride and a bit of homesickness, about 300 soldiers in the British-dominated headquarters of the NATO force in Afghanistan were glued to a live feed of England's opening World Cup win against Paraguay. ... Lance Corporal Glen Steerment said he was “pretty gutted” not to be in the stands in Frankfurt. “It is good that it's on here but it would be better if we were at home with a few beers,” he said.

Lance Corporal Steerment's statement suggests an analysis along the following lines:

- (7.53) **ALT**, the alternatives under consideration, are:

- a. The troops do not watch the World Cup match.
- b. The troops watch the World Cup match in Afghanistan.
- c. The troops watch the World Cup match in Europe while drinking beer.

- (7.54) Alternative (7.53b) is better than  $\cup$ **ALT**, because the latter includes the horrible (7.53a), where they do not watch the game.

- (7.55) However, alternative (7.53c) is still better than (7.53b).

(No word yet on whether watching with beer in Afghanistan would be better or worse than watching without beer in Europe.)

What is most interesting about this example is that, if the alternative set is fixed in advance, it would be self-contradictory on the “*good = best of the alternatives*” semantics. The example suggests that one alternative can be good even while there is another, better alternative, as long as there are other alternatives that are even worse (and sufficiently likely). This is readily possible if we combine an expected-value semantics with the union interpretation of the positive form, though, and it is supported by general theoretical background on the interpretation of relative adjectives in the positive form.

An alternative analysis is available that is compatible with the equation “*good = best of the alternatives*”, though. Example (7.52) could be analyzed as involving a mid-utterance context change. As the account would go, when “It is good that it’s on here” is evaluated, the alternative set originally contains only (7.53a) and (7.53b). That is, the Lance Corporal is initially assuming that watching the game at home (with beer) is not a live option, but later in the utterance, he adds (7.53c) to the relevant alternatives.

This analysis would be able to get the facts right, and it is not implausible in the case at hand. However, when combined with general considerations about domain-shifting it makes a problematic prediction. In general, pragmatic pressures seem to favor domain expansion over domain contraction: it is more natural to add to the possibilities under consideration than to ignore possibilities already under consideration. (Compare von Stechow 2001; Moss 2012.) As a result, the order in which the alternatives are added should matter for this example: once the possibility of watching in Europe is mentioned, it should be difficult to ignore it in the immediately following discourse. Consider in this light a modified version of the example:

(7.56) “It would be better if we were at home with a few beers, but it is good that it’s on here,” he said.

According to my intuition, at least, (7.56) is perfectly natural. But if domain contraction is markedly less natural than domain expansion, the modified example in (7.56) should be correspondingly less natural than the original in (7.52). This is because the possibility of watching in Europe, when it is mentioned first, should continue to be relevant during the evaluation of “it is good that it’s on here”; this should render the latter statement false since it is not the best of the alternatives. If *good* means *best of the alternatives*, then, there should be at least some tendency to read (7.56) as a contradiction and therefore infelicitous. I do not detect any such tendency.

Theoretical and empirical considerations favor the union analysis of positive-form *good* over the superlative and “better than not” treatments, then. This is good news because it relieves us of the need to treat *good* as being unique among relative adjectives in the synonymy of its positive and superlative forms, and in the way its meaning relates to its reference class. The union semantics allows us to maintain a unified semantics for the positive form of relative adjectives.

Note in addition that the intermediacy of the expected-value semantics plays a critical role in making this unification possible. If goodness were Maximal,  $A >_{good} \cup\{B, C\}$  would imply  $A >_{good} B$  and  $A >_{good} C$ . The union semantics for *good* would collapse into the superlative semantics. (Recall, as we noted in chapter 2, that we do not know of any other plausible candidates for maximality among scalar adjectives. If there were any, their positive and superlative forms would presumably be synonymous as well.) Once we admit the possibility that comparative goodness is intermediate, new ways of analyzing monadic deontic concepts become available as well—not only positive form

*good* but, as we will now see, the more popular items *ought* and *should* as well.

## CHAPTER 8

### *Ought and should*

Our study of epistemic vocabulary in chapters 4-6 was unusual in using scalar adjectives as a starting point. In the deontic case as well, we started with a scalar adjective, *good*, even though this item is much less studied than the verbal exponents of deontic modality—*may*, *must*, and especially *should* and *ought*. However, there is a crucial difference: while the linguistic expression *good* has been neglected, there is a long tradition of analyzing the meanings of the deontic modals, including the auxiliary verbs, as expressions of a fundamentally scalar concept.

The chapter begins by problematizing efforts to define *ought/should* directly in terms of comparative goodness (in my own earlier work as well as others'). Next, I discuss semantic connections between *ought*, *should*, and *better* and motivate a different set of constraints on the semantic relationship among these items. A number of empirical advantages over the classical semantics are noted as well. Finally, I will discuss the challenge posed by the grammatical gradability of *ought* and *should*, arguing that a complete theory of the semantics of *ought* and *should* must be able to make sense of their status as scalar verbs.

Two caveats for the reader. First, this chapter cannot be read in isolation from the rest of the book. It presupposes a general understanding of measurement theory (chapter 2), and the detailed theory of scalar goodness developed in chapter 7, including several formal definitions given there and the arguments for identifying scalar goodness with expected value. Readers who have not yet read these chapters would do well to return to them before continuing.

Second, I will not attempt to formulate a complete possible-worlds semantic theory for *ought* here. Instead, in the spirit of the earlier investigations of *likely/probable*, *certain*, and *good*, I will focus on providing empirical motivation for certain constraints on what a full semantic theory should look like. The result—a list of constraints that I will sometimes call a “skeletal theory”—is not the familiar form for a theory of *ought*, but I hope that it can be seen to have theoretical value nonetheless. Among other things, this investigative strategy may help to focus on the empirical content of theories of *ought/should*, abstracting away from representational questions about the format in which these theories are written down.

I will argue that the final theory must conform to several axioms which relate *ought* to *good*, and which rule out a variety of well-known theories of *ought/should*. These axioms will be treated initially as stipulations designed to account for empirical phenomena. However, at the end of the chapter I will argue that *ought* and *should* are gradable verbs, and show that a plausible scalar semantics may be able to derive several of these axioms from structural features of the scale.

#### 8.1 Basic connections among *ought*, *should*, and *better*

*Ought* and *should* seem to be truth-conditionally equivalent. For instance, both of the following are pragmatically bizarre.

- (8.1) a. You ought to leave, though I'm unsure if you should.
- b. You should leave, though I'm unsure if you ought to.

Plausibly, this is because in each case the second clause expresses uncertainty about the truth of the assertion in the first. But if this is true, they the two items are at least contextually equivalent. I will assume something stronger: they they are truth-conditionally equivalent in all contexts.

Note, however, that *ought* and *should* are not equivalent in terms of their patterns of usage and probable stylistic import. In the 100 million word British National Corpus, *should* occurs 107,813 times, or about 19 times as often as *ought* (5770; Davies 2004-). The situation is similar in the 450 million word Corpus of Contemporary American English, where *should* is about 15 times as frequent (Davies 2008-). *Should*—while much less discussed in the linguistic and philosophical literature on deontic modals—appears to be the preferred expression of weak obligation in modern English.

*Ought/should* and *better* are clearly semantically related. For example, both of the following sentences express incoherent claims as long as the relevant source of value is held fixed.

- (8.2) a. You ought to leave, but it's better for you to stay than to leave.  
b. You should leave, but it's better for you to stay than to leave.

(Naturally, the sentences in (8.2) are coherent if *ought/should* and *better* are used to talk about different kinds of value: “You ought to leave (morally), but it's better to stay (given your goal of annoying people)”. Here and throughout the chapter, I will put aside the possibility of such shifts, since they have the general feature of invalidating the semantic connections that are our primary interest here.)

Equally, it would be bizarre to hold simultaneously that *ought/should*  $\phi$  is true, and that it is indifferent whether  $\phi$  or  $\neg\phi$  holds.

- (8.3) a. You ought to leave, but it's just as good if you stay.  
b. You should leave, but it's just as good if you stay.

These observations are explained if we suppose that the following principle is generally valid:

- (8.4) ***Ought/should*, first principle:** If  $\phi$  and  $\neg\phi$  are both relevant options and *ought/should*  $\phi$  is true, then  $\phi$  is better than  $\neg\phi$ .

While this principle is fairly weak, it establishes an important fact that will guide the rest of our discussion: a theory of the semantics of *ought/should* cannot be formulated in isolation from of a theory of scalar goodness. Each will place constraints on the other, and it is possible that facts about one domain can be used to show that an otherwise plausible theory of the other domain is incorrect.

## 8.2 Defining *ought* and *should* in terms of *better*?

Given that *should* and *ought* are semantically related to *better*, an obvious line of attack is to simply define one concept in terms of the other. In fact there is a long tradition of taking betterness to be the more basic concept, and that it can be used to give definitions of *ought* and *should*. This is a key assumption of classical utilitarianism, for example, according to which *ought*( $\phi$ ) indicates that  $\phi$  is associated with the **best** outcomes—according to some antecedently given concept of goodness (Mill 1863).

More recently, this style of analysis has been proposed in the semantics literature in at least three guises. (These are, non-coincidentally, precisely the same analyses that we considered for the positive form of *good* in the previous chapter. However, their plausibility is rather different, I will suggest.) First, we could define *ought/should* as a predicate that holds of a proposition whenever it is better than its own negation (e.g., Lewis 1973: §5).<sup>1</sup>

(8.5) *ought*( $\phi$ ) is true iff  $\phi$  is (significantly) better than  $\neg\phi$ .

Second, we could define *ought/should*  $\phi$  as holding whenever  $\phi$  is the best of some contextually determined set of alternatives (e.g., Sloman 1970; Goble 1996).

(8.6) *ought/should*  $\phi$  is true iff  $\phi$  is (significantly) better than every other element of **ALT**( $\phi$ ), the set of contextually relevant alternatives to  $\phi$ .

Note that, if the alternative set happens to be  $\{\phi, \neg\phi\}$ , then the second semantics reduces to the first.

Third, we might consider a variant according to which *ought/should* compares a proposition to the *union* of an alternative set (Lassiter 2011a).

(8.7) *ought/should*  $\phi$  is true iff  $\phi$  is (significantly) better than  $\cup \mathbf{ALT}(\phi)$ .

As Lassiter (2011a) observes, if scalar goodness is expected value then this proposal also reduces to the first when the alternative set happens to be  $\{\phi, \neg\phi\}$  or any other set whose union is  $W$ . This is because it is a formal property of expected value that  $\mathbb{E}_V(\phi) > \mathbb{E}_V(\neg\phi)$  holds if and only if  $\mathbb{E}_V(\phi)$  exceeds the expected value of a tautology.

I once devoted considerable effort to formulating and defending a definition of *ought/should* in terms of comparative goodness, specifically (8.7) (Lassiter 2011a: §6). This analysis had a number of attractive features, including providing a simple treatment of phenomena that were problematic for the classical semantics, such as the grammatical gradability of *ought* and *should* and their information-sensitivity. However, I am now inclined to think that this style of analysis is wrongheaded. It is surely correct that the truth of *ought/should*  $\phi$  places quite severe restrictions on  $\phi$ 's degree of goodness. However, I will argue here that goodness facts in themselves are not sufficient to determine whether something ought to be the case. The main empirical argument for this conclusion involves the possibility of action beyond the call of *ought*, or what is known in the literature as “supererogation”. As I will show, each of the three ways of deriving *ought* from *good* that we are considering makes problematic predictions here.

### 8.2.1 Supererogation

Suppose that you ought to visit an ailing friend. Suppose also that it would be even better, when you visit, to cook your friend dinner. Does it then follow that you ought to visit *and* cook your friend dinner? No doubt in some cases you should, but this is not a semantic fact: it is clearly possible that the following could hold as well.

---

<sup>1</sup> Note that, if we assume maximality and connectedness, the “better than not” semantics is equivalent to a semantics in the classical style, where *ought* means “true in all of the best worlds”. This is what Lewis seems to intend, but the definition itself is noncommittal in this respect, with different consequences depending on what is assumed about the formal properties of comparative goodness.

- (8.8) a. You ought to visit your friend.  
 b. Visiting and cooking dinner is better than visiting and not cooking dinner.  
 c. However, cooking dinner is strictly optional: it's not the case that you ought to visit *and* cook dinner.

This indicates that the argument in (8.9) is not valid: the truth of an *ought*-claim should not exclude the possibility that refinements of the prejacent could be even better than the bare prejacent.

- (8.9) a.  $\phi$  ought to be the case.  
 b.  $\phi \wedge \psi$  is better than  $\phi \wedge \neg\psi$ .  
 c. So,  $\phi \wedge \psi$  ought to be the case.

The problem of **supererogation**, in a nutshell, is to interpret *ought* and *better* in a way that captures their relationship without validating this inference (Chisholm 1963b).

The problem is perhaps even more vivid with actions that we are inclined to consider heroic. It could easily be that you ought to give a significant portion of your income to charity (say, at least 10%), and that it would be even better for you to give more (say, 20% or 30%). But this could all hold even while *You ought to give 30%* is not true: this would be a good thing to do, but it is beyond the call of duty.

Another way to put the point is this: a theory of *ought/should* must be able to make sense of the idea of doing the least one ought to do, without collapsing this with what one ought to do *simpliciter*. In the examples just discussed, the least one ought to do is visit the friend without cooking dinner, and to give exactly 10%. Yet it is clearly not true that one ought to visit *without* cooking, or to give *exactly* 10%. This is merely the least one should do, and it would be better to do even more.

### 8.2.2 *Ought* as “better than not”

Consider first the “better than not” interpretation of *ought*. (I will drop the “significantly” caveat in (8.5)-(8.7) from now on; as far as I can see, it does not affect anything of substance in the discussion.)

- (8.10) *ought*( $\phi$ ) is true iff  $\phi >_{good} \neg\phi$ .

No matter how *better* is interpreted, the intuitive example of supererogation given above, summarized in (8.8), is inconsistent according to this interpretation of *ought*. The definition in (8.10) renders *ought*( $\phi \wedge \psi$ ) equivalent to  $\phi \wedge \psi >_{good} \neg\phi \vee (\phi \wedge \neg\psi)$ . Whether goodness is maximal or intermediate, this relation will hold if  $\phi \wedge \psi$  is strictly better than both disjuncts. Since we already know from premise (8.9b) that  $\phi \wedge \psi >_{good} \phi \wedge \neg\psi$ , we only have to show that  $\phi \wedge \psi >_{good} \neg\phi$ .

Suppose that goodness is maximal: then  $\phi \wedge \psi$ , being better than  $\phi \wedge \neg\psi$ , is exactly as good as their disjunction—which is equivalent to  $\phi$ . Since *ought*( $\phi$ ) holds,  $\phi$  is better than  $\neg\phi$ , and so  $\phi \wedge \psi$  is better than  $\neg\phi$  as well. So,  $\phi \wedge \psi$  is better than  $\neg\phi \vee (\phi \wedge \neg\psi)$ , i.e., better than  $\neg(\phi \wedge \psi)$ ; so, by (8.10), the conclusion *ought*( $\phi$ ) follows.

Suppose that goodness is intermediate: then, since  $\phi \wedge \psi >_{good} \phi \wedge \neg\psi$ ,

$$\phi \wedge \psi \geq_{good} \phi \geq_{good} \phi \wedge \neg\psi.$$



But since  $\text{ought}(\phi)$  is true,  $\phi >_{\text{good}} \neg\phi$  by definition (8.10). So, by transitivity  $\phi \wedge \psi >_{\text{good}} \neg\phi$ , from which the conclusion  $\phi \wedge \psi >_{\text{good}} \neg\phi \vee (\phi \wedge \neg\psi)$ —i.e.,  $\text{ought}(\phi \wedge \psi)$ —again follows.

As Lewis (1973) notes, a definition of *ought* as “better than not” is, in the presence of the maximality assumption, equivalent to the classical interpretation of  $\text{ought}(\phi)$  as “ $\phi$  holds in all of the optimal worlds”. A defender of the classical semantics is thus faced with a problem: given that supererogation is possible, it is unclear how the semantic connection between *ought* and *better* can be formalized without leading to implausible predictions. (In addition, the classical semantics makes a variety of further predictions that will emerge as problematic later in the chapter: see §8.12 for a summary.)

### 8.2.3 *Ought* as “best of the relevant alternatives”

The second proposal for reducing *ought* to *better* that we saw above was:

(8.11) *ought/should*  $\phi$  is true iff  $\phi$  is better than every other element of  $\text{ALT}(\phi)$ , the set of contextually relevant alternatives to  $\phi$ .

The additional degrees of freedom associated with the introduction of an alternative set come in handy here: supererogation is not ruled out logically, for arbitrary alternative sets. However, if we make some reasonable assumptions about the alternatives to each *ought*-claim in (8.9), the problem re-emerges. Recall that we do not want to validate this argument (where **visit** = “You visit your friend”, and **cook** = “You cook dinner for your friend”).

- (8.12) a.  $\text{ought}(\mathbf{visit})$   
       b.  $\mathbf{visit} \wedge \mathbf{cook} >_{\text{good}} \mathbf{visit} \wedge \neg\mathbf{cook}$   
       c.  $\therefore \text{ought}(\mathbf{visit} \wedge \mathbf{cook})$

Intuitively, the relevant alternatives for **visit** are those in (8.13).

(8.13)  $\text{ALT}(\mathbf{visit}) = \{\mathbf{visit}, \neg\mathbf{visit}\}$

The alternatives for  $\mathbf{visit} \wedge \mathbf{cook}$  may be either those in (8.14a) or those in (8.14b)

- (8.14) a.  $\text{ALT}(\mathbf{visit} \wedge \mathbf{cook}) = \{\mathbf{visit} \wedge \mathbf{cook}, \mathbf{visit} \wedge \neg\mathbf{cook}, \neg\mathbf{visit}\}$   
       b.  $\text{ALT}(\mathbf{visit}) = \{\mathbf{visit}, \neg\mathbf{visit}\}$

There are two cases to consider. Suppose first that scalar goodness is maximal, and consider the first alternative set (8.14a). Since  $\mathbf{visit} \wedge \mathbf{cook}$  is better than  $\mathbf{visit} \wedge \neg\mathbf{cook}$ , maximality implies that  $\mathbf{visit} \wedge \mathbf{cook}$  is exactly as good as  $(\mathbf{visit} \wedge \mathbf{cook}) \vee (\mathbf{visit} \wedge \neg\mathbf{cook})$ . The latter is logically equivalent to **visit**. Since **visit** is better than  $\neg\mathbf{visit}$  by (8.12a) and (8.13),  $\mathbf{visit} \wedge \mathbf{cook}$  is better than  $\neg\mathbf{visit}$  by transitivity of betterness. The conclusion (8.12c) then follows because  $\mathbf{visit} \wedge \mathbf{cook}$  is better than both of its alternatives,  $\mathbf{visit} \wedge \neg\mathbf{cook}$  and  $\neg\mathbf{visit}$ . A similar, but simpler, argument establishes the validity of the argument relative to the alternative set in (8.14b).

On the more plausible assumption that scalar goodness is intermediate, a similar result follows: (8.12b) plus intermediacy gives us

$$\mathbf{visit} \wedge \mathbf{cook} >_{\text{good}} \mathbf{visit} >_{\text{good}} \mathbf{visit} \wedge \neg\mathbf{cook}.$$

Combining this information with the observation that  $\mathbf{visit} >_{good} \neg\mathbf{visit}$ , we can infer the ordering

$$\mathbf{visit} \wedge \mathbf{cook} >_{good} \mathbf{visit} >_{good} \{\neg\mathbf{visit}, \mathbf{visit} \wedge \neg\mathbf{cook}\}$$

(with no known ordering between  $\neg\mathbf{visit}$  and  $\mathbf{visit} \wedge \neg\mathbf{cook}$ ).

$\mathbf{visit} \wedge \mathbf{cook}$  is thus better than both  $\mathbf{visit} \wedge \neg\mathbf{cook}$  and  $\neg\mathbf{visit}$ . If  $ought(\phi)$  means “ $\phi$  is best among the alternatives”, then, (8.12) is also wrongly validated on either a maximal or an intermediate theory of scalar goodness.

Thus, supererogation remains a problem for the alternative-sensitive reduction of *ought* to *better*. If the alternatives for *ought* are the intuitively relevant ones in (8.13), the definition in (8.11) predicts that *You ought to visit and cook* should be necessarily true in this scenario. Now, this result clearly falls short of a proof of the inadequacy of this account: a defender of the inference pattern (8.12) could interpret it as a proof that the alternatives in (8.13) are inappropriate. Still, the alternatives in (8.13) are intuitively operative, and elaborating the story so that they are explicitly given as the relevant choices does little to improve the prediction that (8.12c) should follow inexorably. This suggests that we should look for a better way to account for the semantic connection between *better* and *ought*.

## 8.2.4 *Ought* as “better than the union of the relevant alternatives”

The third definition of *ought* in terms of *better* considered above was (8.15):

$$(8.15) \quad ought/should \phi \text{ is true iff } \phi \text{ is better than } \bigcup \mathbf{ALT}(\phi).$$

If scalar goodness is maximal, this semantics collapses into the “better than all alternatives” semantics just considered (ch.7, §7.9). It is problematic for the same reason: supererogation is ruled out by the meaning of *ought*.

On the other hand, suppose that goodness is intermediate, and the relevant alternatives are again as in (8.13). Then the first premise of (8.12),  $ought(\mathbf{visit})$ , is rendered as  $\mathbf{visit} >_{good} (\mathbf{visit} \vee \neg\mathbf{visit})$ . As we saw above, the second premise implies

$$\mathbf{visit} \wedge \mathbf{cook} >_{good} \mathbf{visit} >_{good} \mathbf{visit} \wedge \neg\mathbf{cook}.$$

So, the best element of  $\mathbf{ALT}(\mathbf{visit} \wedge \mathbf{cook})$  is  $\mathbf{visit} \wedge \mathbf{cook}$ , and not all elements of this alternative set are equally good. Given this (and assuming that all elements of this set are possible), intermediacy implies that  $\bigcup \mathbf{ALT}(\mathbf{visit} \wedge \mathbf{cook})$  is better than its worst element, and worse than its best element. Since the best element is  $\mathbf{visit} \wedge \mathbf{cook}$ , we have the result that  $\mathbf{visit} \wedge \mathbf{cook} >_{good} \bigcup \mathbf{ALT}(\mathbf{visit} \wedge \mathbf{cook})$ . So, the conclusion  $ought(\mathbf{visit} \wedge \mathbf{cook})$  follows. While the derivation does require the auxiliary assumption that all of the alternatives are possible, note that adding this assumption explicitly to the basic story does not render the conclusion more plausible. Visiting *and* cooking dinner can be strictly optional, even when failing to visit and visiting without cooking are both relevant options.

As von Fintel (2012) points out, this definition of *ought* also makes problematic predictions about cases in which there are multiple equally good ways to realize some proposition that ought to be the case: under light assumptions, *ought* holds of each of the ways to realize the best alternative. This problem also affects a definition of *ought* as “better than not”, on the (well-motivated) assumption that goodness is intermediate. See also Hansson 2001, 2004 for related discussion.

### 8.3 Sloman's Principle

There are clear semantic connections between *ought* and *better*. Given this, it is appealing to try to define *ought* from *better*. While some workable definition might still turn up, the previous section showed that it is at least quite difficult to frame such a definition in an intuitively reasonable way without encountering problems with supererogation. I propose that we pursue a different strategy: rather than trying to define *ought* in terms of *better*, we should settle for stating constraints on the relationship between the two.

Hansson (2001, 2004) makes this point convincingly, and follows it up with the interesting suggestion that we could maintain a strong *better-ought* connection by constraining *ought* to be “contranegative” with respect to comparative goodness. In the case at hand, contranegativity corresponds to the validity of the following statement:

$$(ought(\phi) \wedge \neg\phi \succ_{good} \neg\psi) \rightarrow ought(\psi).$$

Hansson motivates this condition by pointing to the intuitive reasonableness of the following inference: “if you ought to work hard, and it is worse to be drunk than not to work hard, then you ought not to be drunk” (2004: 8). (He also shows that the condition generates generally plausible consequences in combination with a certain theory of scalar goodness, but the theory of scalar goodness that Hansson assumes fails to meet several of the desiderata laid out in chapter 7.)

Unfortunately, contranegativity is not very plausible when combined with the richer theory of scalar goodness that I argued for above, built around expected values. The reason is that the expected value of a proposition and that of its negation reflect around the point of indifference, which is the expected value of a tautology (i.e., their disjunction). With some subtleties associated with probability weighting, the condition that  $\mathbb{E}_V(\neg\phi) \geq \mathbb{E}_V(\neg\psi)$  will generally indicate that  $\mathbb{E}_V(\psi) \geq \mathbb{E}_V(\phi)$ . But if we were to substitute the latter, we would again have a semantics that rules out supererogation. Because of this, the conclusions of our earlier investigation of scalar goodness lead me to tentatively reject the contranegativity thesis, and to look instead for a constraint or set of constraints that interact well with our theory of scalar goodness while making room for supererogation.<sup>2</sup>

A different constraint which may work better for us is what I will call “Sloman's Principle”.

(8.16) **Sloman's Principle:**  $ought(\phi) \rightarrow [\forall \psi \in \mathbf{ALT}(\phi) : \psi \neq \phi \rightarrow \phi \succ_{good} \psi]$ .

I've named the principle for Sloman (1970), who proposes a version of the “best alternative” semantics that we discussed and rejected above. The inspiration for it is a weak, one-way implication from *ought* to *better* that Sloman argues for:

<sup>2</sup> Naturally, this could also be construed as an argument against our theory of scalar goodness, if there were powerful independent arguments in favor of contranegativity. Hansson does give several interesting arguments, chief among which is the observation that a permission predicate  $\mathcal{P}$  defined as  $\mathcal{P}(\phi) \leftrightarrow \neg\mathcal{O}\neg\phi$  can be (in his terminology) “positive” only if  $\mathcal{O}$  is contranegative. This positivity condition is different from the notion of Positivity considered at various points earlier in this book: Hansson's corresponds to the plausible claim that  $\mathcal{P}(\phi) \wedge \psi \succ_{good} \phi$  implies  $\mathcal{P}(\psi)$ .

This is all right, but it is beside the point at the moment if we read  $\mathcal{P}$  as “is permitted”, because English *permitted* is not the dual of *ought/should* but of *must*. Otherwise, it would be impossible to be permitted to do something that one should not; but this is a commonplace situation.

[S]omeone who said “ $p$  ought to be the case, but it would be better if  $q$  were the case”, where  $p$  and  $q$  are incompatible, would be contradicting himself.

Of course, this argument—and the principle that it motivates—is only plausible if *ought* and *good* are being interpreted as talking about the same kind of value: moral goodness, teleological goodness relative to some set of goals, etc. When this is satisfied, I will suggest that a slightly strengthened version of this one-way implication can be used to build a bridge between *should/ought* and *better*. This relationship places strong constraints on which propositions *ought/should* can be true of, and increases the falsifiability of a theory of *ought/should* by requiring it to be responsive to independently motivated facts about the relative goodness of alternatives.

Sloman’s Principle needs a bit of unpacking. The first thing to note is that—even in combination with a comprehensive theory of scalar goodness—it is definitely not sufficient to characterize the meaning of *ought/should* fully. Nevertheless, as we will see, Sloman’s Principle constrains the interpretation of *ought/should* in a way that generates reasonable predictions while also maintaining flexibility where it is needed. Later in this chapter we will add some additional constraints on the interpretation of *ought/should*.

Next, using Sloman’s Principle to make concrete predictions requires us to make some assumptions about the form of possible alternative sets. I will assume that, for any  $\phi$ , the alternatives relevant to *ought/should* (a) include  $\phi$ , (b) cover the domain  $D$ , and (c) are mutually exclusive within  $D$ . Condition (b) means, for example, that if  $D$  is the set of epistemically possible worlds, then every epistemically possible world must be contained in some alternative. The effect is that  $\{\psi \cap D \mid \psi \in \mathbf{ALT}(\phi)\}$  is a partition of  $D$ ; alternatives may overlap in logical space, but not within the domain.

For example, consider the sentence *Mary should leave*. A plausible alternative set is:

$$(8.17) \quad \mathbf{ALT}(\text{Mary should leave}) = \{\text{Mary leaves, Mary does not leave}\}$$

Suppose in addition that it is better that Mary leave than that she stay. Because Sloman’s Principle is a one-way implication from *ought* to *better*, it allows that *Mary should leave* could be either true or false here. However, it rules out the possibility that *Mary should not leave* could be true, since in this case *should* would be true of a proposition that is not better than all of its alternatives.

Suppose, on the other hand, that it is indifferent whether Mary stays or leaves. Then Sloman’s Principle implies that both *Mary should leave* and *Mary should not leave* are false. (As we will see below, this holds even if they are evaluated relative to different alternative sets, since  $\{\phi, \neg\phi\}$  is the weakest possible alternative set.)

Now, consider the same sentence—*Mary should leave*—against a more ramified alternative set.

$$(8.18) \quad \mathbf{ALT}(\text{Mary should leave}) = \{\text{Mary sings, Mary dances, Mary leaves, Mary plays cards, ...}\}$$

This alternative set is admissible, as long as it is not possible for Mary to do any two of these things, and the ellipsis in (8.18) is fleshed out so that  $\mathbf{ALT}(\text{Mary should leave})$  covers the domain  $D$ . Suppose, in addition, that *Mary leaves* is better than each of the other propositions in the alternative set. Then, *Mary should leave* could be either true or false. On the other hand, suppose that *Mary leaves* is exactly as good as *Mary sings*, and both are better than all of the other alternatives. Then considered against this alternative set, all of the following are false—*Mary should leave*, *Mary should sing*, *Mary should dance*, ...—because none of them is better than all the alternatives.

Interestingly, a potentially true *ought* statement lies dormant in the example just given. Consider the sentence *Mary ought to leave or sing*, with the comparative goodness facts held fixed. Since *Mary leaves or sings* is required to be in the alternative set, the alternatives might have the structure in (8.19).

(8.19)  $\mathbf{ALT}(\text{Mary should leave or sing}) = \{\text{Mary leaves or sings, Mary dances, Mary plays cards, ...}\}$

Since *Mary leaves* is exactly as good as *Mary sings*, their disjunction has the same degree of goodness as both. This degree is greater than the goodness of *Mary dances, Mary plays cards, ...*, and so Sloman's Principle is compatible with the truth of *Mary ought to leave or sing*. (As always, Sloman's Principle does not tell us that anything actually ought to be the case, only that goodness facts do not rule out that it could be.)

Note, however, that the requirement that the alternatives cover the domain—i.e., that every accessible world appear in some proposition in the alternative set—means that the *weakest possible* alternative set, as far as Sloman's Principle is concerned, will always be the one that compares a proposition to its own negation. For example, suppose that  $\phi$  is strictly better than every other element of  $\mathbf{ALT}(\phi) = \{\phi, \psi_1, \psi_2, \dots\}$ . Then from the intermediacy of  $\mathcal{S}_{good}$  we can infer that  $\phi >_{good} \bigcup \mathbf{ALT}(\phi)$ . Since by assumption  $\bigcup \mathbf{ALT}(\phi)$  contains all of the worlds with positive probability, its expected value is the same as  $\mathbb{E}_V(W)$ ; and, as we already saw above,  $\mathbb{E}_V(\phi) > \mathbb{E}_V(W)$  if and only if  $\mathbb{E}_V(\phi) > \mathbb{E}_V(\neg\phi)$ .

As a result, Sloman's Principle—together with the expected value hypothesis and the constraints on alternatives that we have assumed—entails a one-way version of the *ought/better than not* connection considered above: in any context,

$$\text{ought}(\phi) \rightarrow \phi >_{good} \neg\phi.$$

#### 8.4 Sloman's Principle and supererogation

A key restriction flows from the requirement that the alternatives be disjoint within the domain  $D$ . Sloman's Principle would not be at all plausible if it applied to arbitrary sets of alternatives whose members might overlap within the modal domain under consideration—and this restriction is what makes room for supererogation. Consider, for example, the typical format of the examples use above to motivate the problem: it is given that you ought to visit your sick friend, and that visiting *and cooking her dinner* would be even better. You are then left to ponder the truth of *You ought to visit your sick friend and cook her dinner*. However, Sloman's Principle does not apply here, since there is no way that  $\phi \wedge \psi$  could be a relevant alternative to  $\phi$ : they are not mutually exclusive. When we turn to the question of whether the more specific  $\phi \wedge \psi$  ought to be the case, we must construct a different alternative set – say,  $\{\phi \wedge \psi, \neg(\phi \wedge \psi)\}$  or  $\{\phi \wedge \psi, \phi \wedge \neg\psi, \neg\phi\}$ .

As a result, the facts in the story together with Sloman's Principle do not tell us whether you ought to cook dinner as well. This is as it should be: in some scenarios that fit this description, *You ought to visit and cook dinner* will be true, and in others it will be false.

Another class of relevant cases involve alternative sets which differ in goodness, but for which *ought* is true of none of the alternatives. Modifying an example from Slote (1984), let's suppose that the following is true:

(8.20) Doctors ought to devote significant effort to helping people with insufficient medical care.

(No matter whether you think this is really the case; please grant it for the example's sake.) Now, imagine that a doctor decides to devote a significant part of her career to this activity, and chooses to do so in a particular country (say, India) because she has a personal interest in some irrelevant aspect of the country—its art or languages or mythology, for example. In addition, suppose that she could do even more for the poor by going to Namibia. Then the following is true.

(8.21) It is better that the doctor go to Namibia than to India.

As Slote points out, classical utilitarianism and other theories that tie *ought* directly to *best* predict that the following judgment follows inexorably from (8.21): *the doctor has failed to do what she ought to do*. But this is not correct as a descriptive, linguistic claim about the meaning of *ought/should*. Indeed, it is easy to see how someone could have a judgment along the following lines.

(8.22) The doctor did what she ought to do, even though it was possible for her to do it in a way that would have been even better. (What she did was a commendable sacrifice nonetheless.)

On the other hand, the story is also consistent with a different elaboration: depending on other factors (say, a sudden epidemic), it could easily be that the doctor indeed ought to go to Namibia. But since this elaboration would not affect the comparative goodness facts—going to Namibia was, and remains, the best thing the doctor could do—we can derive an important conclusion: facts about what *ought* to happen can vary, even while the relative goodness of alternatives is held fixed. That is, **comparative goodness facts constrain, but do not determine, what ought to be the case**.

Corroborating this conclusion, suppose that we were deliberating about (8.23) in light of the following alternative set.

(8.23) The doctor ought to go to India.

(8.24) **ALT**((8.23)) includes the following, ordered from best to worst:

- a. The doctor goes to Namibia to help those with insufficient medical care.
- b. The doctor goes to India to help those with insufficient medical care.
- c. The doctor goes to Sri Lanka to help those with insufficient medical care.
- d. The doctor goes to Mozambique to help those with insufficient medical care.
- e. The doctor goes to Paraguay to help those with insufficient medical care.
- f. ...
- g. The doctor goes to Switzerland to help those with insufficient medical care.
- h. The doctor does not go to help those with insufficient medical care.

Sloman's Principle is compatible with the possibility that *ought* is true of none of these alternatives. For example, it might be that the doctor ought to go to some country in the set  $C = \{c_1, c_2, \dots, c_n\}$  of the world's  $n$  most desperate, but there is no country  $c \in C$  such that he ought to go to  $c$  (and no strict subset of  $C' \subset C$  such that he ought to go to some country in  $C'$ ). As long as India is in  $C$ , our kind doctor is in the moral clear despite her extracurricular interests.

For a related case, imagine that Bill should give 10% of his income to charity, and that it would be even better for him to give 15% or 20%. We could set up the alternative sets relevant to this judgment in several different ways, including the following.

(8.25) Bill should give 10% of his income to charity.

(8.26) a. Option 1:  $\text{ALT}((8.25)) = \{\text{Bill gives exactly } n\% \mid n \in [0, 100]\}$

b. Option 2:  $\text{ALT}((8.25)) = \{\text{Bill gives at least 10\%, Bill does not give at least 10\%}\}$

Under option 1, (8.25) would entail (by Sloman's Principle) that giving 10% is better than giving  $n\%$  for all other  $n \in [0, 100]$ —including all larger quantities. While this could be true in some strange scenarios, it is generally much more plausible that (8.25) is interpreted relative to option 2. In the case of option 2, Sloman's Principle does not rule out the possibility that giving exactly 15% could be better, because giving exactly 15% is not and could not be among the alternatives. (It is properly contained in the alternative *Bill gives at least 10%*, and so the alternatives would not be mutually exclusive if it were included.)

The lesson of these examples, then, is that Sloman's Principle will sometimes generate apparently problematic predictions if we focus on alternative sets that include doing the minimum we ought to do *and no more*—such as giving *exactly* 10%. However, it is not generally the case that we ought to do the bare minimum: rather, we ought to do the minimum *or better*. This allows us to make sense of “minimum obligation” readings of *ought* statements, and the fact that there is a way of doing what one ought that is even better.

However, Sloman's Principle does rule out the possibility that *ought* could be true of any proposition which actively **excludes** the best alternative(s)—for example, giving exactly 10% in a scenario where giving more would be better. Similarly in the doctor scenario, if the betterness facts are as in (8.24)—with going to Namibia fixed as better than any other country—Sloman's Principle rules out the possibility that *The doctor ought to go to a country in C* could be true for any set  $C$  which excludes Namibia. This remains true for many, though not all, ways of coarsening the alternative set under consideration. For instance, consider (8.27) in light of the alternative set in (8.28).

(8.27) The doctor ought to go to Mozambique or Paraguay.

(8.28)  $\text{ALT}((8.27)) = \{\text{India, Mozambique or Paraguay, Namibia, Switzerland, Canada, ...}\}$

Relative to this alternative set, the truth of (8.27) would require that *The doctor goes to Mozambique or Paraguay* is better than every alternative, including *The doctor goes to Namibia*. Since goodness is an intermediate quantity, the goodness of *The doctor goes to Mozambique or Paraguay* is somewhere in between the goodness of *The doctor goes to Mozambique* and *The doctor goes to Paraguay*. Since going to Namibia is pairwise better than each of these, it is better than the disjunction. The truth of (8.27) is thus correctly ruled out by Sloman's Principle.

However, there are other ways of coarsening the alternative set that will generate apparently paradoxical predictions. For instance, consider (8.27) again in light of a different alternative set:

(8.29)  $\text{ALT}((8.27)) = \{\text{India, Mozambique or Paraguay, Namibia or Switzerland or Canada, ...}\}$

In this case, the expected value of *Namibia or Switzerland or Canada* could be **lower** than that of *Mozambique or Paraguay*. (Whether it actually is depends on the details of *prob* and *V*.) This is a result of the intermediacy property of expected value: a proposition can have as a proper subpart some proposition that would be very good, and yet fail to be very good if its other parts are less good and sufficiently probable. But then (8.27) would not be ruled out by Sloman's Principle, since there is no alternative in **ALT**((8.27)) that is better than *Mozambique or Paraguay*.

We could try adding principles that would rule out such alternative sets on the basis of their structure: for example, we might try to work out some way to enforce a requirement that the alternatives must form value-continuous portions of the alternative space. On this suggestion, the division {*Namibia or Sri Lanka, Mozambique, Paraguay or India, ...*} would be acceptable, but {*Namibia or India, Sri Lanka or Mozambique, Paraguay, ...*} would not.

I do not think that this is a serious enough problem to warrant such a stipulation, though. First, while we did not state any constraints that would rule it out, it seems likely that such gerrymandered alternative sets are pragmatically irrelevant. If so, we do not need to rule them out by fiat in addition. Specifically, we might suppose that there is a general principle that the alternatives under consideration should correspond to practically relevant divisions in the space of possibilities: for example, that it will be a set of propositions each of which describes a choice that the doctor could make. Her choices in this context are of the form *Go to India, go to Namibia, go to Switzerland, ...*, and do not include disjunctions like *Go to Namibia or Switzerland or Canada*.

A further point is that, if we were to place conditions on the betterness-continuity of the alternatives, we would make it difficult to account for a number of cases in which the natural, practically relevant structure of the alternatives has precisely this non-continuous character. These are the subject of the next several sections. As we will see, with careful attention to the practically relevant sets of alternatives, Sloman's Principle and the theory of scalar goodness proposed earlier in chapter 7 allow us to make sense of a number of empirical problems that have been raised as objections to the classical semantics for *ought/should*.

## 8.5 Uncertainty, choice, and information redux

Earlier I presented the **Juliet** and **Professor Procrastinate** scenarios, as well as the Miners' Puzzle, as problems for a theory in which scalar goodness is maximal. Equally, they are problematic for two kinds of theories of *ought/should*:

- Theories on which goodness is maximal, and *ought/should* is defined directly from *better*.
- Theories on which the meaning of *ought/should* is defined in some other way in terms of the status of some set of "best" worlds.

Recall the **Professor Procrastinate** scenario that we discussed in chapter 7 (§7.7.2), taken from Jackson & Pargetter (1986) and with modifications inspired by Cariani (2013).

Prof. Procrastinate is invited to review a book on which he is the only fully qualified specialist on the planet. Procrastinate's notable character flaw, however, is his inability to bring projects to completion. In particular, if Procrastinate accepts the



request to review the book, he *might* write it, but it is very likely (though not quite certain) that he will forget and not end up writing the review. In the eyes of the editor, and of the whole scientific community, this is the worst possible outcome. If Procrastinate declines, someone else will write the review—someone less qualified than him, but more reliable.

As Jackson & Pargetter observe, (8.30) is intuitively false in this scenario—since it would likely lead to the worst possible outcome.

(8.30) Professor Procrastinate ought to accept.

Nevertheless, many people also find (8.31) compelling:

(8.31) Professor Procrastinate ought to accept and write the review.

On most theories of *ought* it should be impossible, by virtue of their meaning, for (8.30) to be false while (8.31) is true. *ought*( $\phi \wedge \psi$ ) entails *ought*( $\phi$ ), period. This is true of classical utilitarianism, von Fintel & Iatridou's (2008) elaboration of Kratzer's (1991) semantics, and any other theory in which *ought/should* is upward monotone.

**Upward monotonicity of *ought*:** If  $\phi$  entails  $\psi$ , then *ought*( $\phi$ ) entails *ought*( $\psi$ ).

Since *Procrastinate accepts and writes* entails *Procrastinate accepts*, an upward monotone theory of *ought* renders (8.30) an entailment of (8.31).

In contrast, the semantics for *ought/should* that we have been building allows that (8.30) can be false while (8.31) is true. (Absent further constraints, that is. If the affirmation of (8.31) is indeed consistent with the denial of (8.30), then the lesson of these examples is that, whatever further constraints we assume, they had better not render (8.30) an entailment of (8.31).)

In chapter 7's discussion of the *good*-involving Procrastinate cases (§7.7.2) I noted that intuitions about the relative goodness of propositions in this scenario have the following form—paradoxically, if you think goodness is maximal.

- (8.32) a. Procrastinate does not accept  $>_{good}$  Procrastinate accepts  
 b. Procrastinate accepts and writes  $>_{good}$  Procrastinate does not accept  
 c. Procrastinate does not accept  $>_{good}$  Procrastinate accepts and does not write

Now, (8.30) favors an alternative set like that in (8.33).

(8.33) {Procrastinate accepts, Procrastinate does not accept}

Since *Procrastinate accepts* is not better than *Procrastinate does not accept*, Sloman's Principle does not allow for *Procrastinate ought to accept* to be true relative to this alternative set. The same principle allows—but does not require—that *Procrastinate ought not to accept* can be true.

Due to the requirement that *ought*'s argument be among the alternatives, (8.33) cannot be the alternative set relevant to the evaluation of (8.31). The relevant alternatives for (8.31) must include *Procrastinate accepts and writes the review*. This *ought*-statement is most naturally associated with one of the alternative sets in (8.34):

- (8.34) a.  $\{\neg\phi, \phi \wedge \psi, \phi \wedge \neg\psi\} = \{\text{Procrastinate does not accept, Procrastinate accepts and writes, Procrastinate accepts and does not write}\}$   
 b.  $\{\phi \wedge \psi, \neg(\phi \wedge \psi)\} = \{\phi \wedge \psi, ((\phi \wedge \neg\psi) \vee \neg\phi)\} = \{\text{Procrastinate accepts and writes, Procrastinate either (accepts and fails to write, or does not accept)}\}$

Sloman's Principle and (8.32) together allow that *Procrastinate ought to accept and write* could be true here, since it is better than all of its alternatives in either case. Of course, nothing we have said forces the judgments in (8.30) and (8.31) either. However, they do leave room for these intuitively reasonable judgments, depending on what (if any) further constraints are placed on *ought/should*.

This is a marked improvement on classical utilitarianism, which predicts that *Procrastinate ought to accept and write* is true, and *Procrastinate ought not to accept* is false given these betterness facts. It is also preferable to accounts that follow Kratzer's (1991) closely, such as von Fintel & Iatridou 2008. On the latter account, judging (8.30) false while endorsing (8.31) is coherent only if one has subtly shifted the modal domain (cf. von Fintel 2012). Getting the right result for (8.30) requires excluding from its domain the less-likely worlds where Procrastinate accepts and does write. While there is no technical barrier here, the analysis seems ill-motivated given that the story specifically emphasizes that it is not certain that the Professor will not write if he accepts—i.e., that he *might*. In contrast, the skeletal theory being developed here—built around scalar goodness, with Sloman's Principle as a bridge—requires no *ad hoc* maneuvers in order to make room for the right result.

The situation is similar with the Miners' Puzzle and its variants. Focusing on the original puzzle for clarity's sake (chapter 7, §7.7.3), recall that the probability of the miners' being in each shaft is 50%, and that the possible outcomes are distributed as in Table 8.1. The paradox was generated by Kolodny & MacFarlane's (2010) observation that all of the following can be judged true —

- (8.35) We ought to block neither shaft.  
 (8.36) If the miners are in shaft *A*, we ought to block shaft *A*.  
 (8.37) If the miners are in shaft *B*, we ought to block shaft *B*.  
 (8.38) Either the miners are in shaft *A* or they are in shaft *B*.

— even while (8.39), which is entailed by these premises on standard assumptions, is judged false.

- (8.39) Either we ought to block shaft *A* or we ought to block shaft *B*.

---

	block <i>A</i>	block <i>B</i>	block neither
miners in <i>A</i>	10	0	9
miners in <i>B</i>	0	10	9

---

**Table 8.1** Structure of the original Miners' Puzzle, repeated.

---

If we continue to assume a simple value function where the value of a world is equal to the number of miners saved, the expected values of the various actions are:

- $\mathbb{E}_V(\mathbf{block\ A}) = 10 \times .5 + 0 \times .5 = 5$
- $\mathbb{E}_V(\mathbf{block\ B}) = 0 \times .5 + 10 \times .5 = 5$
- $\mathbb{E}_V(\mathbf{block\ neither}) = 9 \times .5 + 9 \times .5 = 9$

I assume further that the alternatives in each case are just descriptions of the available actions:

{We block A, we block B, we block neither}

Sloman's Principle then allows that *We ought to block neither* may be true (since it is better than all other alternatives), and requires that both *We ought to block A* and *We ought to block B* are false.

As expected, this prediction depends on the assumptions we have made about probabilities and values. If the miners were almost certainly in A,  $\mathbb{E}_V(\mathbf{block\ A})$  would be greatest. In this case, *We ought to block neither* would be guaranteed false, but *We ought to block A* would be possibly true. Other manipulations discussed above, affecting the goodness of outcomes, have similarly reasonable effects on predictions about which *ought* claims can and cannot be true. This includes adopting the god's-eye view, which would rule out *ought* being true of any action except blocking the shaft that the miners are actually in. Then, (8.39) is possibly (and plausibly) true.

It remains to account for the conditional statements (8.36) and (8.37). Sloman's Principle does not speak directly to occurrences of *ought/should* in embedded contexts, but a minimal extension taking states of information into account does.

(8.40) **Extended Sloman's Principle:** Let  $D$  be some set of relevant worlds (e.g., the epistemically possible worlds). Then the following holds at all worlds, in all models, and for any state of information  $I$ :

If *ought*( $\phi$ ) is true relative to  $I$ , then  $\forall \psi \neq \phi \in ALT(\phi) : \phi >_{good} \psi$ , where  $>_{good}$  is computed relative to  $I$ .

In the case at hand, we are considering the state of information being used to evaluate the consequent of the conditional *If the miners are in A, we ought to block A*. This is the information state generated by restricting the domain  $D$  to antecedent-worlds, with the effect of conditioning the global state  $prob(\cdot)$  on the proposition that the miners are in A:  $prob(\cdot \mid \mathbf{miners\ in\ A})$  (chapter 7, §7.7.3). Extended Sloman's Principle thus tells us that *ought* can be true of a proposition in this environment—say, *We block A*—only if its expected value relative to this restricted information state is greater than those of all alternatives relative to the same restricted information state. The expected value of a proposition in this environment is just the conditional expected value of the proposition, given *The miners are in A*. We already calculated these above:

- $\mathbb{E}_V(\mathbf{block\ A} \mid \mathbf{miners\ in\ A}) = 10 \times 1 + 0 \times 0 = 10$
- $\mathbb{E}_V(\mathbf{block\ B} \mid \mathbf{miners\ in\ A}) = 0 \times 1 + 10 \times 0 = 0$
- $\mathbb{E}_V(\mathbf{block\ neither} \mid \mathbf{miners\ in\ A}) = 9 \times 1 + 9 \times 0 = 9$

In the information state generated by conditioning on *The miners are in A*, *We block A* is better than both *We block B* and *We block neither*. Extended Sloman's Principle tells us that, in

the consequent of a conditional whose antecedent is *The miners are in A*, *We ought to block A* is possibly true, *We ought to block B* is false, and *We ought to block neither* is false.

This result is encouraging, in particular, because it shows how we can endorse all of (8.35)-(8.38) while rejecting (8.39). Sloman's Principle, together with our semantics for comparative goodness, entails that the former are all at least possibly true, and the latter is guaranteed false.

There are many more examples we could go through to illustrate the information-sensitivity of *ought/should* and their intricate empirical consequences. I hope that these examples illustrate the point to a sufficient extent. The next sections examine some further points of comparison with the classical semantics, and motivate some further constraints on the meaning of *ought*.

## 8.6 Deontic detachment

An argument form that is known as “deontic detachment” in the literature provides another interesting puzzle involving *ought/should* and conditionals. Consider the following argument form:

- (8.41) a. *ought*( $\phi$ ).  
       b. If  $\phi$ , then *ought*( $\psi$ ).  
       c.  $\therefore$  *ought*( $\psi$ ).

This argument is valid according to many variants of the classical semantics, where *ought/should* denote quantifiers over a set of “best” worlds. In particular, it is valid in the theory of Lewis (1973: §5) and of von Fintel & Iatridou (2008), assuming that the latter is combined with a restrictor theory of conditionals (Kratzer 1986, 1991). In brief, this argument form is valid in both theories because, whenever  $\phi$  is true in all of the best worlds, the interpretation of the conditional ( $\psi$  is true in all of the best  $\phi$ -worlds) reduces to the condition that  $\psi$  is true in all of the best worlds, period.<sup>3</sup>

Deontic detachment has been much discussed in the deontic logic literature as part of “Chisholm's Paradox” (Chisholm 1963a). The solution to that paradox is, I suggest, very simple: deontic detachment is not a valid argument form. Consider the following argument:

- (8.42) a. Bill ought to give money to Greenpeace.  
       b. If Bill gives money to Greenpeace, Greenpeace ought to send Bill a thank-you gift.  
       c. So, Greenpeace ought to send Bill a thank-you gift.

This argument is plainly wrong. Supposing that the premises are true, whether Greenpeace ought to send Bill a thank-you gift does not depend on whether Bill *ought* to give them money: it depends

<sup>3</sup> In more detail: In Lewis' case, (8.41a) means that all of best worlds are  $\phi$ -worlds. (We are making the limit assumption for simplicity, but a slightly more complex version of the proof goes through without it.) (8.41b) means that all of the best  $\phi$ -worlds are  $\psi$ -worlds. Clearly, then, all of the best worlds are  $\phi \wedge \psi$ -worlds, and so they are all  $\psi$ -worlds, and *ought*( $\psi$ ) is true. The argument is essentially the same for a Kratzerian theory like that of von Fintel & Iatridou (2008), where the conditional holds just in case *ought*( $\psi$ ) holds in all of the best worlds according to a derived ordering which is created by filtering the original ordering to exclude all  $\neg\phi$ -worlds. Since  $\phi$  holds in all of the best worlds in the initial ordering, the best worlds in the filtered ordering are the same as the best in the original ordering. So,  $\phi \wedge \psi$  is true in all of these worlds, and *ought*( $\psi$ ) holds unconditionally.

on whether he *does* give them money. If Bill ignores his obligation and fails to give money to Greenpeace, Greenpeace is under no obligation to send him anything.<sup>4</sup>

For a more dramatic example, consider cases where what one individual ought to do reverses depending on what another individual does, and the two individuals are choosing independently or even at cross purposes. For example, during World War II the British government evacuated well over one million people—mostly children—from cities to smaller towns as a protective measure against Nazi bombing. Many hundreds of thousands of children were separated from their parents for years. Consider the following reasoning, undertaken in the midst of the war. All three of the premises in (8.44) seem to be true here.

- (8.44) a. The Nazis ought to stop bombing British cities.  
       b. If the Nazis stop bombing British cities, we ought to bring evacuated children back to the cities.  
       c. If the Nazis do not stop bombing British cities, we ought not to bring evacuated children back to the cities.

Here, the theories of Lewis and of von Fintel & Iatridou predict that (8.45) must be true, given the truth of the sentences in (8.44).

- (8.45) So, we ought to bring evacuated children back to the cities.

In this context, it seems clear that what we ought to do depends not on what the Nazis *ought to do*, but on what they *actually do*. If they do stop bombing, we should reunite children with

---

<sup>4</sup> Willer (2015) has recently defended deontic detachment as a kind of default inference rule, which can be defeated—in the case of (8.42)—by the addition of the information that Bill does not give money. This diagnosis still strikes me as incorrect: the inference in (8.42c) is not even reasonable as a default unless we have reason to believe that Bill will probably do what he ought to do. Similarly, in the Nazi bombing example discussed below (8.44), it would be strange indeed to conclude that, in the absence of further information about the bombings, *We ought to bring the children back to the cities* is true as a default. Presumably, what we ought to do instead is to suspend judgment and investigate the current status of the bombing campaign.

While the idea that deontic detachment is a reasonable default is plausible for many of the examples discussed in the literature, I suspect that this is due to pragmatic interference associated with intuitions about control. The examples discussed in the literature on Chisholm's Paradox generally involve sentences with the same subject. In these examples, as in the original Professor Procrastinate example, the first premise of the argument strongly implicates that the subject has control over the truth-value of the antecedent of the second premise. As a result the argument tends to sound somewhat better. Consider for example (8.43), which is intuitively much more compelling than (8.42).

- (8.43) a. Bill ought to give money to Greenpeace.  
       b. If Bill gives money to Greenpeace, he ought to give money to Amnesty International.  
       c. So, Bill ought to give money to Amnesty International.

The reason that these examples differ, I suggest is that (8.43a) implies that Bill has a choice about whether to give money to Greenpeace, and can do it if he chooses. Given this, it is under his control whether the antecedent of (8.43b) is true. It is the outcome of this choice (and not the obligation *per se*) that controls whether (8.43c) comes out true. Still, the fact that Bill is ultimately in control of *whether* he has the obligation described by (8.43c), and that he ought to act in such a way that would activate this obligation, seems to incline us more strongly toward the judgment that the obligation holds *simpliciter*. We can eliminate this confound by considering scenarios where two agents are acting independently or in competition, as I did in (8.42) and (8.44) as well as the **Juliet** and **Disaster Relief** scenarios.

their families. If they do not, we should not reunite children with their families, tragic as the separation may be, because that move would expose the children to mortal danger. The fact that (8.45) follows logically from the information in (8.44) is a *reductio ad absurdum* of the theory of deontic conditionals in Lewis 1973: §5, and of the combination of von Fintel & Iatridou’s (2008) theory with a restrictor theory of conditionals (Kratzer 1986, 1991).

The puzzles around deontic detachment are not a deep mystery: they are theory-internal problems driven by the widespread and incorrect assumption that *ought/should* picks out a universal quantifier over a set of best worlds. The semantic constraints on *ought/should* that we have motivated go a considerable way toward dispelling the puzzle. These constraints do not just fail to validate deontic detachment (a good start), but they specifically predict that the conclusion of the argument cannot be true in a plausible model for (8.44) and similar examples.

Value	bombing	no bombing
stay in countryside	-10	0
return to cities	-100	10

**Table 8.2** A representative value function for the **Nazi Bombing** example.

The model in Table 8.2 is one way to encode the information implicit in the story: staying in the countryside is a sad state of affairs, but returning to the cities is very risky, leading either to an improved outcome (reunion with families) or a terrible tragedy depending on the choices of an agent over whom we have no control. Here, we have the following predictions via extended Sloman’s Principle:

- (8.46) a.  $(8.44a) \Rightarrow \mathbb{E}_V(\text{no bombing}) > \mathbb{E}_V(\text{bombing})$   
b.  $(8.44b) \Rightarrow \mathbb{E}_V(\text{return to cities} \mid \text{no bombing}) > \mathbb{E}_V(\text{stay in countryside} \mid \text{no bombing})$   
c.  $(8.44c) \Rightarrow \mathbb{E}_V(\text{stay in countryside} \mid \text{bombing}) > \mathbb{E}_V(\text{return to cities} \mid \text{bombing})$

These constraints are all satisfied by the model described in Table 8.2, regardless of the assignments of probabilities.  $\mathbb{E}_V(\text{no bombing}) > \mathbb{E}_V(\text{bombing})$  holds no matter what, since **no bombing** dominates **bombing**—it is better regardless of our choice. Returning to the cities (value 10) is better than staying in the countryside (value 0) if there is no bombing, and so (8.46b) is satisfied. Likewise, (8.46c) is satisfied because staying in the countryside is better if there is bombing (−10 vs. −100).

The candidate conclusion in (8.45) will not generally be true here, and in fact Sloman’s Principle will frequently require that it fail. Sloman’s Principle requires that, if (8.45) is true, then  $\mathbb{E}_V(\text{return to cities}) > \mathbb{E}_V(\text{stay in countryside})$ . But the expected value of returning to the cities will exceed that of staying in the countryside if and only if there is a very high probability ( $> .9$ ) that the bombing campaign has ceased. So, unless the probability of being in a good situation is very high, Sloman’s Principle correctly rules out the truth of the counter-intuitive conclusion that detachment would generate.

Given that scenarios like **Nazi bombing** are possible—indeed, that they have occurred in the real world—deontic detachment cannot be a generally valid principle. The perspective that we

have adopted here helps us to understand why. We cannot focus merely on ideal worlds when reasoning about *ought* and *should*. Instead, we must take into account the very real possibility that the best outcomes will not occur. More generally, good deontic reasoning requires weighing values and goals against all of the available information, rather than focusing myopically on the “best” outcomes.

## 8.7 Anti-coordination and agglomeration: An experiment

This section discusses another type of scenario that illustrates the information-sensitivity of *ought/should* and the semantic differences between the present theory and the classical theory. I will argue that the principle of agglomeration is not unrestrictedly valid, and that the theory that we are developing makes reasonable predictions about when it should fail.

**Agglomeration:** *ought*( $\phi$ ) and *ought*( $\psi$ ) are both true iff *ought*( $\phi \wedge \psi$ ) is true.

Anti-coordination contexts—where two agents are trying to act, or should try to act, in such a way that their actions are different—present an under-appreciated challenge to the classical semantics. These scenarios are important because they demonstrate that agglomeration and Sloman’s Principle are in semantic competition. If Sloman’s Principle is correct, then agglomeration cannot be unrestrictedly valid. Conversely, if agglomeration is unrestrictedly valid, then there are sometimes cases where *ought*( $\phi$ ) holds even while  $\phi$  is strictly worse than  $\neg\phi$ .

**Disaster Relief** is one example of an anti-coordination scenario.<sup>5</sup>

Two teams of aid workers, the Red Team and the Blue Team, are trying to help people in the wake of an earthquake. Communications are out, so they can’t contact each other. Two villages are in desperate need of food: Village A and Village B.

- If the teams head toward different villages, they can save all of the residents.
- If they head toward the same village, the people in the other village will starve to death.

When they lost contact, both teams were located close to Village B. So, both of them would by default go to B, since it’s easier to get there—and both of them know that the other will by default go to B.

Thinking of the Red Team’s decision-making, I am inclined to endorse (8.47) here:

(8.47) The Red Team ought to go to A.

<sup>5</sup> This scenario was inspired by Jackson’s (1985) Chicken example. The choice to concentrate on **Disaster Relief** here and in the experiment described below was due to the results of pilot experiments. I first tested a scenario nearly identical to Jackson’s Chicken scenario, but discovered in debrief that many experimental participants had difficulty understanding the scenario, and that those who did were often so outraged by the irresponsible behavior of the participants that their data were of questionable value. (Sample comments: “Both should swerve and have their cars wrecked as a lesson to not do anything like this again.”; “These two are idiots. The gene pool needs to be thinned a bit”.) The **Disaster Relief** scenario was designed to replicate the key features of the Chicken example, while ensuring that the people involved were behaving in a reasonable enough way that the scenario would be intelligible and believable to non-philosophers.

After all, the Blue Team will (by assumption) probably go to B, and the Red Team are trying not to go to the same place. However, the situation is precisely parallel with the Blue Team: wiping the slate clean and thinking of their decision-making, we have all the same reasons to judge the following true.

(8.48) The Blue Team ought to go to A.

(Considering (8.47) and (8.48) together creates interference, presumably because of modal subordination and/or dynamic effects. We have to present them separately, with a palate cleanser in between, to see the problem.)

However, the following is clearly not true here:

(8.49) It ought to be that the Red Team and the Blue Team both go to A.

Since these intuitions are apparently not clear to all, I conducted an experiment on Amazon's Mechanical Turk platform in order to investigate whether non-theorists' patterns of assent and dissent were consistent with them, with the predictions of the classical semantics, or with neither. 600 participants saw the **Disaster Relief** scenario, interspersed with two different deontic judgment scenarios that are not reported here, in a random order. Participants were randomly assigned to a **probability** condition, where they were told either that both teams would by default go to A, or that both teams would by default go to B. Each participant was asked to "Agree" or "Disagree" with a single, randomly assigned prompt from the following list. Participants were asked only to judge only one of the prompts, in order to avoid the possibility of order effects generated (for example) by modal subordination.

- (8.50) a. What ought to happen is this: the Red Team goes to Village A.  
b. What ought to happen is this: the Blue Team goes to Village A.  
c. What ought to happen is this: the Red Team goes to Village A and the Blue Team goes to Village A, too.

The "What ought to happen is this:" syntax was chosen because it seemed to me to be the most natural way to express the intended thought in (8.50c), while ensuring that the entire conjunction remained in the scope of *ought*. Unfortunately this choice rendered both (8.50a) and (8.50b) slightly awkward, but this seemed preferable to introducing an asymmetry among conditions by using a more natural phrasing for the first two prompts.

I also manipulated whether participants saw an additional prompt that emphasized whose decision-making was at issue: "The Red Team has to make a decision.", etc. However, this manipulation had no effect, and I average over it in the results reported.

I discarded data from 21 participants who reported a native language other than English. Data from the remaining 579 participants is summarized in Table 8.3 and Figure 8.1.

Participants demonstrated a clear sensitivity to the probability manipulation. When they were told that both teams would by default go to A, most participants disagreed when asked about either (8.50a) or (8.50b) (agreement 32% and 41%, respectively). This suggests that the participants absorbed the anti-coordination structure of the scenario: each team ought to do the opposite of what they think the other team will. Similarly, when both teams were by default going to B, the large majority of participants who were asked about the Red Team agreed that the Red Team should go to



---

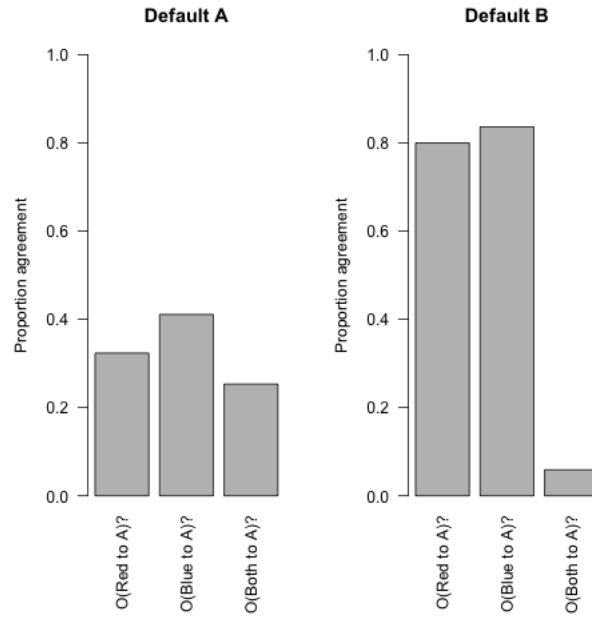
Default A			
	O(Red to A)?	O(Blue to A)?	O(Both to A)?
Agree	44	33	20
Disagree	63	69	59
Total	107	102	79
Proportion	0.32	0.41	0.25
95% Upper	0.41	0.50	0.35
95% Lower	0.24	0.32	0.16

Default B			
	O(Red to A)?	O(Blue to A)?	O(Both to A)?
Agree	87	68	6
Disagree	17	17	96
Total	104	85	102
Proportion	0.80	0.84	0.06
95% Upper	0.88	0.90	0.11
95% Lower	0.72	0.76	0.02

**Table 8.3** Numerical results from the **Disaster Relief** experiment, with bootstrap confidence intervals.

---



**Figure 8.1** Results from the **Disaster Relief** experiment.

A (80%). At the same time, the large majority of those asked about the Blue Team agreed that the Blue Team should go to A (84%).<sup>6</sup>

The key experimental finding is that agreement with (8.50c) in the crucial “Default B” condition was far lower than would be expected, if agglomeration were valid, on the basis of participants’ responses to (8.50a) and (8.50b) in the same condition. Here is why.

Since participants were assigned randomly to conditions, we can use participants’ pattern of responses in one condition to estimate how they *would have responded* if they had been assigned to a different condition. This, in turn, allows us to calculate the expected proportion of participants who would have given any particular pattern of responses in the three conditions (with no memory of having answered similar questions).

Let  $p_a$  be the true underlying proportion of people who would agree with premise (8.50a), and similarly let  $p_b$  and  $p_c$  be the proportion who would agree with premise (8.50b) and the conclusion (8.50c). We can obtain a lower bound on  $p_{ab}$ , the proportion who would agree with both (8.50a) and (8.50b), very simply: no matter what,  $p_{ab}$  cannot be less than  $1 - p_a - p_b$ . This provides an extreme lower bound on  $p_{ab}$ , representing the proportion that would endorse both premises, on the implausibly strong assumption that *everyone* who rejects one premise would have accepted the other. Typically, we expect  $p_{ab}$  to be greater.

If agglomeration is valid, then no one who agrees with both  $p_a$  and  $p_b$  should fail to agree with

<sup>6</sup> An interesting and unexpected finding was that the choice of scenario influenced whether participants agreed with (8.50c) (that both teams should go to Village A). Most participants rejected this prompt in both probability conditions, but significantly more (25%) agreed when Village A was the default location than when Village B was (6%,  $p < .001$ ). I do not know how to explain this result.

$p_c$ . So,  $p_{ab}$ —the proportion who would agree with both (8.50a) and (8.50b)—should be less than or equal to  $p_c$ . Agglomeration gives us a clear lower bound on possible values of  $p_c$ , then:

**Experimental consequence of agglomeration:** the rate of endorsement of (8.50c) should be at least as great as 1 minus the sum of the rejection rates of (8.50a) and (8.50b).

Call this theoretical minimum  $p_{min}^{agg}$ , reflecting the fact that it was derived by using the assumption that agglomeration is valid.

This prediction is strongly falsified by the experimental data in probability condition B. There, 80% agreed with prompt (8.50a), and 84% agreed with prompt (8.50b). Our best estimates  $\hat{p}_a$ ,  $\hat{p}_b$ , and  $\hat{p}_c$  of the true values of  $p_a$ ,  $p_b$ , and  $p_c$  are

- $\hat{p}_a = .80$
- $\hat{p}_b = .84$
- $\hat{p}_c = .06$

Similarly we can estimate  $\hat{p}_{min}^{agg}$ —the predicted minimum possible value for  $p_c$ , calculated from  $\hat{p}_a$  and  $\hat{p}_b$  using the assumption that agglomeration is valid—as

$$\begin{aligned}\hat{p}_{min}^{agg} &= 1 - \hat{p}_a - \hat{p}_b \\ &= (1 - 0.8 - 0.84) \\ &= .64\end{aligned}$$

This is nowhere close to the observed value of  $\hat{p}_c = .06$ . Indeed, the probability that the observed value of  $\hat{p}_c$  would be .06 or lower in an experiment of this size, if the true value were .64, is the same as the probability that a coin with a slight bias towards heads would come up heads 6 or fewer times in 102 tosses—a negligible quantity.

$$\begin{aligned}prob(6 \text{ or fewer "agree" responses in } 102) &= \sum_{i=0}^6 \binom{102}{i} p_c^i \times (1 - p_c)^{102-i} \\ &= \sum_{i=0}^6 \binom{102}{i} .64^i \times .36^{102-i} \\ &\approx 2 \times 10^{-35}\end{aligned}$$

The best possible case for agglomeration would be that our experiment greatly overestimated  $p_a$  and  $p_b$ , and both of the true values fall at the lower edge of the respective 95% confidence intervals. In this case we would still get  $p_a = .72$ ,  $p_b = .76$ . The predicted value of  $\hat{p}_{min}^{agg}$  would then be  $1 - .72 - .76 = .48$ . Even if the gods of experimental chance were strongly aligned against agglomeration, the probability of getting the observed value  $\hat{p}_c = .06$  would be the still-negligible  $2 \times 10^{-20}$ .

The conclusion, in brief, is that our participants do not behave as if agglomeration were valid. In probability condition B, **a majority were inclined to accept both (8.50a) and (8.50b), but to**

**reject (8.50c)**—and this by an analysis that makes a number of assumptions that are extremely generous to the agglomeration principle. The true numbers are probably much greater. The experiment thus indicates that experimental participants do not use *ought* in a way that is even approximately consistent with agglomeration.

## 8.8 Modeling anti-coordination scenarios

The combination of Sloman’s Principle with an expected-value semantics for scalar goodness does rather better. The first thing to note is that this account does not validate agglomeration, trivially, because Sloman’s Principle is a one-way implication. The principle is not capable of validating any *ought*-involving inference, because it only says that *ought* cannot be true of any proposition that is not better than all of its alternatives. However, a theory that meets these conditions can say a bit more about **Disaster Relief**. Sloman’s Principle predicts that agglomeration *could not* succeed in this scenario, and gives a plausible explanation of why. The scenario is constructed to take advantage of a key feature of expected values:  $\mathbb{E}_V(\phi) > \mathbb{E}_V(\neg\phi)$  and  $\mathbb{E}_V(\psi) > \mathbb{E}_V(\neg\psi)$  do not entail  $\mathbb{E}_V(\phi \wedge \psi) > \mathbb{E}_V(\neg(\phi \wedge \psi))$ .

More concretely, consider the case of prompt (8.50a) in the “probably B” condition, where we are reasoning about whether the Red Team ought to go to Village A against the background assumption that going to B is the default action. While we do not know with any precision what probability participants assigned to the Blue Team’s possible actions, their pattern of responses suggests that they thought the Blue Team would probably go to Village B. If this is the case, then the expected value of the possible actions is distributed, very roughly, like this:

$$\begin{aligned}\mathbb{E}_V(\text{RedA}) &= \mathbb{E}_V(\text{RedA} \mid \text{BlueA}) \times \text{prob}(\text{BlueA}) + \mathbb{E}_V(\text{RedA} \mid \text{BlueB}) \times \text{prob}(\text{BlueB}) \\ &= \text{low} \times \text{low} + \text{high} \times \text{high} \\ &= \text{high}\end{aligned}$$

$$\begin{aligned}\mathbb{E}_V(\text{RedB}) &= \mathbb{E}_V(\text{RedB} \mid \text{BlueA}) \times \text{prob}(\text{BlueA}) + \mathbb{E}_V(\text{RedB} \mid \text{BlueB}) \times \text{prob}(\text{BlueB}) \\ &= \text{high} \times \text{low} + \text{low} \times \text{high} \\ &= \text{lowish}\end{aligned}$$

If the sentence under consideration is *Red ought to go to A*, then the alternatives are presumably  $\{\text{RedA}, \text{RedB}\}$ , and *RedA* (*Red goes to A*) is clearly better. Thus Sloman’s Principle allows that *Red ought to go to A* could be true here. By analogous reasoning, *Blue ought to go to A* is compatible with Sloman’s Principle.

However, if we are considering something of the form *It ought to be that Red and Blue both go to A*, the situation is very different. The story clearly indicates the following ordering on combinations of actions, and participants’ responses indicate that they inferred this.

$$\text{RedA} \wedge \text{BlueB} \approx_{\text{good}} \text{RedB} \wedge \text{BlueA} >_{\text{good}} \text{RedA} \wedge \text{BlueA} \approx_{\text{good}} \text{RedB} \wedge \text{BlueB}$$

Since  $\text{RedA} \wedge \text{BlueA}$  is among the worst options, Sloman’s Principle rules out *ought* being true of this proposition under *any* admissible alternative set (given our assumption that the alternatives

must partition the relevant worlds, and include  $RedA \wedge BlueA$ ). This is true even while both *Red ought to go to A* and *Blue ought to go to A* are true. In other words, it is not only logically possible, but required by the particular distribution of probabilities and utilities in **Disaster Relief**, that agglomeration fail.

For example, Figure 8.2 gives a model of this scenario which verifies  $ought(RedA)$  and  $ought(BlueA)$ . In this model, the conclusion that agglomeration would give us— $ought(RedA \wedge BlueA)$ —is ruled out by Sloman’s Principle. This has an important consequence: if anti-coordination scenarios are possible, then unrestricted agglomeration and Sloman’s Principle cannot both hold. Either agglomeration sometimes fails, or there are cases in which  $ought(\phi)$  holds even while  $\phi$  is **worse** than  $\neg\phi$ . Figure 8.3 illustrates the denotation of *ought* in this model.

The **Professor Procrastinate** scenario revealed problems with the classical semantics’ prediction that  $ought(\phi)$  is always true when  $ought(\phi \wedge \psi)$  is. **Disaster Relief** problematizes agglomeration, the property that  $ought(\phi \wedge \psi)$  is true if and only if  $ought(\phi)$  and  $ought(\psi)$  are both true. In certain carefully constructed cases,  $\phi$  and  $\psi$  can both have a higher expected value than their obvious alternatives, even while  $\phi \wedge \psi$  is worse than one or more of its alternatives. The experiment reported in the previous section provided strong evidence that people’s *ought*-judgments are sensitive to these facts: in such cases they are willing to endorse  $ought(\phi)$  and  $ought(\psi)$  while denying  $ought(\phi \wedge \psi)$ . This means that agglomeration does not hold generally. If so, the classical semantics for *ought* is incorrect, as is any other account that validates unrestricted agglomeration.

Note, however, that the solution leaves room for the strongly held—and correct—intuition that there is something right about agglomeration. Exposing the inadequacy of agglomeration as a categorical rule of the semantics required us to set up the elaborate, fairly unusual context of an anti-coordination game. Agglomeration is probably a reasonable default strategy—in the sense that, most of the time, agents are not in the kind of bind that **Disaster Relief** puts them in. Agglomeration yields disastrous predictions in anti-coordination scenarios as a result of agents’ choices interacting in very particular ways, and that they are unable to coordinate actions. In addition, as we will now see, it may be that a restricted form of agglomeration actually is valid.

## 8.9 The Smith argument

Horty (1993; 2003) presents the following challenge to a semantics that denies agglomeration: if  $ought(\phi) \wedge ought(\psi)$  does not entail  $ought(\phi \wedge \psi)$ , how do we explain the intuitive validity of the **Smith argument**?<sup>7</sup>

The argument goes like this. We are to imagine a society where everyone is obligated either to serve in the military or to perform alternative public service. Consider a certain man, Smith. As a member of this society, Smith should either serve in the military ( $M$ ) or perform alternative public service ( $S$ ). But it is also true, of Smith specifically, that he should not serve in the military. (Say, he is a Quaker.) Intuitively, it follows that Smith should perform alternative public service. That is, the following argument is compelling.

(8.51) a.  $ought(M \vee S)$

<sup>7</sup> The name is from Goble (2009), whose discussion I have relied on to a considerable extent.

---

	Blue A	Blue B
Red A	0	1
Red B	1	0

Premises:

- $RedA \in \llbracket \text{ought} \rrbracket^{\mathcal{M},w}_{\{RedA,RedB\}}$
- $BlueA \in \llbracket \text{ought} \rrbracket^{\mathcal{M},w}_{\{BlueA,BlueB\}}$

Consistency of premises with Sloman's Principle:

- $\mathbb{E}_V(RedA) > \mathbb{E}_V(RedB) \quad (.8 > .2)$
- $\mathbb{E}_V(BlueA) > \mathbb{E}_V(BlueB) \quad (.8 > .2)$

Candidate conclusion:

- $RedA \wedge BlueA \in \llbracket \text{ought} \rrbracket^{\mathcal{M},w}_{\{RedA \wedge BlueA, \neg(RedA \wedge BlueA)\}}$

Inconsistency with Sloman's Principle:

- $\mathbb{E}_V(RedA \wedge BlueA) = 0$
- $\mathbb{E}_V(\neg(RedA \wedge BlueA)) = 1 \times \frac{.16}{.36} + 1 \times \frac{.16}{.36} + 0 \times \frac{.04}{.36} = 8/9$
- $\mathbb{E}_V(RedA \wedge BlueA) \not> \mathbb{E}_V(\neg(RedA \wedge BlueA))$

Value information:  $V(w) = \dots$

- 0 if  $w \in (RedA \wedge BlueA)$
- 1 if  $w \in (RedA \wedge BlueB)$
- 1 if  $w \in (RedB \wedge BlueA)$
- 0 if  $w \in (RedB \wedge BlueB)$

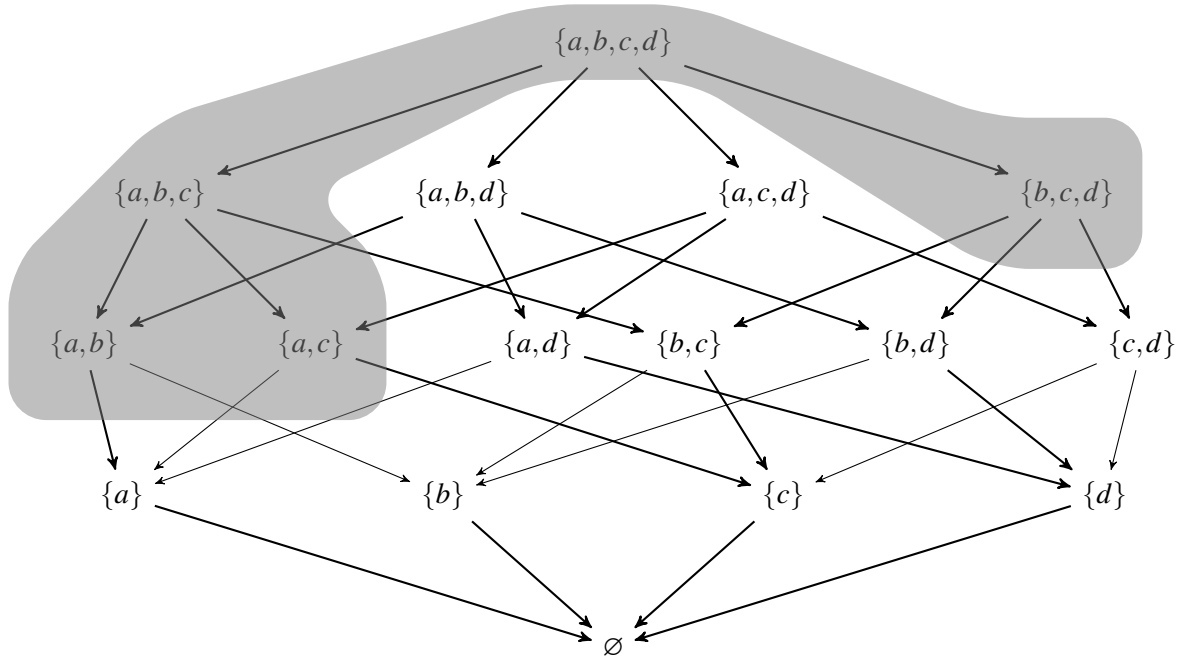
Probability information:

- $P(RedA) = P(BlueA) = .2$
- $P(RedB) = P(BlueB) = .8$
- $P(RedB \wedge BlueB) = .64$
- $P(RedB \wedge BlueA) = .16$
- $P(RedA \wedge BlueB) = .16$
- $P(RedA \wedge BlueA) = .04$

### Figure 8.2

A model of the **Disaster Relief** scenario, for the experimental condition in which the “default” location is *B*. Here Sloman's Principle rules out the truth of the conclusion that agglomeration would yield.

---



$$\begin{array}{llll}
 \{a\} = RedA \wedge BlueA & \{c\} = RedB \wedge BlueA & \{a,b\} = RedA & \{c,d\} = RedB \\
 \{b\} = RedA \wedge BlueB & \{d\} = RedB \wedge BlueB & \{a,c\} = BlueA & \{b,d\} = BlueB
 \end{array}$$

**Figure 8.3**

The gray-shaded area is the denotation of *ought* in **Disaster Relief**, assuming, as the scenario makes clear, that  $ought(RedA \vee BlueA)$  and  $ought(RedB \vee BlueB)$  are true. In a semantics that validates agglomeration, the inclusion of  $\{a,b\}$  and  $\{b,c\}$  would require their intersection  $\{b\}$  ( $= RedA \wedge BlueA$ ) to be in the denotation of *ought* as well. (On the inclusion of  $\{a,b,c,d\}$ , see §8.10 below.)

- b.  $ought(\neg M)$
- c.  $\therefore ought(S)$

However, this argument admits of counter-examples in the very simple semantics we have developed so far—trivially so, since the constraints that we have stated on *ought/should* only tell us what *ought/should* **cannot** be true of, relative to a given comparative goodness relation. Supposing that (8.51) is a valid argument, we should consider some ways of adding constraints to the theory of *ought/should* in order to enforce this consequence.

As Horty (2003); Goble (2009) point out, the problem would not arise if agglomeration were valid: (8.51a) and (8.51b) would entail  $ought((M \vee S) \wedge \neg M)$ , which is logically equivalent to  $ought(S)$ —Smith ought to perform alternative service.

(Note that this statement of the additional assumption does not make reference to alternative sets. For present purposes, I am considering theories that result when we consider the weakest possible alternative sets: evaluating  $ought(\phi)$  relative to  $\mathbf{ALT}(\phi) = \{\phi, \neg\phi\}$ . It makes sense to

concentrate on this interpretation in the present context because it allows us to consider the effects of additional axioms in a maximally context-independent way. More ramified alternative sets will always place stronger restrictions, via Sloman’s Principle, on which sentences *ought* can be true of. It would be interesting to consider stronger theories that could be generated by considering more ramified alternative sets and requiring them to be held fixed when evaluating arguments wherever possible.)

However, we already saw that unrestricted agglomeration generates intuitively incorrect predictions about anti-coordination scenarios, and is in conflict with Sloman’s Principle in these scenarios. One way to get around this problem is to restrict agglomeration in some way that tracks the differences between anti-coordination scenarios and the Smith argument. For example, (8.52b) would do the job. This axiom restricts agglomeration to cases in which the disjunction of the propositions agglomerated is a tautology.

(8.52) Refined constraints on *ought*

- a.  $ought\ \phi \Rightarrow \forall \psi \neq \phi \in \mathbf{ALT}(\phi) : \phi >_{good} \psi$  (Sloman’s Principle)
- b.  $[\phi \vee \psi = W] \wedge ought(\phi) \wedge ought(\psi) \Rightarrow ought(\phi \wedge \psi)$  (Smith Principle)

Figure 8.4 illustrates the consistency of these assumptions with a simple model of the Smith argument. Here,  $M = \text{Smith serves in the Military}$ ,  $S = \text{Smith performs alternative Service}$ , and  $N = \neg(M \cup S)$  abbreviates *Smith does Neither*.

The addition of axiom (8.52b) (the “Smith Principle”) has no effect on anti-coordination scenarios, though. This is because the propositions that we are considering agglomerating in these scenarios—for example, *the Red team goes to B* and *the Blue team goes to B*—are not exhaustive. It is possible for neither to hold (with disastrous consequences for the earthquake victims): see again Figures 8.2 and 8.3.

We could get a stronger theory of *ought* by requiring that agglomeration be valid whenever it would not result in a violation of other constraints on the meaning of *ought*. (For example, we could add to the antecedent of the Smith Principle (8.52b) the condition that Sloman’s Principle is satisfied.)

However, I am inclined to stick with the simple principle (8.52b) for a theoretical reason: later in the chapter I will propose a way to derive this condition from assumptions about the meaning of *ought* that can be given independent motivation from its behavior in comparative and degree modification structures (§8.14). This is a desirable result because it allows us to reduce the number of special meaning stipulations associated with *ought*—perhaps even to prune them down to Sloman’s Principle relating *ought* and *good*, as we will see.

For present purposes, the main takeaway message is that it is probably empirically necessary to validate the Smith argument, as Horty has argued. Fortunately, it is possible to do so without making the mistake of validating agglomeration generally.

## 8.10 Weakening

Cariani (2016) points out that the inference pattern in (8.53), which he calls **Weakening**, is plainly valid.



---

$D =$	$M$	$S$	$N$
-------	-----	-----	-----

Premises:

- (8.51a):  $M \cup S \in \llbracket \text{ought} \rrbracket_{\{M \cup S, -(M \cup S)\}}^{\mathcal{M}, w}$
- (8.51b):  $-M = (S \cup N) \in \llbracket \text{ought} \rrbracket_{\{M, -M\}}^{\mathcal{M}, w}$

Conclusion (by Smith Principle):

- (8.51c):  $\therefore S \in \llbracket \text{ought} \rrbracket_{\{S, -S\}}^{\mathcal{M}, w}$

Requirements of Sloman's Principle:

- $(M \cup S) >_{\text{good}} -(M \cup S)$
- $-M >_{\text{good}} M$
- $S >_{\text{good}} -S$

Sample model:

- $V(w) = 1$  for all  $w \in S$
- $V(w) = 0$  for all  $w \in M$
- $V(w) = 0$  for all  $w \in N$
- $P(M) = P(S) = P(N) = 1/3$

Consistency with Sloman's Principle:

- $\mathbb{E}_V(M \cup S) > \mathbb{E}_V(N)$  (.5 > 0)
- $\mathbb{E}_V(-M) > \mathbb{E}_V(M)$  (.5 > 0)
- $\mathbb{E}_V(S) > \mathbb{E}_V(-S)$  (1 > 0)

**Figure 8.4**

A model for the Smith scenario that conforms to the *ought* axioms in (8.52).

---

- (8.53) a.  $ought(\phi)$   
 b.  $ought(\psi)$   
 c.  $\therefore ought(\phi \vee \psi)$

While it is reasonable in some cases to hold  $ought(\phi)$  and  $ought(\psi)$  while denying  $ought(\phi \wedge \psi)$ —anti-coordination scenarios, for instance, or normative dilemmas—the inference to  $ought(\phi \vee \psi)$  is intuitively reasonable in both kinds of scenarios, and no other counter-examples come to mind.

This inference pattern is interesting because, as Cariani shows, quite a few theories of *ought* fail to validate it—including all of the ones that we considered above as ways to define *ought* in terms of scalar goodness, if the latter is cashed out as expected value. For these theories, Cariani give the counter-model described in Table 8.4.

world	probability	value
$A \wedge B$	.25	100
$A \wedge \neg B$	.25	-50
$\neg A \wedge B$	.25	-50
$\neg A \wedge \neg B$	.25	0

**Table 8.4** Cariani’s counter-model to **Weakening** for theories that equate *ought* with *better than not* or *best among the alternatives*.

Here we have  $\mathbb{E}_V(A) = 25$ , which is greater than  $\mathbb{E}_V(\neg A) = -25$ , and also greater than all of the relevant alternatives (with values  $-50$  and  $0$  respectively). In addition,  $\mathbb{E}_V(B)$  is  $25$ , which is greater than  $\mathbb{E}_V(\neg B) = -25$  and greater than the alternatives (also with values  $-50$  and  $0$ ). So, any of the attempts to define *ought* from expected value that we considered above will render both *ought A* and *ought B* true. Nevertheless, they will render *ought (A or B)* false:  $\mathbb{E}_V(A \vee B) = 0$ , which is the same as  $\mathbb{E}_V(\neg(A \vee B)) = \mathbb{E}_V(\neg A \wedge \neg B)$ .

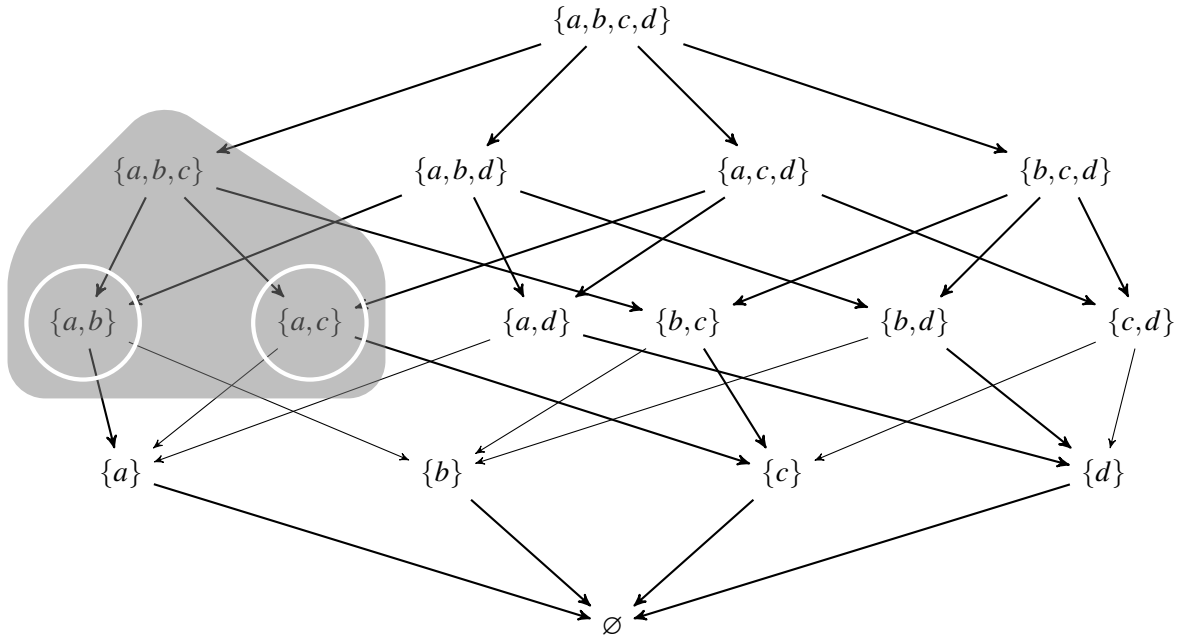
In the case of the account currently under consideration, the situation is different. Since  $\mathbb{E}_V(A \vee B)$  is the same as  $\mathbb{E}_V(\neg(A \vee B))$ ,  $ought(A \vee B)$  cannot be true by Sloman’s Constraint. However, nothing in the theory tells us whether *ought(A)* and *ought(B)* are true or false: any combination of truth-values for these statements is consistent with the two constraints in (8.52).

Since Weakening is intuitively valid and consistent with our other assumptions, we can simply add it as a further constraint on the meaning of *ought/should*. This gives us the refined constraint set in (8.54b).

- (8.54) Further refined constraints on *ought*
- a.  $ought(\phi) \Rightarrow \forall \psi \neq \phi \in \mathbf{ALT}(\phi) : \phi \succ_{good} \psi$  (Sloman)
  - b.  $[(\phi \vee \psi) = W \wedge ought(\phi) \wedge ought(\psi)] \Rightarrow ought(\phi \wedge \psi)$  (Smith)
  - c.  $[ought(\phi) \wedge ought(\psi)] \Rightarrow ought(\phi \vee \psi)$  (Weakening)

Adopting (8.54c) has an interesting consequence for the problematic model described in Table 8.4. Since the truth of  $ought(A \vee B)$  is ruled out by Sloman’s constraint, (8.54c) allows us to infer, by





**Figure 8.6**

An example where *ought* holds of  $\{a, b\}$  and  $\{a, c\}$ . Weakening lets us deduce that it also holds of  $\{a, b, c\}$ , but—since  $\{a, b\} \cup \{a, c\} \neq W$ —the restricted form of agglomeration in (8.54b) does not apply, and *ought* may be false of  $\{a\}$ .

details.

### 8.11 Conflicting *oughts* and *shoulds*

I've promised to pick up my sister at the airport, and I've also promised to go to my friend's concert. I've just discovered that my sister is arriving during the concert. What am I to do? Here we have conflicting *oughts*: assuming that I ought to fulfill my promises, I ought to pick up my sister and I ought to go to the concert. Unfortunately, I can't do both.

Such conflicts are real, and yet standard semantic theories render it an *a priori* truth that they cannot exist. On the standard analysis, *ought*  $A$  is true if and only if  $A$  is true in all of the “best” or “ideal” worlds. On this assumption, there is a simple proof of the non-existence of deontic conflicts. If  $\phi$  is true in all of the ideal worlds, and  $\psi$  is true in all of the ideal worlds, then  $\phi \wedge \psi$  is true in all of them. But since a deontic conflict is (by definition) a situation in which *ought*( $\phi$ ) and *ought*( $\psi$ ) hold but  $\phi \wedge \psi$  is true in no worlds in the domain, *ought*( $\phi \wedge \psi$ ) can't hold in all of the best worlds in the domain. So, there can be no conflicting *oughts*.

More generally, any semantic theory on which agglomeration is unrestrictedly valid will rule out the possibility of conflicting *oughts* on semantic grounds. This is wrong. Since the theory that we have been developing does not validate unrestricted agglomeration, it is worth considering whether it does better.

So far we are assuming a restricted version of agglomeration (the “Smith Principle” (8.54c)), where  $ought(\phi) \wedge ought(\psi)$  necessitates  $ought(\phi \wedge \psi)$  only when  $\phi \vee \psi$  is a tautology. Predictions about conflicting *oughts* follow immediately: they are semantically possible, but only when the options are not exhaustive. For example, in the scenario described above, going to my friend’s concert and picking my sister up are not the only possibilities. I could also, perversely, decide to do neither. As a result, the argument in (8.55) is not valid. However, the argument in (8.56) is valid, via Weakening. This is also intuitively correct: surely I ought to do one or the other.

(8.55) Invalid by (8.54):

- a. I ought to pick up my sister.
- b. I ought to go to my friend’s concert.
- c.  $\therefore$  I ought to (pick up my sister **and** go to my friend’s concert).

(8.56) Valid by (8.54):

- a. I ought to pick up my sister.
- b. I ought to go to my friend’s concert.
- c.  $\therefore$  I ought to (pick up my sister **or** go to my friend’s concert).

While this is interestingly different from the classical semantics, it is not yet a general solution to the problem of conflicting *oughts* (*pace* Lassiter 2011a: §6). The basic problem is that the constraints under consideration continue to predict that there can never be situations in which both  $ought(\phi)$  and  $ought(\neg\phi)$  are true. If they were, the Smith Principle would allow us to derive  $ought(\phi \wedge \neg\phi)$ . This is intuitively absurd. It is also in conflict with Sloman’s Principle, which would then imply that a contradiction—which contains no worlds, and so has no expected value—have greater expected value than a tautology. This prediction seems incorrect. Intuitively, there are cases in which  $ought(\phi)$  and  $ought(\neg\phi)$  are both true: I might have a compelling reason from one source to act to bring about  $\phi$ , and an equally compelling reason from a different source to act to bring about  $\neg\phi$ . (See Horty 2014 for an excellent discussion of such cases with reference to the classical semantics.) Consider, for example, the following passage from a *BBC News* story (5 January 2008):

Francesc Vendrell is the EU’s Special Representative in Kabul and I asked him what Michael Semple was doing? “Quite honestly I am not sure. I had authorised Michael to go to Helmand, I knew he was vaguely going to do some work on reconciliation with the Taleban, but beyond that I had absolutely no idea what he was going to do.”

But as his deputy, should he have known what Michael was doing? ... “**I should and I shouldn’t**. I didn’t feel that someone with the background of Michael could be kept on a short leash.”

In the context of the reporter’s question, Vendrell’s statement seems to indicate that he has two conflicting—and distinct!—sources of *prima facie* obligation. On the one hand, he should be aware of the whereabouts and activities of an employee under his direct supervision. On the other hand, he should not be too heavy-handed in controlling the activities of a person with the background and experience of Michael Semple (“one of the West’s most respected experts on Afghanistan”, according to the same article).

The key question is whether we should attempt to model these conflicting sources of obligation within a single, coherent system of value, or if we should allow for the existence of multiple conflicting systems of value together with some kind of conflict resolution mechanism. This is

closely related to two kinds of familiar issues, one from deontic logic and one from degree semantics. In deontic logic, issues around integrating multiple sources of obligation have been studied in much detail, notably by John Horty in a long line of work (1993; 2003; 2012). Working in a somewhat different formal framework from the intensional semantics assumed here, Horty provides a theory in which reasons function as default rules, and a nonmonotonic logic is used to model the resolution of conflicts among default rules.

There are many ways that we could try to achieve a similar effect within an intensional semantics. One way is to modify the classical definition so that agglomeration is not valid when there are conflicting norms or reasons in force. (See the next section for an evaluation of one such proposal.) Another is to try to defuse the problem by attending to the ways that multiple sources of information—e.g., conflicting values or reasons—can be combined into a coherent whole. Here is one way we might try to handle this issue within the metasemantics, inspired by the treatment of multidimensional adjectives.

The interpretation of multi-dimensional adjectives like *big* and *clever* has received a considerable amount of discussion in the literature on scales (e.g., Kamp 1975; Sassoon 2013). Clearly, people can be clever in many different ways. It would not be surprising if  $x$  were more clever than  $y$  in one respect—say, book smarts—while  $y$  is more clever in another respect—say, street smarts. How, then, can we integrate the multiple conflicting sources of information (themselves typically represented by scalar properties) into a single scale of cleverness?

While it is very plausible that multidimensional adjectives have non-connected scales (cf. Kamp 1975), simply declaring that this is so is not enough. Just as in the deontic case, we need some way to model conflict resolution: when different kinds of cleverness give us different comparative judgments, we need to know how to combine them into a single judgment. A satisfying approach will provide a procedure to construct the all-things-considered cleverness scale from multiple relevant kinds of cleverness. This means that  $\mathcal{S}_P$  will be constructed from a set of subscales  $\mathbb{S}_P = \{\mathcal{S}_P^1, \mathcal{S}_P^2, \dots, \mathcal{S}_P^n\}$ . One straightforward way to do this is to universally quantify over subscales. We can define “ $x$  is at least as  $P$  as  $y$ ” as a relation that holds if and only if  $x$  is at least as  $P^i$  as  $y$  for every subscale  $\mathcal{S}_P^i$ .

$$x \geq_P y \Leftrightarrow \forall \mathcal{S}_P^i \in \mathbb{S}_P : x \geq_{P^i}^i y.$$

This rule gives us a very strong theory of multidimensional adjectives. For example, Mary cannot be as clever as Bill unless she is as clever as he is in *every* relevant respect. If he is more clever in even one respect, *Mary is as clever as Bill* will be false—even if she is much more clever in every other relevant respect. So, this rule will tend to generate scales with a high degree of incomparability.

A second way of integrating multiple sources of information is to construct a weighted sum of the degrees assigned to individuals by each subscale. The weight of a subscale,  $weight(\mathcal{S}_P^i)$ , represents how important that subscale is to the final result.

$$\mu_P(x) = \sum_{i=1}^{|\mathbb{S}|} \mu_P^i(x) \times weight(\mathcal{S}_P^i)$$

Unlike the first account, this method will always generate a connected scale (assuming that the underlying subscales are connected). If Mary is more clever than Bill in every relevant respect save

one, she will usually be more clever than him all things considered—unless the one respect in which he is more clever happens to receive inordinate weight. There are many and more complex ways that we could encode multidimensional scales, and there is much research investigating which of these are psychologically relevant. (See, for example, Markman 1998.)

Crucially, if we take the latter approach we have to allow that the weight function can vary depending on the topic of conversation and other pragmatic factors. For example, if we are in the midst of a conversation about school and you ask me

(8.57) Who is more clever—Mary or Bill?

it would not be unreasonable to answer simply “Mary”, with no hedges about different kinds of cleverness. This is because Mary’s book-smarts are clearly the more interesting kind for current purposes. In contrast, if we were talking about success in door-to-door sales, “Bill” would probably be the appropriate answer. We can model this shift by supposing that the same scalar properties are relevant to *clever* in both contexts, but that the weight function can vary in context. The school context assigns high weight to book smarts and low weight to street smarts. The sales context reverses this relationship.

Now consider a context in which it is not clear what kind of cleverness is relevant. If you were to ask simply

(8.58) Mary is more clever than Bill, right?

I might answer with (8.59)—

(8.59) She is and she isn’t.

—intending to communicate that there is a relevant way of resolving “clever” on which she is indeed more clever, and a relevant way of resolving clever on which she is not.

My suggestion here is that this way of reasoning about multidimensionality may also give us a way of modeling conflicting *oughts*. In examples where *ought*( $\phi$ ) and *ought*( $\neg\phi$ ) are asserted side-by-side, what is conveyed is that there are (at least) two relevant sources of value, and each token of *ought* is responsive to a weighting that privileges one of them. For example, recall the instance of *I should and I shouldn’t* quoted above from a news article. The effect was that Francesc Vendrell *should* be keeping close tabs on Michael Semple, being his boss; but he *should not* be keeping close tabs on him, given Semple’s expertise and experience. There are multiple sources of value here which generate conflicting recommendations, and the question—

(8.60) Shouldn’t you know what Michael is doing?

—does not specify which of these conflicting sources of value is the relevant one. So, it may be possible to explain Vendrell’s response (8.61) in the same way that we accounted for (8.59).

(8.61) I should and I shouldn’t.

When sources of obligation related to being a superior at work receive high weight, “I should” comes out as true. When sources of obligation related to respect for the autonomy of experts receive high weight, “I shouldn’t” comes out as true. The shrug implied by Vendrell’s response suggests that he is unsure which of these sources of obligation should be given more weight.

While I find the intuition behind this comparison compelling, it remains just an intuition at this point: we do not yet have a theory of *ought* in which it makes reference to a scale that can be directly compared to *clever*. However, in §8.13 I will use data involving degree modification and comparison structures to motivate a scalar analysis of *ought/should*. On this account, these are gradable verbs—like *matter*, *like*, and *hate*—and their meaning makes direct reference to the position of a proposition on a scale. If this theory is right, it may be possible to apply the analysis of multidimensionality described in this section directly to account for conflicting *oughts*.

In sum, then, it may be possible to assimilate the treatment of conflicting *ought* sentences like (8.61) to the treatment of apparently conflicting assertions involving multidimensional adjectives, like (8.58) and (8.59). There is little temptation to account for (8.59) semantically, e.g., by allowing that *x is more clever than y* and *y is more clever than x* can be simultaneously true. Instead, the phenomenon seems to involve a subtle shift in meaning, and one that makes pragmatic sense. Similarly, I suggest that we should not weaken the semantics to make room for the simultaneous truth of *ought*( $\phi$ ) and *ought*( $\neg\phi$ ). Instead, we should allow that there are various, possibly conflicting sources of value that are potentially relevant to *ought/should*. Conversational pragmatics plays a key role in determining how these sources are weighted in determining the meaning of a given token of *ought*.

## 8.12 Evaluating a revised classical theory of *ought*

While we are on the topic of conflicting *oughts*, it is worth pausing to consider the prospects for a revised theory of *ought* that has recently been suggested as a solution to this problem within the general framework of the classical semantics.

On the theory of Lewis (1973); Kratzer (1991), *ought*( $\phi$ ) holds iff  $\phi$  holds in every world that is maximal in the ordering, i.e., every world for which there is no world that is strictly better. (Note that we are continuing to simplify intuitions by making the limit assumption.) Based on discussion in this chapter, we can identify numerous problems for this approach:

- The classical semantics validates agglomeration, and so
  - rules out conflicting *oughts*, and
  - makes incorrect predictions about anti-coordination scenarios.

(The problem about conflicting *oughts* could, of course, be massaged using a pragmatic approach analogous to the one that I proposed in the previous section. This would not help with anti-coordination scenarios, though.)

- The classical semantics is upward monotone, and so predicts the impossibility of examples where *ought*( $\phi \wedge \psi$ ) holds while *ought*( $\phi$ ) does not (e.g., **Juliet, Professor Procrastinate**).
- The classical semantics incorrectly validates deontic detachment: the inference from *ought*( $\phi$ ) and *If  $\phi$ , ought*( $\psi$ ) to  $\therefore$  *ought*( $\psi$ ) is valid. This inference was counter-exemplified by several scenarios, most dramatically the **Nazi Bombing** example.



- The classical semantics encounters theoretical difficulties in attempting to account for the Miners' Puzzle and other cases of information-sensitivity. While the theory has room to model the intuitive judgments in some cases, it lacks a well-motivated account of why non-categorical shifts in probability would have the semantic effects that that do.
- The theory of scalar goodness most naturally associated with the classical semantics is empirically incorrect, because it assumes that goodness is maximal (ch.7). This is problematic because there are significant semantic connections between *ought* and *good*, and it is not clear that there is an adequate way to incorporate into the classical semantics in a way that coheres with an empirically adequate theory of scalar goodness.

However, in the case of conflicting *oughts* there may be a way out: an account first described in van Fraassen 1973 has recently been discussed as a conservative revision to the classical semantics which makes room for conflicting *oughts*. This account does not validate agglomeration, and may even help with some of our other objections to the classical semantics.

Since the orderings that the classical semantics invokes are not required to be connected, there is another way to state its truth-conditions: *ought*( $\phi$ ) holds iff, for every set of worlds  $S$  that are maximal (and so tied) within their respective branch of the ordering, every world in  $S$  is a  $\phi$ -world.

The classical semantics rules out conflicting *oughts* as a consequence of validating agglomeration. However, there is a way to modify it that can do better. The idea is to define *ought*( $\phi$ ) as holding whenever there is *some* set of worlds  $S$  that are maximal in some branch of the ordering, and every world in  $S$  is a  $\phi$ -world. (See van Fraassen 1973; Horty 1993, 2014; von Fintel 2012. There are some subtleties about how best to state the definition when we relax the limit assumption: see Swanson 2011.)

This weakening of the classical *ought* leaves room for conflicting *oughts*. Suppose that the ordering source at world  $w$  gives us  $g(w) = \{A, B\}$ , where  $A \cap B \cap f(w) = \emptyset$ . Then the ordering  $\succsim_{g(w)}$  will not be connected: one branch will have a set of maximal  $A$ -worlds, and one branch will have a set of maximal  $B$ -worlds. So, *ought* is true of both  $A$  and  $B$ , even though it is not true of  $A \wedge B$ . The revised theory does not validate agglomeration, and so it does not rule out conflicting *oughts*.

Since agglomeration was the troublemaker in anti-coordination scenarios, we might expect that this solution would resolve that problem as well. For example, consider a model where  $\succsim_{g(w)}$  has two maximal branches—one where *Red to A and Blue to B* is true throughout, and one where *Red to B and Blue to A* is true throughout. The maximal worlds in these branches are incomparable to each other, and both sets dominate worlds where both teams go to A, or both to B. In this model, both *ought*(*Red to A*) and *ought*(*Blue to A*) are true, but *ought*(*both to A*) will be false. The revised semantics thus make room for the crucial experimental result discussed above (§8.7): high rates of agreement, in the condition where the default location was  $B$ , with both *ought*(*Red to A*) and *ought*(*Blue to A*) despite near-zero agreement with *ought*(*both to A*).

However, using the revised semantics to model anti-coordination cases in this way generates a different problematic prediction. In the model described, *ought*(*Red to B*) and *ought*(*Blue to B*) are true as well in the same model—i.e., that both teams ought to go to both locations, regardless of default status. However, this prediction was tested in the experiment, modulo the label given

to the locations: we found much lower rates of agreement with *ought*(Red to A) and *ought*(Blue to A) when A was the default location (on average, 82% for the non-default location vs. 36% for the default location; bootstrap  $p$ -value  $< .000001$ ). But in the revised semantics *ought*(Red to A), *ought*(Blue to A), *ought*(Red to B), and *ought*(Blue to B) are all equally true. The revised theory's judgment apparently does not take into account the crucial probabilistic information which enables participants to distinguish among these options in their *ought*-judgments.

In weakening the classical semantics to make room for conflicting *oughts*, the revised theory seems to go too far by treating all failures of agglomeration as deontic conflicts. What is missing, I suggest, is a semantic hook for non-categorical probabilistic information to influence the judgment. The theory sketched in this chapter, on the other hand, provides just this.

The revised theory of *ought* improves on the classical semantics in one further respect: it does not validate deontic detachment. Consider a model in which there are two branches  $C$  and  $D$ .  $\phi \wedge \neg\psi$  is true in the maximal worlds in  $C$ , and  $\neg\phi \wedge \neg\psi$  is true in the maximal worlds in  $D$ . Then *ought*( $\phi$ ) is true, and *ought*( $\psi$ ) is false. The conditional *If*  $\phi$ , *ought*  $\psi$  can still be true in such a model, if all of the best  $\phi$ -worlds in branch  $D$  are  $\psi$ -worlds. Then, restricting  $\succsim_{g(w)}$  to worlds where  $\phi$  holds would “expose” the  $\phi \wedge \psi$ -worlds in branch  $D$ . Thus *ought*( $\phi$ ) and *if*  $\phi$ , *ought*  $\psi$  can be true while *ought*( $\psi$ ) is false.

While failure to validate deontic detachment is a good result, it is not clear that the model just sketched is a reasonable model for the **Nazi Bombing** scenario. There, the problem was not that there were two sets of “best” worlds that we were unable to compare. Rather, there was a single, clearly best outcome—cessation of bombing, and return of children to their families—and the problem was that our choices interacted with in an unpredictable way with factors outside of our control. It remains a challenge for a proponent of the revised semantics of *ought* to give a model that accounts for intuitive failures of deontic detachment in a plausible way.

In any case, the revised theory of *ought* has similar problems to the original in several other respects. The revised account renders *ought* upward monotone, and so inherits the problems around the **Juliet** and **Professor Procrastinate** cases. In addition, as I noted above in the context of anti-coordination scenarios, the revised semantics remains too close to the classical account in that it lacks a mechanism for incorporating non-categorical shifts in information. This means that the theory will continue to struggle in providing a satisfying account of puzzles involving information-sensitivity, including the Miners' Puzzle as well as several more that were mentioned briefly above—e.g., Jackson's (1991) Medicine puzzle and a deontic analogue of Levinson's (2003) Insurance puzzle (cf. Lassiter 2011a: §5). Finally, the revised version continues to face a challenge in encoding semantic connections between *ought* and *good/better*. Perhaps the improved theory of scalar goodness given in chapter 7 could be linked semantically to the revised classical semantics in some way, but the details are exceedingly unclear at present.

Finally, both the classical semantics and the conservative revision discussed in this section encounter what seems to me an insuperable problem: both treat *ought* as denoting a first-order quantifier ( $\forall$  or  $\exists$ ). Since these quantifiers do not have scalar meanings, there is no obvious way to introduce a degree argument into the proposed interpretations for *ought/should*. So, a semantics with this format seems to predict that *ought* and *should* should be unambiguously non-gradable. However, as we will see in the next section, there is good reason to think that these verbs are indeed

gradable. If so, there is little hope that a conservative modification of the classical semantics will be able to capture the grammatical behavior of *ought* and *should* and the full nuance of their meanings.

### 8.13 Gradability of *ought* and *should*

So far we have been talking about the simple, unmodified forms of *ought* and *should*, which we have been treating as monadic predicates of propositions. While this is customary, I will argue that they are better viewed as scalar predicates. The item that we—and most of the voluminous literature on *ought*—have so far focused on is simply the positive form of this scalar predicate. The reason for making this move is squarely empirical: *ought/should* display the same grammatical behavior as gradable verbs like *like*, *matter*, *care*, *want*, and *need*. In order to model this behavior, we need to treat the basic form of *ought* as a scalar predicate associating propositions  $\phi$  with “the degree to which  $\phi$  ought to hold”.

Gradable verbs, like their adjectival counterparts, can be diagnosed by their participation in degree modification, comparative, and equative structures. Here are some naturalistic examples involving the gradable verbs *like* and *matter*.

(8.62) *Like*:

- a. Tropical Duck Army! Do you like it as much as I do?? (equatives)
- b. Do you occasionally have to send him multiple messages before he decides to answer you back? If so, then it's a sign that you like him more than he likes you. (comparatives)
- c. Juan Uribe doesn't seem to like football very much. (degree modification)

(8.63) *Matter*:

- a. When people matter as much as great food, come to Nanna's. (equatives)
- b. Looks: Why They Matter More Than You Ever Imagined (comparatives)
- c. Peter Carruthers argues that phenomenal consciousness might not matter very much ... for the purpose of determining which nonhuman animals are appropriate objects of moral sympathy ... (degree modification)

The general approach to scalar semantics developed earlier in chapters 1-2 of this book extends readily to these constructions. For example, *matter* denotes a function from individuals (events, etc.) to the degree to which that individual matters. *Like* denotes a function from pairs of individuals ( $x, y$ ) to the degree to which  $x$  likes  $y$ . Both of these have qualitative counterparts:  $S_{matter}$  is a scale of individuals (like *heavy* and *happy*), while  $S_{like}$  is a scale of pairs of individuals, ordering pairs in terms of the degree to which the first member likes the second. This introduces a compositional puzzle in the unmodified form (*Bill likes Mary*, etc.). This puzzle is precisely analogous to the compositional puzzle surrounding positive-form scalar adjectives: if *happy* picks out a function from individuals to heights, some additional mechanism is needed to convert it into a predicate in sentences like *Bill is happy*. The puzzle admits of a range of candidate solutions that we discussed in chapters 1-2: assuming a silent *pos* morpheme, adding some type-shifting operations, or adopting a delineation semantics. The same range of solutions is available in the case of scalar verbs, and any of them will do for current purposes.

*Ought* and *should* are also gradable verbs, participating in equative and comparative constructions and allowing degree modification. Here are some examples involving *ought*:

- (8.64) a. Most opponents of public-private combinations fear the private side, worrying that the cost-cutting and profit-maximizing impulses of the business ethos will hurt the public sector's ability to serve the greater good. Jacobs's warning is different. It holds that commerce ought to avoid guardian values as much as guardians ought to avoid commercial values. (equatives)
- b. [O]nce the damage is done, Constance ought to help George—or, at least, she ought to help him more than she ought to help anyone else similarly situated. (Driver 1997: 853) (comparatives)
- c. British Royals ought to be beaten with sticks now, even more than they ought to have been in 1776, and driven from our shores with condemnation. (comparatives)
- d. Autonomy absolutely does not strike me as “beyond the pale” when it comes to either abortion or pornography. Indeed, there are situations in which concerns of autonomy ought very much to matter (as they did to Justice Brennan). (degree modification)

Of course, it is a *prima facie* possibility that the degree operators in these examples could be associating with the main verbs, rather than the modal verbs. But this hypothesis encounters difficulties. First, compositional problems arise for the comparative and equative examples, since the modal verb appears twice. Even if this can be accounted for, this strategy for explaining away these examples fails because it generates the wrong interpretations. For example, (8.64a) is explicitly equating two *ought*-degrees:

- the degree to which people in commerce ought to avoid adopting “guardian” (politician/civil servant/military/etc.) values; and
- the degree to which the latter group ought to avoid adopting commercial values (competition, thriftiness, etc.).

(all according to urban studies theorist Jane Jacobs). What is definitely *not* under consideration is some kind of “degree of avoiding” (whatever this would mean), or the amount of avoidance behavior that is recommended. (8.64b) is taken from a moral philosophy paper that deals with obligations derived from considerations of personal history rather than impersonal moral reasoning. In this context, the intended comparison involves the amount of *obligation* that Constance has toward various people, given her history of personal interactions with them—rather than the amount of *helping* that she owes them.

Along similar lines, (8.64c) compares the degrees to which a certain group *ought* to be beaten with sticks at various times. An interpretation in terms of amount of beating would make little sense here: the point of the example is that it is even more incumbent on us now to perform this action than it was on our ancestors in 1776. In (8.64d), the interpretation “ought to (matter very much)” would be sensible—but it does not seem to capture the intent of the example, given the previous context.

Gradable uses of *should* are also not too hard to find. Again, a careful reading of these examples should help to rule out the conceivable interpretations in which the degree operators take the main verbs as arguments.

- (8.65) I don't think he [UFC fighter Phil Davis] should be compared to Rosholt as much as he should be to Houston Alexander. (equative)
- a. ✓ "For our purposes, a comparison to H.A. is more useful than one to R."
- b. ?? "A greater amount of comparison is needed"
- (8.66) C G P Grey has made three simple, explanatory videos, which are all worth watching by electoral reformers and each one is followed by on-line debate. ... Non-reformers should watch them even more than reformers should. (comparative)
- a. ✓ "It is more incumbent on non-reformers to watch the videos"
- b. ?? "Non-reformers should spend more time watching the videos"
- (8.67) Oh, and I reverted the change to  $\Pi$ , because it's the same article as  $\pi$ , which should link here, even more than  $\pi$  should link to any of the product related pages. (comparative)
- a. ✓ "It's more important for the article on  $\pi$  to link here"
- b. ?? "The article on  $\pi$  should have a greater degree of linkage"
- (8.68) If you desperately need to change an old post then PM one of the moderators ... This should very much be considered the exception though. The normal edit window usually be enough. (degree modification)
- a. ✓ "It's very important that this be considered the exception"
- b. ?? "It's important that this be (very much/often/???)considered the exception"

Like other gradable verbs, then, *ought* and *should* must relate their overt arguments to a degree—here, a degree of obligation or similar. In qualitative terms, this means that there is a scale  $S_{ought}$  which orders propositions in terms of the degree to which they ought to hold. Furthermore, the unmodified form of *ought/should*—the form that this chapter has discussed in detail—is the positive form. From what we know about other positive-form scalar expressions, we expect unmodified *ought/should* to be derived from the root with some additional compositional help, and that their meaning involves meeting or exceeding a possibly context-sensitive threshold  $\theta_{ought}$ .

What are the degrees underlying the meaning of *ought/should*, then? They could be many things: degrees of obligation for moral *oughts*, degrees of expediency for teleological *oughts*, and so on. But, in keeping with the broader themes of this book, I would like to focus on the interpretation of the positive form (whether it is analogous to minimum, relative, maximum, or extreme adjectives, for example) and the structural aspects of the scale. Are the scales ordinal, interval, ratio, or something else? Upper- and/or lower-bounded? Positive, maximal, intermediate, or something else? Here I will offer only some brief suggestions.

The first thing to note is that *ought*-degrees are definitely not the same as degrees of goodness (*pace* Lassiter 2011a: §6). The possibility of supererogation already shows this: it is possible that I ought to visit my friend, and it would be even better to cook her dinner while visiting, but it's not the case that I ought to visit *and* cook her dinner. So, there can be propositions  $\phi$  and  $\psi$  such

that *ought* is true of  $\phi$  and false of  $\psi$ , even though  $\psi$  is better than  $\phi$ . Assuming that positive-form *ought* means “having a degree of obligation greater than  $\theta_{ought}$ ”, it follows that these two scales are not identical, or even in a monotonic relationship. The relationship is perhaps better analogized to the relationship between *big* and *tall*. While relative height is a good clue to relative size, it is not a perfect predictor:  $x$  can be taller than  $y$  even while  $y$  is bigger.

The next thing to consider is whether positive-form *ought* is analogous to the relative, minimum, maximum, or extreme adjectives (or something else). Extreme status seems doubtful, especially since this role may already be occupied by deontic *must* (cf. Portner & Rubinstein 2015 and §6.6 above). As for maximum status, a straightforward test is the acceptability of the “A but could be A-er” construction (Kennedy & McNally 2005a).

- (8.69) a. ?? This glass is full, but it could be fuller.  
 b. You ought to wash the dishes, but you ought to apologize to your sister even more.

While (8.69a) is quite odd, (8.69b) seems like a straightforward way to communicate that positive-form *ought* holds of both propositions, even while the latter is associated with a greater degree. Here is a naturalistic example of the relevant construction (with *should*):

- (8.70) Hey folks! Join the hilarious Brian and Lisa Anderson ... for their live broadcast of their podcast. You should already do this, but you should do it even more so because... Whoop whoop! That’s right! I’m the guest!!

This leaves minimum and relative, if we are to analogize the positive form of *ought/should* directly to known categories of positive-form adjectives. Of these two possible classifications, I favor a relative classification. First, relative adjectives such as *heavy*, *tall*, *happy* are sensitive to comparison classes, and the relative modal adjectives *likely/probable* and *good* are sensitive to alternatives and focus (§§4.2.2 and 7.9). So, a relative classification for *ought* would predict that the threshold value associated with *ought* should be similarly sensitive to contextual alternatives. If so, *ought* could hold of a proposition  $\phi$  or not in some context depending on the alternatives that  $\phi$  is being compared to, even while  $\mu_{ought}(\phi)$  is held fixed.

Recall the example that we saw in §7.9 showing that, relative to a fixed scenario, the truth-values of *It is good that Clyde MARRIED Bertha* and *It is good that Clyde married BERTHA* can diverge. Similarly, the kind of alternative-sensitivity that we are envisioning would be the result of variability in the value of  $\theta_{ought}$  that the positive form meaning references, induced by changes in the alternative set.

- (8.71) a. Clyde ought to MARRY Bertha.  
 b. Clyde ought to marry BERTHA.

I can easily imagine assenting to (8.71a) while disagreeing with (8.71b). This may support the idea that *ought/should* are relative-standard scalar expressions.

Analogizing *ought/should* to the minimum adjectives may seem to make somewhat better sense on conceptual grounds: after all, if it is clear that you should not do *A* or *B*, it would seem odd to ask which one you ought to do more. This contrasts with typical relative adjectives such as *heavy*, where there is no oddity in asking which of two light objects is heavier. However, this feature of *ought/should* has a counterpart in the minimum adjectives, where it is strikingly odd to ask,

for example, which of two dry towels is wetter. However, I wish to tentatively reject a minimum classification for *ought/should*, simply because it would give *ought* some odd properties when combined with certain properties of  $\mathcal{S}_{ought}$  that I will argue for below. This is a tentative conclusion, however, and worth revisiting.

If *ought/should* is relative-standard, its scale may or may not be lower-bounded (assuming that we were correct to reject Interpretive Economy in chapter 4). Is this scale upper-bounded? In other words, is there some fixed point such that it is conceptually impossible that anything could be more important (deontically, teleologically, etc.)? My hunch is that the answer is negative, but it is hard to be sure.

Another obvious question to ask is how  $\mathcal{S}_{ought}$  interacts with disjunction. If we restrict attention to the three possibilities that we have discussed at length in this book—maximality, positivity, and intermediacy—only intermediacy is plausible. For, if  $\mathcal{S}_{ought}$  were maximal or positive, the following inference would be valid:

- (8.72)    a.  $\mu_{ought}(\phi) > \theta_{ought}$  (scalar interp. of *ought*( $\phi$ ))  
             b.  $\mu_{ought}(\phi \vee \psi) \geq \mu_{ought}(\phi)$  (maximality or positivity)  
             c.  $\therefore \mu_{ought}(\phi \vee \psi) > \theta_{ought}$  ((a), (b), transitivity)

(8.72a) is the scalar interpretation of positive form *ought*. (8.72b) follows because  $(\phi \vee \psi) \geq_P \phi$  holds whenever  $\mathcal{S}_P$  is a positive or maximal scale. But then we would be able to infer  $\mu_{ought}(\phi \vee \psi) > \theta_{ought}$ , which is the interpretation of *ought*( $\phi \vee \psi$ ). As a result, positivity or maximality would validate the inference in (8.73)—that is, it would render the positive form of *ought* upward monotone.

(8.73) Upward monotonicity of *ought*:

- a. *ought*( $\phi$ )  
 b.  $\phi \subset \chi$   
 c.  $\therefore \textit{ought}(\chi)$

If *ought* were upward monotone, it would be impossible to account for the *ought*-involving **Juliet** and **Professor Procrastinate** scenarios, which involved  $\phi$  and  $\psi$  such that *ought*( $\phi \wedge \psi$ ) was intuitively true even while *ought*( $\phi$ ) was false. In addition, unrestricted agglomeration would be valid.<sup>8</sup>

---

<sup>8</sup> von Fintel (2012) observes that, if we tie NPI licensing to downward monotonicity, the claim that *ought/should* is non-monotone is problematic in light of its failure to disrupt NPI licensing in downward contexts.

- (8.74)    a. I doubt that Mary brought anything/a red cent to the party.  
             b. I doubt that Mary should bring anything/a red cent to the party.

But the literature already contains numerous examples of NPIs in non-monotone and even upward monotone environments, where licensing apparently has to do with the “negativity” of the context, “negative implicatures”, or something of this form (e.g., Linebarger 1987; Horn 2001; Israel 2004). Here are a few examples. ((8.75a) is from Horn 2001. (8.75b) and (8.75c) are naturalistic examples that I have collected.)

- (8.75)    a. Exactly 4 people in the world have ever read that dissertation.  
             b. Getting anybody to lift a finger would be really tough.  
             c. She was old enough to know the risk. Asking for a red cent from McDonalds was too much.

I conclude that  $\mathcal{S}_{ought}$  is probably intermediate. When  $\phi >_{ought} \psi$ ,

$$\phi \succcurlyeq_{ought} \phi \vee \psi \succcurlyeq_{ought} \psi.$$

We can thus rule out the possibility that  $\mathcal{S}_{ought}$  is a ratio scale, since the latter are (by definition) positive. It also cannot be ordinal, since ordinal scales fail to support even weak quantitative comparisons like those in (8.76) (modified from naturalistic examples given above).

- (8.76) a. Constance ought to help George much more than she ought to help anyone else similarly situated.  
b. Non-reformers should watch them much more than reformers should.

The interpretability of quantitative comparisons does not necessitate an interval scale classification, but this is a strong possibility.

#### 8.14 Deriving the Smith and Weakening principles from the scalar semantics

An nice consequence of adopting an intermediate scalar semantics for  $\mathcal{S}_{ought}$  is that it allows us to derive the validity of the Weakening rule that we stipulated above as a constraint on the meaning of *ought*. Weakening corresponds to the validity of the argument in (8.53). This corresponds in scalar terms to the valid argument in (8.78), where (8.78c) follows from intermediacy.

- (8.77) a.  $ought(\phi)$   
b.  $ought(\psi)$   
c.  $\therefore ought(\phi \vee \psi)$
- (8.78) a.  $\mu_{ought}(\phi) > \theta_{ought}$  (premise)  
b.  $\mu_{ought}(\psi) > \theta_{ought}$  (premise)  
c.  $\mu_{ought}(\phi \vee \psi) \geq \min(\mu_{ought}(\phi), \mu_{ought}(\psi))$  (intermediacy)  
d.  $\therefore \mu_{ought}(\phi \vee \psi) > \theta_{ought}$  ((a), (b), (c), transitivity)

This derivation raises the interesting question of whether any of our other assumptions about the meaning of *ought* could be derived from the scalar meaning of *ought* in a similar fashion. Sloman's Principle is not a candidate, since it does not merely constrain  $\mathcal{S}_{ought}$ , but relates this scale to  $\mathcal{S}_{good}$ . However, the Smith Principle—the restricted form of agglomeration I proposed above, which applies only when the propositions embedded under *ought* are logically exhaustive—can be derived from structural features of scalar *ought*, plus one additional assumption. The necessary assumption is what I will call “*ought*-exclusivity”: the requirement that  $ought(\phi)$  implies  $\neg ought(\neg\phi)$ . If this implication is valid, we can derive the Smith principle as follows.

As Horn (2001: 177) summarizes, “operators that are not downward entailing in Ladusaw’s sense, such as the classically non-monotonic *exactly n*, may trigger NPIs in the presence of a discourse-plausible negative inference”. (8.74b) plausibly has the same kind of “negative” character of these examples, though it is challenging to spell out what this amounts to. In any case, these examples indicate that the distribution of NPIs in the scope of an expression is not a totally reliable test for its monotonicity properties. At most, it is suggestive.



Consider two logically consistent propositions whose disjunction is a tautology, and suppose that *ought* holds of both. For example, the original Smith argument had  $M \cup S$ —“Smith serves in the military or performs alternative service”—and  $\neg M$ —“Smith does not perform in the military”. The latter is equivalent to  $S \cup N$ , “Smith performs alternative service or does neither”. From the premises  $\text{ought}(M \cup S)$  and  $\text{ought}(\neg M) = \text{ought}(S \cup N)$ , the intermediate scalar semantics for *ought* allows us to derive the conclusion that  $\text{ought}(S)$ , as desired.

- (8.79) a.  $\mu_{\text{ought}}(M \cup S) > \theta_{\text{ought}}$  (premise)  
 b.  $\mu_{\text{ought}}(S \cup N) > \theta_{\text{ought}}$  (premise)  
 c.  $M \cup S \cup N = W$  (premise)  
 d.  $\mu_{\text{ought}}(M) \leq \theta_{\text{ought}}$  (8.79b, 8.79c, *ought*-exclusivity)  
 e.  $\mu_{\text{ought}}(M \cup S) > \mu_{\text{ought}}(M)$  (8.79a, 8.79d)  
 f.  $\mu_{\text{ought}}(S) \geq \mu_{\text{ought}}(M \cup S) > \mu_{\text{ought}}(M)$  (8.79e, intermediacy of  $S_{\text{ought}}$ )  
 g.  $\mu_{\text{ought}}(S) > \theta_{\text{ought}}$  (8.79a, 8.79f)

The conclusion that  $\mu_{\text{ought}}(S) > \theta_{\text{ought}}$  is, on the present proposal, equivalent to the target conclusion  $\text{ought}(S)$ . The intermediate scalar semantics for *ought* allows us to derive the conclusion “Smith ought to perform alternative service”, using only the assumption that *ought* cannot be true of a proposition and its negation at the same time. The plausibility of the derivation depends, of course, on the plausibility of this additional assumption, which rules out the possibility of semantically conflicting *oughts*. See §8.11 above for discussion. If the analysis floated there is viable, then this principle may be compatible with the appearance of conflicting *oughts*.

This derivation of the Smith principle also helps to explain why agglomeration is only valid when the propositions of which *ought* is true are logically exhaustive. Recall that this property was crucial in accounting for the difference between the **Smith** scenario and the **Disaster Relief** scenario, where agglomeration would have produced incorrect results. Suppose that there were not just three options  $M, S, N$ , but also a fourth option  $Q$ . In this case, the derivation would proceed as follows:

- (8.80) a.  $\mu_{\text{ought}}(M \cup S) > \theta_{\text{ought}}$  (premise)  
 b.  $\mu_{\text{ought}}(S \cup N) > \theta_{\text{ought}}$  (premise)  
 c.  $M \cup S \cup N \cup Q = W$  (premise)  
 d.  $\mu_{\text{ought}}(M \cup Q) \leq \theta_{\text{ought}}$  (8.79b, 8.79c, *ought*-exclusivity)  
 e.  $\mu_{\text{ought}}(M \cup S) > \mu_{\text{ought}}(M \cup Q)$  (8.79a, 8.79d)  
 f. ???

We could fill in the step marked “???” with various further consequences—for example, that  $\mu_{\text{ought}}(S) > \mu_{\text{ought}}(Q)$ . But this would not be sufficient to derive the desired result that  $\mu_{\text{ought}}(S) > \theta_{\text{ought}}$ , since we do not know the relative order of  $\theta_{\text{ought}}$  and  $\mu_{\text{ought}}(Q)$ .

The failure of this derivation of  $\text{ought}(S)$  thus sheds light on a key feature of the Smith Principle—its restriction to exhaustive alternatives. While this feature of the Smith Principle was originally motivated as an effort to account for competing intuitions about agglomeration in two

different scenarios, it may turn out to have a direct derivation from independently motivated facts about the structure of  $\mathcal{S}_{ought}$ .

In sum, the scalar perspective on *ought* may shed considerable light on the semantics of its monadic form, and provide a principled derivation of several features of its semantics which we were originally forced to take on as mere stipulations. This discussion has, of course, just scratched the surface of these issues.

## 8.15 Summary

In this chapter we have approached the meaning of *ought/should* from a different perspective than the usual one in formal semantics. Rather than starting with strong assumptions about the *form* of these expressions' denotations—for example, that they must be interpreted as some kind of quantifiers—we started by asking about which inferences the meaning of *ought/should* should validate. Because the meaning of *ought/should* is closely related to that of *good*, and because we developed a fairly detailed theory of the latter item in chapter 7, we paid special attention to the constraints that comparative goodness facts place on the interpretation of *ought*. We also considered a number of questions about the relationships among *ought*-statements, concluding that the right theory of *ought/should* will not license a number of inferences that are associated with the classical semantics: agglomeration, validity of  $ought(\phi \wedge \psi) \Rightarrow ought(\phi)$  (upward monotonicity), and deontic detachment among others. In addition, I argued that a good theory of *ought/should* must explain (rather than merely making it possible to encode) the information-sensitivity of *ought/should*. The classical semantics does not meet the desideratum. However, the relationship between *ought* and information-sensitive scalar goodness encoded by Sloman's Principle takes us a long way.

While the bulk of the chapter motivated constraints on the meaning of *ought/should* and discussed inferences that it should not validate, it avoided committing to a specific semantic proposal for the meaning of these items. The last main section explains why: I presented evidence for the gradability of *ought* and *should*, concluding that the focus of our earlier discussion—the unmodified forms of these verbs are merely the positive forms of two scalar verbs. While much remains to be said about the scalar semantics of these items, we were able to isolate several aspects of the scale structure of  $\mathcal{S}_{ought}$  and to use them to explain one of the properties of the positive form (Weakening) that we had previously been forced to treat as axiomatic.

This concludes our discussion of deontic modals. There is, needless to say, much more to be done on this front. For example, we have not discussed the deontic interpretations of the non-gradable (?) items *must* and *may*. Our discussion does have considerable consequences for them, though: in particular, if *must* asymmetrically entails *ought/should* (Horn 1972; von Fintel & Iatridou 2008), then *must* validates all the inferences that positive-form *ought* does—e.g., weakening, Sloman's Principle, and the Smith Principle—and then some. It may turn out that agglomeration is unrestrictedly valid with *must*, and perhaps even that *must* is upward monotonic. Significant empirical and formal work will be needed in order to determine what kind of interpretation deontic *must* can have, subject to these constraints. All of this will, in turn, have consequences for deontic *may*, at least on the usual assumption that *must* and *may* are duals.

Another important set of questions that we not address here involves the scalar properties of the deverbal adjectives *permissible* and *obligatory*. To my knowledge, no one has yet investigated these items in any depth from the perspective of scalar semantics, or considered their relationship to scalar goodness. This is a potentially rich area for future empirical and theoretical work.

## CHAPTER 9

### Concluding remarks

In an introductory semantics text, it is not uncommon to introduce the study of modality with a summary along the following lines: “The basic idea of the possible worlds semantics for modal expressions is that they are quantifiers over possible worlds” (von Fintel & Heim 2011: 30). After reading this book, I hope you will agree that the study of modal expressions is a broader enterprise than this gloss would suggest, and that the formal tools needed in a possible-worlds semantics for English modals go far beyond first-order quantification over sets of possible worlds. Many modal expressions are gradable, and are better thought of on the model of scalar expressions than quantifiers: *very likely*, *better*, *very much should*, *totally certain*, and many more. Many puzzling data points involving the grammatical and inferential behavior of modals can be understood by drawing explicit connections between graded modality and graded non-modal concepts, and between gradable modals and gradable adjectives and verbs. The scalar analyses of various modals given in this book rely on the standard formal tools of intensional semantics, but the meanings that these modals express tend to be semantically more open-ended and grammatically more flexible than what the traditional analysis would lead us to expect. This flexibility allows us to use tools developed for the analysis of scalar meaning in general to get clearer insights into the structure of epistemic and deontic concepts.

Even though the basic theoretical assumptions and the development of ideas in this book remain within a quite conservative semantic tradition—essentially that of Montague (1973), as refined by Gallin (1975)—the phenomena have pushed us toward a theoretical analysis of epistemic and deontic meaning that make use of formal tools less familiar to many linguists. In fact, these concepts—probability and Bayesian decision theory—play a key role in our best current frameworks for modeling the cognitive domains that epistemic and deontic modals describe: human reasoning and action. Bayesian models have been shown to provide a promising unifying perspective on human cognition, encompassing many areas of learning, reasoning, perception, and action (Oaksford & Chater 2007; Kemp & Tenenbaum 2009; Tenenbaum, Kemp, Griffiths & Goodman 2011). Bayesian models have also been argued to shed light on on-line language processing (Jurafsky 1996; Levy 2008) as well as linguistic pragmatics (Frank & Goodman 2012; Goodman & Lassiter 2015; Franke & Jäger 2016).

The perspective developed in this book suggests that the language of uncertainty and value has a close relationship to the cognitive representations that are involved in reasoning about these same domains. This would be a surprising coincidence if meaning and cognitive representation were not closely connected. However, the convergence of semantics and cognitive science in this domain makes a certain amount of sense if we suppose that modal expressions get their meanings by reference to our mental representations of the relevant domains—in other words, if modal semantics and reasoning is deeply enmeshed with our intuitive theories of the world and of each others’ psychology. Given the practical importance of communicating and reasoning about the beliefs and deliberative processes of (approximately) Bayesian agents such as ourselves, it would be most unfortunate if we were bound to an interpretation scheme which is unable to accurately represent crucial features of these agents’ psychology. Fortunately, it appears that we are not in this situation:

epistemic and deontic expressions are semantically well-adapted to mirror the structure of those aspects of cognition that are their subject matter.

## References

- Bach, Emmon. 1986. The algebra of events. *Linguistics and Philosophy* 9(1). 5–16.
- Bale, Alan C. 2006. *The universal scale and the semantics of comparison*: McGill University dissertation.
- Bale, Alan C. 2008. A universal scale of comparison. *Linguistics and Philosophy* 31(1). 1–55.
- Bale, Alan C. 2011. Scales and comparison classes. *Natural Language Semantics* 19. 169–190.
- Barker, Chris. 2002. The dynamics of vagueness. *Linguistics and Philosophy* 25(1). 1–36.
- Barker, Chris. 2009. Clarity and the grammar of skepticism. *Mind and Language* 24(3). 253–273.
- Bartsch, Renate & Theo Vennemann. 1973. *Semantic Structures: A Study In the Relation Between Semantics and Syntax*. Athenäum.
- Beaver, David & Brady Clark. 2008. *Sense and Sensitivity: How Focus Determines Meaning*. Wiley-Blackwell.
- Beddor, Robert. 2015. Certainty first. Ms., Rutgers University.
- Bierwisch, Manfred. 1989. The semantics of gradation. In M. Bierwisch & E. Lang (eds.), *Dimensional adjectives: Grammatical structure and conceptual interpretation*, 71–261. Springer-Verlag.
- Bledin, Justin. 2015. Modus ponens defended. *The Journal of Philosophy* 112(2). 57–83.
- Bresnan, Joan. 1973. Syntax of the comparative clause construction in English. *Linguistic inquiry* 4(3). 275–343.
- Bresnan, Joan. 2007. Is syntactic knowledge probabilistic? Experiments with the english dative alternation. In Sam Featherston & Wolfgang Sternefeld (eds.), *Roots: Linguistics in search of its evidential base*, 75–96. Mouton de Gruyter.
- Bresnan, Joan, Anna Cueni, Tatiana Nikitina, R Harald Baayen et al. 2007. Predicting the dative alternation. In Irene Kramer Gerlof Bouma & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: KNAW.
- Bresnan, Joan & Marilyn Ford. 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of english. *Language* 86(1). 168–213.
- Brogan, Albert P. 1919. The fundamental value universal. *The Journal of Philosophy, Psychology and Scientific Methods* 16(4). 96–104.
- Buckmaster, Robert A. 2014. *The grammar of english ideas: Distance, meaning and conventions. a re-imagining of english grammar for language teachers*. English Ideas.
- Burnett, Heather. 2014. A Delineation Solution to the Puzzles of Absolute Adjectives. *Linguistics and Philosophy* 37(1). 1–39.
- Burnett, Heather. 2016. *Gradability in natural language: Logical and grammatical foundations*. Oxford University Press.
- Cariani, Fabrizio. 2013. “Ought” and resolution semantics. *Noûs* 47. 534–558.
- Cariani, Fabrizio. 2016. Consequence and contrast in deontic semantics. To appear in *Journal of Philosophy*.
- Cariani, Fabrizio, Stefan Kaufmann & Magdalena Kaufmann. 2013. Deliberative modality under epistemic uncertainty. *Linguistics and Philosophy* 36. 225–259.
- Carr, Jennifer. 2012. Deontic modals without decision theory. In *Proceedings of Sinn und Bedeutung 17*, 167–182.

- Carroll, Lewis. 1866. *Alice's Adventures in Wonderland*. Macmillan.
- Champollion, Lucas. 2013. Man and woman: The last obstacle to boolean coordination. In *Proceedings of the 19th amsterdam colloquium*, 83–90.
- Charlow, Nate. 2013. What we know and what to do. *Synthese* 1–33.
- Charlow, Nate. 2015. Decision theory: Yes! Truth conditions: No! To appear in M. Chrisman and N. Charlow (eds.), *Deontic Modals*. Oxford University Press.
- Chisholm, Roderick M. 1963a. Contrary-to-duty imperatives and deontic logic. *Analysis* 33–36.
- Chisholm, Roderick M. 1963b. Supererogation and offence: a conceptual scheme for ethics. *Ratio* 5. 1–14.
- Chisholm, Roderick M. 1975. The intrinsic value in disjunctive states of affairs. *Noûs* 295–308.
- Chisholm, Roderick M. & Ernest Sosa. 1966. On the logic of “intrinsically better”. *American Philosophical Quarterly* 3(3). 244–249.
- Chomsky, Noam. 1977. On *wh*-movement. In Peter Culicover, Thomas Wasow & A. Akmajian (eds.), *Formal syntax*, vol. 132, Academic Press.
- Clark, Eve V & Herbert H Clark. 1979. When nouns surface as verbs. *Language* 55(4). 767–811.
- Constantz, Brent & Annemarie Meike. 1989. Calcite centers of calcification in *mussa angulosa* (scleractinia). In Rex E. Crick (ed.), *Origin, evolution, and modern aspects of biomineralization in plants and animals*, 201–207. Springer.
- Copley, Bridget. 2004. So-called epistemic should. *Snippets* 9. 7–8.
- Cover, Thomas M. & Joy A. Thomas. 1991. *Elements of information theory*. Wiley.
- Cresswell, Max J. 1976. The semantics of degree. In B. Partee (ed.), *Montague Grammar*, 261–292. Academic Press.
- Cruse, D. Alan. 1986. *Lexical semantics*. Cambridge University Press.
- Danks, David. 2014. *Unifying the mind: Cognitive representations as graphical models*. MIT Press.
- Davidson, Donald. 1973. Radical interpretation. *Dialectica* 27(3-4). 313–328.
- Davies, Mark. 2004-. BYU-BNC. (Based on the British National Corpus from Oxford University Press). Available online at <http://corpus.byu.edu/bnc/>.
- Davies, Mark. 2008-. The Corpus of Contemporary American English: 450 million words, 1990-present. Available online at <http://corpus.byu.edu/coca/>.
- De Finetti, Bruno. 1989. Probabilism: A critical essay on the theory of probability and on the value of science. *Erkenntnis* 31(2). 169–223.
- DeRose, Keith. 1991. Epistemic possibilities. *The Philosophical Review* 581–605.
- Dretske, Fred. 1972. Contrastive statements. *The Philosophical Review* 411–437.
- Driver, Julia. 1997. The Ethics of Intervention. *Philosophy and Phenomenological Research* 57(4). 851–870.
- Egré, Paul & Mikaël Cozic. 2011. If-clauses and probability operators. *Topoi* 1–13.
- Fara, Delia Graff. 2000. Shifting sands: An interest-relative theory of vagueness. *Philosophical Topics* 20. 45–81.
- Feduzi, Alberto, Jochen Runde & Carlo Zappia. 2011. De finetti on the insurance of risks and uncertainties. *The British Journal for the Philosophy of Science* 63(2). 329–356.
- Fine, Kit. 1975. Vagueness, truth and logic. *Synthese* 30(3). 265–300.
- Finlay, S. 2010. What ought probably means, and why you can't detach it. *Synthese* 177(1). 67–89.

- von Fintel, K. 2012. The best we can (expect to) get? Challenges to the classic semantics for deontic modals. Paper presented at the Central meeting of the American Philosophical Association, February 17. <http://mit.edu/fintel/fintel-2012-apa-ought.pdf>.
- von Fintel, Kai. 2001. Counterfactuals in a dynamic context. In Michael Kenstowicz (ed.), *Ken Hale: A life in language*, 123–152. MIT Press.
- von Fintel, Kai & Anthony Gillies. 2008. CIA leaks. *The Philosophical Review* 117(1). 77.
- von Fintel, Kai & Anthony Gillies. 2010. Must... stay... strong! *Natural Language Semantics* 18(4). 351–383.
- von Fintel, Kai & Irene Heim. 2011. Intensional semantics. Ms., Massachusetts Institute of Technology. [http://web.mit.edu/fintel/www/intens\\_sem.pdf](http://web.mit.edu/fintel/www/intens_sem.pdf).
- von Fintel, Kai & Sabine Iatridou. 2008. How to say *ought* in foreign: The composition of weak necessity modals. In Jacqueline Guéron (ed.), *Time and Modality*, 115–141. Springer.
- Fox, Craig R. & Gülden Ülkümen. 2011. Distinguishing two dimensions of uncertainty. In G. Kirkeboen W. Brun, G. Keren & H. Montgomery (eds.), *Perspectives on thinking, judging, and decision making*, Universitetsforlaget.
- Fox, Danny & Martin Hackl. 2006. The universal density of measurement. *Linguistics and Philosophy* 29(5). 537–586.
- van Fraassen, Bas C. 1966. Singular terms, truth-value gaps, and free logic. *Journal of Philosophy* 63(17). 481–495.
- van Fraassen, Bas C. 1968. Presupposition, implication, and self-reference. *Journal of Philosophy* 65(5). 136–152.
- van Fraassen, Bas C. 1973. Values and the heart's command. *Journal of Philosophy* 70(1). 5–19.
- Frank, M.C. & N.D. Goodman. 2012. Predicting pragmatic reasoning in language games. *Science* 336(6084). 998–998.
- Franke, Michael & Gerhard Jäger. 2016. Probabilistic pragmatics, or why Bayes' rule is probably important for pragmatics. *Zeitschrift für Sprachwissenschaft* 35(4).
- Gallin, Daniel. 1975. *Intensional and higher-order modal logic: With applications to Montague semantics*. Elsevier.
- Gillies, Anthony S. 2007. Counterfactual scorekeeping. *Linguistics and Philosophy* 30(3). 329–360.
- Glass, Arnold L & James Lau. 2003. Grammatical intuitions about irregular verb inflections. *American Journal of Psychology* 116(1). 51–70.
- Goble, Lou. 1996. Utilitarian deontic logic. *Philosophical Studies* 82(3). 317–357.
- Goble, Lou. 2009. Normative conflicts and the logic of “ought”. *Noûs* 43(3). 450–489.
- Goodhue, Daniel. 2016. Epistemic must is not evidential, it's epistemic. To appear in *Proceedings of NELS* 46.
- Goodman, Noah D. & Daniel Lassiter. 2015. Probabilistic semantics and pragmatics: Uncertainty in language and thought. In Shalom Lappin & Chris Fox (eds.), *Handbook of Contemporary Semantic Theory*, Wiley-Blackwell 2nd edn.
- Grice, H. Paul. 1975. Logic and conversation. In P. Cole & J. Morgan (eds.), *Syntax and semantics* 9: *Pragmatics*, 41–58. Academic Press.
- Groenendijk, Jeroen & Martin Stokhof. 1984. Studies in the Semantics of Questions and the Pragmatics of Answers. Ph.D. thesis, University of Amsterdam.



- Hájek, Alan. 2003. What conditional probability could not be. *Synthese* 137(3). 273–323.
- Hájek, Alan. 2007. Most counterfactuals are false Manuscript, Australian National University.
- Halmos, Paul & Steven Givant. 1998. *Logic as algebra* 21. Cambridge University Press.
- Hamblin, Charles L. 1959. The modal “probably”. *Mind* 68(270). 234–240.
- Hansson, Bengt. 1968. Fundamental axioms for preference relations. *Synthese* 18(4). 423–442.
- Hansson, Sven Ove. 1990. Defining “good” and “bad” in terms of “better”. *Notre Dame Journal of Formal Logic* 31(1). 136–149.
- Hansson, Sven Ove. 2001. *The structure of values and norms*. Cambridge University Press.
- Hansson, Sven Ove. 2004. Semantics for more plausible deontic logics. *Journal of Applied Logic* 2(1). 3–18.
- Harrison-Trainor, Matthew, Wesley H Holliday & Thomas F Icard. 2015. From almost-representability by a probability measure to imprecise-representability by a set of measures. Ms., University of California at Berkeley and Stanford University.
- Hatzivassiloglou, Vasileios & Kathleen R McKeown. 1993. Towards the automatic identification of adjectival scales: Clustering adjectives according to meaning. In *Proceedings of the 31st annual meeting on association for computational linguistics*, 172–182. Association for Computational Linguistics.
- Hawthorne, John. 2004. *Knowledge and lotteries*. Oxford University Press.
- Hay, Jennifer, Chris Kennedy & Beth Levin. 1999. Scalar structure underlies telicity in ‘degree achievements’. In *Semantics and Linguistic Theory* 9, 127–144.
- Heim, Irene. 2001. Degree operators and scope. In Fery & Sternefeld (eds.), *Audiatur vox sapientiae: A festschrift for Arnim von Stechow*, 214–239. Akademie Verlag.
- Helmholtz, Hermann von. 1887. Zahlen und Messen erkenntnistheoretisch betrachtet. *Philosophische Aufsätze Eduard Zeller gewidmet* 356–391.
- Herburger, Elena & Aynat Rubinstein. 2014. Is ‘more possible’ more possible in German?
- Hintikka, Jaako. 1962. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Cornell University Press.
- Hölder, Otto. 1901. Die Axiome der Quantität und die Lehre vom Mass. *Ber. Verh. Kgl. Sächsis. Ges. Wiss. Leipzig, Math.-Phys. Klasse* 53. 1–64.
- Holliday, Wesley H. & Thomas F. Icard. 2013. Measure semantics and qualitative semantics for epistemic modals. In Todd Snider (ed.), *Semantics and Linguistic Theory (SALT) 23*, Ithaca, NY: CLC Publications.
- Horn, Laurence. 1972. *On the Semantic Properties of Logical Operators in English*: UCLA dissertation.
- Horn, Laurence R. 2001. Flaubert triggers, squattive negation, and other quirks of grammar. In Jack Hoeksema, Hotze Rullmann, Víctor Sánchez-Valencia & Ton van der Wouden (eds.), *Perspectives on negation and polarity items*, 173–200. Benjamins, Amsterdam.
- Horty, John. 2014. Deontic modals: Why abandon the classical semantics? *Pacific Philosophical Quarterly* 95(4). 424–460.
- Horty, John F. 1993. Deontic logic as founded on nonmonotonic logic. *Annals of Mathematics and Artificial Intelligence* 9(1-2). 69–91.
- Horty, John F. 2003. Reasoning with moral conflicts. *Noûs* 37(4). 557–605.

- Horty, John F. 2012. *Reasons as defaults*. Oxford University Press.
- Ichikawa, Jonathan. 2011. Quantifiers, knowledge, and counterfactuals. *Philosophy and Phenomenological Research* 82(2). 287–313.
- Israel, Michael. 2004. The pragmatics of polarity. In Laurence R. Horn & Gregory Ward (eds.), *The handbook of pragmatics*, 701–723. Blackwell.
- Jackson, Frank. 1985. On the semantics and logic of obligation. *Mind* 94(374). 177–195.
- Jackson, Frank. 1991. Decision-theoretic consequentialism and the nearest and dearest objection. *Ethics* 101(3). 461–482.
- Jackson, Frank & Robert Pargetter. 1986. Oughts, options, and actualism. *The Philosophical Review* 95(2). 233–255.
- Jeffrey, Richard C. 1965. *The logic of decision*. University of Chicago Press.
- Jurafsky, Daniel. 1996. A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science* 20(2). 137–194.
- Kahneman, Daniel & Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society* 47(2). 263–291.
- Kahneman, Daniel & Amos Tversky. 1982. Variants of uncertainty. *Cognition* 11(2). 143–157.
- Kamp, Hans. 1975. Two theories about adjectives. In E. Keenan (ed.), *Formal semantics of natural language*, 123–155. Cambridge University Press.
- Kamp, Hans & Barbara Partee. 1995. Prototype theory and compositionality. *Cognition* 57(2). 129–191.
- Karttunen, Lauri. 1972. Possible and must. In John Kimball (ed.), *Syntax and semantics*, vol. 1, Seminar Press.
- Katz, G., P. Portner & A. Rubinstein. 2012. Ordering combination for modal comparison. In Anca Chereches (ed.), *Semantics & Linguistic Theory (SALT) 22*, CLC Publications.
- Keenan, Edward L. & Leonard M. Faltz. 1985. *Boolean semantics for natural language*. Kluwer Academic Publishers.
- Kemp, Charles & Joshua B. Tenenbaum. 2009. Structured statistical models of inductive reasoning. *Psychological Review* 116(1). 20–58.
- Kennedy, Chris. 1997. *Projecting the adjective: The syntax and semantics of gradability and comparison*. U.C., Santa Cruz dissertation.
- Kennedy, Chris. 2007. Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy* 30(1). 1–45.
- Kennedy, Chris & Louise McNally. 2005a. Scale structure, degree modification, and the semantics of gradable predicates. *Language* 81(2). 345–381.
- Kennedy, Christopher & Louise McNally. 2005b. The syntax and semantics of multiple degree modification in english. In *Proceedings of the 12th international conference on head-driven phrase structure grammar*, 178–191. CSLI Publications.
- Keynes, John Maynard. 1921. *A Treatise on Probability*. Macmillan.
- Klecha, Peter. 2012. Positive and conditional semantics for gradable modals. In Anna Chernilovskaya Ana Aguilar Guevara & Rick Nouwen (eds.), *Sinn und bedeutung* 16, vol. 2, 363–376. MIT Press.
- Klecha, Peter. 2014. *Bridging the Divide: Scalarity and Modality*. University of Chicago disserta-

- tion.
- Klein, Ewan. 1980. A semantics for positive and comparative adjectives. *Linguistics and Philosophy* 4(1). 1–45.
- Klein, Ewan. 1982. The interpretation of adjectival comparatives. *Journal of Linguistics* 18(1). 113–136.
- Klein, Ewan. 1991. Comparatives. In A. von Stechow & D. Wunderlich (eds.), *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung*, Walter de Gruyter.
- Koller, Daphne & Nir Friedman. 2009. *Probabilistic graphical models: Principles and techniques*. The MIT Press.
- Kolodny, Niko & John MacFarlane. 2010. Ifs and oughts. *Journal of Philosophy* 107(3). 115–143.
- Krantz, David H., R. Duncan Luce, Patrick Suppes & Amos Tversky. 1971. *Foundations of Measurement*. Academic Press.
- Kratzer, Angelika. 1981. The notional category of modality. In Eikmeyer & Rieser (eds.), *Words, Worlds, and Contexts*, 38–74. de Gruyter.
- Kratzer, Angelika. 1986. Conditionals. In *Chicago linguistics society*, vol. 22 2, 1–15.
- Kratzer, Angelika. 1991. Modality. In Arnim von Stechow & Dieter Wunderlich (eds.), *Semantics: An international handbook of contemporary research*, 639–650. de Gruyter.
- Kratzer, Angelika. 2012. *Modality and Conditionals: New and Revised Perspectives*. Oxford University Press.
- Krifka, Manfred. 1989. Nominal reference, temporal constitution and quantification in event semantics. *Semantics and contextual expression* 75. 115.
- Krifka, Manfred. 1990. Four thousand ships passed through the lock: Object-induced measure functions on events. *Linguistics and Philosophy* 13(5). 487–520.
- Krifka, Manfred. 1998. The origins of telicity. *Events and grammar* 197. 235.
- Krifka, Manfred. 2007. Basic notions of information structure. In Caroline F’ery, Gisbert Fanselow & Manfred Krifka (eds.), *Working papers of the sfb632, interdisciplinary studies on information structure*, vol. 6, 13–36. Universitätsverlag Potsdam.
- Kripke, Saul. 1963. Semantical considerations on modal logic. *Acta philosophica fennica* 16(1963). 83–94.
- Kyburg, Henry. 1961. *Probability and the logic of rational belief*. Wesleyan University Press.
- Lagnado, David A & Steven A Sloman. 2004. Inside and outside probability judgment. *Blackwell handbook of judgment and decision making* 157–176.
- Laserson, Peter. 1999. Pragmatic halos. *Language* 522–551.
- Lassiter, Daniel. 2010. Gradable epistemic modals, probability, and scale structure. In Nan Li & David Lutz (eds.), *Semantics & Linguistic Theory (SALT) 20*, 197–215. CLC Publications.
- Lassiter, Daniel. 2011a. Measurement and Modality: The Scalar Basis of Modal Semantics. Ph.D. thesis, New York University.
- Lassiter, Daniel. 2011b. Vagueness as probabilistic linguistic knowledge. In R. Nouwen, R. van Rooij, U. Sauerland & H.-C. Schmitz (eds.), *Vagueness in communication*, 127–150. Springer.
- Lassiter, Daniel. 2014. Modality, scale structure, and scalar reasoning. *Pacific Philosophical Quarterly* 95(4). 461–490.
- Lassiter, Daniel. 2015a. Adjectival modification and gradation. In Shalom Lappin & Chris Fox

- (eds.), *Handbook of Contemporary Semantic Theory*, Wiley-Blackwell 2nd edn.
- Lassiter, Daniel. 2015b. Epistemic comparison, models of uncertainty, and the disjunction puzzle. *Journal of Semantics* 32(4). 649–684.
- Lassiter, Daniel. 2015c. Linguistic and philosophical considerations on Bayesian semantics. To appear in M. Chrisman & N. Charlow (eds.), *Deontic Modals*. Oxford University Press.
- Lassiter, Daniel. 2016. *Must*, knowledge, and (in)directness. *Natural Language Semantics*.
- Lassiter, Daniel & Noah D. Goodman. 2013. Context, scale structure, and statistics in the interpretation of positive-form adjectives. In Todd Snider (ed.), *Semantics & Linguistic Theory (SALT)* 23, 587–610. CLC Publications.
- Lassiter, Daniel & Noah D. Goodman. 2015a. Adjectival vagueness in a Bayesian model of interpretation. *Synthese*.
- Lassiter, Daniel & Noah D. Goodman. 2015b. How many kinds of reasoning? Inference, probability, and natural language semantics. *Cognition* 136. 123–134.
- Lawson, Tony. 1988. Probability and uncertainty in economic analysis. *Journal of Post Keynesian Economics* 38–65.
- Levinson, D. 2003. Probabilistic model-theoretic semantics for *want*. In Robert B. Young & Yuping Zhou (eds.), *Semantics and Linguistic Theory (SALT)* 13, 222–239. CLC Publications.
- Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106(3). 1126–1177.
- Lewis, David. 1973. *Counterfactuals*. Harvard University Press.
- Lewis, David. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic* 8(1). 339–359. doi:10.1007/BF00258436.
- Lewis, David. 1981. Ordering semantics and premise semantics for counterfactuals. *Journal of Philosophical Logic* 10(2). 217–234.
- Lewis, David. 1996. Elusive knowledge. *Australasian Journal of Philosophy* 74(4). 549–567.
- Lewis, Karen S. 2016. Elusive Counterfactuals. *Nous* 50(2). 286–313.
- Linebarger, Marcia C. 1987. Negative polarity and grammatical representation. *Linguistics and Philosophy* 10(3). 325–387.
- Link, Godehard. 1983. The logical analysis of plurals and mass terms: A lattice-theoretical approach. In R. Bäuerle, C. Schwarze & A. von Stechow (eds.), *Meaning, use and interpretation of language*, vol. 21, 302–323. Walter de Gruyter.
- Link, Godehard. 1998. *Algebraic Semantics in Language and Philosophy*. CSLI Publications.
- Løhre, Erik & Karl Halvor Teigen. 2015. There is a 60% probability, but I am 70% certain: Communicative consequences of external and internal expressions of uncertainty. *Thinking & Reasoning* 1–28.
- Luce, R. Duncan. 1959. *Individual choice behavior: A theoretical analysis*. John Wiley.
- Luce, R. Duncan & Louis Narens. 1985. Classification of concatenation measurement structures according to scale type. *Journal of Mathematical Psychology* 29(1). 1–72.
- Macchi, Laura, Daniel Osherson & David H. Krantz. 1999. A note on superadditive probability judgment. *Psychological Review* 106(1). 210.
- MacFarlane, John. 2014. *Assessment Sensitivity: Relative Truth and its Applications*. Oxford University Press.
- MacKay, David J.C. 2003. *Information theory, inference, and learning algorithms*. Cambridge

- University Press.
- Mandelkern, Matthew. 2016. A solution to Karttunen's problem. To appear in *Proceedings of Sinn und Bedeutung 2016*.
- Markman, Arthur B. 1998. *Knowledge representation*. Lawrence Erlbaum.
- Meek, Christopher & Clark Glymour. 1994. Conditioning and intervening. *The British journal for the philosophy of science* 45. 1001–1021.
- Merchant, Brian. 2013. Things scientists are less sure of than climate change. *Motherboard* <http://motherboard.vice.com/blog/things-scientists-are-less-sure-of-than-climate-change>.
- Mill, John Stuart. 1863. *Utilitarianism*. Longmans, Green, Reader, and Dyer.
- von Mises, R. 1957. *Probability, statistics, and truth*. Allen and Unwin.
- Montague, Richard. 1973. The proper treatment of quantification in ordinary English. In J. Hintikka, J. Moravcsik & P. Suppes (eds.), *Approaches to natural language*, vol. 49, 221–242. Reidel.
- Morzycki, Marcin. 2012. Adjectival extremeness: Degree modification and contextually restricted scales. *Natural Language & Linguistic Theory* 30(2). 567–609.
- Morzycki, Marcin. 2015. *Modification*. Cambridge University Press.
- Moss, Sarah. 2012. On the pragmatics of counterfactuals. *Noûs* 46(3). 561–586.
- Moss, Sarah. 2015. On the semantics and pragmatics of epistemic vocabulary. *Semantics and Pragmatics* 8. 1–81.
- Narens, Louis. 1985. *Abstract Measurement Theory*. MIT Press.
- Narens, Louis. 2007. *Theories of probability: an examination of logical and qualitative foundations*. World Scientific Publishing Company.
- Nerbonne, John. 1995. Nominal comparatives and generalized quantifiers. *Journal of Logic, Language and Information* 4(4). 273–300.
- von Neumann, Jon & Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton University Press.
- Nouwen, Rick. 2008. Upper-bounded *no more*: the exhaustive interpretation of non-strict comparison. *Natural Language Semantics* 16. 271–295.
- Oaksford, Mike & Nick Chater. 2007. *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press.
- Palmer, Frank Robert. 1979. *Modality and the English modals*. Longman London.
- Pearl, Judea. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.
- Pearl, Judea. 2000. *Causality: Models, reasoning and inference*. Cambridge University Press.
- Popper, Karl R. 1959. The propensity interpretation of probability. *The British journal for the philosophy of science* 10(37). 25–42.
- Portner, Paul. 2009. *Modality*. Oxford University Press.
- Portner, Paul & Aynat Rubinstein. 2015. Extreme and non-extreme deontic modals. To appear in M. Chrisman and N. Charlow (eds.), *Deontic Modals*. Oxford University Press.
- Potts, Christopher. 2008. Interpretive Economy, Schelling points, and evolutionary stability. Ms., University of Massachusetts at Amherst.
- Pustejovsky, James. 1995. *The Generative Lexicon*. MIT Press.
- R Core Team. 2015. *R: A language and environment for statistical computing*. R Foundation for

- Statistical Computing Vienna, Austria. <https://www.R-project.org/>.
- Regan, Donald. 1980. *Utilitarianism and Co-operation*. Oxford University Press.
- Roberts, Craige. 1996. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Working Papers in Linguistics-Ohio State University Department of Linguistics* 91–136.
- Roberts, Fred S. 1979. *Measurement Theory with Applications to Decisionmaking, Utility, and the Social Sciences*. Addison Wesley Publishing Company.
- Roelofsen, Floris & Donka Farkas. 2015. Polarity particle responses as a window onto the interpretation of questions and assertions. *Language* 91(2). 359–414.
- van Rooij, Robert. 2003. Questioning to resolve decision problems. *Linguistics and Philosophy* 26(6). 727–763.
- van Rooij, Robert. 2004. Utility, informativity and protocols. *Journal of Philosophical Logic* 33(4). 389–419.
- van Rooij, Robert. 2009. Up and down the scale: Adjectives, comparisons, and measurement. Ms., ILLC, Universteit van Amsterdam .
- van Rooij, Robert. 2011. Measurement, and interadjective comparisons. *Journal of Semantics* 28(3). 335–358.
- Rooth, Mats. 1992. A theory of focus interpretation. *Natural language semantics* 1(1). 75–116.
- Rothschild, Daniel. 2011. A note on conditionals and restrictors. Ms., Oxford University .
- Rotstein, Carmen & Yoad Winter. 2004. Total adjectives vs. partial adjectives: Scale structure and higher-order modifiers. *Natural Language Semantics* 12(3). 259–288.
- Rubin, Rebecca B, Alan M Rubin, Elizabeth Graham, Elizabeth M Perse & David Seibold. 2009. *Communication research measures II: A sourcebook*. Routledge.
- Rudin, Deniz. 2015. Deriving a variable-strength *Might*. In *Proceedings of Sinn und Bedeutung* 20.
- Rullmann, Hotze. 1995. *Maximality in the Semantics of wh-constructions*: University of Massachusetts, Amherst dissertation.
- Sapir, Edward. 1944. Grading: A study in semantics. *Philosophy of Science* 11(2). 93–116.
- Sassoon, Galit. 2007. *Vagueness, Gradability and Typicality, A Comprehensive semantic analysis*: Tel Aviv University dissertation.
- Sassoon, Galit. 2010a. Measurement theory in linguistics. *Synthese* 174(1). 151–180.
- Sassoon, Galit. 2010b. The degree functions of negative adjectives. *Natural Language Semantics* 18(2). 141–181.
- Sassoon, Galit. 2014. A *slightly* modified economy principle: Stable properties have non-stable standards. In Evan Cohen (ed.), *Proceedings of the Israel Association of Theoretical Linguistics (IATL) 2011*, MIT Working Papers in Linguistics.
- Sassoon, Galit W. 2013. A typology of multidimensional adjectives. *Journal of Semantics* 30(3). 335–380.
- Savage, Leonard J. 1954. *The Foundations of Statistics*. Wiley.
- Scha, Remko & Yoad Winter. 2015. Plurals. In Shalom Lappin & Chris Fox (eds.), *Handbook of Contemporary Semantic Theory*, Wiley-Blackwell 2nd edn.
- Schulz, Moritz. 2010. Epistemic modals and informational consequence. *Synthese* 174(3). 385–395.
- Schütze, Carson T. 1996. *The empirical base of linguistics: Grammaticality judgments and linguistic*

- methodology*. University of Chicago Press.
- Schwarzschild, Roger & Karina Wilkinson. 2002. Quantifiers in comparatives: A semantics of degree based on intervals. *Natural Language Semantics* 10(1). 1–41.
- Scott, Dana & Patrick Suppes. 1958. Foundational aspects of theories of measurement. *Journal of Symbolic Logic* 23(2). 113–128.
- Shannon, Claude E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27. 379–423.
- Sheinman, Vera, Christiane Fellbaum, Isaac Julien, Peter Schulam & Takenobu Tokunaga. 2013. Large, huge or gigantic? identifying and encoding intensity relations among adjectives in wordnet. *Language resources and evaluation* 47(3). 797–816.
- Silk, A. 2012. Modality, weights, and inconsistent premise sets. In Anca Chereches (ed.), *Semantics & Linguistic Theory (SALT)* 22, CLC Publications.
- Sloman, Aaron. 1970. ‘Ought’ and ‘Better’. *Mind* 75(315). 385–394.
- Sloman, Steven A. 2005. *Causal models: How we think about the world and its alternatives*. Oxford University Press.
- Slote, Michael. 1984. Satisficing consequentialism, i. *Proceedings of the Aristotelian Society, Supplementary Volumes* 139–163.
- Solt, Stephanie. 2011. Notes on the comparison class. In Rick Nouwen, Robert van Rooij, Uli Sauerland & Hans-Christian Schmitz (eds.), *Vagueness in communication*, 189–206. Springer.
- Solt, Stephanie. 2012. Comparison to arbitrary standards. In *Proceedings of sinn und bedeutung*, vol. 16, 557–570.
- Stalnaker, Robert. 1975. Indicative conditionals. *Philosophia* 5(3). 269–286.
- Stanley, Jason. 2008. Knowledge and certainty. *Philosophical Issues* 18(1). 35–57.
- von Stechow, Arnim. 1984. Comparing semantic theories of comparison. *Journal of Semantics* 3(1). 1–77.
- Stevens, S.S. 1946. On the theory of scales of measurement. *Science* 103(2684). 677–680.
- Suppes, Patrick. 1959. Measurement, empirical meaningfulness, and three-valued logic. *Measurement: definitions and theories* 129.
- Suppes, Patrick. 2002. *Representation and invariance of scientific structures*. CSLI Publications.
- Suppes, Patrick & Joseph L. Zinnes. 1963. Basic measurement theory. In R. D. Luce, R. R. Bush & E. Galanter (eds.), *Handbook of mathematical psychology*, vol. 1, 1–63. Wiley.
- Swanson, Eric. 2006. Interactions With Context. Ph.D. thesis, MIT.
- Swanson, Eric. 2011. How not to theorize about the language of subjective uncertainty.
- Swanson, Eric. 2014. Ordering supervaluationism, counterpart theory, and ersatz fundamentality. *The Journal of Philosophy* 111(6). 289–310.
- Swanson, Eric. 2015. The application of constraint semantics to the language of subjective uncertainty. *Journal of Philosophical Logic* 1–26.
- Syrett, Kristen, Chris Kennedy & Jeffrey Lidz. 2010. Meaning and context in children’s understanding of gradable adjectives. *Journal of Semantics* 27. 1–35.
- Teigen, Karl. 1988. When are low-probability events judged to be ‘probable’? Effects of outcome-set characteristics on verbal probability estimates. *Acta Psychologica* 68. 157–174.
- Tenenbaum, Joshua B., Charles Kemp, Tom L. Griffiths & Noah D. Goodman. 2011. How to grow

- a mind: Statistics, structure, and abstraction. *Science* 331(6022). 1279–1285.
- Turri, John. 2015. Epistemic modals and alternative possibilities. Manuscript, University of Waterloo.
- Tversky, Amos. 1977. Features of similarity. *Psychological Review* 84(4). 327.
- Tversky, Amos & Derek J. Koehler. 1994. Support theory: A nonextensional representation of subjective probability. *Psychological Review* 101(4). 547–566.
- Uegaki, Wataru. 2015. Content nouns and the semantics of question-embedding. To appear in *Journal of Semantics*.
- Ülkümen, Gülden, Craig R Fox & Bertram F Malle. 2015. Two dimensions of subjective uncertainty: Clues from natural language. To appear in *Journal of Experimental Psychology: General*.
- Unger, Peter. 1971. A defense of skepticism. *The Philosophical Review* 80(2). 198–219.
- Unger, Peter. 1975. *Ignorance*. Oxford University Press.
- Verney, Kevern J. 2013. *The art of the possible: Booker t. washington and black leadership in the united states, 1881-1925*. Routledge.
- Villalta, Elisabeth. 2008. Mood and gradability: An investigation of the subjunctive mood in Spanish. *Linguistics and Philosophy* 31(4). 467–522.
- Von Wright, Georg Henrik. 1972. The logic of preference reconsidered. *Theory and Decision* 3(2). 140–169.
- Wasow, Thomas. 2002. *Postverbal behavior*. Center for the Study of Language and Information.
- Wasow, Thomas & Jennifer Arnold. 2005. Intuitions in linguistic argumentation. *Lingua* 115(11). 1481–1496.
- Wedgwood, Ralph. 2015. Objective and subjective *ought*. To appear in M. Chrisman & N. Charlow (eds.), *Deontic Modals*. Oxford University Press.
- Willer, Malte. 2010. *Modality in flux*: University of Texas dissertation.
- Willer, Malte. 2013. Dynamics of epistemic modality. *Philosophical Review* 122(1). 45–92.
- Willer, Malte. 2015. Dynamic foundations for deontic logic. To appear in M. Chrisman and N. Charlow (eds.), *Deontic Modals*. Oxford University Press.
- Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford University Press.
- Windschitl, Paul D. & Gary L. Wells. 1998. The alternative-outcomes effect. *Journal of Personality and Social Psychology* 75(6). 1411–1423.
- Winter, Yoad. 2001. *Flexibility Principles in Boolean Semantics: The Interpretation of Coordination, Plurality, and Scope in Natural Language*. The MIT Press.
- Yalcin, Seth. 2005. A puzzle about epistemic modals. *MIT Working Papers in Linguistics* 51. 231–272.
- Yalcin, Seth. 2007. Epistemic modals. *Mind* 116(464). 983–1026.
- Yalcin, Seth. 2010. Probability operators. *Philosophy Compass* 5(11). 916–937.
- Yalcin, Seth. 2011. Nonfactualism about epistemic modality. In A. Egan & B. Weatherson (eds.), *Epistemic modality*, 295–332. Oxford University Press.
- Yalcin, Seth. 2012a. Bayesian expressivism. *Proceedings of the Aristotelian Society* 112. 123–160.
- Yalcin, Seth. 2012b. A counterexample to modus tollens. *Journal of Philosophical Logic* 41(6). 1001–1024.
- Yalcin, Seth. 2015. Modalities of normality. To appear in M. Chrisman and N. Charlow (eds.),



- Deontic Modals*. Oxford University Press.
- Yoon, Youngeun. 1996. Total and partial predicates and the weak and strong interpretations. *Natural Language Semantics* 4(3). 217–236.
- Zadeh, Lofti A. 1965. Fuzzy sets. *Information and Control* 8(3). 338–353.
- Zadeh, Lofti A. 1978. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems* 1(1). 3–28.