

Bridging the Divide: Scalarity and Modality

Peter Klecha
University of Chicago

Contents

1	Background	8
1.1	Defining Modality	8
1.1.1	Parameters of Variation	10
1.2	The Basic Framework	13
1.2.1	The Standard Model	13
1.2.2	Conditionals	16
1.3	Modal Strength Variability	16
1.3.1	Gradable Modality	17
1.3.2	Imprecise Modality	18
1.3.3	Modified Modality	19
2	Gradability and Modality	20
2.1	Gradability	21
2.1.1	Degree Modification	22
2.1.2	Scale Structure	24
2.1.3	Positivity	28
2.1.4	Coercion and Metalinguistic Comparison	31
2.1.5	Scalarity and Gradability	32
2.2	Gradability of Modal Adjectives	33
2.2.1	Corpus Study of Adjectives and Degree Modifiers	34
2.2.2	Acceptability Judgment Task	47
2.3	Scale Structure of (Gradable) Modal Adjectives	53
2.3.1	Gradable Modals	53
2.4	Conclusion	65
3	The Semantics of Gradable Modality	69
3.1	Lewis and Kratzer	71
3.1.1	Proximity to Ideal	73

3.1.2	The Divide between Scalarity and Modality	74
3.1.3	Core Kratzer	75
3.2	Gradable Epistemic Modality	76
3.2.1	Important Properties of the Ordering	77
3.2.2	Epistemic Entailments on Probabilistic Accounts	81
3.3	Gradable Deliberative Modality	85
3.3.1	Polysemy	86
3.3.2	Modality of Importance	87
3.3.3	Weak Teleological Necessity	88
3.3.4	Deriving Entailment Relations	94
3.4	Comparisons	96
3.4.1	Comparative Possibility	96
3.4.2	Mathematical Probability	98
3.5	Conclusion	102
4	Imprecision and Modality	103
4.1	Contextual Variability in Modals	105
4.2	Imprecision	111
4.2.1	Semantic Context and Pragmatic Context	114
4.2.2	Pragmatic Slack	115
4.2.3	The Pragmatic Model	118
4.2.4	Implicision	124
4.2.5	Precisification	127
4.2.6	Previous Discussion of Imprecision	130
4.2.7	Interim Summary	135
4.3	Imprecise Modality	136
4.3.1	Strength of Necessity in Context	136
4.3.2	Possibility and Precisification	140
4.4	Conditionals	144
4.4.1	Two Kinds of Sobel Sequence	147
4.4.2	Similarity and Causality	153
4.4.3	Applying Imprecision to Conditionals	158
4.4.4	Previous Discussion	160
4.5	Conclusion	167
5	Modification and Modality	169
5.1	Modifiers of Modal Auxiliaries	172
5.1.1	What <i>easily</i> Is Not	174

5.1.2	Degree Semantics	175
5.1.3	Composition	176
5.1.4	Context Sensitivity	181
5.1.5	Interim Conclusion	182
5.2	Non-Degree Intensification	183
5.2.1	Adjectival Modification	183
5.2.2	Auxiliary Modification	186
5.3	Alternative Manipulators	189
5.3.1	Precisifiers	189
5.3.2	Approximators	194
5.4	Conclusion	197
6	Conclusion	200
6.1	Semantic Scalarity	202
6.2	Pragmatic Scalarity	203
	Bibliography	204

Acknowledgments

Dissertations are hard. I am not actually capable of writing one. But I did, thanks to a lot of help from a lot of people. I hope to acknowledge a reasonably large fraction of them here.

First I would like to thank the many many people who I talked with about this dissertation, or some subcomponent of it, who shared their time and energy with me despite no official arrangement of any kind. This is surely not a complete list, but here goes: Fabrizio Cariani, Ezra Cook, Kai von Fintel, Thony Gillies, Michael Glanzberg, Wesley Holliday, Thomas Icard, Graham Katz, Stefan Kaufmann, Greg Kierstead, Dan Lassiter, Sarah Moss, Barbara Partee, Paul Portner, Kyle Rawlins, Jessica Rett, Craige Roberts, Aynat Rubinstein, Ken Safir, Roger Schwarzschild, Stewart Shapiro, Seth Yalcin, and audiences at the University of Chicago Workshop on Semantics and Philosophy of Language, the Northwestern University Philosophy of Language workshop, the University of Massachusetts Semantics Reading Group, the University of Ottawa Workshop on Modality, the Michigan State University Undergraduate Linguistics Colloquium, an Indiana University colloquium, and a Rutgers University colloquium.

But most of all, I would like to thank my committee and especially my chair, Chris Kennedy. If it is not impossible to list every way in which he has supported me throughout my time in Chicago, it would certainly be tedious to do so. So I'm not going to even try. CK is the best; he is constantly and relentlessly awesome, and his mentorship has touched every aspect of my academic life, and some aspects of my personal life, too. Thanks as well to the rest of my dissertation committee: Itamar Francez, Anastasia Giannakidou, and Malte Willer. I cannot count the hours I have spent trying to convince Itamar of things I had already convinced him of, and in the process, usually unconvincing myself of them; Itamar has been a great advisor and a good friend. Anastasia has ever been an invaluable resource to me,

especially as I wander into the territory of things like polarity, tense, and mood. And Malte has always been kind as he reminded me of my ignorance of the work of countless philosophers of language. I had a fantastic committee, and wouldn't trade any of them for anything. Thanks as well to Angelika Kratzer, who hosted me at UMass, for tons of invaluable perspective, insight, and generosity.

I had other fantastic teachers at the University of Chicago who may not be officially recognized but contributed immensely to my dissertation and/or growth as a scholar, including Karlos Arregi, Amy Dahlstrom, John Goldsmith, Lenore Grenoble, Greg Kobele, Jason Merchant, Jason Riggall, Jerry Sadock, Ming Xiang, and Alan Yu. Thanks especially to Greg and Ming for their input and help with this dissertation, and to Karlos for many years of advice and many pitchers of beer.

My linguistics education began at Michigan State University, with Cristina Schmitt's Intro to Linguistics class. The next semester I took Intro to Syntax with her and one homework she wrote "You are CLEARLY a linguist." I can never thank Cristina enough for encouraging me to major, to apply for graduate school, and for all around being my guiding light throughout undergrad. Likewise Alan Munn, who taught me my first graduate linguistics course, and helped me tremendously with the application process, and to this day permits me to bother him for LaTeX advice. I was also delighted that Alan bookended my linguistics education by attending my dissertation defense in Chicago. Finally, Marcin Morzycki, who introduced me to semantics, who, like Cristina and Alan, took a very special and encouraging interest in the unformed lump of human clay that I was, who talked with me for long hours in his office about much more than just linguistics, and who I have aspired to emulate for the past decade or so.

My intellectual growth really began all the way back in high school, so I would like to acknowledge several teachers who, though they will probably never read this, had a profound influence on me: Vicki Granger, Mike Karas, David Kirck, Rich Kopas, Xavier Pankovits, Rudolph Saenz, and Mike Sza-tkowski.

I might have gotten through graduate school without my friends, but it would have been a mostly pointless, miserable experience. Thanks to my cohort; Ryan Bochnak, Jasmin McCrory, Alice Rhomieux, Susan Rizzo, and Christina Weaver; they taught me an incredible amount and continue to do so. #! Thanks as well to what may as well be my second cohort; Rebekah Baglini, Tim Grinsell, Martina Martinović, and Julia Swan. Without them,

nothing. Thanks to my roommates over the years during graduate school; Andrew Bednar, Volodja Borshev, Bill Donahue, Jesse Harris, Tom Houser, Barbara Partee, Graham Rosby, Ben Schapiro, Bryan Tebeau, Ashly Will, as well as Julia, Rebekah, and Ryan, who tolerated me so well they decided to put up with me at school *and* home. Thanks to the linguistics pub trivia team, the Jerry Sadock Experience, whose core members include Ryan Bochnak, Jackie Bunting Reber, Erik Levin, Jett McAlister, Michael Meeuwis, Patrick Midtlyng, Chris Straughn, the man himself, on a few occasions, and many, many others. Thanks to all the students I have had over the years, who are too many to name, but especially to those who have become my friends, and who have generally taught me as much as or more than I ever taught them; Julia Goldsmith-Pinkham, Lelia Glass, Graham Rosby, and Steve SanPietro. Thanks to my party, Gallagher Flinn, Ben Schapiro, Mike Pham, and Jason Riggle; to the Dead CKs, Julian Grove, Steve SanPietro, and Karen Ye, and to the members of Fight Club, who shall remain anonymous. Thanks to all my other friends in the alarmingly fantastic linguistics department, of whom the following is surely only a partial list: Carissa Abrego-Collier, Helena Aparicio, Max Bane, Andrea Beltrama, Juan Bueno Holle, Tasos Chatzikonstantinou, Katie Franich, Tom Grano, Aidan Gray, Emily Hanink, Doro Hoffmann, Arum Kang, Jon Keane, Ed King, Niko Kontovas, Jackson Lee, Yaron McNabb, Prerna Nadathur, Asia Pietraszko, Diane Rak, Adam Singerman, Morgan Sonderegger, Theano Stavrinos, Matt Teichman, and Tamara Vardomskaya; thanks as well to neighbors to the north, Fabrizio Cariani, Ezra Cook, Magda Kaufmann, and Stefan Kaufmann. Thanks to those few individuals outside of the department who have managed to befriend me; Lobke Aelbrecht, Ben Cannon, Ezra Cook, Maggie Fritz-Morkin, Sarah Ingraham, Eric McCrory, Russell Rhomieux, John Sanders, and Miranda Sklaroff. Thanks going way back to my high school and college friends who will never read this but have continued to be a great presence in my life: Andrew Bednar, Derrick Burton, Tom Castellani, Jason Cormier, Andrés Galarza, Devin Iler, Lou Leskoviansky, Colin Nemchik, and Ed Simpson. And thanks again to those who have touched my life tremendously and been my very best friends; Julia Goldsmith-Pinkham, Rebekah Baglini, Andrew Bednar, Ryan Bochnak, Tom Castellani, Jason Cormier, Gallagher Flinn, Julian Grove, Ed King, Martina Martinović, Jasmin McCrory, Prerna Nadathur, Steve SanPietro, Julia Swan, and Bryan Tebeau.

Of course, I would like to acknowledge my family, who more than anyone made me the person I am. My mom Patty, and my dad John, who sacrificed

greatly on my behalf. My siblings Dave, Joe, Mary Lou, and Charlie, who I have looked up to and been inspired by my whole life. My quasi-siblings, Tarri, Liz, Julie, and Julie, who have supported me as a brother. And finally to my nieces and nephews, Alexa, Call, Hannah, and Tony, who are everything to me.

And however it may seem, I would like to lastly acknowledge, and dedicate this dissertation to, myself. Above everything this has been a personal journey. For every dear friend I have made, and for every friendship that has deepened, I have learned more about myself and grown to love myself better. Go me! I did it!

Abstract

This dissertation interrogates the boundary between scalarity and modality, two phenomena which have lengthy but largely independent histories in natural language semantic analysis. Particularly this dissertation examines cases in which the strength of a modal expression may vary across lexical or discourse contexts, where that variation can be analyzed as correlating with a conceptual scale of some kind. The upshot of this is that this kind of scalarity can occur in many different ways; this dissertation examines three.

First, gradable modality, in which scalarity is an inherent semantic property of the modal. Second, imprecise modality, in which scalarity is a pragmatic property of use of the modal in communication. Third, modified modality, in which scalarity is induced extrinsically by another expression. None of these kinds of scalarity is special to modality; this is the major point of this thesis. A given modal may, or may not, exhibit these properties.

Gradable modality has seen a recent focus in the literature, with many challenges to traditional models of modality. Against this background I argue that the existence of gradable modality does not fundamentally challenge the prevailing view of non-gradable modality. Imprecision, on the other hand, has not previously been argued to have an effect of modality, so this dissertation makes a novel claim: That modals are in fact subject to imprecision, which should be modeled as a pragmatic phenomenon, and that moreover this can be used to explain several long contested puzzles about modality. Finally, modification of modality has received very little treatment; I argue that modification of a modal can induce scalarity despite a lack of inherent scalarity in said modal, both by semantic and pragmatic means – a microcosm of the larger picture of scalarity in modality.

The analysis therefore is heterogeneous – and this is just as we should expect. There is nothing about modals that should make us think that they are especially likely to be gradable or ungradable, precise or imprecise, mod-

ifiable or unmodifiable. But there is much insight to be gained by examining this variation.

Chapter 1

Background

This dissertation comprises an investigation into the variety of ways in which the strength of a modal expression can vary in different linguistic or pragmatic contexts. Before discussing the data that this dissertation is primarily focused on, I first review the basic concepts and terminology relating to modality as well as establish the theoretical foundations and assumptions I build on in later chapters.

1.1 Defining Modality

If you ask a typical semanticist to define *modal* they will probably tell you a modal is a quantifier over possible worlds. And while this is indeed a popular theory, it is not a universally adopted one. What's more, the notion that modals are quantifiers over worlds is an analysis, but not a description, of modality. Another answer you might get is that modals are used to talk about non-real things.¹ There might be some sense in which this is right, but this definition is problematized by what [Kratzer \(1981a\)](#) calls realistic modals, like ones that denote mathematical necessity.

- (1) It's mathematically necessarily true that two plus two is four.

The definition that I adopt here is that modal expressions are those which involve modal operators, which are relations between two pieces of information. This might seem a bit odd, especially if we use the term *information* in

¹[Portner \(2009\)](#), for example, very tentatively adopts a definition like this.

a colloquial sense. Colloquially, the term *information* is used to describe *true* information. But there can be non-true information too – a work of fiction can be thought of as a large body of information, which does not actually describe reality. Things like goals, desires, rules, and (potentially false) beliefs, can likewise be thought of as information. The common definition of modals as being non-real things can be traced to the large bevy of modals which relate a proposition to such non-real information.

Following [Kripke \(1959\)](#) and many others the way I choose to represent information is with possible worlds. A piece of information is a proposition, i.e., a set of possible worlds; the set of worlds described truthfully by that information. If a proposition is true, then the actual world we live in is among the set of worlds which represents that information; thus though the proposition is true, it contains worlds which are not the actual world. Simply asserting a proposition, on a very simple view, means becoming committed to inclusion of the actual world in the set of worlds which that proposition represents.

So as [Portner \(2009\)](#) points out, the existence of modals, which relate propositions, allow one to say things about or on the basis of situations (worlds) which need not be real. Because a modal allows us to take a proposition, which we otherwise might just assert, and simply relate to some other information, which we may also not be committed to.

In formal terms I analyze modal operators as relations between two sets of possible worlds. According to Generalized Quantifier theory ([Barwise and Cooper, 1981](#)), a quantifier is simply a relation between two sets. So according to the theoretical choices I make here, this definition lines up with what most semanticists will tell you: A modal is a quantifier over worlds. But this is a theoretical choice.

Modal logic, which underlies possible worlds, and therefore many analyses of modal expressions, has its roots in philosophy and logic, with too many contributions and sources to cite here. But the application of this logic to the analysis of certain natural language expressions begins with [Stalnaker \(1968\)](#), [Hintikka \(1969\)](#), [Lewis \(1973\)](#), and [Kratzer \(1977\)](#), along with much other work by those authors and others. While Stalnaker and Lewis were concerned most with counterfactual conditionals, Hintikka analyzed attitude verbs (like *believe*, *want*, and *know*) and Kratzer focused initially on modal auxiliaries (like *may*, *might*, and *must*) as well as a semantics for integrating them with the semantics of conditionals, including counterfactuals, as discussed by Stalnaker and Lewis.

There are other modal expressions as well, and they can come in nearly any syntactic packaging; we can often get clued in to their modal nature by finding a paraphrase that includes a classic, commonly acknowledged modal term, like *possible* or *necessary*. The degree modifier *too* means something like “the relevant degree is higher than what is possible” (where *possible* is interpreted relative to something like a rule or a goal; more on this below.) Even the determiner *some* can have a modal kind of meaning; when I say (2), I am committed to the notion that there is no one guy who I *know* to be the guy who called.

(2) Some guy called earlier today.

The syntactic variation among modal expressions is enough to show that modal expressions themselves cannot be defined in terms of their semantic type, but I argue that they do share a common core: A modal operator of type $\langle\langle s, t \rangle, \langle\langle s, t \rangle, \alpha \rangle\rangle$, a relation between two sets of worlds. A modal *expression* is thus an expression whose meaning has to be stated in terms of such an operator.

Of the modal operators of type $\langle\langle s, t \rangle, \langle\langle s, t \rangle, \alpha \rangle\rangle$, by far the most commonly discussed are those of type $\langle\langle s, t \rangle, \langle\langle s, t \rangle, t \rangle\rangle$, i.e., those that map two sets of worlds to a truth value. We can call these modals *semantically categorical*. The “standard” theory of modality as it has been built up from the starting point of Kratzer (1981a) relies exclusively on modals having (essentially) this type.

What this dissertation is concerned with is ways in which modals may not be totally categorical; i.e., there may be some gradable or scalar element to certain modals. This will require some divergence from the “standard” theory although much of it is well within the spirit of the analysis laid out by Kratzer (1981a) and supported by many others. I begin by first reviewing this standard theory and pointing out several crucial points of departure.

1.1.1 Parameters of Variation

As a quantifier over possible worlds, the modal relates two sets of worlds asymmetrically. One set, which serves as the domain, I call the **modal domain**,² while the other set, which feeds the nuclear scope, is called the **pre-**

²This is often called the *modal base* elsewhere. See below for a discussion of this terminological confusion.

jacent. So there are essentially three ingredients to any modal construction: the quantifier itself, the domain, and the prejacent. For most well-studied modal expressions the prejacent is determined by the syntactic complement of that expression. The other two ingredients may vary lexically, and are discussed in detail below.

1.1.1.1 Force

The first parameter of lexical variation for modals is their force, or strength. Two basic types of force are necessity and possibility. At least in English, force is always lexically specified.

- (3) a. Ashly must be asleep. (NECESSITY)
- b. Ashly might be asleep. (POSSIBILITY)

Since (3a) entails (3b), it is the stronger of the two. (3a) is a necessity modal, modeled as a universal quantifier over worlds; (3b) is a possibility modal, modeled as an existential quantifier over worlds.

The main point of this dissertation is that modal strength varies much more substantially than this, and that even within a given lexical modal like *must* or *might* there can be significant variation in the force due to both linguistic and discourse context. Two other varieties of force which have received attention in the literature, and which illustrate this point a bit, are ‘weak’ necessity modals and gradable modals.

- (4) a. Ben should be at work by now. (WEAK NECESSITY)
- b. Ben is likely to be at work by now. (GRADIENT FORCE)

It is fairly intuitive that *should* is weaker than *must*, and thus is weaker than ‘regular’ necessity. But the analysis of weak necessity has been essentially that it involves universal quantification over a more restricted domain; thus it in some sense has the same force as other necessity modals. Gradable modals, which have received comparatively little discussion in the literature until very recently, will be discussed at length later.

1.1.1.2 Flavor

The second parameter of variation is in terms of the modal domain, and what kind of worlds populate it; this is often called *flavor*. Kratzer (1981a) proposes that this is determined by the kind of **conversational background**

that a modal takes. Conversational backgrounds are contextually determined (and therefore implicit) elements which help to fix the domain of a modal. There are many kinds of conversational backgrounds and thus modal flavors, but a few large families of flavors are ubiquitous and so are worth reviewing.

Epistemic conversational backgrounds are those which fix a set of worlds consistent with a set of information known by a particular agent or agents at a given time.

(5) *Epistemic Conversational Background*

- a. Chuck must have left Chicago by now.
- b. Chuck might have left Chicago by now.

Metaphysical or historical conversational backgrounds are those which fix a set of worlds consistent with a set of information that describes the real world, up to a given point in time.

(6) *Metaphysical Conversational Background*

- a. Drew is bound to be here next weekend.
- b. Drew might be here next weekend.

Deontic conversational backgrounds are those which fix a set of worlds consistent with a set of idealized information stipulated by some authority, i.e., rules.

(7) *Deontic Conversational Background*

- a. Erik has to submit his taxes soon (according to the law).
- b. Erik can submit his taxes soon (according to the law).

Teleological conversational backgrounds are those which fix a set of worlds consistent with a set of information describing the achievement of a particular individual's goals.

(8) *Teleological Conversational Background*

- a. Felicia has to take the F Train (to get to Brooklyn).
- b. Felicia could take the F Train (to get to Brooklyn).

Ability readings of modals seem to involve a distinct conversational background type, though there is significant debate as to what that is and what force ability modals have.

(9) *Ability Conversational Background*

Graham can solve a crossword puzzle in under 15 minutes.

This does not exhaust the list of possible modal flavors, but most discussion in this dissertation will be limited to these.

1.2 The Basic Framework

According to [Kratzer \(1977\)](#) a given modal lexical item may take different conversational backgrounds; she proposes that modal auxiliaries (in English and German) are lexically specified for force but underspecified for what kind of conversational background they may take. However, at least some modals must also carry some presuppositions about what kinds of conversational backgrounds they may take, since not all modals can take any conversational background.³ Consider the judgments below for Colloquial Standard American English (the stigma % indicates acceptability in some registers).

	EPI	MET	DEO	TEL
Helena must be in Mallorca.	✓	#	%	%
Itamar has to have two forms of ID.	✓	#	✓	✓
Jasmin should not be drinking.	✓	#	✓	✓
Katie is bound to have a license.	✓	✓	#	#
Lelia can be standing by the door.	#	#	✓	✓
Max could be the drummer.	✓	✓	✓	✓
Nick might write short stories.	✓	✓	#	#
Octavia may be in the country on a student visa.	✓	✓	%	#

Table 1.1: Compatibility of modals with four basic flavor types.

A fully complete analysis of the meaning of any modal expression must include these presuppositions (at least for modals that are not unrestricted in terms of what conversational background they may take). However, I will elide them in this dissertation for expository ease.

³It is also worth pointing out that there are significant restrictions on the co-occurrence of certain flavors with certain temporal orientations; see [Condoravdi \(2002\)](#) and [Klecha \(to appear\)](#). To be sure that a given modal cannot combine with a given modal flavor it is important to check for past, present, and future temporal orientations, e.g., by using different temporal adverbials.

1.2.1 The Standard Model

According to Kratzer (1981a), the flavors discussed in Section 1.1.1.2 (and others) arise from the interaction of two distinct sorts of conversational background, the **modal base** and the **ordering source**. The modal base is a set of propositions which are intersected to form what I call the **preliminary modal domain**. In the case of an epistemic reading, the modal base would be a set of propositions known at a given time. In the case of a circumstantial reading, the modal base would be a set of propositions which describe the circumstances that hold at a given time. (Concerns relating to time are glossed over in the formalisms.) By taking the intersection of, say, an epistemic modal base, we get an **epistemic preliminary modal domain**: The set of worlds consistent with some body of knowledge.

An ordering source is a second set of propositions which serves to narrow the modal domain further. A deontic ordering source, for example, is a set of rules. A teleological ordering source is a set of goals. What distinguishes modal bases from ordering sources is that latter establish a *gradient* notion of consistency, while the former only a categorical one. A world is either consistent or not consistent with the objective circumstances, but a world can satisfy someone's goals better or worse. This gradient notion of consistency is of central interest to this dissertation.

Using the ordering determined by the ordering source (see below for a more detailed discussion of how it is determined), the preliminary modal domain is then narrowed to just its best subset. Kratzer proposes three types of modal bases and many more varieties of ordering source. Modal bases may be epistemic, circumstantial, or empty, while ordering sources may be stereotypical, deontic, teleological, or empty, among other possibilities. Thus an epistemic modal domain arises from an epistemic modal base with a stereotypical or empty ordering source (more on this in Chapter 4) while a deontic modal domain arises from an epistemic or circumstantial modal base with a deontic ordering source.

If we have an ordering source g and a preliminary modal domain m , and a world of evaluation w , we can use the formula $\text{BEST}_{g(w)}(m)$ to stand in for the modal domain determined by applying an ordering source. Then, on a fairly strict Kratzerian view, we can imagine a denotation for *must* looks like this (10). In this formula p is a variable over propositions, type $\langle s, t \rangle$; f and g are variables over functions from worlds to sets of propositions, type $\langle s, \langle \langle s, t \rangle, t \rangle \rangle$; and w is a variable over worlds, type $\langle s \rangle$.

$$(10) \quad \llbracket \text{must} \rrbracket = \lambda p \lambda f \lambda g \lambda w [\forall v \in \text{BEST}_{g(w)}(\cap f(w))][p(v)]$$

It is worth pointing out that there is significant terminological variation in the literature here. For example, with respect to the formula above, any of the following components could be called the *modal base*.

- (11) a. f : a function from worlds to sets of propositions
- b. $f(w)$: a set of propositions
- c. $\cap f(w)$: a set of worlds
- d. $\text{BEST}_{g(w)}(\cap f(w))$: a set of worlds

To avoid confusion I elect to use distinct terms for each of these components and adhere to these terminological choices throughout this dissertation. I refer to the set of propositions (11b) as the *modal base*, since this appears to be Kratzer’s original usage. I refer to the function from worlds to modal bases (11a) as **modal base intensions**. As noted above I refer to the set of worlds gotten by doing intersection over a modal (11c) as the *preliminary modal domain*. The final set of worlds after restriction by the ordering source is the *modal domain*.

Likewise I reserve *ordering source* for a set of propositions used to generate an ordering, and **ordering source intension** for a function from worlds to ordering sources. Thus *must* in (10a) denotes a (curried) function from a proposition, a modal base intension, an ordering source intension, and a world to a truth value. This isn’t quite the type of a modal operator, but it clearly involves one. We can abbreviate a true necessity operator as follows:

$$(12) \quad \text{NEC}(x, y) = \forall w \in x [w \in y]$$

where x and y are sets of worlds

We can then state the meaning of *must* in terms of it:

$$(13) \quad \lambda p \lambda f \lambda g \lambda w [\text{NEC}(\text{BEST}_{g(w)}(\cap f(w)), p)]$$

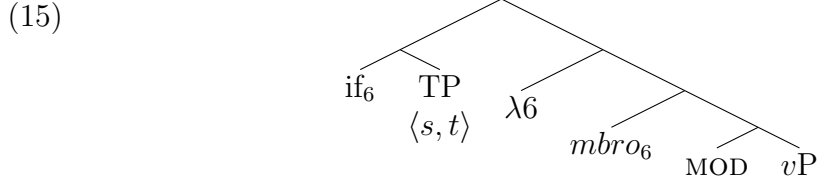
Because the first argument, the domain, is determined in such a complex way, the type of *must* is rather complicated. But it can very clearly be seen as a *lift* of the basic type of modal operators. The best studied modal expressions have always been the expressions whose type are closest to the fundamental type of modal operators, i.e., expressions like *must*. But other expressions can have even more complicated types, like *too* and *some*. Semanticists are generally less comfortable calling these expressions ‘modal expressions’

or ‘modals’, especially in the case of an expression like *some*, where the modal component of the meaning may be seen as secondary or backgrounded. But it is not clear how to draw the line. Fortunately, it is not a matter of great theoretical import, and so, following tradition, will not choose to define ‘modal’, though I will use ‘modal expression’ to have the meaning discussed above, i.e., the very permissive sense.

1.2.2 Conditionals

I follow Kratzer (1986) in taking conditionals to be simple domain restrictors, which serve to narrow the *preliminary* modal domain intersectively. A few assumptions allow for a simple compositional account along these lines. First, we can assume that modal base intensions are represented in the syntax as silent pronouns, which we can call *mbro* (part of Chapter 5 will depend on this assumption as well; see there for more details). If *if* can then be given a semantics which allows it to bind the modal base intension index carried by *mbro*.

$$(14) \quad \llbracket \text{if}_6 \rrbracket = \lambda p \lambda f \lambda w [f(\lambda v [g(6)(v) \cup \{p\}])(w)]$$



As discussed by Klecha (2014), conditionals are a crucial diagnostic for modality. Because, as discussed above, modals are relations between a domain and a prejacent, but only the prejacent is ‘visible’. Thus the modal status of some expressions is more controversial than others. Klecha (2014) discusses the case of the English predictive expressions *will* and *gonna*, showing that they are modals by showing that their domains may be restricted by conditionals, both overt and covert.

Thus any theory of modality must either adopt a version of Kratzer’s theory of conditionals and take modals to be relations between domains and prejacentes, or else offer an alternative by which conditionals do not derive from domain restriction.

1.3 Modal Strength Variability

The central topic of this dissertation is variability in the strength that modals display. In other words, a given modal *lexical item* may display differing strength in differing contexts. Already we have the tools to discuss two specific ways in which this could occur. Both involve the more general strategy of modulating the size of a modal’s preliminary modal domain.

One method for varying the strength of a given modal is through variation in its modal base. Depending on the discourse context, a larger or smaller modal base may be selected. The fewer propositions included in the modal base, the more worlds are included in the preliminary domain. In the case of a possibility modal like *can* this means the assertion is weaker. Thus even controlling for the ‘flavor’ of the modal, we can have larger and smaller modal bases.

A second method involves keeping the modal base constant, while modifying the preliminary modal domain through conditionalization as shown in Section 1.2.2. Here the domain may be contracted, resulting in a stronger assertion in the case of a possibility modal, and weaker in the case of a necessity modal.

What distinguishes these two cases of modal strength variability from the ones that occupy the central focus of this dissertation is that these are not scalar. In other words we cannot identify a (nontrivial) conceptual scale which correlates to the scale of strength displayed by these modals. Both of these methods of strength modulation simply involve adding or subtracting information from the modal domain in discrete, categorical chunks.

The focus of this dissertation is the way strength may vary in non-categorical ways, along a continuous scale. This dissertation identifies at least three ways this may occur, and argues for a conservative approach which allows for the basic theory of categorical modals sketched above to be retained.

1.3.1 Gradable Modality

Gradable modals are modal expressions which combine with degree modifiers, a class of expressions whose function is to map a scalar predicate into a categorical one. An example, discussed at length in this dissertation, is *likely*.

- (16) Scenario A is more likely than Scenario B. Scenario B is more likely than Scenario C.

Assuming that the NPs *Scenario A*, *Scenario B*, and *Scenario C* have propositional meanings, each of these propositions is related to some modal domain (say, an epistemic one) to differing degrees. (16) entails something stronger about Scenario A than about Scenario B, and likewise something stronger about Scenario B than about Scenario C. We could imagine that *likely* is a necessity modal of a certain kind, with progressively smaller domains for each of the scenarios. But the apparently continuous nature of the scale associated with *likely* suggests this isn't the case – we can't keep coming up with discrete propositions to add or subtract to the domain indefinitely. Moreover, we need a way to relate the semantics of *likely* to some reasonable kind of semantics for degree modifiers, like Kennedy's (1999).

In **Chapter 2** I examine modal adjectives and argue that while some of them are gradable, some are not, which means that we must have a semantics which is compatible both with non-scalar and scalar modals. In **Chapter 3** I then provide a semantics for such gradable modal adjectives which does allow for a conservative view of non-categorical modals.

1.3.2 Imprecise Modality

Gradable modality essentially is a case of scalarity being introduced by the semantics; i.e., it is a conventional property of the modal *likely* that it introduces a scale. But such scales may also arise pragmatically. Consider a sentence like (17).

- (17) It's raining outside, so you have to wear a raincoat.

The strength of this claim is semantically fixed, assuming a given modal domain (say, a given circumstantial modal base with a given teleological ordering source). But as I argue in **Chapter 4**, pragmatics may affect the modal strength that is *communicated*. In other words, the strength of the above may vary along a scale that corresponds to the scale of imprecision which governs the permissible 'looseness' of utterances in a discourse. If we are in a very imprecise context, we may ignore all but the most reasonable possibilities, in which case (17) is relatively weak, meaning something more like the strict meaning of (18).

- (18) It's raining outside, so you have to wear a raincoat unless you wanted to do something weird.

In a less precise context, we may ignore some very outlandish possibilities, but not less outlandish ones. In such a context, (17) has a stronger meaning. And it has the strongest meaning in a context where we cannot ignore any possibilities, i.e., a highly precise context. In this context, (17) is very strong.

1.3.3 Modified Modality

The last way to vary the strength of a modal along a scale is to extrinsically introduce variation with a modifier. In this case a categorical modal is modified by an expression which introduces a scale. Two examples are below.

- (19) a. You could (easily) have fallen.
b. You (absolutely) have to wear a raincoat.

In each case the presence of the modifier in parentheses alters the strength of the categorical modal it adjoins. In the case of (19a), this involves domain restriction like that discussed above, but by means of a gradable expression. In the case of (19b), the modifier gives rise to pragmatic precisification, inducing variation of the variety described in Chapter 4.

In **Chapter 5** I analyze these expressions compositionally, explaining how a categorical modal can be related to a scalar semantics or pragmatics through modification.

Chapter 2

Gradability and Modality

The purpose of this chapter is to determine the range of semantic variation among modal adjectives with regard to gradability and scale structure. To this end, I argue that modal adjectives may be gradable or non-gradable. Moreover I argue that modals may vary in their scalar structure, consistent with the notion that there is no single, universal ordering which has central importance to modality; rather variation among gradable modals in scale structure is no different than variation among gradable properties of individuals.

This is in part a rejection of [Kratzer's \(1981a\)](#) implicit claim that all modals can be modeled as either universal or existential quantifiers over worlds. But it also supports a more fundamental aspect of Kratzer's theory, namely, that modals are relations between a modal domain and a prejacent.

If modals are just expressions which relate a prejacent and a modal domain, it follows naturally that this relation may or may not be a gradable one. Nothing about this definition demands the modality be uniformly categorical and therefore non-gradable, or that it be uniformly scalar. Of course, if indeed there are both scalar and nonscalar modals, we must have a way of relating them; this is the work of the next chapter. But in this chapter, I simply lay out what it means to be gradable and scalar, and show that while some modals are, some are not; and that among gradable modals, there is variation in scalar structure as well. This is contrary to what is claimed in [Lassiter \(2011\)](#).

The tests I will propose for gradability and scale structure follow from an extensive literature on the topic. I deem an expression gradable if it co-occurs with a robust sample of degree modifiers, expressions which manipulate the

scale introduced by gradable predicates. Likewise, particular degree modifiers diagnose particular scale structures, as is explained below. While this is fairly straightforward, the empirical landscape is littered with difficulties, not the least of which is disagreement among authors as to the acceptability of various expressions in English. Thus in this chapter I attempt to navigate said landscape with the aid of two quantitative studies.

Lassiter (2011) argues that all modals, including modal auxiliaries, are fundamentally scalar, but that such auxiliaries do not combine with degree modifiers (and therefore do not satisfy my definition of *gradable*) for purely syntactic reasons. Implicitly he therefore predicts that all modal adjectives should be gradable, since adjectives freely combine with degree modifiers (at least in certain syntactic positions). I therefore constrain my focus to modal adjectives.

This chapter sets the stage for what could be called a conservative theory of gradable modality; i.e., one which proposes re-thinking neither our semantics for gradability, nor for modality. In the first section I lay out some basic concepts relating to gradability and scalarity and argue for certain diagnostics for gradability and scale structure in adjectives. In the second section I present the findings of two empirical studies showing that modal adjectives including *possible* are not gradable. In the third section I discuss the implications of these findings for the scalarity of modal adjectives, and conclude in the fourth section. In the following chapter, I propose unifying semantics for gradable and nongradable modals.

2.1 Gradability

There are a number of ways in which natural language expressions can have a scalar interpretation. One way is for their conventional meanings to be specified as such; in other words, the semantics itself is scalar. This can be seen in expressions like gradable adjectives as in (1); compare these to non-gradable adjectives as in (2).

- (1) a. Chris is (very) tall.
b. Itamar is taller than Tim.
- (2) a. Julia is (*very) vegan.
b. Martina is (*more) Balkan.

I will assume a theory of gradability along the lines of Cresswell (1976), von Stechow (1984), Kennedy (1999) and Kennedy and McNally (2005). This theory takes *degrees* to be a basic type which the interpretation of gradable expressions depends upon. Degrees are abstract objects which can form a dense, linearly ordered scale, where the ordering is determined by a dimension (height, beauty, etc.). On this approach, gradable adjectives like *tall* denote measure functions, or functions from entities to degrees (type $\langle e, d \rangle$)¹. In other words, a predicate like *tall* in (1a) maps an entity like Chris to a degree – his height.

$$(3) \quad \llbracket \text{tall} \rrbracket = \lambda x [\text{HEIGHT}(x)]$$

This contrasts with the denotation of a simple non-gradable adjective, which maps an individual to a truth value. The difference between gradable and non-gradable adjectives is therefore their type. It is their type which in turn constrains the types of expressions they may combine with; it is because of this that we can readily diagnose gradability in adjectives.²

Below I review the semantics for the expressions which may combine only with gradable adjectives: degree modifiers.

2.1.1 Degree Modification

Degree modifiers are complex functions which take a measure function, possibly in addition to other arguments, and return a simple non-gradable predicate. Thus *six feet* combines with *tall* to produce a non-gradable predicate, *six feet tall*.³

$$(4) \quad \llbracket \text{six feet} \rrbracket = \lambda g_{\langle e, d \rangle} \lambda x [\max\{d : g(x) \geq d\} \geq 6 \text{ feet}]$$

¹Alternatively, they could denote relations between degrees and entities, $\langle d, et \rangle$, but this choice doesn't really matter for the present study.

²There are alternatives to the degree-based approach which do not take gradability to be a matter of type, e.g., Klein (1980). The discussion here is not actually sensitive to this concern; it is widely acknowledged that only gradable adjectives may combine with degree modifiers, whatever the analysis of gradability may be. For example, on Klein's analysis, gradable adjectives are the ones which are sensitive to a comparison class, while non-gradable (non-vague) adjectives are not. Thus the analysis of gradability itself is not a question under discussion here, and nothing in this chapter or the next hinges on this choice.

³Most proposals argue that this is mediated by a null functional head, but this isn't important here.

More complicated expressions like the comparative take an additional argument which denotes the *standard of comparison*, formally a degree; in (1b) above, the standard is denoted by *than Tim*. The individual who is denoted by the subject of a predicative construction is called the target, while the standard is the other degree involved in the comparison. On the degree view, gradable predicates are always used to denote some kind of comparison.

$$(5) \quad \llbracket \text{more than} \rrbracket = \lambda g_{\langle e, d \rangle} \lambda d \lambda x [\max\{d' \in g(x)\} > d]$$

That the semantics of gradable adjectives, and their degree modifiers, rely crucially on degrees, allows for a straightforward explanation of the incompatibility of degree modifiers with non-gradable adjectives, as shown in (2), which are assumed to denote simple properties, $\langle e, t \rangle$. A crucial result of this is that degree modifiers constitute a diagnostic for gradability. Throughout this study, I will rely on robust compatibility of an expression with a wide range of degree modifiers as the primary diagnostic for gradability.

Below are the acceptability judgments for *big* with degree modifiers.

- (6) *big* with degree modification
- a. The ball is **bigger** (than the block).
 - b. The ball is **very big**.
 - c. That ball is the **biggest** (in the room).
 - d. **How big** is the ball?
 - e. The ball is **too big** (to fit in the box).
 - f. The ball is **so big** (that it won't fit in the box).
 - g. The ball is **big enough** (to do some damage).
 - h. The ball is **as big** (as the block).
 - i. The ball is **pretty big**.
 - j. The ball is **that big**.
 - k. The ball is **this big**.
 - l. The ball is **quite big**.
 - m. The ball is **rather big**.
 - n. The ball is **less big** (than the block).

Non-gradable adjectives, in turn, should be robustly unacceptable with degree modifiers. Consider a classic example, *dead*.⁴

⁴ Kennedy and McNally (2005) actually argue that *dead* is gradable on the basis of (i).

- (i) The cat is **{almost/nearly/totally}** dead.

- (7) *dead* with degree modification
- a. *The linguist is {**deader/more dead**} (than the psychologist).
 - b. *The linguist is **very dead**.
 - c. *That linguist is the {**deadest/most dead**} (in the room).
 - d. ***How dead** is the linguist?
 - e. *The linguist is **too dead** (to give his talk).
 - f. *The linguist is **so dead** (that he can't give his talk).
 - g. *The linguist is **dead enough** (to be disregarded).
 - h. *The linguist is **as dead** (as the stiffs at my feet).
 - i. *The linguist is **pretty dead**.
 - j. *The linguist is **that dead**.
 - k. *The linguist is **this dead**.
 - l. The linguist is **quite dead**.
 - m. *The linguist is **rather dead**.
 - n. *The linguist is **less dead** (than the psychologist).

This diagnostic confirms that *dead* is not gradable.

2.1.2 Scale Structure

The set of degree modifiers listed above can be considered the most general degree modifiers, in that they do not impose many restrictions on what kinds of adjectives they may combine with. But there are still many such restrictions. For example, extreme adjectives (Morzycki, 2012) cannot combine with *very* or other non-extreme intensifiers.

Moreover, some degree modifiers take on different meanings when combined with certain kinds of adjectives. For example, *a little* in (8a) means “to a small degree” but in (8b) means “in excess, to a small degree”.

However, the use of *almost* and *nearly* as diagnostics for gradability is questionable. While these expressions are certainly scalar, they apply very widely, without respect to gradability:

- (ii) {Nearly/Almost} every student failed.

Moreover *totally* (and some other maximizing degree modifiers) can be used with a non-degree semantics.

- (iii) A: Seven isn't prime.
 B: What? Seven is *totally* prime.

- (8) a. The photograph is a little crooked.
- b. The photograph is a little dark.

Thus degree modifiers can be used not only to determine gradability in the first place, but also to highlight parameters of variation among gradable adjectives.

One such parameter is scale structure. According to [Kennedy and McNally \(2005\)](#) there are four basic kinds of scales which gradable adjectives can be associated with; upper-closed scales, lower-closed scales, fully-closed scales, and fully-open scales.

- (9) a. FULLY CLOSED SCALE: $[0,1]$
- b. UPPER-CLOSED SCALE: $(0,1]$
- c. LOWER-CLOSED SCALE: $[0,1)$
- d. FULLY OPEN SCALE: $(0,1)$

The parameters of variation here are: whether there is a highest-ordered element on the scale, and whether there a lowest-ordered member on the scale. While there is sometimes an intuitive scale type to be associated with a given gradable adjective, the only reliable diagnostic of scale type is compatibility with scale-specific degree modifiers, as discussed in [Rotstein and Winter \(2004\)](#) and [Kennedy and McNally \(2005\)](#). The semantics of these modifiers depend crucially on a particular scale structure; thus compatibility with them is diagnostic of that scale structure. This will be important for the discussion of gradable modal adjectives as it will crucially affect the semantics that we give for them.

Below I discuss each of these scalar possibilities in turn, using the adjectives *full*, *straight*, *tall*, and *dirty* to exemplify.

2.1.2.1 Maximizing Degree Modifiers

Maximizing degree modifiers are acceptable only with upper-closed scales, because their semantics pick out a maximum point on the relevant scale. Thus if there is no maximum point, these expressions will result in infelicity. The maximizing degree modifiers I examine are *totally*, *completely*, *perfectly*, and *absolutely*.

- (10) *full* with maximizing degree modifiers
- a. The cup is **totally full**.

- b. The cup is **completely full**.
 - c. The cup is **perfectly full**.
 - d. The cup is **absolutely full**.
- (11) *straight* with maximizing degree modifiers
 - a. The line is **totally straight**.
 - b. The line is **completely straight**.
 - c. The line is **perfectly straight**.
 - d. The line is **absolutely straight**.
- (12) *tall* with maximizing degree modifiers
 - a. #The boy is **totally tall**.
 - b. #The boy is **completely tall**.
 - c. #The boy is **perfectly tall**.
 - d. #The boy is **absolutely tall**.
- (13) *crooked* with maximizing degree modifiers
 - a. #The photo is **totally crooked**.
 - b. #The photo is **completely crooked**.
 - c. *The photo is **perfectly crooked**.
 - d. #The photo is **absolutely crooked**.

Crucially, compatibility must be assessed with respect to the *maximizing reading* of the modifiers. For example, *totally* can have non-degree reading, as mentioned in footnote 4 above. Moreover, *totally* and *completely* can have partitive readings; e.g., (14) can mean something like “every part of the room is dirty” but not “the room is dirtier than it can otherwise possibly be”. It is the absence of this reading in (13) and (14) that diagnoses a lack of a scalar maximum.

- (14) The room is totally dirty.

The possible presence of a partitive reading can be diagnosed with an overt partitive construction. If such a construction is possible, a partitive reading may be present even without the overt partitive.

- (15)
 - a. All of the room is dirty.
 - b. *All of the photo is crooked.

Thus, *full* and *straight* are associated with upper-closed scales; *tall* and *crooked* are not.

2.1.2.2 Proportional Degree Modifiers

Proportional degree modifiers are only acceptable with fully closed scale adjectives. The semantics of these expressions involves take a proportion of a scales, requiring both a minimum and maximum. The proportional degree modifiers I examine are *n percent* (for any value of *n*) fractions like *three-quarters* and *half* (Bochnak, 2013), complex modifiers like *all-the-way* and *most-of-the-way*, and simple modifiers *mostly* and *partially*. As with maximizing modifiers, partitive uses must be excluded.

- (16) *full* with proportional modifiers
 - a. The cup is **{10/50/100}% full**.
 - b. The cup is **{three quarters/half/one third} full**.
 - c. The cup is **{all/most/half}-(of)-(the)-way full**.
 - d. The cup is **mostly/partially full**.
- (17) *straight* with proportional modifiers
 - a. #The line is **{10/50/100}% straight**.
 - b. #The line is **{three quarters/half/one third} straight**.
 - c. #The line is **{all/most/half}-(of)-(the)-way straight**.
 - d. #The line is **mostly/partially straight**.
- (18) *tall* with proportional modifiers
 - a. *The boy is **{10/50/100}% tall**.
 - b. *The boy is **{three quarters/half/one third} tall**.
 - c. *The boy is **{all/most/half}-(of)-(the)-way tall**.
 - d. *The boy is **mostly/partially tall**.
- (19) *crooked* with proportional modifiers
 - a. *The photo is **{10/50/100}% crooked**.
 - b. *The photo is **{three quarters/half/one third} crooked**.
 - c. *The photo is **{all/most/half}-(of)-(the)-way crooked**.
 - d. *The photo is **mostly/partially crooked**.

While maximizing degree modifiers showed that both *straight* and *full* are upper-closed, only *full* is lower-closed as well.

2.1.2.3 Minimizing Degree Modifiers

Finally, minimizing modifiers like *slightly* and *a little* are compatible only with lower-closed scale adjectives (which are not also upper-closed). These modifiers crucially can also take an excessive reading with any kind of adjective (Bylinina, 2012, Solt, 2012) as shown above; this reading must be excluded.

- (20) *full* with minimizing degree modifiers
 - a. The cup is **slightly full**. (#min.)
 - b. The cup is **a little full**. (#min.)
- (21) *straight* with minimizing degree modifiers
 - a. The line is **slightly straight**. (#min.)
 - b. The line is **a little straight**. (#min.)
- (22) *tall* with minimizing degree modifiers
 - a. The boy is **slightly tall**. (#min.)
 - b. The boy is **a little tall**. (#min.)
- (23) *crooked* with minimizing degree modifiers
 - a. The photo is **slightly crooked**.
 - b. The photo is **a little crooked**.

Thus, *crooked* is associated with a lower-closed scale.

Open-scale adjectives do not comport with any of these diagnostics. Thus, these diagnostics show that *tall* is associated with a fully open scale.

2.1.3 Positivity

The view of gradable expressions discussed so far in a way makes degree modified forms of gradable adjectives seem like the basic case, while the unmarked form of the adjective, called the positive, requires further explanation.⁵

- (24) Erik is tall.

Since the gradable adjective denotes a measure function, it cannot combine directly with an entity-denoting expression like *Erik* to produce a proposition; rather, it must undergo some kind of type-shift. von Stechow (1984)

⁵But not on a non-degree semantics.

accomplishes this through the mediation of a silent operator, which he terms *pos*. On the analysis articulated there, *pos* is essentially a degree modifier as well, combining with the adjective to produce a non-gradable predicate. It also introduces a standard of comparison, however, and this standard varies depending upon the adjective, and for some adjectives, upon the context.

Kennedy and McNally (2005) establish a contrast between absolute and relative adjectives; those whose positive meaning is fixed are absolute, those whose positive meaning varies contextually are relative. Contrast *straight*, a canonical absolute adjective, with *tall*, a typical relative one.

- (25) a. That bamboo shoot is straight.
b. That skyscraper is straight.
- (26) a. That bamboo shoot is tall.
b. That skyscraper is tall.

For the cases in (25), *straight* has the same, absolute, positive meaning; namely, in each case it means “having the maximum degree of straightness”. But in (26), it is clear that the minimum possible target height associated with a true utterance of each example is different; perhaps ten feet in the first case, and five hundred feet in the second. External context may alter the standard even further.

External context may alter the standard at play in (25a-b) as well, but not nearly to the same extent. This alteration of the standard is a fully pragmatic effect which does not implicate *straight* as gradable, since it may affect non-gradable expressions as well. For example, (27) can be uttered even when there were only 99 people in the theater, at least in some contexts.

- (27) There were 100 people in the theater.

This phenomenon is known as *imprecision* and is discussed in depth in Chapter 4.

Within absolute adjectives, there is another distinction, between maximal and minimal adjectives. Maximal adjectives’ positive meanings are associated with the maximum point on the scale they are associated with (modulo imprecision). Synonymy between the positive form of the adjective and the same adjective with a maximizing degree modifier (on its maximizing reading) is the primary diagnostic for maximal adjectives, as in (28). Another more paraphrastic diagnostic is in (29).

- (28) *Synonymy of positive and maximizing modifier*
- a. The cup is full. \models The cup is totally full.
 - b. The ruler is straight. \models The ruler is totally straight.
 - c. The countertop is dry. \models The countertop is totally dry.
- (29) *Infelicity of positive with could be more A*
- a. The cup is full, #but it could be more full.
 - b. The ruler is straight, #but it it could be straighter.
 - c. The countertop is dry, #but it could be drier.

Maximal adjectives include *full*, its antonym *empty*, *straight*, *dry*, *clean*, *transparent*, and *pure*.

Minimal adjectives' positive meanings are associated with the minimum point on the scale they are associated with. As the converse of maximal adjectives, they should display the opposite behavior, the negative form being infelicitous with *could be less*, as well as failing the tests associated with maximal adjectives.

- (30) *Infelicity of negative with could be more A⁻¹*
- a. The ruler is not bent, # but it could be straighter.
 - b. The countertop is not wet, # but it could be drier.
 - c. The tablecloth is not dirty, # but it could be cleaner.

Minimal adjectives include *bent*, *open*, *spotted*, *bumpy*, *wet*, *dirty*, and *impure*.

Relative adjectives' positive meanings are associated neither with a minimum or maximum point on the scale they are associated with; as seen above, they can vary greatly depending on the context they are used in; they are also vague (Kennedy, 2007). Thus, failure of the diagnostics for minimum and maximum adjectives is one diagnostic for relative adjectives; contextual variation is another; vagueness is yet another. Vague predicates are those which give rise to the Sorites Paradox, where the (apparently true) premises in (31a-b) give rise to the infelicitous conclusion in (31c).

- (31) *Sorites Paradox*
- a. A man who is 6 foot five is tall.
 - b. A man who is one centimeter shorter than a tall man is also tall.
 - c. #A man who is 4 feet tall is tall.

Vague predicates also give rise to borderline cases; a borderline case is one in which the positive and negative forms of the adjective are false when applied to the target. Observe that such cases cannot arise for absolute adjectives.

- (32) *#A and #not A*
- a. I wouldn't say he's tall, but I wouldn't say he's not tall either.
 - b. #I wouldn't say it's full, but I wouldn't say it's not full either.
 - c. #I wouldn't say it's clean, but I wouldn't say it's not clean either.

Relative adjectives include antonymic pairs *tall* and *short*; *big* and *small*; and *bright* and *dim*.

2.1.4 Coercion and Metalinguistic Comparison

There is still the concern that degree modification with *dead* and other non-gradable adjectives is not totally unattested – moreover it seems unlikely that attestations of NGAs with degree modification can all be explained as speech errors. These can be explained, however, as cases of coercion, where NGAs behave like GAs so long as some kind of scale can be readily associated with their conventional, non-gradable interpretations. Consider, for example, (33).

- (33) *Alice:* The patient has died.
Bryan: Can he be resuscitated?
Alice: ?No, he's too dead.

Here it can be imagined that Alice means something like “it has been too long since he died”, or “his body is too damaged”. Conversely, Alice might mention a patient who is not too dead to resuscitate, i.e., the patient has no heartbeat but could be defibrillated. And while such sentences can be, and are, uttered, they are degraded relative to cases of truly gradable adjectives combining with degree modifiers. See also [Todorova et al. \(2000\)](#) who show that coercion results in processing difficulty.

A second option for explaining small-scale attestations of non-gradable adjectives is metalinguistic comparison. (34) can be understood as metalinguistic if it means something like “it would be more appropriate to say he's dead than to say that he's comatose”; see, e.g., [Morzycki \(2011\)](#).

- (34) He's more dead than comatose.

Since coercion and metalinguistic comparison can muddle the picture of gradability, the possibility of apparent gradability being due to one these phenomena must be carefully controlled for.

2.1.5 Scalarity and Gradability

Lassiter (2011) introduces an important distinction, between expressions which are *scalar* and those which are *gradable*. Lassiter argues that scalar expressions are those whose meaning depends on some kind of scale; this includes gradable expressions. Gradable expressions, however, are those whose scales are syntactically manipulable. Thus, an expression may be scalar without being gradable. A very basic example is *six feet tall* versus the bare adjective *tall*; clearly *tall* is gradable, being associated with a scale which can be manipulated by degree modifiers. But the complex expression *six feet tall*, while clearly built on *tall* and its scale, is itself not gradable since it can not (any longer) have its scale be manipulated by an external expression. Another example is non-gradable adjectives derived from gradable adjectives by the prefix *non-*.

- (35) a. *Chris is more six feet tall than Itamar.
b. *Tim is a little six feet tall.
- (36) I am only talking about (*very) non-full cups.

Defining *scalar* formally is a bit tricky since it is a metasemantic distinction; what if an expression can be translated into a formal language in a way that invokes a scale, but it can also be translated in a way that does not? However, *gradable* is a much more tractable term; it involves just the type of the relevant expression.

Henceforth I will use *scalar* in a very general way, alluding to any variety of scalarity, be it semantic or pragmatic; an expression is scalar if its meaning supports an ordering on the objects in its domain. And while I will still make use of the notion of scalarity as it is called by Lassiter, I will refer to it as *descalarity*. Thus a descscalar property P of type $\langle \alpha, t \rangle$ is one which for which there is some measure function δ , degree d , and degree relation $R \in \{>, \leq, =\}$ such that P can be characterized with the function $\lambda o_\alpha[\delta(o)Rd]$. Similarly, a descscalar relation Q of type $\langle \alpha, \langle \beta, t \rangle \rangle$ is one for which there is some relational measure function δ , degree d , and degree relation $R \in \{>, \leq, =\}$ such that Q can be characterized by $\lambda o_\alpha[\lambda o'_\beta[\delta(o, o')Rd]]$. A descscalar

expression therefore is one that can be characterized by a descalar function. But one point which I will pursue in Chapter 2 is that, at least among the domain of modal expressions, there are no *necessarily* descalar expressions – even if the meaning of an expression like *must* can be characterized in terms of a scale, it need not be.

As discussed above, basic gradable expressions are taken to be of type $\langle e, d \rangle$; however, it is easy to see that, especially for the purposes of this study, this is insufficiently general. There can clearly be gradable properties of objects of any type; moreover, degree modifiers are not picky about what kinds of gradable properties they combine with, so long as they are gradable.

- (37) a. This party is too soon.
b. That outcome is extremely unlikely.

I will therefore define gradable expressions as any expression of type $\langle \alpha, d \rangle$; i.e., an expression is gradable just in case it denotes a measure function (or a function from something to a measure function). Correspondingly, the semantics of degree modifiers must be relaxed a bit to account for this.

$$(38) \quad \llbracket \text{more than} \rrbracket = \lambda g_{\langle \alpha, d \rangle} \lambda d \lambda o_{\alpha} [g(o) > d]$$

This concludes the background discussion of gradability.

2.2 Gradability of Modal Adjectives

Lassiter’s claim that the non-gradability of auxiliaries is due to their syntactic properties makes the prediction that among adjectives, which are clearly syntactically compatible with gradability, all modals should be gradable. Lassiter explicitly argues for gradability of *possible* and *necessary* among many others. And while it is still coherent in Lassiter’s view that some modal adjectives may not be gradable, the motivation for the existence of adjectives which are descalar but not gradable is rather unclear.

To this is end, I argue that in fact modal adjectives are often non-gradable, including *possible* and *necessary*. As discussed in Chapter 1, the primary diagnostic for gradability is compatibility with degree modifiers.

- (39) a. It’s (more) likely that the ring was stolen.
b. It’s (very) important that the house be kept very clean.

- (40) a. It's (*more) possible that Johnson took it.
 b. It's (*very) necessary that the United States import oil.

Represented above are the author's native speaker judgments. However, while Lassiter does not disagree about the viability of the diagnostic, his own presented judgments differ from those above. Focusing on *possible*, and in light of the conflicting intuition reports, Lassiter goes so far as to claim that *possible* should be taken to be gradable on the basis of robust corpus attestation of its co-occurrence with degree modifiers.

I agree with Lassiter that since intuition reports are conflicting, we must turn to quantitative measures to assess this question about the grammar. Below I present two such studies. The first is a cluster analysis of co-occurrence frequencies culled from the Corpus of Contemporary American English (Davies, 2008), henceforth COCA, in which I show that Lassiter's claim that expressions like *more possible* are robustly attested cannot be reproduced. In the second study I conduct an acceptability judgment task using Amazon Mechanical Turk, and show that *possible* cannot combine with degree modifiers.

2.2.1 Corpus Study of Adjective/Degree Modifier Co-Occurrence

The reliability of attestation, especially through non-specialized corpora like Google, as a method for determining grammaticality/acceptability is shaky at best. The ideal method for determining this is acceptability judgments. However, while reports of personal observation of acceptability are sufficient for the "low-hanging fruit" of language data, here is a case where the observational reports of various authors conflict; thus, there is no choice but to turn to quantitative methods.

Certainly even when carefully searching a specialized corpus like COCA, there are limits to what simple frequency data can tell. If the co-occurrence of two expressions is very infrequent, what does this tell us about grammaticality? Possibly nothing. For example, the corpus search discussed below turned up no hits for *as grateful*, but we would never want to take this as an argument for the ungrammaticality or unacceptability of *as grateful*. Clearly it is acceptable.

- (41) I know John didn't look that happy when you gave him that gift,

but believe me, he was as grateful as I was.

Similarly, a non-zero frequency does not necessarily mean that a given expression is perfectly acceptable. Besides speech errors, it is well known that people sometimes produce utterances which are not judged as perfectly acceptable, like certain island violations, or cases of coercion. However, it is true that a very high frequency of co-occurrence probably does tell us that a given expression is robustly grammatical. This would be especially important if intuition judgments are conflicting, because it might tell us that there is significant dialectal variation.

In this study, I examined relative frequencies of co-occurrence of degree modifiers with adjectives in the Corpus of Contemporary American English (Davies, 2008), and performed a cluster analysis on the resulting data. Below I first describe the data collection, and second the cluster analysis.

2.2.1.1 Data Collection

All data was collected using COCA, a 450 million word corpus spanning language data from 1990 to 2012, tagged for part of speech, as well as for medium (spoken, magazine, academic, etc.). The search was restricted to spoken contexts to prevent the possibility of variation between grammars of spoken and written or literary registers from muddling the data.

The study was also restricted to predicative adjectives since many degree modifiers are unacceptable in attributive position, thus likely skewing their attestation frequencies.

(42) #The tall enough guy came over last night.

This restriction was accomplished by only searching for adjectives or degree modifiers preceded by a form of the copula. A first search was conducted to determine the 150 most common adjectives in immediately post-copular position. Several of these were deemed to be variants of each other (e.g., *better* and *best*) and one (*not*) was deemed not be an adjective, so the final list comprised 142 adjectives.

A search was then conducted for copula–degree modifier–adjective sequences, restricting just to the list 142 common adjectives. No restriction was placed on the degree modifiers, so every expression deemed by the COCA tagging regime to be a degree modifier was included. The status of some of the expressions so tagged as degree modifiers was questionable and were thus

ignored. One additional degree modifier not tagged as such by COCA, *really*, was included manually. Number of hits per adjective for the following degree modifiers was recorded:

- (43) Degree Modifiers: *more/-er, very, so, as, most/-est, too, pretty, quite, enough, this, that, less, least, rather, really*

Separate searches were conducted to determine the frequencies of synthetic comparative and superlative forms; if an adjective had hits for both synthetic and paraphrastic forms of the comparative or superlative, these hits were combined to give a single value.

The number of hits for the positive and each degree modifier were summed for each adjective to produce a total frequency for that adjective. Then the number of hits for the positive and for each degree modifier were relativized to the overall frequency of the adjective. A sample of the table produced can be seen below.

Adjective	Total Hits	Positive	Comparative	<i>very</i> + A	<i>so</i> + A
<i>angry</i>	2403	0.7412	0.0112	0.1165	0.0712
<i>appropriate</i>	1400	0.8957	0.0271	0.0379	0.0043
<i>available</i>	2481	0.9879	0.0052	0.0016	0.0012
<i>aware</i>	2631	0.8955	0.0194	0.0582	0.0027
<i>bad</i>	5706	0.4939	0.2161	0.0471	0.0782

Table 2.1: Selection of corpus data.

This should be read as saying that *angry* appeared 2,403 times, with 74.12% of its occurrences being in the positive, 1.12% being in the comparative (either *more angry* or *angrier*), 11.65% being preceded by *very* and 7.12% being preceded by *so*.

Note that these counts exclude a significant portion of the data. Any data with a word intervening between the copula and the adjective which was not a degree modifier (according to the COCA transcription) was excluded, except *really*. A number of expressions which I consider to be degree modifiers were therefore excluded, including all maximizing, minimizing, and proportional (i.e., scale-type specific) degree modifiers as well as others, like *extremely* and *incredibly*. This also means that any cases of non-degree modifiers intervening were excluded, including negation or adverbs like *definitely*. Finally multiple

degree modifiers were also not captured by the search, such as *Nick is very very tall*.

2.2.1.2 Cluster Analyses

The idea here is that gradable adjectives should be good with degree modifiers and nongradable adjectives should not. As discussed above, though, it's not clear what a given frequency can tell us about acceptability. If an adjective appears with a certain degree modifier 1% of the time, does that tell us that that pairing is acceptable? Where is the cut-off?

By running a cluster analysis, the analyst need not pick a particular cut-off. Cluster analysis is a statistical method for determining natural groupings among a set of items based on their similarity across an arbitrary number of dimensions. In this case there are 142 items which have values between 1 and 0 along 16 dimensions (one for the positive and one for each of the 15 degree modifiers). This can be thought of as involving arranging the 142 adjectives in 16-dimensional space, with each value providing the distance along a particular axis, then “seeing” where the groupings lay.

2.2.1.2.1 Predictions First, if relative frequencies of co-occurrence of adjectives with degree modifiers is at all indicative of gradability, this analysis should provide two clusters, with one group generally having higher relative frequencies for all the degree modifiers, and lower relative frequencies for the positive, and with one group generally have lower relative frequencies for degree modifiers, and higher relative frequencies for the positive.

Second, if there are two clusters which are determined by the gradability distinction, I predict that there should be modal adjectives to be found in both clusters, including *possible* and *necessary* in the “non-gradable” cluster, if the intuitive acceptability judgments provided above are right.

Lassiter, on the other hand, implicitly predicts that while a division can be made among adjectives by their co-occurrence with degree modifiers, that all modal adjectives should cluster in together as uniformly gradable.

I used two different methods for clustering, which provided very similar results: K-means and hierarchical clustering. I explain each below.

2.2.1.2.2 Analysis 1: K-means The first method used was K-means. The K-means method involves picking n random points in the 16-dimensional space and dividing the items into groups according to which of the points

they are closest to. Once these groups are determined, the average point for each grouping is determined and the items are re-grouped according to which of *these* points they are closest to. This process is iterated until the groupings become stable.

Importantly, which points serve as the starting point can affect the final outcome – i.e., algorithm does not necessarily converge on the same clustering regardless of the starting point. Therefore it is important to run the K-means analysis repeatedly.

Also crucially, the analyst must pick the number of clusters ahead of time. The standard method for doing this is to plot the within group sum of squares by number of clusters extracted and finding the “bend” in the plot. This was done for the adjective/degree modifier data in the statistical analysis program R, producing the plot below.

There is some subjectivity in determining where the bend in the plot lies, but here it can at least be narrowed to 2, 3, or 4; i.e., the ideal number of clusters is 2, 3, or 4. Thus a K-means analysis was run in R 100 times for each of these values.

When dividing into 2 clusters, K-means produced the same clustering 100 out of 100 times. When dividing into 3 and 4 clusters, however, K-means did not always produce the same clustering. K-means divided the data into three clusters three different ways; one 61 times, one 38 times, and another one just once.

This is taken to show that two is the ideal number of clusters. This is a hopeful sign for the idea that this analysis would cluster adjectives according to the gradable/nongradable distinction.

The clusters produced are given in Tables 2 and 3. The first thing to note about the clustering is that it can be reduced just to frequency of the positive. In other words, the division can be summarized as follows: If an adjective appeared in the positive form less than 77% of the time, it was put in cluster A; if an adjective appeared in the positive more than 77% of the time, it was put in cluster B. This should not be too surprising since the positive frequency in a sense summarizes the rest of the data.

Before discussing the implications of these findings, I discuss the second cluster analysis method.

2.2.1.2.3 Analysis 2: Hierarchical Clustering The second method that was used was the agglomerative hierarchical clustering method. This

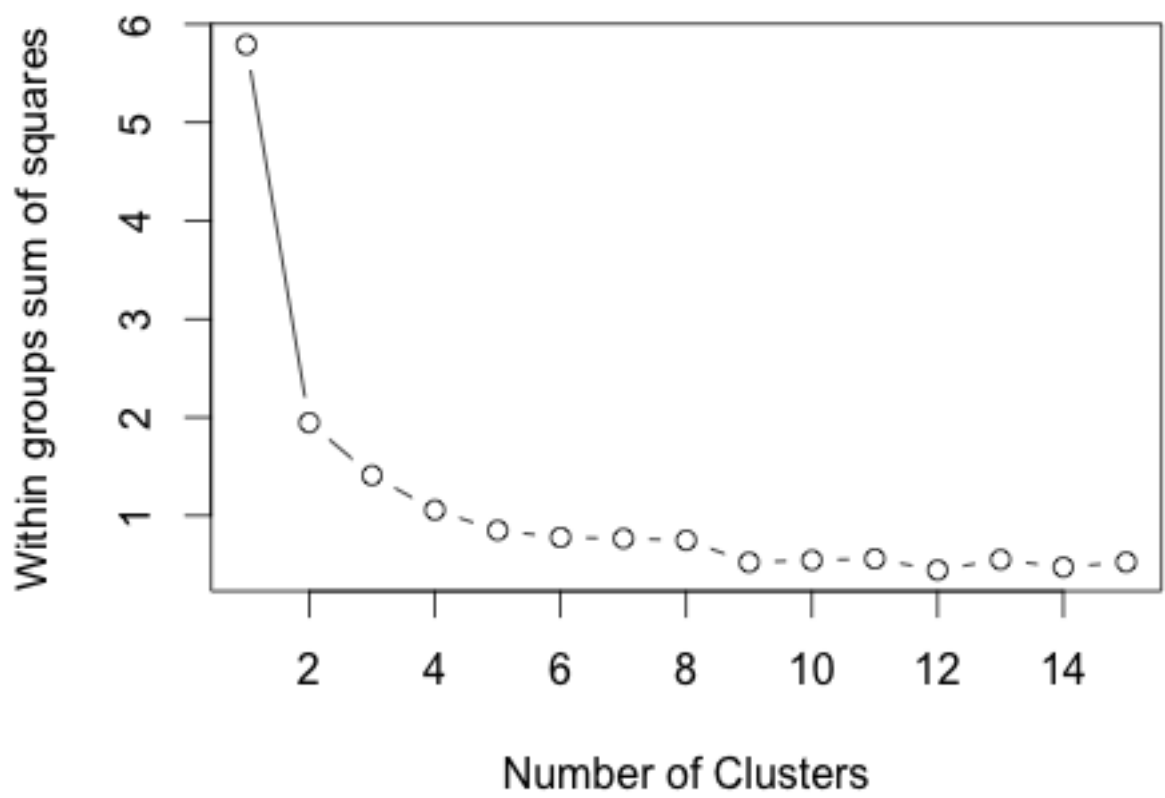


Figure 2.1: Within group sum of squares by number of clusters extracted

Adjective	Positive	Adjective	Positive
close	0.3653	worried	0.6325
simple	0.3735	grateful	0.6342
high	0.3794	lucky	0.6372
old	0.4181	clear	0.6422
small	0.4349	hard	0.6460
little	0.4442	hot	0.6500
difficult	0.4586	helpful	0.6570
strong	0.4606	happy	0.6623
big	0.4917	tough	0.6703
bad	0.4939	significant	0.6761
young	0.5255	serious	0.6831
smart	0.5282	nice	0.6854
easy	0.5367	successful	0.7121
good	0.5456	special	0.7125
important	0.5551	obvious	0.7185
busy	0.5686	disappointed	0.7220
cool	0.5736	healthy	0.7226
sad	0.5789	funny	0.7276
well	0.5812	confident	0.7287
pleased	0.5827	interesting	0.7306
effective	0.5951	positive	0.7356
proud	0.6113	critical	0.7398
comfortable	0.6200	angry	0.7412
dangerous	0.6208	familiar	0.7447
careful	0.6244	interested	0.7615
likely	0.6275	different	0.7644
nervous	0.6281		

Table 2.2: Cluster A

Adjective	Positive
beautiful	0.7884
concerned	0.7923
unlikely	0.7925
accurate	0.7962
safe	0.8047
fascinating	0.8061
shocked	0.8112
famous	0.8146
scared	0.8182
surprised	0.8188
tired	0.8260
only	0.8289
glad	0.8309
open	0.8317
fun	0.8325
mad	0.8387
sick	0.8459
unfair	0.8476
consistent	0.8522
just	0.8541
amazing	0.8590
great	0.8598
honest	0.8747
welcome	0.8770
curious	0.8793
normal	0.8865
incredible	0.8894
american	0.8950
wonderful	0.8951
aware	0.8955

Adjective	Positive
appropriate	0.8957
used	0.9005
political	0.9027
full	0.9043
like	0.9046
crazy	0.9102
terrible	0.9103
huge	0.9122
convinced	0.9145
new	0.9168
capable	0.9178
possible	0.9193
afraid	0.9251
certain	0.9257
real	0.9275
sorry	0.9277
human	0.9289
fair	0.9299
fantastic	0.9312
ridiculous	0.9338
perfect	0.9361
unbelievable	0.9399
terrific	0.9417
sure	0.9442
white	0.9465
still	0.9541
responsible	0.9566
true	0.9568
involved	0.9573
fresh	0.9612

Adjective	Positive
black	0.9623
willing	0.9624
free	0.9633
innocent	0.9654
necessary	0.9678
present	0.9687
wrong	0.9693
ready	0.9732
guilty	0.9751
alone	0.9766
impossible	0.9789
fine	0.9822
correct	0.9833
available	0.9879
alive	0.9882
right	0.9888
dead	0.9901
illegal	0.9910
gay	0.9912
due	0.9922
legal	0.9935
other	0.9942
unable	0.9942
okay	0.9962
pregnant	0.9962
able	0.9973
live	0.9975
married	0.9976
supposed	0.9988

Table 2.3: Cluster B

involves computing the differences between every pair of points (i.e., adjectives) and then clustering them into groups of two based on similarity, i.e., creating a tree from the bottom up. This does not require specifying the desired number of clusters ahead of time, rather it requires interpreting the graph to determine the ideal number of clusters.

This analysis was done using the Ward method in R, which produced the dendrogram in Figure 2; two clusters are indicated in red.

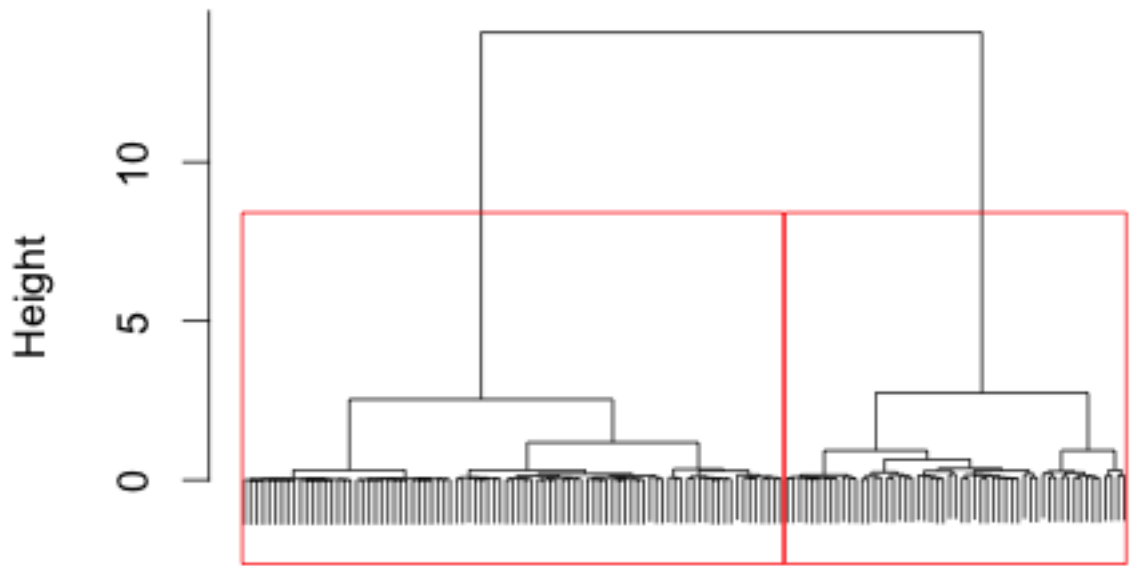


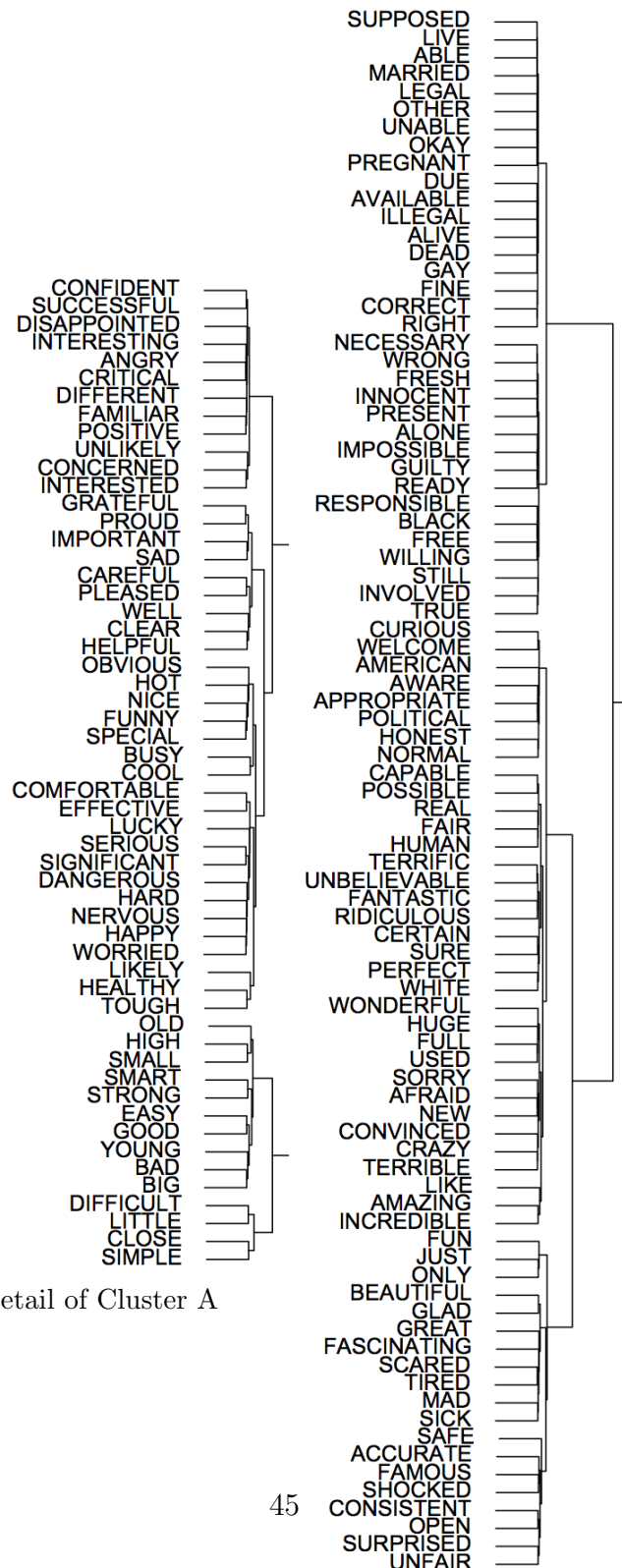
Figure 2.2: Agglomerative Hierarchical Cluster Dendrogram, with two clusters indicated

These clusters can be seen in detail below.

2.2.1.3 Does Clustering Derive the Gradability Distinction?

The first thing to note about these clusterings is that they are almost identical. The two methods only disagree on two adjectives; *unlikely* and *concerned* were put into Cluster A by the hierarchical analysis, where they were in Cluster B in the K-means analysis.

Thus the hierarchical method largely confirms the findings of the K-means



(a) Detail of Cluster A

(b) Detail of Cluster B

method. However, it also provides additional information which is not provided by a simple execution of the K-means analysis for two clusters, in that it groups the adjectives all the way down, providing for potentially interesting sub-clusters. For example, one subcluster of Cluster B contains the following adjectives, including every extreme adjective (Morzycki, 2012) included in the study except for *fascinating*:

- (44) *capable, possible, real, fair, human, **terrific, unbelievable, fantastic, ridiculous**, certain, sure, perfect, white, **wonderful, huge, full, used, sorry, afraid, new, convinced, crazy, terrible, like, amazing, incredible***

The apparent naturalness of these clusters is a good sign that the analysis is making worthwhile distinctions.

The two clusters that are produced by each of these analyses clearly correlate with the gradable/nongradable distinction. Cluster A, which had generally low occurrence in the positive form and high occurrence with degree modifiers, consists entirely of adjectives which are perfectly acceptable with a wide range of degree modifiers.

Cluster B is more of a mixed bag. It contains adjectives which seem to be acceptable with degree modifiers, like *beautiful, unlikely, mad, and sick*, including some with very high frequencies of positive attestation, like *fresh*, which appeared without a degree modifier 96.12% of the time. Cluster B also contained adjectives which are unacceptable with degree modifiers, like *dead, legal, and pregnant*. An important third category also appeared within Cluster B, of homophonous pairs of gradable and nongradable adjectives. Since the search was conducted just on the basis of spelling and syntactic category, ambiguities could not be controlled for. For example, *certain* has both gradable and nongradable uses, which are similar in meaning to *sure* and *particular*, respectively.

- (45) a. I'm (very) certain Obama will be re-elected.
b. She has a (# very) certain way about her.

Given the drawbacks of the corpus approach, this division is what we should expect. As discussed above, a low frequency of attestation of a particular expression doesn't tell us much – it could be that the expression is ungrammatical, or it could be that it simply isn't used very often. On the other hand, a very high degree of attestation strongly suggests that the ex-

pression is grammatical.

Thus we have two groups: the expressions which occur very frequently with degree modifiers, which we can thus be sure are gradable just from frequency of attestation, and the expressions which occur less frequently with degree modifiers, which we therefore cannot be sure about, purely on the basis of frequency of attestation.

Returning to the subcluster in (45), note also that extreme adjectives could sensibly be explained as being a part of Cluster B since they do not combine as readily with certain degree modifiers that most other gradable adjectives do combine with, like *more* and *very*, although they are compatible with many others (*pretty*, *so*, *too*). Extreme adjectives also combine with extreme degree modifiers (Morzycki, 2012), which were not included in the study, thus inflating the relative proportion of positive occurrences for these adjectives. The same can be said for maximum standard adjectives, several of which also appear in this subcluster (*full*, *certain*, *sure*), which can appear with maximizing degree modifiers like *completely*, which were not included.

2.2.1.4 (Non)gradable Modal Adjectives

Lassiter’s claim that *more possible* is robustly attested could not be reproduced. Out of 3,793 hits for *possible*, *more possible* only occurred 0.05% of the time, i.e., twice.

positive	comparative	<i>very</i>	<i>so</i>	<i>as</i>	superlative	<i>too</i>	<i>pretty</i>
0.9193	0.0005	0.0380	0.0000	0.0011	0.0000	0.0000	0.0000
<i>quite</i>	<i>enough</i>	<i>this</i>	<i>that</i>	<i>less</i>	<i>least</i>	<i>rather</i>	<i>really</i>
0.0369	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0042

Table 2.4: Relative frequencies of *possible* in COCA (Davies, 2008)

However, the clustering analysis cannot be taken to show definitively that *possible* is not gradable; in general, it appears that this method can only provide positive evidence for gradability, but not for nongradability.

Even looking closer at the three largest subclusters within Cluster B does not clarify the picture. One such subcluster contains the most clearly nongradable adjectives, including *pregnant* and *dead*. One contains a number of adjectives which appear robustly gradable, like *accurate* and *sick*. The

middle subcluster, however, is the one that contains the subcluster in (44), including *possible*, as well as *curious*, *welcome*, *American*, *aware*, *appropriate*, *political*, *honest*, and *normal*. In other words *possible* is grouped right in the middle of the subcluster which most clearly contains both gradable and nongradable adjectives.

However, while the cluster assignment itself may not tell us whether *possible* is gradable, a closer look at the individual frequencies for *possible* may. Notably, for most degree modifiers, *possible* had no occurrences; almost all of its occurrences of degree modification were concentrated in *quite* and *very*. The placement of *possible* into the intermediate subcluster of Cluster B can be attributed entirely to these two degree modifiers.

Indeed, relative frequency of *possible* with *quite* is higher than all but four of the 142 adjectives studied. Moreover, *quite possible* is clearly grammatical. The status of *very possible* is more mixed, but it is clearly better than any other degree modifier besides *quite*, though the relative frequency of *possible* with *very* is still much closer to the average for Cluster B (1.96%) than for Cluster A (14.87%).

- (46) a. It's quite possible that Obama will nominate Susan Rice.
 b. ?It's very possible that Obama will nominate Susan Rice.

While the co-occurrence of *possible* with *very* may be analyzed as coercion, the same cannot be said for *quite*. There could be a few explanations for this, however. One is that *quite possible* is simply entirely idiosyncratic and a compositional account need not be provided. Another is that *quite* has a meaning which is like that of an intensifier but not dependent upon a degree-based analysis, so that it may combine with nongradable elements. I pursue this analysis in detail in Chapter 5.

So a strong case can be made for *possible* as nongradable on the basis of these corpus findings. But given the caveat above, that low frequencies are not informative, this still cannot be considered conclusive. Ultimately acceptability judgments are the more important evidence, and are pursued in the next subsection.

However, even if the question as it relates to *possible* is not settled definitively by the corpus study, the larger question seems to be. Namely, are there nongradable modal adjectives? Considerable doubt is raised by the hierarchical cluster analysis's placement of several modal adjectives, including *possible*'s antonym *impossible*, as well as *necessary*, *legal*, *illegal*, *able*, and

unable, which are categorized into the ‘least gradable’ subcluster of Cluster B, with other adjectives which had very very few attestations with degree modifiers. Even if *possible* can be shown to be gradable, the burden of proof is on defenders of Lassiter’s theory to show that all these adjectives are gradable as well.

2.2.2 Acceptability Judgment Task

In order to further clarify the status of the acceptability of *possible* with degree modification, an acceptability judgment task was conducted using Amazon Mechanical Turk, which has been shown to be a valid experimental method with results comparable to laboratory experiments (Sprouse, 2011). If Lassiter is right that *possible* is gradable, its acceptability with a variety of degree modifiers should be comparable to that of *likely*, an uncontroversially gradable modal adjective.

2.2.2.1 Participants

Participants were 30 self-reported native speakers of English, using the Amazon Mechanical Turk (AMT) online marketplace. Participants were paid \$1.20 for their participation. Participants were filtered by AMT to only IP addresses in the United States, and were asked to identify their native language; all participants were native speakers of English.

2.2.2.2 Materials

Participants were presented with 54 sentences, each containing an adjective. There were two conditions (“positive” and “degree modified”) for each item, for a total of 108 sentences. For most items, the item varied between conditions only in whether there was a degree modifier.

- (47) a. So it’s possible that this whole contest could be over on the 31st.
(positive)
- b. So it’s very possible that this whole contest could be over on the 31st.
(degree modified)
- (48) a. It’s likely that we will know soon whether Stevens is retiring.
(positive)

- b. It's very likely that we will know soon whether Stevens is retiring. (degree modified)

For items which contained a degree modifier which itself subcategorized for a clause, this clause was kept in the positive condition, with whatever other alterations necessary to keep the sentence grammatical/felicitous, and to keep sentence length comparable between conditions.

- (49) a. That job is demanding; you shouldn't be working there. (positive)
- b. That job is too demanding for you to be working there. (degree modified)
- (50) a. It's important that we have enough food, and that the apartment be pristine. (positive)
- b. It's more important that we have enough food than that the apartment be pristine. (degree modified)

Four degree modifiers were tested: the comparative (*more* or *-er*), *very*, *too*, and *pretty*. For many cases of *more* and *too*, their respective complements (indicated below in strikethrough) were left implicit. Items also frequently contained pronouns, discourse connectives, and other expressions that required participants to reconstruct the context.

For each degree modifier, there were four items with *possible* and four with *likely* in each condition. Additionally there were six other fillers for each degree modifier in each condition (three adjectives considered gradable and three considered non-gradable). A full list of items is given in the appendix.

2.2.2.3 Task

Participants were asked to rate the naturalness of sentences on a 1 to 7 scale. They were given the following instructions:

- (51) "Each page will consist of a sentence, whose naturalness you will need to judge on a 1 to 7 scale (where 1 is completely unnatural, and 7 is completely natural). We are looking for your judgments about how natural these sentences could sound in a conversation. **Feel free to imagine that the sentence is a part of a larger conversation in order to help it make sense.**" (bold original)

Subjects saw a mix of items from the positive condition and the degree modified condition, for a total of 54 items. The items were distributed randomly between two lists so that subjects only ever saw one condition for any given item. The lists were constructed using software by [Gibson et al. \(2011\)](#).

2.2.2.4 Results

Degree modification of *possible* was found to have a significant negative effect on acceptability, beyond what would be expected if *possible* were gradable.

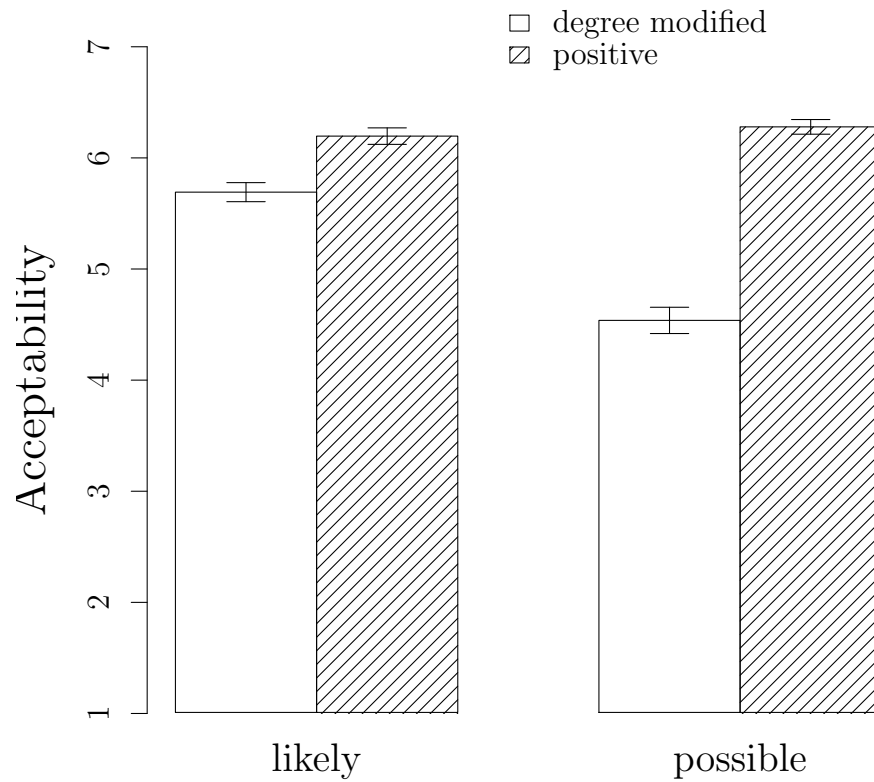


Figure 2.4: Acceptability of *possible* and *likely* with and without degree modifiers

The average acceptability for all the data, excluding *possible* and excluding cases of nongradable adjectives with degree modifiers (which were expected to be ungrammatical) is 6.16 out of 7. This represents a baseline for acceptability. The average acceptability for *possible* without degree modifiers was 6.28, while for *likely* it was 6.20. Average acceptability for *possible* with degree modifiers was 4.54, while for *likely* it was 5.69.

A t-test showed that acceptability of *possible* was significantly less with degree modifiers than without ($p < .01$), with an effect of 1.48–2.01 points difference (95% CI). Note that while 4.54 is still above the midpoint of 4, there was a strong bias toward the top half of the scale in the study. Even the worst sentences had scores higher than 3.

- (52) a. I’m more able to solve this problem than she is. (3.20, n=15)
 b. I’m not going to try; it’s too illegal. (3.40, n=15)

The positive versions of (52a-b) were rated 6.53 (n=15) and 6.47 (n=15) on average, respectively.

A t-test showed that *likely* also had significantly worse acceptability ratings with degree modifiers ($p < .01$), though the effect was much smaller, at 0.28–0.73 (95% CI).

Breaking down degree modification by degree modifier also shows interesting contrasts. In 2.5 below, average acceptability judgments for each degree modifier paired with either *likely* or *possible* are paired with acceptability judgments for their respective positive conditions. In other words, the values listed as “positive” in each degree modifier column are for the positive versions of the sentences containing that degree modifier.

The acceptability of *very possible* was 5.57, which is higher than some ratings for sentences which were expected to be entirely acceptable, like *too likely* (5.23)⁶, having an average effect on acceptability of just -0.60. A t-test showed that this difference was significant, however ($p = 0.009$), with the effect ranging from 0.15–1.05 (95% CI). And while *very* did also produce a slight negative effect on *likely*, a t-test showed that this was not significant ($p = .54$).

⁶Note that across all gradable adjectives, *too* had the biggest negative effect on acceptability, at -0.44. This could be because *too*, which itself has a modal semantics, requires significant contextual enhancement to determine its modal standard.

	comparative	<i>very</i>	<i>too</i>	<i>pretty</i>
<i>likely</i>				
–positive	6.05	5.95	6.12	6.67
–degree modified	5.77	5.82	5.23	5.95
–difference	-0.38	-0.13	-0.88	-0.72
<i>possible</i>				
–positive	6.30	6.17	6.47	6.18
–degree modified	4.48	5.57	3.78	4.32
–difference	-1.82	-0.60	-2.68	-1.87

Table 2.5: Acceptability judgments for *likely* and *possible* by degree modifier.

2.2.2.5 Discussion

The results here support the hypothesis that *possible* is not a gradable adjective. While degree modification did produce a significant negative effect on acceptability for *likely*, the effect was much smaller than the effect on *possible*.

The fact that *very* shows a smaller, but still significant, effect, is not too surprising given the findings of the previous study which showed that attestation frequency of *possible* was much higher than for any other degree modifier (except *quite*). This could indicate that *possible* can be coerced into having a gradable meaning by the presence of *very*, but that coercion results in slightly degraded acceptability judgments.

What is slightly surprising about this is that while we should expect adjectives to vary in terms of how easily they are coerced (say, *pregnant* is easier than *prime*) it's less obvious why different degree modifiers should vary in terms of how easily they give rise to coercion.

I argue that the relatively high coerceability of *possible* with *very* is due to analogy with *quite*. As I argue in Chapter 5, *quite* has a semantics which allows it to have a intensifier-like meaning, but without appealing to a degree semantics. This means that when combined with a gradable adjective its meaning aligns with that of *very*, but this meaning can be extended to non-gradable adjectives.

Non-gradable adjectives which can semantically combine with *quite* to have an intensifier-like meaning prime speakers to more easily coerce *very* into having a *quite*-like meaning when it combines with said adjectives. Since

possible readily combines with *quite*, *very* is more easily coerced when they combine. This predicts that, among non-gradable adjectives, acceptability or attestation with *quite* should correlate with acceptability or attestation with *possible*.

The latter prediction can be tested with the data presented here. Looking only at adjectives in the ‘less gradable’ cluster identified by the k-means clustering discussed in the previous section, a Pearson’s product-moment correlation test showed a correlation between frequency of attestation of *very* and *quite* ($r=0.397, p<0.001$). The same test showed no correlations between the comparative or *too* and *quite* or *very*.

Moreover, there was no correlation between frequency of co-occurrence with *quite* and with *very* among adjectives in the ‘gradable’ cluster ($r=0.009, p=0.947$). This shows that among gradable adjectives, frequency of co-occurrence with one of these intensifiers does not predict frequency of co-occurrence with the other. This lays the groundwork for a satisfying explanation for why, of all the degree modifiers sampled in the corpus study, *possible* only has non-trivial co-occurrence with *very* and *quite*. The finer details of the semantics of *quite* are left for Chapter 5.

This in turn allows for an explanation of *possible* as a non-gradable adjective. While some gradable adjectives may show low- or zero-frequency attestation or slight degradation of acceptability judgments with certain degree modifiers, widespread zero-frequency attestation and significant degradation of acceptability judgments indicates nongradability. A few cases of acceptability or high-frequency attestation are easier to explain as idiosyncrasies than as following from gradability.

Moreover, the status of *possible* itself is not all that crucial if it can be shown that other modal adjectives are non-gradable; and indeed, the corpus study suggests this. Adjectives like *able*, *legal*, and *illegal* almost never appear with degree modifiers in the corpus study and are grouped with the ‘least gradable’ subcluster.

2.3 Scale Structure of (Gradable) Modal Adjectives

Having established the distinction between gradable and non-gradable modals, I turn now to gradable modals. While it has been shown that *possible* (among

many other modal adjectives) is not gradable, it can still be argued that all modals share a common scalar core, which may or may not express as gradability. Thus we could imagine the following semantics for *possible*; this is along the lines of what Lassiter (2011) proposes for *might*.

$$(53) \quad \llbracket \text{possible} \rrbracket = \lambda p[\lambda w[Pr(p)(w) > 0]]$$

This expression is not gradable in my terminology because it is not of type $\langle \alpha, \langle s, d \rangle \rangle$, but it is still ‘de-scalar’ because it can be derived from a simple measure function. In this section I argue against this view by examining the scale structure of two gradable modals, *likely* and *certain*. I argue, contrary to Lassiter (2011), that the probability scale, in its typical mathematical formulation, does not underlie the meaning of either of these expressions, undermining the view that probability is central to modal expressions, and that they should therefore be modeled in terms of probabilistic measure functions.

2.3.1 Gradable Modals

Lassiter (2011) argues that *likely* is associated with the probability scale, a fully closed scale.⁷ As Lassiter would predict, *likely* is bad with minimizing modifiers; it is not a lower-closed, upper-open scale.

- (54) *likely* with minimizing modifiers
- a. #It is slightly likely that Herman Cain will lose.
 - b. #It is a little likely that Herman Cain will lose.

However, consider *likely* with maximizing and proportional degree modifiers, which are diagnostic of upper- and fully-closed scales, respectively.

- (55) *likely* with maximizing degree modifiers
- a. #It is totally likely that Herman Cain will lose.
 - b. #It is completely likely that Herman Cain will lose.

⁷As far as I can tell, *probable* and *likely* are totally synonymous; the former is just a higher-register/less colloquial expression. I won’t discuss *probable* any further. The adverb *probably* does not appear to be gradable at all.

- (i) #John more probably left than Bill.

- c. #It is perfectly likely that Herman Cain will lose.
 - d. #It is absolutely likely that Herman Cain will lose.
- (56) *likely* with proportional modifiers
- a. ?It is {10/50/100}% likely that Herman Cain will lose.
 - b. #It is {three quarters/half/one third} likely that Herman Cain will lose.
 - c. #It is {all/most/half}-(of)-(the)-way likely that Herman Cain will lose.
 - d. #It is mostly likely that Herman Cain will lose.

First, *likely* is bad with maximizing degree modifiers, contrary to what is expected if it is a fully closed scale adjective. Lassiter (2010) actually points this out as well, but argues that maximizing degree modifiers do not diagnose scales with maxima (more on this claim below).

Furthermore, *likely* is also clearly bad with most proportional modifiers. One potential exception is *n%*, which Lassiter argues is acceptable with *likely* and is in fact crucially diagnostic of a fully-closed scale. Intuition judgments vary, but there are indeed attestations of *n% likely*.⁸ See below for discussion of this potential exception.

Lassiter (2010), in attempting to explain the badness of (55) and (56b-d), argues that degree modifiers select not only for scale type but for positive meaning – thus *n%* is the only true diagnostic of closed scales because it selects only for a closed scale adjective; the other proportional degree modifiers, and apparently all the maximizing ones, select for fully-closed-scale adjectives *whose positive meanings are maximal*. Since *likely* idiosyncratically

⁸A corpus search could not conclusively decide this matter. A search for “percent likely” yielded 2 hits, out of 8,402 total hits for *likely*; however, a search for “percent full” yielded only 12 hits out of 11,760 total hits for *full*. This means that *percent full* is about 5 times more common than *percent likely*; however, a statistical analysis has not been run on these terms and the overall numbers are quite small, so gauging significance is difficult.

In order to increase the size of the compared values, a search was conducted which included written sources as well; on this search 74 of 82,541 instances of *full* were preceded by *percent*, whereas only 3 of 68,260 instances of *likely* were. This could be a product of register distinctions, however, which is why written sources were excluded in all other searches. These searches, in addition to the mixed results for acceptability judgements elicited, leave me skeptical about the grammaticality of *n% likely*. However, in the absence of conclusive evidence against it, I will continue on the assumption that *n% likely* may be acceptable, at least in some grammars, and must be explained. If it can be shown that *n% likely* is not acceptable, this would only simplify my case.

has a relative positive meaning (all positive meanings are determined idiosyncratically on Lassiter’s take, which follows from [Kennedy and McNally \(2005\)](#)) it cannot combine with any other maximizing or degree modifiers outside of *n%*.

This raises a lot of questions, though. First, why is *n%* the only such modifier? It would strengthen Lassiter’s claim if even one other such modifier could be found. Moreover, *likely* appears to be the only adjective with a fully-closed scale but which has a relative positive meaning. Moreover, is there an external motivation for claiming that *n%*, but no other proportional modifiers, have this property? Below I explore an alternative explanation for *n%*, while concluding that these diagnostics affirm that *likely* has an open scale. I will refer to the measure function associated with *likely* as *lhood*, and provide a semantics for this measure function in the next chapter.

2.3.1.1 *certain*

[Lassiter \(2010\)](#) argues that *certain* is a gradable modal adjective, again relying on a probability scale. Lassiter in fact argues that *certain*, *likely*, and *possible* all share a common scale. While it has been shown that *possible* is not gradable and that *likely* does not have a probability scale in the strict sense, Lassiter’s argument for the coscalarity of *certain* and *likely* still stands, namely, entailment relations:

- (57) a. It’s certain that Obama will nominate Jack Lew. → It’s likely that Obama will nominate Jack Lew.
 b. It’s not likely that Obama will mint the platinum coin. → It’s not certain that Obama will mint the platinum coin.

It has been shown that if *certain* were a nongradable modal, these entailments could be derived rather easily. However, *certain* is clearly (and uncontroversially) gradable.

- (58) *certain* with degree modification
 a. It is more certain that the ball is in his left hand.
 b. It is very certain that the ball is in his left hand.
 c. ...

Given this, Lassiter simply argues that *certain* and *likely* involve the same scale, with their positive forms picking out different points on those scales.

Lassiter argues that while *likely* picks out a contextually determined midpoint in its positive form, *certain* picks out a maximum in its positive form. This seems right.

(59) #It is certain that the Jets will win, but it could be more certain.

But consider the following.

- (60) a. Obama is more certain to nominate Hagel (for Defense) than Brenner (for the CIA).
b. Obama is more likely to nominate Hagel (for Defense) than Brenner (for the CIA).

These are not synonymous. If *likely* and *certain* denoted the same measure function, they should be synonymous when paired with the same degree modifier. Clearly the difference cannot just be a matter of what point on the scale is picked out. Consider also (61).

- (61) a. Obama's reelection couldn't be less certain \rightarrow Obama's defeat couldn't be less certain
b. Palin's election couldn't be less likely \rightarrow Palin's defeat couldn't be more likely

If *couldn't be less* is a kind of minimizing modifier⁹ which picks out a very low point on a scale, this suggests that there are differing entailments determined by the low points on the scale. If *couldn't be less* was represented as \downarrow , a formal representation of the two entailments above might be (62).

- (62) a. $\downarrow (\text{certain})(p) \rightarrow \downarrow (\text{certain})(\neg p)$
b. $\downarrow (\text{likely})(p) \nrightarrow \downarrow (\text{likely})(p)$

I propose that *certain* employs a distinct scale, which I will call *cert*. While *cert* is upper bounded, it must be lower open, given its incompatibility

⁹Lassiter (2011) rightly points out that *couldn't be less* is not a perfect device for getting at the minimum value on a scale, based on data like the following:

- (i) He couldn't be less friendly.

It is not obvious that *friendly* has a lower bound. And in fact, it has already been argued that *likely* does not have a lower bound. However, the intuition that (61) seems to get at is sound, namely that *certainty* and *likelihood* do not perfectly correlate.

with proportional modifiers.

- (63) *certain* with proportional modifiers
- a. #Herman Cain is {10/50/100}% *certain* to lose.
 - b. #Herman Cain is {three quarters/half/one third} *certain* to lose.
 - c. #Herman Cain is {all/most/half}-(of)-(the)-way *certain* to lose.
 - d. #Herman Cain is mostly *certain* to lose.

The judgment in (63a) is contrary to Lassiter’s claim (but see further discussion of $n\%$ below). However, it is worth pointing out here that Lassiter considers *certain* with an expletive subject and a finite complement, which I find degraded, rather than in the raising frame I have shown. However, a limited corpus search returns only nine instances of *certain* in such a frame, none of them with a degree modifier. There are, however, many attestations of *certain* taking a genuine subject and a finite complement (as in (64c)), a handful of them with degree modifiers.

- (64)
- a. Herman Cain is *certain* to lose.
 - b. %It is *certain* that Herman Cain will lose.
 - c. I am *certain* that Herman Cain will lose.

As with *likely*, I assume no semantic difference between the raising and expletive-subject versions of *certain* (64a-b), but this is an assumption since I do not have judgments about the latter.

The question then, is, how can *cert* relate to *hood*? I explore this question in the next chapter. Lassiter (2011), conceding that *certain* may take a different scale from *likely* in response to the claims made in an earlier version of this chapter (Klecha, 2012) provides one possible strategy, which is that the measure function that *certain* denotes is itself defined in terms of the measure function that *likely* denotes; in Lassiter’s case the information-theoretic notion of ‘entropy’ and probability, respectively.

2.3.1.2 Scale Structure and Positivity

In addition to claiming that the entailment properties of the various modal adjectives motivate an entirely scalar conception of modality, Lassiter argues that the triplet *possible/likely/certain* constitute a counterexample to Kennedy’s (2007) generalization about the relationship between positivity and scale structure.

Given that positive meanings vary between relative, minimal, and maximal adjectives, the default hypothesis would be to suggest that there are three *pos* morphemes, each syntactically selecting for a different set of adjectives, and each yielding a different meaning; maximal, minimal, and relative, respectively. Kennedy (2007) notes that this misses a crucial generalization, namely that scale structure correlates with positive meaning; relative adjectives are open-scale adjectives, maximal adjectives are upper-closed adjectives, and minimal adjectives are lower-closed adjectives. On this basis, Kennedy argues for a unified *pos* which determines the standard of comparison with the potentially context-sensitive function *s*.

$$(65) \quad \llbracket \text{pos} \rrbracket = \lambda g_{\langle e,d \rangle} \lambda x [s(g)(x)]$$

According to Kennedy, *s* is a function which takes a measure function and returns a degree on the scale associated with the measure function which “stands out” relative to the other degrees on the scale. The *s* function is underdefined in the way that any other context-sensitive expression might be. However, in conjunction with this proposal, he adopts the pragmatic principle of Interpretive Economy, which guides how this expression’s meaning is fixed.

Interpretive Economy says this: satisfy contextual variables in a way that is context-insensitively determinable from the conventional meaning of the relevant expression if possible; only if this is not possible should the variable be filled in a truly context-sensitive way.

Thus the meaning of “stand-out” is not always truly contextually determined. In the case of adjectives whose scales have upper or lower bounds, there is a unique function which returns a degree that stands out; in the case of upper-bounded scales, this function is the property of having the maximal degree; in the case of lower-bounded scales, this function is the property of exceeding the minimum degree. Only for scales with no bounds is there no such conventionally determinable stands-out-function; thus speakers must turn to context.

$$(66) \quad s = \begin{array}{ll} \text{a.} & \lambda g \lambda x [g(x) = \max(g)] & (\text{maximal adjectives}) \\ \text{b.} & \lambda g \lambda x [g(x) > \min(g)] & (\text{minimal adjectives}) \\ \text{c.} & \lambda g \lambda x [g(x) \geq C(g)] & (\text{relative adjectives}) \end{array}$$

Lasiter’s claims about *possible*, *likely*, and *certain* challenge this. But by showing that *possible* is in fact not gradable, and that *certain* and *likely* are

indeed not associated with the same scale, this particular counterexample to Kennedy (2007) is defused.

By the diagnostics given, *likely* takes an open-scale; given that it is a relative adjective, this too has the benefit of according with Kennedy (2007). However, there is the apparent drawback that the proposed semantics for *likely* seems to not relate to the intuitive, albeit mathematically sophisticated notion of probability which involves a closed scale; indeed, Lassiter seems to take this intuitive notion of probability as being the primary motivator for pursuing a closed-scale analysis of *likely*.

But ultimately the “intuitive scale” associated with an adjective does not always align with its lexical scale, nor should we expect it to. An adjective may have as its basic scale something fairly intuitive, but then build arbitrary (or non-arbitrary) presuppositions or constraints which alter the scale (structure). Or an adjective’s meaning may simply be unintuitive.

Particularly, consider *short* and *inexpensive*, which Lassiter (2010) cites as further counterexamples to Kennedy (2007). Lassiter argues that *short*’s scale has a maximum value, since there is obviously (intuitively) a maximum to shortness (likewise a minimum for tallness; in other words there is a minimum height, namely 0). Since these terms are relative, not minimal and maximal respectively, these are counterexamples to Kennedy (2007). Likewise, *inexpensive* has an obvious maximum, namely a cost of 0, but is not a maximum adjective.

However, just because these intuitive scales are closed, we should not conclude that the lexical scales are. First of all, maximizing modifiers do not work with *short* or *inexpensive*.

- (67) *short* with maximizing degree modifiers
 - a. #The boy is **totally short**.
 - b. #The boy is **completely short**.
 - c. #The boy is **perfectly short**.
 - d. #The boy is **absolutely short**.
- (68) *inexpensive* with maximizing degree modifiers
 - a. #The car is **totally inexpensive**.
 - b. #The car is **completely inexpensive**.
 - c. #The car is **perfectly inexpensive**.
 - d. #The car is **absolutely inexpensive**.

Certainly these cannot mean “the boy has a height of 0” or “the car is free”, respectively. This is troubling for an account which says these adjectives’ scales have maxima.

Second, elements that occupy the apparent endpoints of these scales are not obviously admissible with these expressions. In the case of *short*, it is very clear that elements with 0 height are not ordered; after all, these items do not take up space.

- (69) #That blade of grass is very short, but it is taller than dignity (since dignity does not have physical extent).

While *short* may be upper-bounded, it does not have a maximum element, which is the crucial factor.¹⁰ This requires a qualification of Kennedy and McNally’s (2005) typology of scales; while scales may be upper-closed or upper-open, for example, upper-open scales may correspond to scales which an upper bound but no upper maximum, or scales which truly have no upper bound at all.

As far degree modification is concerned, this does not affect the basic four-way split that Kennedy and McNally (2005) proposed, since there are no expressions which distinguish between open and bounded endpoints. Maximizing modifiers like *completely* crucially pick out the maximum degree on a scale; thus if a scale is not upper-closed, even if it is upper-bounded, it will be incompatible with maximizing modifiers. Incompatibility with maximizing modifiers is thus a diagnostic for lacking a maximum element, but does not distinguish between upper-bounded and upper-open scales.

- (70) a. FULLY CLOSED SCALE: $[0,1]$
b. $\left\{ \begin{array}{l} \text{UPPER-CLOSED/LOWER-BOUNDED SCALE: } (0, 1] \\ \text{UPPER-CLOSED/LOWER-OPEN SCALE: } (-\infty, 1] \end{array} \right\}$
c. $\left\{ \begin{array}{l} \text{UPPER-BOUNDED/LOWER-CLOSED SCALE: } [0, 1) \\ \text{UPPER-OPEN/LOWER-CLOSED SCALE: } [0, \infty) \end{array} \right\}$
d. $\left\{ \begin{array}{l} \text{FULLY BOUNDED SCALE: } (0, 1) \\ \text{UPPER-BOUNDED/LOWER-OPEN SCALE: } (-\infty, 1) \\ \text{UPPER-OPEN/LOWER-BOUNDED SCALE: } (0, \infty) \\ \text{FULLY OPEN SCALE: } (-\infty, \infty) \end{array} \right\}$

However, higher-order entailment patterns judgments may be able to dis-

¹⁰Similar things can be said about any gradable adjective relating to spatial extent: *long*, *wide*/*narrow*, *big*/*small*, etc.

tinguish among these conceptual scale types. For example, while *tall* is upper-open, *short* is upper-bounded, and this is shown by the ability to move up the scale in arbitrary increments.

- (71) a. Itamar is 6 feet tall and that sign is 10 feet taller than him.
 b. #Itamar is 6 feet tall and that sign is 10 feet shorter than him.

Clearly, *short* is upper-bound but not upper-closed.

As for *inexpensive*, compared to (70), it doesn't seem quite as bad to use it to refer to something which has no cost.¹¹

- (72) ?Breathing air is inexpensive; in fact, it's free.

One could argue that the oddness of using *inexpensive* to describe something free may be a scalar implicature, due to competition with the lexical item *free*.¹² However, this wouldn't be such a bad turn; since Kennedy's (2007) Interpretive Economy is a pragmatic constraint, it seems reasonable that it is sensitive to pragmatic considerations – i.e., positive readings are determined by lexical scale type *modulo scalar implicature*.

One problem with pursuing this line of analysis is that if *inexpensive* is associated with a legitimately closed scale, i.e., $(-\infty, 1]$, with the maximal positive reading ruled out by scalar implicature, we would still expect maximizing degree modifiers to work. One way to rescue this would be to propose that these scalar implicatures are strong enough to rule out even cases where the conventional meaning of the expression can only be the maximal one.

I will pursue an alternative analysis, however: *inexpensive* really does exclude free objects from its ordering, but examples like (72) are really cases of coercion – i.e., scales can be coerced to include end points when there is an intuitive way to include the end points. This can also explain the contrast between *Dignity is short* and *Breathing air is inexpensive*; it is far less of

¹¹Thanks to Dan Lassiter (p.c.) for pointing this out.

¹²Lassiter argues that *free* is also gradable, and thus claims that *inexpensive/free* thus constitute a scalemate pair like *likely* and *certain*, and so provide yet further evidence against Kennedy (2007). However, it only seems plausible that *free* is gradable on the *free speech* reading, not on the *free beer* reading; i.e., the gradable *free* does not have a cost scale.

- (i) a. How free is Libya now that Gaddafi is gone?
 b. #How free is this beer?

a stretch to imagine *breathing air* as being subject to cost than to imagine dignity as taking up space, thus, coercion is easier in the first case than the second. The coercion of non-gradable adjectives into gradable adjectives is similarly subject to a cline of difficulty: *prime* is very hard¹³ to coerce, *pregnant* relatively easy.

- (73) a. ?Carissa is very pregnant (in fact, she's almost due.)
 b. ??This patient is not very dead (he could easily be resuscitated.)
 c. ???6 is less prime than 8. ($p(6)/(6-1) < p(8)/(8-1)$)

So if the scale *likely* relies upon is fully-closed scale with no minimum or maximum (i.e., a scale from 0 to 1 exclusive, or (0,1)) we can preserve the notion that *likely* uses something like a probability scale but whose extremes are excluded from the ordering. This predicts the weirdness of (74).

- (74) a. ??It is likely that $2+2=4$ (in fact, it's certain).
 b. ??It is more likely that $2+2=4$ than that Obama will be reelected.
 c. ??It is more likely that Obama will be reelected than that $2+2=5$.
 d. ??It is not likely that $2+2=5$ (in fact, it's impossible).

As with *inexpensive*, the expression can be coerced into taking on a fully closed scale structure.

2.3.1.3 Probability Measure Phrases

As noted above, the acceptability of $n\%$ with *likely* and *certain* is unclear due to conflicting judgments and limited corpus evidence. If we can categorically say that $n\%$ is bad with these expressions, then there is nothing to be concerned about. However, if the argument can be made that these are good or at least better than other proportional modifiers, there is a problem. I argue that, to the extent to which these modifiers are good, they are not proportional modifiers but in fact measure phrases which denote units of likelihood or certainty.

Assuming for the moment that $n\%$ is acceptable with these adjectives,

¹³But not impossible; Max Bane (p.c.) points out a possible meaning for gradable-*prime*: “Euler’s totient function $p(n)$ equals the number of integers less than n that are co-prime with n ; a prime number has $p(n) = n - 1$. [Thus,] % primeness could be defined as something like $p(n)/(n-1)$.” On this usage of *prime*, (72a) is, according to Wikipedia, true.

why don't other proportional modifiers work? e.g., why shouldn't (75a-c) and (76a-c) have the same acceptability?

- (75) a. ?It's 50% likely.
 b. *It's half likely.
 c. *It's halfway likely.
- (76) a. ?It's 75% likely.
 b. *It's three-quarters likely.
 c. *It's three-quarters-of-the-way likely.

Lassiter argues that *n%*'s distribution is actually different from other proportional modifiers, and works this into his claim about the scale structure of *likely*. He argues that *n%* is the only degree modifier truly diagnostic of fully closed scales, while other proportional modifiers select just for adjectives whose positive meaning is maximal.

Since *n%* should be utterly synonymous with corresponding fractional modifiers, the difference in selectional properties of *n%* and proportional modifiers proposed by Lassiter cannot be semantic, but rather truly arbitrary, which makes the account seem unmotivated. Coercion seems like an unlikely solution given that we should expect fractional modifiers to be coerced just as easily as *n%*.

The explanation I pursue here is that in the case of *likely*, *n%* is a measure phrase rather than a proportional modifier. In other words, *n%* denotes some number of a unit of likelihood or certainty, rather than denoting a proportion. This means that *n%* is not a diagnostic for a closed scale. Of course, the measure phrase *n%* is taken to be homophonous with a proportional modifier *n%* which has the usual semantics.

Such an account has to be right to account for the ambiguous use of *n%* in non-adjectival contexts like (77).

- (77) The Red Wings' odds of winning the Stanley Cup went up 10%.
 a. $\diamond \rightarrow$ The odds went from 40% to 50%.
 b. $\diamond \rightarrow$ The odds went from 40% to 44%.

The reading in (77a) is the measure phrase use, sometimes also paraphrased as *n percentage points*. The reading in (77b) is proportional use. Both are entirely felicitous in standard American colloquial English.

This explanation might also be applicable to *certain*, which displays a

similarly puzzling pattern with regard to $n\%$ modification. Consider that as the values for $n\%$ decrease, so does acceptability with *certain*.

- (78) a. It is 95% certain.
 b. ??It is 60% certain.
 c. *It is 30% certain.

It is not even clear what (78c) is intended to mean.¹⁴ It is unclear to me what the correct analysis of this phenomenon is; but it does at least throw into question *certain*'s status as a fully-closed scale adjective. Compare with *full*, which is clearly acceptable with any value for $n\%$.

- (79) a. It's 95% full.
 b. It's 60% full.
 c. It's 30% full.

If $n\%$ is actually a measure phrase denoting some unit of probability, it stands to reason that it could be used with *certain*, which perhaps employs a scale related to likelihood, but which only orders propositions which have high likelihoods. This allows for a satisfying explanation of the apparently idiosyncratic $n\%$ while maintaining the natural correspondence between proportional modifiers and closed scales. Finally, this could also explain why judgments about $n\%$ are mixed: most adjectives do not accept measure phrases, with adjectives in physical extent domain being a notable exception.

- (80) a. He is five feet tall.
 b. #He is thirty pounds heavy.

Uses of measure phrases with adjectives like *heavy* may be readily coerced, however, resulting in a mixed judgment. This picture accounts for the status of $n\%$ with *likely* and *certain*.

¹⁴A limited corpus search seems to confirm this; a search was conducted for *percent certain*, which yielded 44 hits. Because the total number of hits for the sense of *certain* we are concerned with is unknown, it is hard to evaluate the meaning of this. However, of the 44 hits, all but one were for values between 90 and 100, and the last remainder was for *80 percent certain*. In other words, there were no hits for *sixty percent certain*, *forty percent certain*, etc.

2.4 Conclusion

In light of the finding of the empirical studies presented above, it is clear that some modal adjectives are gradable and others aren't. This undercuts Lassiter's argument that all modals are inherently scalar, and that gradability is a partially syntactically determined property. Lassiter can still of course argue that whether a scalar term is overtly gradable or not is an idiosyncratic property, but this is a less convincing argument.

Lassiter does argue that even in light of idiosyncrasy in the modal domain in terms of what is gradable and what isn't, modals should be considered scalar across the board for uniformity's sake. In other words, given that a purely scalar semantics is clearly needed for such gradable modal adjectives as *likely* and *certain*, and that modality must be modeled in a way that allows for scalarity, modal expressions should be modeled as such across the board.

However, uniformity has no obvious motivation. Lassiter contends that having all modals be uniformly scalar would be simpler, but this kind of uniformity would in fact be rather extraordinary. As discussed above, properties of all types may be gradable or nongradable, scalar or nonscalar. Properties of individuals, for example, are sometimes scalar (*tall*, *flat*, *dirty*) and sometimes categorical (*man*, *dead*, *prime*); and after all, why shouldn't they? Likewise with properties of events, or relations over individuals, or modifiers.

There is one domain in which we would expect to find uniform scalarity, and that is the temporal domain. Times are subject to a single, unifying ordering, making any temporal expression seemingly necessarily scalar. But it is not obvious that modality shares this property. One could argue in light of Lassiter's proposals that *probability* represents such a scale. But in Section 2.3 I show that a view of *likely* and *certain* as having a probabilistic basic meaning does not capture the right facts with regard to degree modification.

And if there is nothing *inherently* scalar about modality, then uniformity is certainly the *less*-well-motivated option. Surveying the landscape of modal expressions, there is almost nothing universal to them, except their definitional property. Modals may be seemingly any syntactic category; there are auxiliaries like *must*, verbs like *have to*, adjectives like *possible* or *likely*, strictly attributive adjectives like *alleged*, nouns like *goal*, degree modifiers like *enough*, adverbs like *probably*, and so on. It has been said that modals uniformly undergo modal subordination, but Klecha (2011) shows that modal subordination is also an idiosyncratic, lexically varying property. Modals may be individual-anchored or not. Modals may have different temporal

properties. Modals may be assessment-sensitive or not.

Ultimately, some of these properties may be shown to derive from more systematic facts, e.g., the kind of modal domain employed may determine the temporal properties of the prejacents. The same may go for scalarity. But the argument for uniformity cannot motivate an underlying scalar semantics for non-gradable adjectives absent some positive, particular evidence; if we are to have any “default” assumption about the scalarity/non-scalarity of modals, it should be that some might be scalar and some might be non-scalar, rather than uniformity in either direction.

Lassiter does claim to have one such piece of evidence, namely entailment relations between modals. Below are some basic entailments that many authors have been interested in predicting (Yalcin, 2010; Lassiter, 2011; Holli-day and Icard, 2013).

- (81) a. He {must, has to, is bound to} be in the kitchen.
b. \models He is likely to be in the kitchen.
- (82) a. He is likely to be in the kitchen.
b. \models He {may, might, could} be in the kitchen.
- (83) a. He cannot be in the kitchen.
b. \models He is not likely to be in the kitchen.
- (84) a. He is not likely to be in the kitchen
b. \models He is not bound to be in the kitchen.

Another, provided by Lassiter (2011), is the entailment from (85a-b) to (85c).

- (85) a. John must be in the kitchen.
b. Roger is as likely as John to be in the kitchen.
c. \models Roger must be in the kitchen.

However, it’s not clear that (84) is a natural language entailment to be overly concerned about. Consider a natural context in which such an entailment.

- (86) *Jefferson commutes between Columbus and Cincinnati. He is in a windowless room in Columbus, talking to his wife, in Cincinnati on the phone. The connection cuts out and Jefferson says to Greg, “It must be raining in Cincinnati.” by way of explaining why he suddenly stopped talking on the phone. Craig then walks in the room, all wet.*

Greg says,

#It's at least as likely to be raining here as in Cincinnati.

(86) is quite odd, making it difficult to ascertain whether it entails (87).

(87) It must be raining in Columbus.

Part of the issue here may be that, as discussed by [von Fintel and Gillies \(2010\)](#), the expression *must* is not a 'vanilla' epistemic necessity operator (if such a thing even exists) but rather, includes information about the evidentiality of the prejacent, namely that it can be inferred from (only) indirect evidence. No such claim has ever been made about *likely*. Given this, it's not clear whether any entailments between *must* and *likely* are really needed to fully capture natural language usage and judgments. Consider also an attempt to exemplify the entailment in [\(23\)](#).

(88) *Ming and Andrea are wondering where Julian is. After talking about it for a bit, Ming says to Andrea, "He must be in the bathroom." Jackson overhears Ming say this. A moment later, Carissa, who did not hear Ming and Andrea's conversation, asks Jackson, "Is it likely that Julian is in the bathroom?"*

Jackson: Yes, it's likely Julian is in the bathroom.

Jackson's response might be acceptable, if underinformative, but notice that Carissa's initial question is itself extremely odd. So again it's not clear that entailments between *must* and *likely* inform our understanding of the natural language distributions of either term. Nonetheless, we may want to preserve the entailments discussed above and write off the oddness of discourses like (88) and (87) to separate issue.

If all modals are scalar, entailment relations between gradable and nongradable modals are easily and naturally derived. On the other hand, if some modals are scalar and some are not, it is not as easy. However, this does not mean that this cannot be done; it only means that the right scalar semantics must be given for gradable modals, and the right semantics for the non-gradable modals as well.

Thus, in the next chapter, I provide semantics for gradable modality which allows for (at least) two paths to capturing both the scalar structural facts discussed in [Section 2.3](#) and the entailment relations discussed just above.

Chapter 3

The Semantics of Gradable Modality

In this chapter I provide a semantics for gradable modals, focusing on *likely* and *important*. The semantics I do provide is, necessarily, underspecified to a certain extent – being totally specified would require me to commit to a particular view about human psychological categories like likelihood or goodness. It is not to say that a totally specified semantics is not desirable; after all, if we are to correctly and completely explain the patterns of usage of these expressions, we ultimately must understand how humans understand things like likelihood and goodness; likewise, we must understand how humans understand the categories FRUIT and VEGETABLE so that we may explain the patterns of usage of the expressions *fruit* and *vegetable*.

However, in the interest of making the biggest point I can in the least time and space, I will focus only on the broader questions that semanticists in tradition of Montague grammar tend to focus on: First, what are the *types* of these expressions, which allow us to broadly predict their combinatoric possibilities, and what is the simplest set of constraints we can place on the qualitative meanings of these expressions so as to explain the facts about them that we deem relevant to the determination of their types? Second, what ontological categories and primitive notions must be invoked to describe their meanings?

The desiderata of the proposal, given these guidelines and the empirical observations made in Chapter 2, are as follows: In answering the first question above, i) gradable modals should denote measure functions, i.e., functions of type $\langle\alpha, \langle s, d \rangle\rangle$, to account for their gradability; and ii) they

should also denote modal operators, i.e., functions of type $\langle\langle s, t \rangle, \langle\langle s, t \rangle, \alpha \rangle\rangle$ or a lift thereof, to account for their compatibility with conditionals. Moreover, these measure functions should have a scalar structure consistent with the observations about compatibility with degree modifiers; e.g., the measure function associated with *likely* should be totally open.

In service of the second question, the answers to the first entail at least this: Analyses of gradable modals must invoke the ontological categories of *degree* and *world*. But further questions remain, especially surrounding the role of probability. Does a Kolmogorovian, mathematical notion of probability need to be invoked in the analysis of *likely*? Does such a notion, or any gradient notion of likelihood, need to be invoked in the analysis of modal auxiliaries like *must*, especially in light of the entailment relations between such modals and the gradable modal *likely*? Does any notion of likelihood need to be invoked in the analysis of deliberative modals like *important* or *should*?

As for *likely* I will argue that the answers to the first guiding question determines that such a notion is at least not the most obvious solution. Since Kolmogorovian probability is associated with a fully closed scale, *likely* cannot denote such a measure function directly. However, it is not to say that such a notion may not be employed somehow, and I will present, but not endorse, such a possibility below. But see [Holliday and Icard \(2013\)](#) for more discussion of this question; they propose a qualitative probability measure for *likely*.

As for auxiliaries like *must*, I will argue that a gradient notion of likelihood need not be invoked to explain their meaning, as long as the measure function *likely* denotes is constrained in such a way as to derive the entailment relations described at the end of Chapter 2. To this end I will argue that two natural and well-motivated constraints achieve this: *Modal Additivity* and *Scalar Conservativity*. On the other hand, as I will discuss below, the gradable modal *important* has entailment relations with another modal auxiliary, *should*, but this must be derived another way: By appealing to the denotation of *should* and arguing that it contains a scalar modal component, in the spirit of many previous accounts, like [Kratzer \(1981a\)](#); [Kolodny and MacFarlane \(2010\)](#); [Lassiter \(2011\)](#); [Cariani et al. \(2012\)](#) and others. Thus my answer to the second guiding question above varies between epistemic strong necessity/possibility modals like *must* and deliberative weak necessity modals like *should*, and informs us generally about the extent of variation that may be seen within the category ‘modal’.

The chapter proceeds as follows. In Section 1 I recapitulate what it is to be a modal on a Kratzerian view and argue for a treatment of gradable modals as modal measure functions, i.e., expressions denoting functions of type $\langle\langle s, t \rangle, \langle\langle s, t \rangle, d \rangle\rangle$. In Section 2 I explore *likely* in detail, providing the minimal constraints on the measure function that it denotes to capture compatibility with degree modifiers and allow for capturing entailment relations. I then discuss an illustrative example of a measure function so constrained, derived from Yalcin’s (2007) proposal. In Section 3 I turn to *important* and *should*, providing a semantics for both which are united by a common scalar component, which derives the scale structure of *important* and the entailment relations between these two expressions.

3.1 Lewis and Kratzer

The first foray into non-categorical modal meaning was Lewis (1973), in his account of counterfactuals as what he called “variably strict conditionals”. He develops a notion of *similarity* which can be applied to worlds; this notion is, crucially, a scalar one. On his definition, a counterfactual like (1a) has the truth-conditions in (1b),¹ where S denotes the set of accessible worlds (the modal domain), and \geq_w is the “at least as similar to w as” operator.²

- (1) a. If kangaroos had no tails ($=\phi$), they would topple over ($=\psi$).
 b. $\lambda w[\exists v \in S(w) \cap \phi[\forall \geq_w v[[\phi \rightarrow \psi](u)]] \vee \neg \exists v \in S(w) \cap \phi]$

From this graded notion of similarity, Lewis develops comparative possibility, introducing the operator \succsim ³, which he says is read as “*as least as possible that ... as it is that ...*”. As will be discussed in Chapter 2, *possible* is not actually a gradable predicate in English, and it is unclear if Lewis meant to suggest that this operator actually be translated into natural language this way, or if this was simply a logic-ese translation. In any case, Lewis does not go into great empirical detail trying to analyze natural language semantic

¹Lewis first develops a semantics for variably strict conditionals which relies on nesting sets of spheres which represent similarity, but he subsequently develops the scalar-relation notion and shows that it is interchangeable with the sphere-based approach.

²Somewhat confusingly, Lewis’s original terminology has \leq meaning “at least as similar as” – this could be understood as, “less dissimilar than”. I adopt the opposite notation to avoid confusion.

³Actually \preccurlyeq

phenomena with his semantic machinery.

Lewis does, however, establish a starting point for integrating modal strength variability into an otherwise traditional, categorical semantics for modals; namely, by proposing that a gradable property of worlds like similarity be employed in the restrictor of an otherwise traditional quantificational modal.⁴ Kratzer (1981a) further develops Lewis’s insight, and points out that a scalar restrictor like the one initially proposed by Lewis is applicable, and in fact necessary to understanding, a wide range of modal phenomena, including gradable modal adjectives, weak necessity modals, and the deontic paradox. It is Kratzer (1981a) who introduces the notion of ordering source which was briefly introduced in Chapter 1.

The ordering source is used to determine a function which orders a given set of worlds (namely those of the preliminary modal domain) with respect to some world of evaluation. So depending on the make up of this ordering source, the set of worlds which ‘survive’ from the preliminary modal domain into the proper modal domain may vary. Consider first the application of this notion to deontic modals, explored by Kratzer (1981a).

(2) If someone is murdered, the killer has to go to jail.

If the domain of *have to* was simply the set of deontically accessible worlds, i.e., the set of worlds where the rules are obeyed, (2) would be nonsense. Since murder is against the rules, there simply are no deontically accessible worlds where someone is murdered. Simple restriction by the conditional makes for an empty domain, and trivial truth of (2).

What Kratzer proposes is that deontic necessity modals have a domain which includes all the circumstantially accessible worlds which are *most* deontically accessible; i.e., the domain results from a circumstantial modal base and a deontic ordering source. By postulating a deontic ordering source, rather than a deontic modal base, Kratzer commits herself to the notion

⁴One noteworthy proposal by Lewis, however, is to reanalyze the traditional necessity and possibility operators as being themselves scalar rather than quantificational.

- (i) a. $\Diamond\phi := \phi \succ \perp$
b. $\Box\phi := \neg\phi \approx \top$

Here \top stands for any tautological proposition, and \perp any contradiction. This suggestion by Lewis would involve upending the traditional, categorical semantics for modals and is echoed in Lassiter’s (2011) proposal, as will be discussed below.

that deontic accessibility is a gradient notion, and we can always shift our threshold for deontic accessibility depending upon the circumstances.

Thus the *if*-clause adds the proposition *someone is murdered* to the modal base, i.e., the stock of circumstances being weighed, as depicted in Section 1.2.2. Then the worlds of the preliminary modal domain determined by that modal base are ranked with respect to a deontic ordering source. The worlds which best match the rules or moral code represented by this ordering source are then included in the modal domain. Thus the problem of empty modal domains does not arise in cases like these.

3.1.1 Proximity to Ideal

Kratzer’s original ordering semantics depends on the basic comparative operator \geq , which she defines as “closeness to an ideal”.

- (3) **Kratzer’s (1981a) Ordering**
 For $w, v \in W$, and $A \subseteq \mathcal{P}(W)$,
 $w \geq_A v$ iff $\{p : p \in A \ \& \ v \in p\} \subseteq \{p : p \in A \ \& \ w \in p\}$

So w is closer⁵ than v to the ideal A iff the set of propositions in A that are true in w is a superset of those that are true in v . Note here that Kratzer’s ordering involves closeness to an ideal, i.e., a set of propositions (something like a modal base), rather than similarity to a particular world, a la Lewis; but we can then compare two worlds’ closeness to an ideal to derive an ordering over worlds.

Kratzer then applies this ordering notion to the definition of what she calls “human necessity”. As opposed to “simple necessity”, which corresponds to the traditional logical necessity, unrestricted by an ordering source, human necessity is necessity with respect to a modal domain that has been limited to just the best worlds.

Kratzer provides an algorithm for determining the best worlds given an ordering source, which Portner (2009) simplifies; my version of this is in (4). Portner’s simplification has only the disadvantage of requiring the Limit Assumption (Lewis, 1973); see Portner (2009) or Kratzer (1981a) for the more complicated version.

⁵Kratzer also uses Lewis’s notational directionality; again, I adopt the more intuitive notation at the risk of muddling the historical record.

- (4) **Best Worlds**
 For a preliminary modal domain m , ordering source g , and world w ,
 $\text{BEST}_{g(w)}(m) = \{v \in m : \neg \exists u \in m [u >_{g(w)} v]\}$

Thus the Lewis-Kratzer tradition of dealing with modal strength variability is to allow for restriction of the modal domain by a scalar restrictor, be it a primitive scalar property like Lewis's similarity or one derived from a conversational background, as in Kratzer's ordering source. This has been a very successful way to treat the gradience of otherwise non-gradable modals, such as auxiliaries in English, but it is not the only method; gradable modals cannot be accounted for in this way.

3.1.2 The Divide between Scalarity and Modality

Kratzer (1981a) proposes that her notion of ordering could also be applied to gradable modals like *likely*. She introduces her notion of *better possibility*.

- (5) *Better Possibility*
 p is a g -better R -possibility than q in world w iff
 $\forall u \in (\cap R(w)) \cap q . \exists v \in (\cap R(w)) \cap p . v \geq_{g(w)} u$ &
 $\exists u \in (\cap R(w)) \cap p . \neg \exists v \in (\cap R(w)) \cap q . v \geq_{g(w)} u$

In other words, p is better than q with respect to the accessibility relation R and the ordering source g iff for every accessible q -world there is some g -better accessible p -world, and, there is some accessible p -world which is g -better than all the other q -worlds.

Although the basic insight that an ordering semantics is needed for the analysis of some (if not all) modal expressions, is correct, the original formulation proposed by Kratzer is problematic; given the analyses of degree modifiers discussed in Chapter 2, this analysis is not compositional. As a gradable adjective, an expression like *likely* should denote a measure function.

Yalcin (2010) and Lassiter (2011) also point out more serious flaws in the theory with regard to a number of predictions it makes. For example, both Yalcin and Lassiter point out that Kratzer falsely validates the disjunctive inference.

- (6) *The Disjunctive Inference*
 a. ϕ is at least as likely as ψ

- b. ϕ is at least as likely as χ
- c. ϕ is at least as likely as $\psi \vee \chi$

Consider that i) for ϕ to be likelier than ψ there must only be one world in ϕ better than all the ψ worlds; ii) for ϕ to be likelier than χ there must only be one world in ϕ better than all the χ worlds; and iii) if both of these things hold, it must be that there is some world in ϕ which is better than all the $\phi \vee \chi$ worlds. However, Yalcin and Lassiter argue that the disjunctive inference is clearly false, since probabilities are additive, whereas Kratzer predicts that the disjunction of two propositions is just as likely as the more likely of the two constituent propositions.

(7) *Additivity*

$$\text{prob}(p \vee q) = \text{prob}(p) + \text{prob}(q)$$

Although the formulation of Kratzer (1981a) has been showed to be flawed, I will not pass judgment on the project of deriving a likelihood measure from a non-quantitative basis, i.e., Better Possibility. Rather, my goal will be to analyze *likely* from the top down, arguing from the empirical landscape established in Chapter 2 for a basic outline of the meaning of the expression.

3.1.3 Core Kratzer

The general theory of modality outlined by Kratzer (1981a, 1986) and summarized in Chapter 1 says the following: Modals are relations between a modal domain and a prejacent proposition; the modal domain is determined by several contextually variable parameters (Relative Modality); one of these parameters is scalar in nature (Ordering); and conditionals are derived by restricting the modal domain (Conditionals as Restrictors). These core proposals have formed the backbone of significant research in modality.

Some aspects of Kratzer's (1981a) original proposal have since been shown inadequate.

Nevertheless, the existence of gradable modals is not a problem for the core Kratzerian approach to modality. To this end, I argue for the following schematic denotation for gradable modals in (8a). This is parallel to the basic schema for non-gradable modals in (8b). Below the variable p has type $\langle s, t \rangle$; the variables m and g have the type $\langle s, \langle \langle s, t \rangle, t \rangle \rangle$.

- (8) a. $\llbracket \text{gradable modal} \rrbracket = \lambda p \lambda m \lambda g \lambda w [meas(\text{BEST}_{g(w)}(\cap m(w)))(p)]$
 b. $\llbracket \text{non-gradable modal} \rrbracket = \lambda p \lambda m \lambda g \lambda w [Q(\text{BEST}_{g(w)}(\cap m(w)))(p)]$

Both kinds of expressions take a prejacent p , a modal base intension m , an ordering source intension g , and an evaluation world w , where $\text{BEST}_{g(w)}(\cap m(w))$ forms the modal domain;⁶gradable modals return a degree, non-gradable modals a truth value. Non-gradable modals relate the modal domain and prejacent by a generalized quantifier; gradable modals, on the other hand, denote a measure function, which maps modal domain-prejacent pairs to a degree on a scale. Gradable modals may differ in terms of what exactly the ordering is, and in terms of what domain-prejacent pairs are included in the ordering.

This is the major proposal of this chapter: Modals are just relations over sets of worlds; gradable modals are just gradable relations over sets of worlds. Clearly, relations can be gradable just like 1-place properties:

- (9) a. John is very similar to Mary.
 b. John is more similar to Mary than Nancy is.

(9b) says that the pair $\langle \text{John}, \text{Mary} \rangle$ is higher on the similarity scale than the pair $\langle \text{Nancy}, \text{Mary} \rangle$. Any arbitrary relations can be gradable; it follows very naturally that modals can be too. Of course, in order to maintain this view, it must be shown that such modals can enter into entailment relations with nongradable modals. In this chapter I provide analyses of two gradable modals which do just this: *likely* and *important*.

3.2 Gradable Epistemic Modality

Given the schematic proposed above, *likely* should have a denotation like (10).

$$(10) \quad \llbracket \text{likely} \rrbracket = \lambda p \lambda m_{\langle s, \langle \langle s, t \rangle, t \rangle \rangle} \lambda g \lambda w [lhood(\text{BEST}_{g(w)}(\cap m(w)))(p)]$$

⁶There is nothing crucial this particular way of determining the domain, and I will not make an extensive argument for it. One may argue there is something intuitively redundant about using an ordering source in a gradable modal, since *meas* itself provides an ordering, although this ordering is over domain-prejacent pairs, not worlds. However it simply must be asked of each particular modal how its domain is determined; this goes for non-gradable modals as well.

The question, then is, what is *lhood*? To give an uninformative answer, *lhood* takes a prejacent p and a domain D and returns the likelihood of p given D . But of course, the precise ordering denoted by *lhood* is difficult to assess, and in fact the exact ordering may vary contextually, and may even be subjective. It is not within the scope of this project to determine *what likelihood is* as a concept; no more than it is the job of natural language semanticists interested in degree terms to know what exactly the ordering is denoted by something like *beautiful* or *smart*. We only need to know enough about *lhood* to be able to say that its semantics can derive the crucial entailment relations, repeated below.

- (11) a. He {must, has to, is bound to} be in the kitchen.
 \models He is likely to be in the kitchen.
- b. He is likely to be in the kitchen.
 \models He {may, might, could} be in the kitchen.
- c. He could not be in the kitchen.
 \models He is not likely to be in the kitchen.
- d. He is not likely to be in the kitchen \models He is not bound to be in the kitchen.
- (12) a. John must be in the kitchen.
- b. Roger is as likely as John to be in the kitchen.
- c. \models Roger must be in the kitchen.

Let's call the relations seen in (11) Necessity Entails Likelihood (equivalent to Non-Likelihood Entails Non-Necessity), and Likelihood Entails Possibility (equivalent to Impossibility Entails Non-Likelihood). Let's call the relation in (12) Likelihood of Necessity Entails Necessity; it is essentially saying that if one proposition is a necessity, then whatever likelihood we associate with it (or any higher likelihood) cannot be associated with any non-necessary propositions.

We must provide sufficient definition to the ordering *lhood* to derive these entailment patterns. This is relatively straightforward, however; two properties, *Modal Additivity* and *Scalar Conservativity*, will suffice.

3.2.1 Important Properties of the Ordering

As discussed above, *lhood* orders pairs of a modal domain and a prejacent proposition. The data from degree modification, discussed in Chapter 2, tells

us that this must also be an open scale; this represents a serious departure from the common ways of thinking about probability.

This gives the basic type of *lhood*; an unbounded ordering of modal domain-prejacent pairs. On this view, I take *lhood* to be like a Popper function,⁷ wherein conditional probability is taken as basic, rather than as a special case of basic unconditional probability (see Hájek, 2003). If we adhered to an view like that of unconditional probability, the Kratzerian view of modals as relations would be lost, and the uniform modifiability of modals by *if*-clauses would as well.

Two properties that unconditional probability functions are often taken to have are Necessity Entails Maximality and Additivity.

- (13) a. *Necessity Entails Maximality*
 $Pr(W) = 1$
 b. *Additivity*
 If $\phi \cap \psi = \emptyset$, then $Pr(\phi \vee \psi) = Pr(\phi) + Pr(\psi)$

Additivity says that if two propositions do not overlap, then the measure of their disjunction (union) is just the sum of their individual measures. This is a general property of many measure functions, including non-propositional measure functions; the definition here is just the definition adapted particularly for a propositional measure function.

However the definition of Additivity above does not apply to *lhood* because it is not the right type; it orders pairs of sets of worlds rather than single sets. But this is an easy fix: Additivity should hold when the modal domain is held constant. I call this Modal Additivity.

- (14) *Modal Additivity.* For relational measure function G , modal domain M , propositions p, q ; if $G(M, p)$ and $G(M, q)$ are defined:
 if $p \cap q \cap M = \emptyset$, then $G(M, p \vee q) = G(M, p) + G(M, q)$

This says that a relational measure function is Modal Additive if, *holding the modal domain constant*, it is Additive. Additionally, the precondition is not that p and q be absolutely disjoint, but just disjoint in the modal domain. This is due to the intuition that worlds outside the modal domain don't matter; this intuition is formalized in (i) below.

Modal Additivity is a property of Popper functions⁸ (see, e.g., Égré and

⁷But crucially *not* a Popper function, as is explained below.

⁸Égré and Cozic (2011) state the first axiom of Popper functions to be (i).

Cozic, 2011). However, unlike Popper functions, *lhood* must be stipulated to not have a maximal or minimal element, given the data on degree modification presented in Chapter 2. Thus in numerical terms it is a scale $(0, 1)$, rather than $[0, 1]$; this distinguishes it from traditional probability definitions, both conditional and unconditional. This has a crucial implication, which is that the second of the definitional properties of probability functions above, Necessity Entails Maximality, cannot be applied to *lhood*. Note, however, that Necessity Entails Maximality is not identical to the entailment relation we wish to derive, Necessity Entails Likelihood. This relation can be achieved if we stipulate just one other property of *lhood*, which I call *Scalar Conservativity*.

- (15) *Scalar Conservativity*. A relational measure function G over two sets P, Q is scalar conservative iff:

$$G(P, Q) = G(P, Q \cap P)$$

A scalar relation over two sets is Scalar Conservative if the degree returned by the measure function for sets P and Q is the same as for P and $Q \cap P$; i.e., for the purposes of the ordering, elements in Q which are not also in P do not matter. This is an intuitive move – the likelihood of a given proposition p given some conversational background presumably does not depend on the p worlds inconsistent with that background. Moreover Scalar Conservativity is just the scalar equivalent of Conservativity, an important property of natural language quantifiers. If *likely* is just a gradable quantifier, as I propose, it is very natural that it should be Scalar Conservative. The gradable quantifier *many* (Hackl, 2000) is Scalar Conservative.

- (16) Many aspens are white. \models Many aspens are white aspens.

Scalar Conservativity is therefore a very natural constraint to place on *lhood*. Scalar Conservativity is also a property of Popper functions; it can be derived from Necessity Entails Maximality and Additivity.

These two constraints are enough to derive everything we need. First, consider Likelihood Entails Possibility. Since *lhood* is Modal Additive, this is relatively easy. If $lhood(M, \phi)$ is positive, and ψ is impossible in M , (i.e.,

-
- (i) for all E_1 , $Pr(\cdot; E_1)$ is a probability function.

Since Additivity is a property of probability functions, Modal Additivity is a property of Popper functions.

$M \cap \psi = \emptyset$), then $lhood(M, \phi \vee \psi) = lhood(M, \phi) + lhood(M, \psi)$; by Scalar Conservativity, $lhood(M, \phi \vee \psi) = lhood(M, \phi \vee \psi \cap M) = lhood(M, \phi)$ i.e., the impossible proposition contributes no probability.

- (17) For ϕ, ψ, M , where $lhood(M, \phi) > 0$ and $\psi \cap M = \emptyset$;
- a. $lhood(M, \phi \vee \psi) = lhood(M, \phi) + lhood(M, \psi)$ (Modal Additivity)
 - b. $lhood(M, \phi \vee \psi) = lhood(M, \phi \vee \psi \cap M)$ (Scalar Conservativity)
 - c. $lhood(M, \phi \vee \psi \cap M) = lhood(M, \phi \cap M)$ (premise: $\phi \vee \psi \cap M = \phi \cap M$)
 - d. $lhood(M, \phi \cap M) = lhood(M, \phi)$ (Scalar Conservativity)
 - e. $lhood(M, \phi) = lhood(M, \phi) + lhood(M, \psi)$

Note that we cannot take the next logical step and show that $lhood(M, \psi) = 0$ by subtraction, since $lhood$ does not map to zero, as indicated by the degree modification data. This can only mean that $\lambda x[lhood(M, x)]$ is undefined for propositions which are impossible in M . But still, if $lhood(M, p) > s$ for some standard s , $lhood(M, p)$ is defined, and therefore, p is possible in M . Thus, Modal Additivity and Scalar Conservativity derive Likelihood Entails Possibility.

Second, consider Necessity Entails Likelihood. Consider any arbitrary proposition ϕ which is likely in M , i.e., $lhood(M, \phi) > s$ for some standard of comparison; as argued in [Lassiter \(2011\)](#), s is contextually variable, so let s be any number in $(0, 1)$. Next, take ψ , which is necessary in M , i.e., $M \cap \psi = M$. By Scalar Conservativity, $lhood(M, \psi) = lhood(M, M \cap \psi) = lhood(M, M)$. Also by Scalar Conservativity, $lhood(M, \phi) = lhood(M, M \cap \phi)$. Clearly, $M \cap \phi \subseteq M$, therefore, $lhood(M, \phi) \leq lhood(M, M)$. Thus, for any arbitrary standard, and any arbitrary proposition whose likelihood exceeds that standard, a necessary proposition must have at least as much likelihood.

- (18) For ϕ, ψ, M , where $lhood(M, \phi) > 0$, and $M \subseteq \psi$;
- a. $lhood(M, \psi) = lhood(M, M \cap \psi)$ (Scalar Conservativity)
 - b. $lhood(M, M \cap \psi) = lhood(M, M)$ (premise: $M \cap \psi = M$)
 - c. $lhood(M, \phi) = lhood(M, M \cap \phi)$ (Scalar Conservativity)
 - d. $lhood(M, M) = lhood(M, M \cap \phi) + lhood(M, M \cap \bar{\phi})$ (Modal Additivity)
 - e. $lhood(M, M) \geq lhood(M, M \cap \phi)$

$$\text{f. } lhood(M, \psi) \geq lhood(M, \phi)$$

In other words, (Modal) Additivity essentially says that bigger sets get bigger likelihoods (or at least as big). Scalar Conservativity says that, for the purposes of the ordering, sets as big as the domain M are the biggest sets you can get. Thus, sets as big as M have the biggest likelihoods you can get. Note that this does *not* say that a necessary proposition will be mapped to a likelihood of 1; just that nothing can be mapped to a bigger likelihood. Thus while *lhood* does not have Necessity Entails Maximality, it does have Necessity Entails Likelihood.

Consider now the entailment discussed by Lassiter, Likelihood of Necessity Entails Necessity. If we take the concerns raised in Chapter 2 seriously and reject this entailment, then we need go no further. But suppose we were to adopt Lassiter’s entailment as well. What would it take for the theory sketched so far to make the ‘wrong’ prediction? It would have to predict that ϕ and ψ receive equal likelihoods when paired with domain M , even though $\psi \cap M \subseteq \phi \cap M$. Given Modal Additivity, this would mean that the $(\phi - \psi) \cap M$ worlds do not contribute any likelihood whatsoever, even though $\Diamond_M(\phi - \psi)$.

So a third constraint is called for; *lhood* should map any pair (M, ϕ) to a (non-zero) likelihood, so long as $M \cap \phi \neq \emptyset$. In other words, Possibility Entails Some Likelihood. With this constraint, Likelihood of Necessity Entails Necessity is derived.

Scalar Conservativity and Modal Additivity are both extremely natural and well-motivated. Possibility Entails Some Likelihood is a less general kind of constraint, but is intuitive as well. It is enough to show that these derive the desired entailments, but an example of such an ordering may bolster this point. Indeed, the probability measure proposed by Yalcin (2007), if adapted to derive probability relative to a modal domain rather than all worlds, satisfies both Scalar Conservativity and Modal Additivity, and also derives the Likelihood of Necessity Entails Necessity.

3.2.2 Epistemic Entailments on Probabilistic Accounts

The constraints given above do not rule out an account of *likely* as ultimately derived from a mathematical notion of probability. Many authors have voiced their concern about basing the semantics of the natural language expression *likely* in a mathematically sophisticated notion like Kolmogorovian probabil-

ity (Hamblin, 1959; Kratzer, 2012; Holliday and Icard, 2013), and I likewise I do not wish to endorse one. But by considering an account based in such a notion, like Yalcin’s (2007), we can see that the desired entailment relations are already present, and remain present even after modifying the account slightly to make the account relational, and therefore a truly modal account.

In Yalcin’s model, an epistemic modal applied to a prejacent p is interpreted with respect to a probability space, or a triple $\langle \Pi, \pi, Pr \rangle$. Π_p here is a partition over the space of all possible worlds, such every partition in Π is homogenous with respect to p ; i.e., for each partition, p is either true throughout it or false throughout it. Pr is a probability function which assigns to every a partition a probability value between 0 and 1 such that the probabilities of the partitions sum to 1. In Yalcin’s account, π is defined as the set of partitions in Π with a non-zero probability, i.e., the live possibilities. Note that this means Yalcin rejects Possibility Entails Some Likelihood. This is not a problem, however; he still captures Likelihood of Necessity Entails Necessity, because for him simple epistemic modals quantify only over worlds in $\cap \pi$. In other words, Yalcin rejects ‘Logical Possibility Entails Some Likelihood’ but still endorses ‘Natural Language Possibility Entails Some Likelihood’.

I adapt Yalcin’s formula slightly so that Π instead partitions the modal domain. This allows for the fact that *likely* may take a different modal base (namely a metaphysical one), but does not dramatically alter the basic story here. On this slightly more Kratzerian take on Yalcin, the probability space can be seen as a kind of ordering source; it makes use of some scalar concept to narrow the domain further.

- (19) *Homogeneity*. For prejacent p , set of worlds ι :
 $\eta_p(\iota) = 1$ iff $\iota \subseteq p \vee \iota \cap p = \emptyset$

This defines homogeneity, indicated by η ; a set of worlds ι is homogenous with respect to p iff p is true throughout (i.e., necessary) or false throughout.

- (20) $\langle \Pi, \pi, Pr \rangle$ is a probability space iff, for modal base M , prejacent p :
- a. $\Pi(M, p) \in Pow(M)$ and $\forall w \in M [\exists! \iota \in \Pi(M, p) [w \in \iota]]$
 - b. $\forall \iota \in \Pi(M, p) [\eta_p(\iota)]$
 - c. $\sum_{\iota \in \Pi(M, p)} Pr(\iota) = 1$

(20a) says that for a modal domain M and prejacent p , Π partitions the modal domain M . (20b) says that every partition-cell in $\Pi(M, p)$, is homogenous with respect to p . (20c) says that the sum of the probabilities of partition-

cells is 1.

- (21) *Live Possibilities*. For probability space g , modal base M , prejacent p :
- $$\pi_g(M, p) = \{\iota \mid Pr(\iota) > 0 \ \& \ \exists X \subseteq \Pi_g(M, p)[\eta_p(X) \ \& \ \iota = \bigcup X]\}$$

(21) defines π as the set of partition-cells with non-zero probability, including partition-cells which are the union of smaller partition-cells in Π (so long as homogeneity is maintained).

Given this set-up, a definition of the measure function *lhood* can be given.

$$(22) \quad \text{a.} \quad lhood_g(M)(p) = \sum_{\iota \in \Pi_g(M, p) \cap Pow(p)} Pr_g(\iota)$$

Since probabilities are only assigned to partitions of M , *lhood* is clearly Scalar Conservative. And since *lhood* is based on the Additive measure Pr , it is Modal Additive.

Recall also the constraint that *lhood* should denote an open but bounded scale, i.e., a scale isomorphic to $(0,1)$, rather than $[0,1]$. This is a more serious problem for the Yalcin-style approach. This narrow problem, of course, can be achieved by stipulating the constraint on Π that there be at least one partition for each truth value of p with non-zero probability. I'll refer to the modification to Yalcin's proposal described above as Yalcin_{2,0}, and the further stipulation ensuring scale openness as Yalcin_{3,0}.

The problem is that while Yalcin_{2,0} surely maps necessary propositions to 1, on Yalcin_{3,0}, $lhood(p, M)$ is undefined if $M \subseteq p$. This in turn would amount to rejecting the entailment relation of Necessity Entails Likelihood.

- (23) a. He {must, has to, is bound to} be in the kitchen.
 b. \models He is likely to be in the kitchen.

This might be doable. First of all, as discussed in Chapter 2, natural language discourses meant to test this *must* and *likely* are ill-formed, possibly due to the mismatch between the flavors of *must* and *likely*. Second, it certainly is odd, although not terrible, for someone to say (23b) if they are committed to (23a). Thus, as argued in Chapter 2 for conceptually closed but lexically open adjectives like *inexpensive*, (23b) might be acceptable as the result of coercing away the constraint against necessary or impossible propositions. Thus, the entailment relation in (23) is validated, as long as coercion is accounted for.

Another option would be to modify our assumptions about the meaning

of *must* and *might*. If *must* is actually weaker than strong necessity, this concern does not arise; i.e., if $\text{must}(M, p)$ does not entail that $M \subseteq p$, but rather $X \subseteq p$, for X a subset of M . For example, *must* could quantify over just the set of worlds in M consistent with certain defeasible rules of inference (e.g., wait rain gear *usually* means rain). Since, by definition, these rules are very unlikely to be falsified (i.e., it is very unlikely that wet rain gear is due to a sprinkler), the probability mass excluded from the domain of *must* is very small, and thus its prejacent is assured to exceed any contextually determined threshold of *hood*.

A third option would be to concede that *must* is strong as von Fintel and Gillies (2010) argue, but take a closer look at the discourse dynamics surrounding the entailment in (23). Consider again the (rather odd) discourse discussed in Chapter 2.

- (24) *Ming and Andrea are wondering where Julian is. After talking about it for a bit, Ming says to Andrea, “He must be in the bathroom.” Jackson overhears Ming say this. A moment later, Carissa, who did not hear Ming and Andrea’s conversation, asks Jackson, “Is it likely that Julian is in the bathroom?”*
Jackson: Yes, it’s likely Julian is in the bathroom.

But consider a slight variant where Carissa instead asks *how likely*.

- (25) *Ming and Andrea are wondering where Julian is. After talking about it for a bit, Ming says to Andrea, “He must be in the bathroom.” Jackson overhears Ming say this. A moment later, Carissa, who did not hear Ming and Andrea’s conversation, asks Jackson, “How likely is it that Julian is in the bathroom?”*
Jackson: #He must be in the bathroom.

It’s very odd, on Lassiter’s (2011) account, that (25) is not a perfect answer given its preceding context. In fact, this is odd even if we accept that *likely* and *must* are not scale mates but do share a particular entailment relation. Even if Jackson, before Carissa asked her question, is committed to the truth of (25), it is odd for him to remain so afterwards. This seems to be because Carissa raises doubt in Jackson about the certainty of his conclusion. Because the expression *likely* admits such fine-grained delineations in likelihoods, we can think of it as raising the ‘resolution’ or precision with which Jackson considers the proposition.

This approach to precision interacting with modality will be discussed in more detail in Chapter 4, but it is enough to say here that it may account for why (24) is acceptable. What happens in (25) is that Jackson actually backs off of the commitment to the necessity of the prejacent, instead believing it to be a (very very) likely proposition. So this is not really an entailment relation.

Finally, it may simply be that Yalcin’s account is not the way to go, and its reliance on a mathematical notion of probability gives us reason to believe that this is the case. I do not endorse the particular Yalcin-style definition of *lhood* given above, but only sketch it to show that it has the properties of Modal Additivity, Scalar Conservativity, and Natural Language Likelihood Entails Natural Language Possibility without stipulation. This is enough to show that a mixed scalar/nonscalar modal semantics can indeed be provided. The burden then shifts to the proponent of a uniformly scalar modal semantics to show why nongradable modals, especially nongradable modal adjectives, should be taken to have an underlying scalar semantics, when neither compatibility with degree modifiers, nor entailment relations make such a semantics necessary, or even detectable.

3.3 Gradable Deliberative Modality

Not all gradable modal adjectives are epistemic/metaphysical – some belong to the root category as well. One such adjective is *important*.

- (26) It’s more important that you continue your studies (than that you join the war effort).

Crucially there is an entailment relation between *important* and the deliberative modal auxiliary *should*.

- (27) (26) \rightarrow Given the option, you should continue your studies.

The solution given for *likely* won’t work here, because *important* is seems very clearly not additive.

In this case, I argue that the entailment comes from the fact that *should* has a scalar component, namely its domain is determined by a priority ranking, which is the ranking that *important* makes use of. In this case I am arguing that *should* is a nongradable modal, with a universal quantifier ranging

over a set of worlds determined by an ordering, very much along the lines of Kratzer (1981a), but drawing insight from Lassiter (2011) and deriving interaction with a gradable modal nonetheless.

3.3.1 Polysemy

The first item to note is that there are (at least) two *important*s which I will discuss. The first I will call *important*₀, which simply takes a subjunctive-mood CP headed by *that*; the second, which I will call *important*_{exp}, additionally takes a PP.

- (28) a. It's important that he put that vase over there.
b. It's important to him that that vase be over there.

There seems to be a significant semantic contrast. For *important*_{exp}, it is only acceptable for the complement clause to be a desire of the experiencer indicated by the PP, while the complement of *important*₀ need not be a desire of anyone.

- (29) a. It's important to me that she be happy.
b. It's important that she be happy.
(30) a. #It's important to me that I do some things I don't want to do.
b. It's important that I do some things I don't want to do.

Consider also the unacceptability of *important*_{exp} with an overt indicator of goals, like the prejacet in (31).

- (31) If you want to get a good job, it's important (#to you/me) that you write a strong research statement.

Moreover, *important*₀ presupposes that its complement is for a given purpose, while *important*_{exp} is acceptable discourse initially.

- (32) *Discourse Initial Context*
a. It's important to me that I get up early every morning.
b. ?It's important that I get up early every morning.

Note that (32b) is not necessarily unacceptable, but it does at least give rise to an inference that there is a particular purpose in mind and would likely,

in a discourse initial setting, give rise to the response question *Why?* Such a response to (32a) would be odd.

Finally, consider that it is odd (or at least unlikely to be informative) to use the addressee as the experiencer in a matrix context, similar to many attitude predicates.

- (33) a. #It's important to you that you eat a good breakfast this morning.
b. It's important that you eat a good breakfast this morning.
- (34) a. #You want to eat a good breakfast this morning.
b. #You believe you'll eat a good breakfast this morning.

The crucial entailment relations I want to consider seem to apply only to *important*₀.

- (35) a. It's more important to Bill that he play video games every morning than that he eat a good breakfast.
b. #?So he should play video games every morning.

3.3.2 Modality of Importance

I argue that the scale associated with *important*_{exp} is simply a priority or preference scale, ranking propositions by their desireableness or their priority for a given experiencer (indicated by the PP). However the scale associated with *important*₀ is less straightforward. Whereas *important*_{exp} is a bouletic modal, *important*₀ appears to be teleological; its compatibility with *If you want...* as seen in (31) is a strong argument for this. This is also sensible given the aforementioned entailment relations with the teleological weak necessity modal *should*.

- (36) If you want to get a good job, you should spend a lot of time writing a research statement.

Whereas *should* seems to relativize to the goals of the subject, the individuals whose goals are at issue is contextually variable in the case of *important*; consider that the goals in question in (36) are the addressee, but in (37) they seem most easily to be associated with the speaker.

- (37) It's important the flowers be brightly colored and aromatic.

Note that the above sentence sounds a lot like a case of *important*_{exp} and could even be ambiguous, but note that (37) need not be a case of the speaker announcing her first-order preferences, rather, it could simply be a statement of a means to a desired end, e.g., *in order to impress the guests....*

However, despite the differences between *important* and *should*, the entailment relation discussed in the previous subsection remains. Below I show how these properties can be derived from a common scalar source.

3.3.3 Weak Teleological Necessity

Consider again the relevant entailment.

- (38) a. It's more important that you continue your studies than that you join the war effort.
b. \rightarrow Given the option between the two, you should continue your studies.

First, consider a possible semantics for *should*, focusing only on the deliberative, teleological reading.

A simple distillation of the standard analysis for teleological *should* is to say that it universally quantifies over the worlds which best meet the goals of a given individual. A somewhat simplified denotation is given below, adapted from Portner (2009).

$$(39) \quad \llbracket \text{should} \rrbracket_{\text{simple}} = \lambda p \lambda m \lambda g \lambda w [\forall v \in \text{BEST}(m(w), g(w)) [v \in p]]$$

Lassiter (2011) raises a number of concerns about this quantificational denotation. While he actually does not consider entailment relations between *should* and gradable adjectives like *important*, he does point out that problems like the Miner's Paradox require that *should* be sensitive to probability, which is of course a scalar notion.

The Miner's Paradox is a problem with the deliberative readings of *should* and *ought* which has been discussed extensively in the literature, and has enjoyed a great deal of attention recently (Kolodny and MacFarlane, 2010; Lassiter, 2011; von Fintel, 2012; Willer, 2012; Carr, 2012; Charlow, 2013; Silk, 2013). The paradox can be summarized as follows: 10 miners are trapped in a mine. Floodwaters are coming which will flood the mine if it is not blocked. The mine has two shafts, only one of which can be blocked. All ten miners are in one of the two shafts, but we don't know which one. Crucially, leaving

both mines unblocked will actually result in neither shaft being completely flooded, such that only one miner will die. Given that we do not know which shaft the miners are in, and have no hope of finding out, (40) is true. Yet, (41a-b) are also true.

- (40) We should block neither shaft.
- (41)
 - a. If the miners are in Shaft A, we should block Shaft A.
 - b. If the miners are in Shaft B, we should block Shaft B.

The truth of (41a-b), on a Kratzerian view of *should* and the conditional, should entail that (42) is true, but this is not the case; so, paradox.

- (42) (The miners are either in Shaft A or Shaft B, so) either we should block Shaft A or we should block Shaft B.

Cariani et al. (2012) clarify the issue somewhat by separating out the true *deliberative* reading from the *objective* reading. On the objective reading, (42) may be true. However the deliberative reading only concerns the possible actions of the non-omniscient individual who must decide what to do. Keeping to this reading for all of the above sentences, there is an apparent conflict.

An appropriate paraphrase of *should* ϕ on the deliberative reading is “ ϕ is the best course of action”. The upshot of the paradox is this: if an agent is deciding between various courses of action, some of which have uncertain consequences, then what is the best course of action seems to depend on the likelihoods of the various possible outcomes of each of the various possible courses of action. If there were a 99% chance that the miners were in Shaft A, and a 1% chance they were in Shaft B, (40) might be deemed false. In other words, even if the globally best possible outcome is in a ϕ -world, it may be very unlikely given ϕ , and so ϕ may not be the best course of action to take for a non-omniscient agent. On a Kratzer (1981a)-style view of *should*, the globally best worlds are ones where all miners live, i.e., the set of worlds where either i) the miners are in Shaft A and we block Shaft A, or ii) the miners are in Shaft B and we block Shaft B. But this wrongly predicts the falsity of (40).

Lassiter convincingly argues that some scalar notion like probability needs to be factored in to the meaning of *should*. He additionally argues that this requires that the meaning of *should* be descalar in my terminology, i.e., that its meaning be derived from a measure function in the way given in (43),

where E is a function which gives the “expected utility” of a proposition, or, its average goodness weighted against the likelihood of each possible outcome.

$$(43) \quad \llbracket \text{should} \rrbracket_{\text{Lassiter}}(\phi) = 1 \text{ iff } E(\phi) > s$$

The first point to make about this denotation is that it gives a relative semantics rather than a superlative one. This seems like a wrong move. After all, (44) seems quite bad.

- (44) (If you want to get to Hyde Park,) you should take the Red Line, then the 55. #But taking the 2 is an even better idea.

Lassiter’s *should* has a semantics like *tall*, which clearly allows for discourses like (44).

- (45) John is tall. But Bill is even taller.

Kratzer (1981a) made the original observation that a superlative semantics seems best for *should*, supported by the fact that (46b) is a very good paraphrase of (46a).

- (46) a. You should take the 2.
b. Taking the 2 is your best option.

Lassiter could alter the way his standard is determined such that the semantics he gives guarantees the superlative meaning as well, but this seems to draw an unnecessary parallel to positive relative adjectives, where the more obvious one is with superlative adjectives.

More crucially, Lassiter’s bigger point, that a descender denotation for *should* is required in light of its sensitivity to probability, is an overreach. Nothing about factoring in probability requires a descender view. While it is clear that *should* involves a scalar component, this is nothing new; there was already a scalar component in (45), as there was in the original proposal by Kratzer (1981a). There is nothing inconsistent about having a quantificational semantics for a modal which also incorporates a scalar component, and Kratzer (1981a) lays out just how to do this: by using this scalar component to restrict the domain of the modal.

What Lassiter’s objection makes clear, however, is that we must be a bit more sophisticated about how to do this. Individual worlds cannot be ranked by their probability-weighted-goodness since probability must be calculated

over sets of worlds; instead we must rank outcomes rather than worlds; i.e., cells in a partition of the modal domain according to what outcome is achieved in each.

Cariani et al. (2012) move toward this as well. They define a ‘decision problem’ parameter, δ , which when combined with the modal domain, yields a partition of the modal domain in terms of what actions the relevant agent can take.⁹ Their denotation of *should* involves a typical-looking ordering on worlds, but in fact what it orders are the partition cells which contain that world.

Their basic picture is outlined below, where c is a context, which determines (at least) a modal base f , an ordering source g , and a partition function δ .

- (47) Cariani et al.’s (2012) deliberative modality
- a. $\llbracket \text{should} \rrbracket = \lambda p \lambda \langle c, w \rangle [\forall v \in O(c, w) [p(c, v)]]$
 - b. $O(c, w) = \{u \in \bigcap f_c(w) : \neg \exists v \in \bigcup f_c(w) [v <_{c,w} u]\}$
 - c. $u \leq_{c,w} v$ iff $\blacktriangle_{c,w}(v) \subseteq \blacktriangle_{c,w}(u)$
 - d. $\blacktriangle_{c,w}(v) = \{p \in g_c(w) : [v]_{c,w} \subseteq p\}$
 - e. $[v]_{c,w}$ = the set of worlds in the same δ_c -cell as v

In a more traditional view of *should* (or any modal involving an ordering source), (47b) would invoke the ordering $g_c(w)$ directly. Instead, however, worlds are ordered by the relative ordering of their associated partitions. So while (47b) makes it look like worlds are being ordered directly, (47d) shows that what is really being ordered are the partitions that contain those worlds. Here they are being ordered by how many propositions in the ordering source they entail, but Cariani et al. (2012) note that the subset relation in (47d) could be replaced with some other relation. This is an important point: Cariani et al. (2012) put forward their ordering as an ordering over worlds, but really it is an ordering over partitions, i.e., sets of worlds. Thus by my terminology it can be thought of as a modal itself.

The intuition is that *should* partitions the modal domain in terms of what a given agent can do, then ranks the cells of those partitions by how well each cell conforms to the agent’s goals; *should* ϕ is true iff ϕ is true in all the worlds of the highest ranked cell(s). Below I sketch another version of

⁹In fact Cariani et al.’s (2012) δ is not strictly a partition because it allows that some of the worlds in the modal domain are not sorted, i.e., doing nothing is an option but not part of the decision problem. For my purposes I will just assume it is a partition.

Cariani et al.’s (2012) account, which is the same in the important details, but differs in it that overtly invokes a modal ordering, called **fit**.

The first ingredient that *should* takes is a metaphysical modal base, giving the set of worlds compatible with what is true up to now, or worlds that branch from now, a la Cariani et al. (2012).

The next ingredient is a partition over the worlds of the modal base based on what a contextually given individual’s possible actions are (presumably given some contextual restriction to a small set of alternative actions) at branching time (the time argument below is suppressed for simplicity). To avoid confusion with Cariani et al. (2012) slightly distinct δ , I call this Γ .

- (48) $\Gamma_x(\bigcap m(w))$ is a partition over $\bigcap m(w)$...
- a. $\forall p, q \in \Gamma_x(\bigcap m(w)) [p \cap q = \emptyset]$
 - b. $\bigcup \Gamma_x(\bigcap m(w)) = \bigcap m(w)$
- (49) ...according to what actions x can take.
- $$\Gamma_x(\bigcap m(w)) \subseteq \{f(x) \subset \bigcap m(w) : \forall v[f(x)(v) \rightarrow \exists e[\mathbf{Ag}(e, v) = x]]\}$$

The third ingredient is a measure function, the crucial component to deriving scalarity. Call this measure function **fit**. This function, also anchored to the subject, ranks domain-prejacent pairs according to how well the prejacent relative to the domain fits the goals of the subject, denoted by $g(x)(w)$. Following von Fintel and Iatridou (2008) and Rubinstein (2012) g could be a complex set of ordering sources, which represent a potentially open-ended number of goals/priorities/preferences. I will leave open the question of how exactly goodness of fit is calculated given a set of priorities or goals – see Kratzer (1981a), Kratzer (2012), von Fintel and Iatridou (2008), Rubinstein (2012), and Katz et al. (2012). But crucially one component that could go into **fit** is some notion of probability. For example, **fit**($\bigcap m(w), \phi$) could partition ϕ in terms of an issue which is epistemically unsettled (e.g., are the miners in shaft A or shaft B?) then assign to each partition a value weighted against the probability of that partition in $\bigcap m(w)$, then return their sum. This has the effect of giving a weighted goodness of fit, something like Lassiter’s expected utility. This is only possible if **fit** ranks *pairs* of a prejacent and its modal domain. If it had access only to the prejacent (or only to a world), it could not calculate probability given the backgrounded information in $m(w)$.

The fourth ingredient is a superlative operator, Δ , which takes a gradable property G and a comparison class C and returns the set of objects in C such

that no other objects outrank them in terms of G . Note crucially that this version of the superlative returns a *set* – the set of objects in C which are highest ranked by G . In many cases this could be a singleton set, but crucially this allows for ties.

$$(50) \quad \Delta(G)(C) = \{x \in C : \neg \exists y \in C [G(y) > G(x)]\}$$

The idea here is to apply this to $\mathbf{fit}_{g(x)(w)}$ to get the best cell(s) in the partition created by $\Gamma_x(\bigcap m(w))$. However, a complicating factor is the fact that \mathbf{fit} orders pairs of a modal domain and a prejacent, not just prejacentes. This makes the math a bit messy but the intuition is still preserved.

The comparison class for Δ should be a set of ordered pairs of a domain and a prejacent, but for our purposes we want every pair to have the same domain, say $\bigcap m(w)$, while the prejacent varies over cells in $\Gamma_x(\bigcap m(w))$. Thus the comparison class is in (51), which I will abbreviate with Γ^* .

$$(51) \quad \Gamma_x^*(\bigcap m(w)) := \{\langle \bigcap m(w), p \rangle : p \in \Gamma_x(\bigcap m(w))\}$$

So we have a set of domain-prejacent pairs, but the domain for the purposes of the larger modal should just be the prejacent of the \mathbf{fit} function, i.e., the best cell in the partition. So this must be extracted, giving the domain for teleological *should*.

$$(52) \quad \mathbf{best-fit}(g, x, m, \Gamma, w) = \bigcup \{p : \langle \bigcap m(w), p \rangle \in \Delta(\mathbf{fit}_{g(x)(w)})(\Gamma_x^*(\bigcap m(w)))\}$$

(52) denotes the set of worlds in one of the partition cells p such that the pair $\langle \bigcap m(w), p \rangle$ is ranked best among the comparison class $\Gamma_x^*(\bigcap m(w))$ by the scale $\mathbf{fit}_{g(x)(w)}$.

This gives us the set of worlds where a given individual takes the best course of action (of which there may be more than one). This is the domain for *should*.

$$(53) \quad \llbracket \text{should} \rrbracket = \lambda P \lambda m \lambda g \lambda x \lambda w [\forall v \in \mathbf{best-fit}(g, x, m, \Gamma, w) [v \in P(x)]] \\ \text{iff } m \text{ is metaphysical and } g \text{ is a teleological ordering source}$$

This says that *should* takes a property and an individual (which put together form the prejacent), a metaphysical modal base m , a (possibly complex) goal-based ordering source g , and gives back true iff all the worlds in the union of the best partition-cells of $m(w)$ are worlds at which the prejacent is true.

The biggest difference with [Cariani et al. \(2012\)](#), and the biggest departure from [Kratzer \(1981a\)](#), is the use of **fit**, which is, by my definition, itself a modal. Whereas Kratzer simply restricts the domain of *should* with an ordering source, I restrict it with a full-fledged gradable modal operator, which itself depends on an ordering source. In this way, the proposal is similar to Lassiter’s in that it is a categorical modal which is derived from a gradable one.

Here, the similarities to Lassiter are greater than the differences. The biggest contrast is not in terms of whether the modal is ultimately derived from a scalar notion – on that we agree (and in fact, so does [Kratzer \(1981a\)](#) and all the work which follows from it). Rather, the biggest contrast is that my proposal incorporates a modal domain, and is thus, by my definition, a modal. If there were a way to re-write my semantics which did not invoke the classical logical \forall , it would still be a modal on my view as long as it related a modal domain and a prejacent.

3.3.4 Deriving Entailment Relations

The route now to generating the entailment relation between *important* and *should* should be clear: by defining the measure function denoted by *important* in terms of the **fit** scale. It may be tempting to simply let *important* denote **fit**, but a few key empirical facts militate against this.

The first is that actions which do not advance the relevant goals cannot be used with *important*.

- (54) #If you want to get a job, it’s (not) important that you make a fool of yourself.
- (55) #If you want to get a job, it’s more important that you give a good impression in your interview than that you submit an incoherent statement.

Therefore I assume that *important* presupposes that its prejacent is ordered higher than the ubiquitous option of doing nothing. Let $\Lambda_g(\Gamma_x(\bigcap m(w)))$ denote the unique cell in Γ which involves x doing nothing relevant to their implicit goals.

The second is that importance can be weighed among two options which are *not* mutually exclusive, but comparison gives rise to the inference that if they *were*, one would be a better course of action than the other.

- (56) a. You should mow the lawn and water the flowers. But you might not have time to do both. So if you can only do one, you should water the flowers. After all, they could die in this heat.
 b. It's more important that you water the flowers than that you mow the lawn.

Therefore I argue that *important* takes into consideration the same deliberative partition that *should* does, but orders only partitions in which exactly one salient course of action is taken. In the above case, assume that Γ partitions the modal domain into four partitions: the set of worlds where you mow the lawn but do not water the flowers, the set of worlds where you water the flowers but do not mow the lawn, the set of worlds where you do both, and the set of worlds where you do neither. Assume that the implicit goals include making the yard look as good as possible, but also include preventing irreparable damage to the lawn (i.e., letting the flowers die is worse than letting the lawn go un-mowed for a day).

What it means to say that (56b) is that the partition-cell where you *just* water the flowers is better than the partition-cell where you *just* mow the lawn. To account for this I introduce the exclusivity operator ∇ .

$$(57) \quad \nabla(X)(Y) = \{Z \subseteq X : Z \in Y \text{ \& no alternatives to } X \text{ overlap with } Z\}$$

If $\phi = \textit{you water the flowers}$, the partition-cell in (56b) in which you just water the flowers is therefore $\nabla(\phi)(\Gamma_x(\bigcap m(w)))$. The denotation for *important* is therefore as in (58).

$$(58) \quad \llbracket \textit{important} \rrbracket(g)(m)(p)(w) \text{ is defined iff } \mathbf{fit}_{g(x)(w)}(\nabla(p)(\Gamma_x(\bigcap m(w))))(\bigcap m(w)) > \mathbf{fit}_{g(x)(w)}(\Lambda_g(\Gamma_x(\bigcap m(w))))(\bigcap m(w))$$

$$(59) \quad \llbracket \textit{important} \rrbracket = \lambda g \lambda m \lambda p \lambda w [\mathbf{fit}_{g(x)(w)}(\nabla(p)(\Gamma_x(\bigcap m(w))))(\bigcap m(w))]$$

Consider once again the entailment relation at issue.

- (60) a. It's more important that you water the flowers than that you mow the lawn.
 b. \rightarrow If you can only do one, you should water the flowers.

I assume that the clause *given the option between the two* has the effect of narrowing the modal base of *should* to include only worlds where the addressee waters the flowers, mows the lawn, or neither, but not both. Assume that that modal domain for (60a) is the set M , that $\llbracket \textit{water the flowers} \rrbracket = \phi$,

that $\llbracket \text{mow the lawn} \rrbracket = \psi$. The translation of (60a) in a given context could then be (61).

$$\begin{aligned}
 (61) \quad & \text{a. } \llbracket (60a) \rrbracket = \mathbf{fit}(\nabla(\phi)(\phi \& \psi \& M, \phi \& \neg \psi \& M, \neg \phi \& \psi \& M, \neg \phi \& \neg \psi \& M)(M)) > \\
 & \quad \mathbf{fit}(\nabla(\psi)(\phi \& \psi \& M, \phi \& \neg \psi \& M, \neg \phi \& \psi \& M, \neg \phi \& \neg \psi \& M)(M)) \\
 & \text{b. } = \mathbf{fit}(\phi \& \neg \psi \& M)(M) > \mathbf{fit}(\neg \phi \& \psi \& M)(M)
 \end{aligned}$$

Since importance of just- ϕ is presupposed to be greater than that of doing nothing, of the options of doing just one, the other, or neither, it follows from (61) that just- ϕ is the best cell. Since the conditional antecedent in (60b) restricts the modal domain to exclude the possibility of doing both, it then follows that (60b) must be true as well. The entailment relation is thus derived.

What this shows is another route for entailment relations between a categorical and a gradable modal. In this case, we do not rely on additivity/scalar conservativity, but rather, a common scalar core that underlies both modals. This is in spirit very close to the original proposal of Kratzer (1981a) which argued for scalarity derived from an ordering source which could underly both modal auxiliaries and modal adjectives like *likely*. How this proposal differs is that it invokes an ordering not over worlds but over partitions of the modal domain, i.e., it has an ordering which is itself modal. This is in response to work by Lassiter (2011) and Cariani et al. (2012) who show that deliberative modality requires the consideration of whole outcomes globally rather than individual worlds.

3.4 Comparisons

This chapter proposes a measure function analysis of gradable modals like *likely* and *important*, and a categorical quantificational account of non-gradable modals like *must* and *should*. These accounts have the following crucial features: First, they invoked measure functions and thus capture compatibility with general-purpose degree modifiers like *more*, *too*, *so*, etc. Second, they are compatible with the appropriate scale structures; e.g., *likely* is not limited to a Kalmogorovian notion of probability, which would entail a closed scale and wrongly predict compatibility with maximizers and proportional modifiers. Third, they invoke modality on my own narrow definition, and therefore compatibility with *if*-clauses on a Kratzerian restrictor analysis. Fourth, they are sufficiently specified to capture entailment relations with

modal auxiliaries. The two accounts given here suggest two ways of capturing these entailments: With additive, conservative measures as in the case of *likely*, or by building a (potentially non-additive) measure into the domain restriction of the relevant categorical modal, as in the case of *important* and *should*.

Still, there is much to be said about both *likely* and *important*; the measure functions *lhood* and *fit* are left highly underspecified. In the case of *likely*, many more specific accounts have been given, which are compatible with the broad notion of a ‘measure function analysis’. However, these accounts all have shortcomings which make them inconsistent with the broad outlines established here.

3.4.1 Comparative Possibility

Accounts like Kratzer (2012) or Katz et al. (2012) seek to derive gradient modal notions from fundamentally categorical notions of consistency, i.e., through the use of ordering sources. We can call these **comparative possibility** accounts, and contrast them with **mathematical probability** accounts, like Yalcin (2007, 2010) and Lassiter (2011). Katz et al. (2012), for example, propose an algorithm for combining multiple ordering sources asymmetrically, so that one ordering source is given priority over another. They then give the analysis for comparative possibility in (61). Here $\leq_g(w)$ is the world ordering provided by Kratzer (1981a) and discussed in Chapter 1, while $q \uparrow_{g(w)} p$ refers to the set of q -worlds which are not ranked equal to any p -worlds by $\leq_{g(w)}$.

- (62) A proposition p is **at least as good a possibility** as a proposition q in w with respect to f and g iff there is no accessible world in q which is both (i) not equivalent [with respect \leq_g] to any world in p and (ii) more ideal than every accessible world in p which is not equivalent to any world in q :
- $$\neg \exists v \in \bigcap f(w) \cap (q \uparrow_{g(w)} p) [\forall z \in \bigcap f(w) \cap (p \uparrow_{g(w)} q) [v \leq_{g(w)} z]]$$
- (Katz et al., 2012, p. 500)

This captures all the right facts on the modal side of things – entailment relations with modal auxiliaries, and compatibility with the restrictor view of conditionals. And while this proposal, like Kratzer’s (1981a) original, does not fully explain how the theory can be related to a degree semantics, it does

lay the foundation for such a semantics, since it is stated in terms of an ordering.

The proposal runs into trouble with scalar structure, however. Since there is not obviously an inherent minimum or maximum in to the scale provided, until (62) is fully integrated into a degree-based theory, it is unclear what the predictions about scale structure really are. But more importantly, whatever the scalar structural properties are, it seems like they will not vary from case to case. [Katz et al. \(2012\)](#) propose (62) as a general schema for all gradable modals – this presumably includes upper-closed gradable modals like *certain* as well as open ones like *likely*. This is a significant concern – thus while their proposal may work for the gradable modal adjective they focus most on (*good*), and perhaps also (*likely*) it does not seem to be appropriate for a general schematic – the landscape of gradable modality is more diverse than a single unified gradient modal semantics may suggest.

3.4.2 Mathematical Probability

Mathematical probability accounts, on the other hands, very naturally fit into a degree-based approach. The approaches of [Yalcin \(2007\)](#) and [Lassiter \(2011\)](#) both have shortcomings relating to scalar structure and conditionals, however. Either could be modified to lexically presuppose that the scale does not map to the endpoints; this would maintain the Kolmogorovian core but derive the degree modification data. But neither account is relational, incorporating a modal domain. This is crucial because the presence of a modal domain is vital to a Kratzerian account of conditionals.

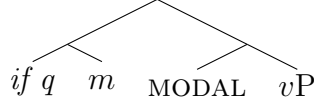
Any theory of gradable modals must account for their behavior in the conditionals, like in (63).

(63) If Drew picks black, he is likely to win.

On my analysis, which takes conditional probability (likelihood) as basic, this can be accounted for within the Restrictor Theory of Conditionals. Crucial to this fact is the presence of the modal domain in my analysis, which is targeted by *if*. Assuming the modal domain is present as a silent indexical in the tree, the denotation for *if* in (64) suffices, where m ranges over accessibility relations.

(64) $\llbracket \text{if} \rrbracket = \lambda q \lambda m \lambda w [m(w) \cap q]$

(65)



if as defined above combines with its complement, then with the modal-base-denoting variable in the syntax, intersecting the modal base with the antecedent clause,¹⁰ giving a conditional reading. Both Yalcin (2010) and Lassiter (2011) attempt to derive conditional probability in similar ways; I will address each proposal in turn.

Yalcin (2010) argues that Kratzer's (1986) Restrictor Theory must be amended slightly to account for *likely*. He argues that the evaluation function, for all expressions, is indexed to a modal domain and to a probability measure (not unlike the probability space of Yalcin (2007)). The function of *if* is to restrict this modal domain in the usual Kratzerian way, but to also conditionalize the probability measure on the truth of the antecedent.

The issue with this approach is that it applies the solution to *conditional probability* to *if* itself; this is a case of overfitting, since this conditional probability measure is vacuous for all expressions except *likely* and perhaps a few others which involve probability, as Yalcin (2010) himself points out. Given the approach presented here, of conditional probability as basic, this second function for *if* is not needed. The proposal here allows for a totally painless analysis of conditionals under the Restrictor Theory. If q is true in all $M \cap p$ -worlds, i.e., if *if p then q in w* is true, then by Scalar Conservativity, $lhoud(M \cap p)(q)$ will have as high a value as $lhoud(M \cap p)(q')$ for any other q' – the desired result. The probability measure itself need not be conditionalized; rather, probability is always relative to a domain.

As for the particular variation suggested above, based on Yalcin's (2007) model, restricting the modal domain via the conditional means that Pr will assign probabilities to the partitions of the *restricted modal domain* in such a way that the sum of these partitions' probabilities is 1, establishing the upper bound as desired.

The contrast with Yalcin (2010) is a subtle one, but it is important to show that conditionals work the same for gradable and nongradable modals, rather than stipulating an additional effect for *likely*; this is due to the role played by the modal domain. In doing so it can be shown that gradable and nongradable modals, despite their differences, share the basic properties of

¹⁰This simplifies the picture somewhat, by excluding discussion of the ordering source and its potential effect on which antecedent worlds are included in the modal domain.

modals established in Kratzer’s work.

Lassiter (2011) does not invoke modal domains at all, although he does claim to adopt a Kratzer-style conditional semantics. For Lassiter, the “domain” that is restricted by the conditional is not a modal domain in the sense used here, but rather, the domain *of the scale itself*. The scale orders propositions, but for the purpose of discussing conditionals, Lassiter considers the scale to order worlds instead. This sort of scale is isomorphic to the order on propositions as long as plural worlds are included, i.e., the concatenation of worlds, which are isomorphic to sets of worlds. On Lassiter’s view, the *if*-clause then removes from the ordering any worlds which the antecedent is false; moreover, any plural worlds *containing* such a world are removed as well. Otherwise the ordering remains the same.

Of course, on my view, this fails to account for conditionals involving non-de-scalar modals. But of course Lassiter (2011) argues for a universally de-scalar account of modality. But even if we were to concede this claim, Lassiter still fails to make the right predictions about comparative conditionals.

On Lassiter’s approach the conditional serves as a function from orders to orders. It acts first on an expression like *likely*, and maps it to a new measure function. Then this new measure function combines with the prejacent just as an unconditionalized measure function would do. By contrast, on my account, there is a single measure function, *hood*, which orders domain-prejacent pairs. Thus on my account, pairs with distinct domains are still ordered, whereas on Lassiter’s they are not. Lassiter’s account thus fails to predict the acceptability of (66).

- (66) High school students are more likely to retake the SAT if they score just below a round number, such as 1290, than if they score just above it.¹¹

Conditionals do not play a major role in Lassiter’s (2011) theory, but any theory of modality must account for them.

Turning finally to entailment relations, it is worth pointing out that Lassiter’s analysis of *likely*, actually already derives the desired entailment relations even when modal auxiliaries are given a classical analysis. Lassiter provides two constraints for the measure function *Pr* which he takes *likely* to denote.

¹¹<http://www.psychologicalscience.org/index.php/news/releases/students-are-more-likely-to-retake-the-sat-if-their-score-ends-with-90.html>

- (67) a. *Necessity Entails Maximality*
 $Pr(W) = 1$
b. *Additivity*
If $\phi \cap \psi = \emptyset$, then $Pr(\phi \vee \psi) = Pr(\phi) + Pr(\psi)$

Note that the first of these properties, which is included in most definitions of probability, actually directly stipulates the first entailment relation, namely that necessity entails a high probability (in fact, the maximum probability), guaranteeing the entailment of likelihood by necessity.

The second entailment relation is derived by Additivity, similar to the way it is derived for *hood*. Additivity says that the probabilities of two disjoint propositions add up to the probability of their disjunction. This means that if any proposition entails another, (i.e., is a superset of the other) then its probability must be at least as great as the entailed proposition.

- (68) a. $Pr(\psi) = n$ (Premise)
b. $\psi \subset \phi$ (Premise)
c. $\phi = (\phi - \psi) \cup (\phi \cap \psi)$ (Set Theory)
d. $\phi \cap \psi = \psi$ (Set Theory)
e. $Pr(\phi) = Pr(\phi - \psi) + Pr(\psi)$ (Additivity)
f. $Pr(\phi) \geq Pr(\psi)$

In other words, bigger sets mean bigger probabilities. Naturally, the smallest possible proposition, the empty set (i.e., a contradiction), has the smallest probability.

- (69) a. $Pr(\phi) = n$ (Premise)
b. $\phi = \phi \cup \emptyset$ (Set Theory)
c. $Pr(\phi) = Pr(\phi) + Pr(\emptyset)$ (Additivity)
d. $Pr(\emptyset) = 0$

Since Pr always assigns 0 to impossible propositions, any proposition which is assigned a non-zero number must not be impossible. Thus, Likelihood Entails Possibility is derived.

This still leaves the entailment Lassiter himself points out, Likelihood of Necessity Entails Necessity. In fact, it is this entailment that Lassiter holds up as the reason for pursuing a de-scalar account of modal auxiliaries, by showing that Kratzer's (1991) account of epistemic *must*, even in conjunction with his own semantics for *likely*, fails to derive the relation. However, this

is predicated on assuming a Kratzer-style weak necessity account of *must*, as opposed to a strong necessity account, like the one put forward by von Fintel and Gillies (2010).¹² Given a strong necessity account of *must*, we only need add Possibility Entails Some Likelihood as a constraint to Lassiter’s *Pr* to derive the desired entailment.

Thus, even on Lassiter’s own account, there is no reason to extend a scalar analysis to modal auxiliaries like *must*.¹³

3.5 Conclusion

This chapter provides an outline of the basic properties gradable modal expressions like *likely* or *important* must have to predict their basic distribution and to explain what kind of primitive notions we must appeal to explain their meaning (as well as the meanings of modal auxiliaries). The biggest point here is that gradable modal expressions must be given a degree semantics, but must also maintain a relational semantics in order to properly account for phenomena like conditionals. In the case of epistemic gradable modals, we must only add a few constraints on what kind of ordering *likely* denotes in order to capture entailment relations; otherwise there is not inconsistency between having gradable and non-gradable epistemic modals. In the case of teleological modals, the connection must run deeper – we must argue (as it already has been argued) that non-gradable teleological modals themselves embed a modal ordering which is common to the adjectival orderings. This derives entailment relations in that domain.

Ultimately, the core of what is it to be a modal is this: Relating a domain to a prejacent. The issue of gradability and scalarity does not threaten that

¹²Further support for this account of *must* can be found in Chapter 4, where I argue that the weakening effect often attributed to stereotypical ordering sources in modal auxiliaries can instead be attributed to the general pragmatic phenomenon of imprecision.

¹³Holliday and Icard (2013) put forward a qualitative approach, which differs from the quantitative approach in that it is not additive, but instead is qualitatively additive. Qualitative additivity is given below for a measure function μ is given below.

- (i) *Qualitative Additivity.* A measure μ is qualitatively additive iff for any w, p, q
 $\mu_w(p) \geq \mu_w(q)$ iff $\mu_w(p - q) \geq \mu_w(q - p)$

Removing additivity as a constraint potentially undermines the derivation of entailment relations I provide above. However, as Holliday and Icard (2013) show, the relevant entailments are derived independently from other properties of qualitatively additive orderings.

notion.

Chapter 4

Imprecision and Modality

The primary concern of this dissertation is with the varying strength of modals. In Chapter 2 and Chapter 3, I examined how this may occur with gradable modals, which denote measure functions and may map to any of a range of degrees on a scale. One possibility is that different degree modifiers may pick out different points on the modal scale.

- (1) That outcome is {extremely likely > very likely > likely}.

Another possibility is that the positive form (or another such context dependent degree modifier) may itself vary from context to context.

- (2) It's { pos_{hi} likely > pos_{mid} likely > pos_{low} likely } he'll nominate a Republican.

These both represent *semantic* strength variation since what varies is semantic. In (1) what varies is the choice of degree modifier; in (2) what varies is the valuation of the semantically determined contextual variable, which is a part of the meaning of the lexical item *pos*. While pragmatics may affect how the anaphor in *pos* is resolved, ultimately its presence is semantically mandated, and its value is fixed by the context of utterance.

This chapter focusses instead on non-gradable modals and how these may have varying strength depending upon the pragmatic context. The purpose of this chapter is to propose and explore a different variety of scalarity in modality, namely imprecision, which I take to be a purely pragmatic phenomenon. I provide a theory of imprecision following from the work of Lewis (1979), Lasersohn (1999), Krifka (2007) and Lauer (2012) and show that

treating modality with this pragmatic theory allows for a simplification of the semantics and broader empirical coverage.

It has never been proposed that modals may be subject to imprecision. In this chapter I argue not only that a general theory of imprecision such as the one provided here predicts that modals should be subject to imprecision, but in fact that a theory of imprecise modality is essential to understanding certain modal phenomena as well as more generally to reconciling a possible world semantics for modality with the real constraints on humans' ability to consider an infinitude of possible worlds.

I focus on two empirical phenomena which I argue can really be treated as one. In both of these cases we see that modals may vary in their strength contextually. Thus, in one context, a modal like *have to* may be stronger than in another context, but with all the anaphoric properties of the modal remaining constant (i.e., the same modal base and ordering source). The first phenomenon is the case of modal domain shifting as observed by [Lewis \(1979\)](#) in cases like (3); for expository ease I will refer to sequences like these as “Lewis sequences” (LSs).

- (3) a. *Yaron*: This must be a pen; I'm looking right at it.
- b. *Itamar*: Not so, you could be the victim of deceiving demon.

The second phenomenon is exemplified by (4), also called a Sobel Sequence ([Sobel, 1970](#), [Lewis, 1973](#)).

- (4) a. If Karlos had come to the party, it would have been a good time.
- b. If Karlos had come to the party and jumped off the balcony, it would not have been a good time.

In both cases I argue that the first sentence is felicitous if there is a low standard of precision at work in the discourse, i.e., speakers are “speaking loosely”. This allows them to ignore certain possibilities if those possibilities are especially unlikely or unexpected. This leads to necessity statements, including counterfactual conditionals like (4), having weaker force than they semantically ought to. What Lewis and Sobel Sequences show is that this standard of precision can be raised through various conversational means, leading to a strengthening of the same necessity/conditional claims.

In Section 1 I discuss Lewis Sequences in more detail and argue they cannot be accounted for by appeal to ‘shifting’. In Section 2 I address Sobel Sequences and make the empirical argument that there actually two varieties

of Sobel Sequences; I show that one of these varieties can be accounted for with a conservative semantics, while the other behaves much like Lewis Sequences, and with them, deserve a pragmatic treatment. In Section 3 I lay out a theory of imprecision and precisification. In Section 4 I argue explicitly that modals are subject to imprecision, and that treating Lewis and Sobel Sequences as cases of precisification provides a dramatic improvement over prior accounts. In Section 5 I conclude.

4.1 Contextual Variability in Modals

A teleological or deliberative necessity modal like *have to* may vary in strength from context to context. Consider two similar contexts: one is a science olympiad where teams compete to solve engineering problems, such as getting a ball from one table to another using a given set of materials; the other is a Rube Goldberg device olympiad, which has the same rules but encourages participants to solve their problems in creative, roundabout ways.

- (5)
- a. *At a science olympiad...*
In order to get the ball across this gap, we have to lay down a bridge. = TRUE
 - b. *At a Rube Goldberg device olympiad...*
In order to get the ball across this gap, we have to lay down a bridge. = FALSE

Since building a complicated Rube Goldberg device is a *live option* (in the sense of [Willer 2013a; 2013b](#)) in the context of (5b) but not (5a), building a bridge is the only (live) option in the context of (5a) but not (5b).

What this shows is that the same sentence containing *have to* may vary in truth in these two different contexts, even when those contexts are identical in terms of the construction materials available and the wherewithal of the participants. What does vary here is more subtle: In the latter context, we are more willing to entertain more outlandish possibilities than in the first context.

It may be argued that what varies in (5a-b) is anaphoric; i.e., (5a-b) vary in terms of how their semantically context dependent content is valued. Particularly, they may vary either in their modal base or ordering source. I assume both cases share the same circumstantial modal base, since we have stipulated that the material circumstances are the same between the two

conditions.

But perhaps the ordering source is not the same in the two cases. In both examples we are dealing with a teleological ordering source, i.e., an ordering determined by the goals of a relevant individual. We could imagine different goals leading to different truth values; here, perhaps, the utterer of (5a) has a goal of efficiency or simplicity which the utterer of (5b) does not (or perhaps explicitly has the goal of eschewing efficiency). As discussed in Chapter 1, differing modal bases or ordering sources may give rise to differing modal domains, and thus, different strengths.

However, consider a possible response to (5a).

- (6) No, we don't *have* to lay down a bridge. We *could* build a complicated Rube Goldberg device to get the ball across the gap.

There may be some specification of the context in (5a) in which this sentence is actually false; e.g., projects are graded in part on efficiency, or there isn't enough time or resources to construct such a device. But crucially there are some specifications in which it is true, e.g., one in which there is in fact enough time to construct such a device and projects are graded only on successful completion of the task. Moreover, in exactly those contexts, (5a) is still judged true, at least until (6) is uttered.

It is the variation in strength of *have to* and other modals *across* different discourses, i.e., (5a) versus (5b), that is central to this chapter, but this point is perhaps easier to see when variation in strength occurs *within* a single discourse. This can be seen in cases pointed out by Lewis like (3) or (7).

- (7) *Bryan is helping Alice move out of her apartment. They have gotten nearly everything out and into the moving van parked on the street below. The only things remaining in the apartment, besides some dust on the floors, Bryan and Alice themselves, and their clothing and personal affects, is the contents of a single closet: A raincoat, a thick roll of duct tape, and a hammer and a few nails. It is raining outside.*
- a. *Alice:* I want to go outside, but I don't want to get wet.
 - b. *Bryan:* You have to wear the raincoat.
 - c. *Alice:* No, I don't have to. I could cover every inch of my skin in duct tape.
 - d. *Bryan:* I suppose that's technically true, but you get my drift.

The modal of primary concern is *have to* in (7b). (7a) establishes the context, particularly which goals the teleological ordering source includes. In (7b), Bryan utters a teleological necessity claim which at first seems true. In (7c), Alice rejects this claim and makes a contradictory teleological possibility claim which also seems true and thus falsifies (7b). In (7d), Bryan concedes the truth of (7c). Here, (7b-c) correspond to (3).

When initially uttered in (7b), *have to* is much weaker and seems only to capture less outlandish possibilities. But after Alice’s rebuttal in (7c), it seems stronger, capturing more outlandish possibilities.

Since this discourse is fundamentally a kind of disagreement, it is a bit odd to characterize sentences as true or false objectively; another way to look at it would be to say that (7b) has an indeterminate truth value when first uttered, but that (7c) establishes its truth value as false, at least in Alice’s eyes. However, we could just as easily imagine Alice *not* uttering (7c), and accepting (7b) as true even if her knowledge of the circumstances (and thus the viability of the duct tape strategy) was exactly the same.

- (8) *Bryan is helping Alice move out of her apartment. They have gotten nearly everything out and into the moving van parked on the street below. The only things remaining in the apartment, besides some dust on the floors, Bryan and Alice themselves, and their clothing and personal affects, is the contents of a single closet: A raincoat, a thick roll of duct tape, and a hammer and a few nails. It is raining outside.*
- a. *Alice:* I want to go outside, but I don’t want to get wet.
 - b. *Bryan:* You have to wear the raincoat.
 - c. *Alice:* Indeed.

In other words she could felicitously ignore the duct tape option, or not, and whether or not she does determines the truth value (i.e., the strength) of (7b); in both (8) and (7) both participants end up agreeing.

It is this fact which falsifies the analysis sketched above, wherein an anaphoric component of the modal, say the ordering source, is what is varying. Any theory of anaphora states that the value of an anaphoric component is fixed by context at utterance and is not negotiable; e.g., see Klecha (2011). If Bryan utters *have to* he *means* a particular ordering source and modal base which then fixes the meaning of *have to* forever. Its value should not change retrospectively. As with any case of anaphoric resolution, Alice may misunderstand Bryan’s meaning. But if that were the case, Bryan would not be

required to concede as in (7d) – in fact it would be very odd if he did. Simple examples can show this:

- (9) a. *Ed*: (I just saw him go in there, so) Graham has to be in the kitchen.
b. *Freddy*: #No, he doesn't. He's allowed to be anywhere he wants.
c. *Ed*: #I suppose you're technically right, but you get my drift.

(9b) is infelicitous, but –and this is an important point– the infelicity of one sentence within a particular discourse does not make that discourse infelicitous on a whole. There are more and less felicitous ways to respond to infelicitous discourse actions. Many native speakers may not make a terribly fine distinction between (7c) and (9b), although I argue that there is one. But what should be very clear is the distinction between (7d) and (9d).

Here Ed means a particular modal base/ordering source (epistemic), but Freddy simply wrongly interprets him as having meant a circumstantial/deontic one, as indicated by Freddy's contrastive use of the modal *allowed to*, which can only have a deontic interpretation. This is bad but not unimaginable – Freddy might have been not quite paying attention to what Ed said, or misunderstood some aspect of the context. But Ed surely cannot make any concession to Freddy unless he himself forgot what he said.

Another example can illustrate this just for variation in ordering source and not modal base.

- (10) a. *Helena*: (In order to keep the bomb from exploding) the bus has to go exactly 60 miles per hour.
b. *Gallagher*: #That's not true! The bus is allowed to go as slow as 45 miles per hour.
c. *Helena*: #I suppose you're technically right, but you get my drift.

Here the modal base is constant, a circumstantial one. But while Helena means a teleological ordering (the salient goal being the prevention of the explosion of the bus) Gallagher interprets it as a deontic one (subject only to the laws governing that road). Helena surely cannot felicitously concede as in (10c).

What (9) and (10) have in common is that discourse participant (b) acts as if a different ordering source was used than what was intended by the speaker. What these examples are intended to show is that this results in

a strong enough infelicity to crash the discourse; in other words, the first speaker in each case is entitled to reject the response of the second speaker, and in fact cannot felicitously do otherwise. These are all simply cases of two individuals disagreeing on how to resolve an anaphor, which is to say, they disagree on what was *said*. A non-anaphoric example of this might be (11).

- (11) a. *Bryan*: You have to wear a raincoat.
 b. *Alice*: #How would a condom help me in this situation?
 c. *Bryan*: #You're technically right, it wouldn't. But you get my drift.

What this shows is that anaphor resolution is not subject to retrospective changes; no aspect of semantic content is, except perhaps on assessment accounts (e.g., MacFarlane's (2008) approach to predictive expressions). This means that when Alice makes the domain of Bryan's *have to* wider, it cannot be because she forces a shift to a different ordering source. Ordering sources are determined anaphorically and thus are fixed at utterance. Nor is it the case that Bryan is simply mistaken, as indicated by his reluctant concession (7d). If Bryan was simply factually mistaken, this would not be felicitous.

- (12) *A context just like (7) except there is an umbrella in the closet.*
 a. *Bryan*: You have to wear a raincoat.
 b. *Alice*: No I don't. I could use an umbrella.
 c. *Bryan*: # I suppose that's technically true, but you get my drift.

Thus we have two different kinds of domain restriction on modals. On the one hand, the anaphoric kind, e.g., that which determines that the domain in (10) should contain only worlds consistent with the interlocutors' goals, and on the other hand the non-anaphoric kind, e.g., that which determines that the domain in (8) excludes crazy worlds where people wrap themselves in duct tape. We know they are distinct because they have different behaviors – only the latter kind can shift within a discourse, like in (7).

Lewis's (1979) original analysis does not differentiate among the kind of domain restriction that is effected anaphorically (and thus permanently by the context) and the kind of domain restriction that may shift. He argues that there is simply a single accessibility relation (comparable to Kratzer's modal base and ordering source) which is "a part of the conversational score".

This accessibility relation on Lewis’s view is subject to accommodation; thus if a world which exists outside the modal domain is brought up, the modal domain shifts to accommodate and include this world. Lewis does not explain why the kind of shifting in (9) and (10) is not acceptable and cannot be accommodated, nor does he provide a formal semantics for accommodation. However, besides observing the phenomenon in the first place, he does raise a number of interesting and valuable points, among them his observation that the sort of domain shifting that goes on in genuine Lewis Sequences is unidirectional and cannot be (easily) resisted.

“[The first sentence in a Lewis Sequence] would be false if the boundary between relevant and ignored possibilities remained stationary. But it is not false in its context, for hitherto ignored possibilities come into consideration and make it true. And the boundary, once shifted outward, stays shifted. If [the first speaker in a Lewis Sequence rejects the shifted-rebuttal], he is mistaken.” (Lewis, 1979, pp. 354–355)

“[The first speaker in (3)] brings into consideration possibilities hitherto ignored, else what he says would be false. The boundary shifts outward so that what he says is true. Once the boundary is shifted, the commonsensical epistemologist [“Yaron” in (3)] must concede defeat. And yet he was not in any way wrong when he laid claim to infallible knowledge. What he said was true with respect to the score as it then was.” (Lewis, 1979, p. 355)

I agree with Lewis that sentences like (3a) and (7b) are true with respect to the “score” as the utterers of those sentences *believed it to be when they uttered them*, but what the rebuttals in (3b) and (7c) (and their subsequent successful uptake) show is that there is a disagreement about the score of the conversation in the first place. What distinguishes these from cases like (9) is that this disagreement is more or less *automatically* resolved by convergence to the score as it was thought to be by the utterers of (3b) and (7c). On this new score, the original utterances of (3a) and (7b) are false, even in retrospect.

As shown above, this cannot be the case if what is shifting is the mapping of an anaphor. A disagreement about that aspect of the “score” cannot be

resolved in this way. For this reason I argue that this “shift” is entirely pragmatic in nature, and that the (intended) domain restriction which occurs in the original utterance of (7b) and which may vary from context to context as illustrated in (5). Particularly, what is “shifting” is the standard of precision, and important parameter in the pragmatic evaluation of utterances. Below, I develop a pragmatic theory of imprecision and show that it can account for Lewis Sequences.

4.2 Imprecision

In this section I lay out the empirical facts surrounding imprecision, explain its crucial defining features, and provide a formal analysis building on Lasersohn (1999), Krifka (2007), and Lauer (2012), which accounts for these defining features.

As pointed out by Lewis (1979), certain expressions can be used loosely.

- (13) a. France is hexagonal.
 b. This table is flat.

Lewis points out that France is not strictly a hexagon, and that almost certainly no table is actually completely flat, down to the molecule. Lewis argues that the *standard of precision* is a crucial conversational index of evaluation – in other words it factors into the “conversational score” along with things like discourse referents and the time of utterance. Lewis argues that contexts may vary in their standards of precision, and that this in turn determines whether or not sentences like (13) can be judged as true or not.

However, the standard of precision does not behave like other elements of the conversational score. As Lewis pointed out, this standard is subject to shifting within a given discourse, in a way that, e.g., discourse referents are not. To illustrate this, consider the anatomy of a discourse in which the standard of precision is shifted.

- (14) *Katie and Lelia stand around a table made by humans.*
 a. *Katie:* This table is flat. LOOSE CLAIM
 b. *Lelia:* Not *really*. Nothing made by humans is actually flat.
 REBUTTAL
 c. *Katie:* Well, okay, whatever. But you get my drift. CONCESSION

There are three crucial properties to this discourse. The first is that the rebuttal, (14b), comes with an intuition of **pedantry**. I define pedantry as a kind of intuition of mild uncooperativity. On the one hand, Lelia's rebuttal seems not entirely appropriate for the context at hand¹ and may annoy the speaker of the loose claim. On the other hand, Lelia's rebuttal is not *totally* inappropriate, as would be in a case where she misinterprets the speaker in a more serious way.

- (15) a. *Katie*: This table is flat.
 b. *Lelia*: #Well sure, but have you ever seen a carbonated table?
 c. *Katie*: What are you talking about? I meant *flat* like *level*.

A response like (15b), where Lelia takes *flat* to mean *not carbonated*, as it is usually applied to beer or pop, is entirely infelicitous; these are the words of a lunatic, not a pedant. They could not be given a response like (14c). Thus, pedantry is neither totally cooperative nor totally uncooperative.

The second important property of this discourse is what I call **inessential disagreement**. This could be thought of as similar to faultless disagreement in the sense that both Kate and Lelia are in some sense right. The discourse in (14) differs from a case of faultless disagreement involving predicates of personal taste like (16) in that there is a concession.

- (16) a. *Raphael*: This is delicious.
 b. *Sterling*: No it isn't.

But, as indicated by the reluctance of the concession in (16c) and the subsequent defense of the original claim (...*But you get my drift*) we can see that Katie is not mistaken about the facts. The table is just as flat as she thought it was, but Lelia has raised the standard of precision. Thus, their disagreement is an inessential one. I will use such *partial concessions* as diagnostic of inessential disagreement.

The third important property, originally pointed out by Lewis (1979) is **unidirectionality**: raising the standard of precision is easier than lowering it. On the one hand, Katie seemingly has to go along with Lelia's shift in (14). But if we try in the other direction, this won't work.

- (17) *Alice*: Michigan is not shaped *precisely* like a hand; after all, there

¹Of course, depending on how we flesh out this context, it may be entirely appropriate. But crucially there are plenty of contexts in which it is not.

are weird bumps all over, and the thumb is too big.
Bryan: # Sure, but France is hexagonal.

In (17), Alice establishes a high standard of precision, one which does not allow Michigan to be characterized as a hand. If Bryan then tries to assert that France is roughly hexagonal by saying “France is hexagonal”, it does not go through. Alice is not forced to a lower standard of precision by Bryan; rather, what he says is just false.

Additionally, consider that discourses just like (14), having all three crucial properties, can occur not just with absolute gradable adjectives, but also with numerals (18) and quantified expressions (19).

- (18) *The facts: Julian arrived at 2:59.*
- a. *Helena*: Julian arrived at 3.
 - b. *Itamar*: No, he arrived 2:59.
 - c. *Helena*: Well, okay, whatever.
- (19) *Of the 30 grad students enrolled in the department, 29 are present.*
- a. *Max*: All the grad students are here.
 - b. *Nick*: Not all of them. Octavia is in Peru doing fieldwork.
 - c. *Max*:Well, yeah, but you know what I mean.

While the semantic context-dependence of quantified expressions may provide some purchase for an attempt to explain this variability as anaphoric, numerals are not obviously context dependent at all, and so cases like (18) cannot be given such an explanation without significantly altering the meaning of these expressions.

And even if we could say of the expressions that are subject to imprecision that they have context dependence in common, we cannot say that imprecision is due to anaphoric variability. As argued in Section 1, anaphors simply do not shift. The semantic context, or conversational score, is responsible for determining the truth values of utterances when they are uttered, and after that point there is no negotiating over the score for the purpose of evaluating that utterance. In other words, even if the score does change, we should still evaluate the original utterance against the score *as it was when the utterance was made*. But all of these examples involve cases where the rebuttal has the effect of shifting the score, not just for the purposes of that utterance, but for the previous one as well. Moreover it comes with an intuition of pedantry, which should not occur with a simple, *totally licit* shift of the conversational

score.

On the basis of this, I argue with Lasersohn (1999) and Lauer (2012) that imprecision is a pragmatic effect,² since i) it is very general, affecting a wide range of expressions; ii) its cannot be accounted for just with a standard treatment of anaphora; and iii) a very intuitive pragmatic motivation can be given for it.

4.2.1 Semantic Context and Pragmatic Context

Before introducing the framework for imprecision, I want to distinguish between two notions of context. One we can call the ‘semantic context’ i.e., the aspect of the context which is used to determine truth conditions, or on a more dynamic view, the aspect of the context which is updated in virtue of the arbitrary conventional meanings of natural language expressions. We could also call this the ‘conversational score’, since it is updated automatically in a rule-governed way. On a dynamic view, this context can be represented as an information state, perhaps a set of world-assignment pairs, where each assignment is a function mapping syntactic indices to semantic objects. On such a view, anaphors are valued by these assignments (Heim, 1982; Condoravdi and Gawron, 1996).

$$(20) \quad \llbracket \text{the dog}_7 \text{ sat down} \rrbracket^\sigma = \lambda \langle w, g \rangle [\text{sat-down}(g(7))(w)] \text{ iff } \forall \langle w, g \rangle \in \sigma [\mathbf{dog}(g(7))(w)]$$

Modals may combine with hidden indexicals, which themselves come loaded with such indices, to determine their domains.

$$(21) \quad \llbracket \text{it must } m_5 \text{ be raining} \rrbracket^\sigma = \lambda \langle w, g \rangle [\forall u \in \cap g(5)(w) [\mathbf{raining}(u)]] \\ \text{iff } \forall \langle w, g \rangle \in \sigma [g(5) \text{ is epistemic}]$$

While in both of these cases the question of which silent index has been used may give rise to a kind of indeterminacy of meaning, the recoverability of which may depend on pragmatic factors, the valuation of and reference to such indices is purely semantic, i.e., stated in the conventional meanings of the expressions.

There is another kind of context which affects interpretation but not in the same way, which I call pragmatic context; this is referred to as the domain goals by Roberts (2012). This level of context may specify, for example,

²Lauer actually argues that Lasersohn’s treatment is not truly pragmatic.

what the goals of the various participants are, which are in turn crucial to the computation of Gricean inferences. This pragmatic context is more or less stable over the course of a discourse; utterances do not update our conversational goals, at least not in an automatic, rule governed way. It is this level of context which I argue encodes the standard of precision as well, since the standard of precision is heavily related to the goals of the interlocutors.

Importantly, this separation of semantic and pragmatic context is not intended to distinguish between kinds of context that should be treated formally and kinds that should not. The pragmatic context deserves a formal treatment as well. However since this requires a formal treatment of goals and human reasoning that would take us well beyond the matter at hand, I will not address this here. Moreover while many authors refer to issues surrounding what I call the ‘semantic context’ as pragmatic, I will not; I reserve that term only for the computation of Gricean inferences and the sorts of contextual factors which feed that computation.

4.2.2 Pragmatic Slack

Lasersohn (1999) represents the first serious attempt to formalize a notion of imprecision. According to Lasersohn, an utterance is imprecise if its semantic interpretation is strictly false, but if its meaning is close enough to something true that the difference can be ignored. What counts as close enough depends entirely on the goals of interlocutors; in one context a small difference in the hue of a paint may not matter, while in another context, it may.

Lasersohn argues that expressions are interpreted with respect to a “halo” of alternative denotations which differ from the semantically specified meaning in only ignorable ways; i.e., the halo of an expression is the set of meanings that are close enough to its lexical meaning that the difference doesn’t matter in the current pragmatic context. Halos are calculated for each individual expression in an LF and then are composed pointwise. Below is a formula for the calculation of a halo given a semantic context σ and a pragmatic context ρ .

- (22) a. If α is a terminal node:
 $H_\rho(\alpha)(\sigma) = \{x : \mathbf{dis}(x)(\llbracket \alpha \rrbracket^\sigma) < s_{\mathbf{dis}}(\llbracket \alpha \rrbracket^\sigma)(\rho)\}$
 b. If α has two daughters, β and γ and $\llbracket \alpha \rrbracket^\sigma = R(\llbracket \beta \rrbracket^\sigma)(\llbracket \gamma \rrbracket^\sigma)$:
 $H_\rho([\alpha\beta\gamma])(\sigma) = \{R(x)(y) : x \in H_\rho(\beta)(\sigma) \ \& \ y \in H_\rho(\gamma)(\sigma)\}$

(22a) says that the halo of an expression α is the set of objects such that the distance between those objects and the semantic meaning of α is less than the threshold or standard (s_{dis}) for distance established by the pragmatic context ρ , given the original semantic meaning. Thus s_{dis} is a function which takes the original semantic content, and the goals of the interlocutors specified in ρ , and returns the maximum degree of distance from or dissimilarity to that original semantic content which is tolerable given those goals.

(22b) specifies that the halos of individual expressions then compose pointwise. The result of this composition is a set of propositions. An utterance is ‘pragmatically true’ then if one of the propositions in this set is true, or alternately, if the disjunction of these propositions is true. This latter view may be more amenable to a dynamic take.

To consider an example, take the halo for an expression like *three o’clock*.

$$(23) \quad H_{\rho}(\text{three o’clock})(\sigma) = \{\dots \llbracket \text{two fifty-nine} \rrbracket^{\sigma}, \llbracket \text{three o’clock} \rrbracket^{\sigma}, \llbracket \text{three oh one} \rrbracket^{\sigma} \dots\}$$

When the expression composes with other elements in the sentence, the result is a set of propositions.³

$$(24) \quad H_{\rho}(\text{Julian arrived at three o’clock})(\sigma) = \\ \{\dots \llbracket \text{Julian arrived at two fifty-nine} \rrbracket^{\sigma}, \\ \llbracket \text{Julian arrived at three o’clock} \rrbracket^{\sigma}, \\ \llbracket \text{Julian arrived at three oh one} \rrbracket^{\sigma} \dots\}$$

Work on imprecision often refers to a ‘standard of precision’ (Lasersohn, 1999, Kennedy, 2007, Morzycki, 2011, Lauer, 2012), with Morzycki going as far as to suggest a parameter of evaluation which is the degree of similarity required for inclusion in a halo, which he calls a ‘standard of precision’. I will not go so far as to posit such a uniform standard given that the level of precision required may be different across different domains; e.g., consider that chemists in a lab may require a very high standard of precision with regard to quantities of certain chemicals, but perhaps may not care enough to distinguish Pepsi from Coke.

³This example ignores the effect of imprecision for all other expressions in the sentence. This may matter in the case of *arrive* where certain events may not strictly constitute arrival (at an implicitly specified place) but might be treated as arrival in the context. Other expressions, like the name *Julian*, as well as functional morphemes, may only have trivial halos, since e.g., there is no individual similar enough to Julian that they can be considered interchangeable in the context ρ .

(25) I bet you a Coke that there is less than .0005 grams of precipitate.

However, I will continue to informally refer to a ‘standard of precision’ as the threshold of similarity *given a particular domain*.

The fact that this view of imprecision takes lexical semantic composition into account should not be taken as contrary to the view that it is a pragmatic phenomenon; clearly pragmatic computation at the lexical level is required in view of scalar implicature. This just means the speakers have some conscious awareness of the individual lexical items they are using and can reason on the basis of them.

Lasersohn motivates pragmatic computation on a word-by-word basis due to the presence of slack regulators, expressions which in his analysis modify the halos of individual expressions.

(26) I bet you a Coke that there is exactly .0005 grams of precipitate.

Even though .0005 is to be interpreted precisely, it may not be the case that *Coke* is. Thus the expression *exactly* seems to effect only the level of precision for the interpretation of .0005.

However, as indicated above, the standard of precision may vary from domain to domain – there could be contexts where the difference between brands of cola is not important but very small differences of weights may matter. Thus, the effect of *exactly* need not be to act directly on the halo as it is computed, but rather to give rise to a metalinguistic inference that differences between alternatives to .0005 may matter (but still not affecting the standard of precision across the board).

Another problem with Lasersohn’s (1999) account is that at the level of single lexical items, the proper standard of precision may not be clear. Consider the following case.

(27) There should be fifty place settings. And we should have two hundred bottles of beer on hand.

How imprecisely these number terms are interpreted could be sensitive to fine details of context, but we can imagine at least one context in which the former is interpreted more precisely than the latter.⁴ If tables have already

⁴This is not a contradiction of unidirectionality. Unidirectionality only says that the standard of precision for a particular domain cannot be lowered. But the point here is that we have two different domains – number of place settings and number of bottles of

been set up, the guest list is known, and a particular aesthetic is desired by the speaker of (27), then presumably *fifty* should be interpreted very precisely, in fact, exactly. However, with *two hundred* there may be room for error. And in fact, the error may be disproportionate –perhaps 199 or 198 beers would be acceptable, though 197 may be too few. But erring on the side of too many may be more acceptable. Thus, the ‘pragmatic halo’ for *two hundred* may extend from 198 to 210. Thus, it cannot be that the halo for *fifty* or *two hundred* must be calculated with no other information besides their strict semantic denotations.

All this being said, a view of halos as being built from the nodes up is not necessarily doomed in light of this evidence. Post hoc halos may be constructed after the hearer has already internalized the strict semantic information, and thus has an idea of what each lexical item’s contribution is to the bigger picture. But in order to simplify my view of the pragmatics of imprecision, I will assume that halos are constructed only for whole sentences, and that slack regulators like *exactly* can be accounted for despite this; see Chapter 5 for some relevant discussion.

4.2.3 The Pragmatic Model

Crucially, the standards of precision are determined entirely by the conversational goals of the interlocutors in a Gricean fashion (Grice, 1957, 1989). One might imagine that the Maxim of Quality (“be truthful”) rules out the possibility of imprecision, but this is balanced by the Maxim of Manner, which encourages brevity and clarity. Not only are approximations generally easier to utter (*thirty* as opposed to *thirty-one*; *flat* as opposed to *almost flat*), they are also easier to process (Sperber and Wilson, 1986a; Solt et al., 2013b,a).

Moreover, while there is a sense in which imprecise expressions may be false (namely, the literal sense), the maxims *do not govern what is literally said*, but rather what is communicated. This is what gives rise to implicatures: semantic meaning which falls short of a given Maxim is enriched pragmatically to bring it in line. The same is true here: imprecise expressions do communicate true things, even if they are literally false. This should be enough to satisfy those who argue that such a theory of imprecision makes everything false.

beer on hand.

Moreover, as noted by Lasersohn (1999, p. 526 fn. 5), imprecision can be motivated by ignorance as well as convenience, due to the limits of human perception. This is a crucial property of imprecision since humans can never know exact quantities in natural continuous domains.

- (28) a. Prerna is 5 feet 5 inches tall.
 b. #Prerna is 5.43342100364.... feet tall.

Imprecision therefore can be thought of as a feature, rather than a bug, of natural language. It is necessary to getting around our natural human deficiencies, and satisfying the second clause of the Maxim of Quality: do not say more than you have evidence for.

Krifka (2007) proposes a theory of why round numbers (like *forty*) are taken to have approximate or exact meanings whereas non-round numbers (like *forty-one*) are taken to have only exact meanings. His theory is couched in bidirectional Optimality Theory and relies on the interaction between a constraint against complex expressions (Manner) and a constraint requiring the interpretation of a given expression to be true (Quality). Krifka’s model takes candidates to be form-meaning pairings, with candidates compared on the basis of each possible actual true state of affairs.

My model is a competition-based account reminiscent of Optimality Theory (Prince and Smolensky, 2002), which has been applied extensively to phonological theory in linguistics, as well as syntax and semantics. On this view, there is a pragmatic algorithm for determining the ‘optimal’ implicature to derive from a given utterance (which has a given literal semantic value). This algorithm proceeds by considering each possible belief state that the utterer might have. Then, for each belief state, the algorithm compares all possible utterances (out of a set of reasonable candidates) and subjects them to particular constraints – in my analysis these constraints are essentially formalized versions of Grice’s Maxims. For each belief state an optimal utterance is then determined; the one which violates the highest ranked constraints the least. The implicature for a given utterance is then the disjunction of all the possible belief states for which that utterance is considered optimal.

Consider three basic constraints: QUALITY, QUANTITY, and MANNER.⁵ QUANTITY assigns violations on the basis of comparative informativeness of

⁵For the purposes of this discussion I exclude RELEVANCE, but obviously a fully-fleshed out pragmatic model should include it.

expressions; MANNER assigns violations on the basis of comparative ease of use (for both speaker and hearer).

First, consider QUALITY.

- (29) Given form F , semantic context σ , pragmatic context ρ , and information state B :

$$\text{QUALITY} := *F \text{ unless } \forall w \in B[\exists u \in \llbracket F \rrbracket^\sigma[w \approx_\rho u]]$$

This constraint takes as its input a linguistic form, F , semantic and pragmatics contexts, σ and ρ , and an information state B , which represents a possible belief state of the speaker. This constraint is violated unless every world in the information state B has a counterpart in the denotation of F at σ , where ‘counterpart’ means a world which is similar given the pragmatic context ρ . The definition of QUALITY does not require that every world in the knowledge state of the speaker be in the denotation of what the speaker utters; i.e., it need not be strictly true. Rather, it simply must be the case that every world in the speaker’s model be similar enough to a world in the denotation of the expression that for the purposes of ρ they are interchangeable. Thus, imprecision is built into the definition of QUALITY.

This is similar to the proposal made by [Lauer \(2012\)](#) who proposes that false expressions can be uttered as long as they give rise to the same actions that would have been given rise to by an utterance of the strictly true expression. It is also not dissimilar from the proposal made by [Sperber and Wilson \(1986a\)](#), who argue a proposition can be uttered if it ‘resembles’ a true proposition; their account is couched in the decidedly non-Gricean Relevance Theory of [Sperber and Wilson \(1986b\)](#).

The next constraint is QUANTITY.

- (30) For given forms F_1 , F_2 , semantic context σ :

$$\text{QUANTITY} := F_1 \geq F_2 \text{ iff } \forall w \in \llbracket F_1 \rrbracket^\sigma[\exists u \in \llbracket F_2 \rrbracket^\sigma[w \approx_\rho u]]$$

One form F_1 is judged superior to another F_2 if it asymmetrically entails it, given context σ ; this is the traditional definition of ‘stronger than’. One modification is that, as with QUALITY, entailment is softened by imprecision. Thus where a traditional account would have equality $=$, the similarity relation \approx_ρ appears. Unlike in the case of QUALITY, nothing significant hinges on this choice, but it is more elegant; see footnote XX below.

Finally, consider MANNER.

- (31) For given forms F_1, F_2 :
 $\text{MANNER} := F_1 > F_2$ iff $\text{ease}(F_1) > \text{ease}(F_2)$

This constraint says that F_1 is preferred to F_2 if F_1 is easier for both speaker and hearer. This of course glosses over a considerable amount; what constitutes ease? The crucial fact here is just that simpler expressions constitute easier ones to process, and thus easier for the purposes of this constraint. See [Sperber and Wilson \(1986b\)](#) and [Solt et al. \(2013a\)](#) for discussion of this.

The ranking of the constraints is as follows:

- (32) $\text{QUALITY} > \text{QUANTITY} > \text{MANNER}$

As in OT ([Prince and Smolensky, 2002](#)), any candidate which violates the highest ranked constraint is eliminated unless every candidate does.

This account is competition-based, but not Optimality Theoretic, strictly speaking.⁶ There is no predicted crosslinguistic variation in terms of constraint rankings. Moreover this is not an algorithm for determining a surface form from an underlying form in any sense of ‘surface’, ‘underlying’, or ‘form’. This takes as its input an utterance and a context, compares form-belief state pairs, and gives a proposition as an output. However, the visual language of Optimality Theory is a useful and familiar one, so I will borrow it here.

The table below presents each candidate in a row and each constraint in a column. A violation of a constraint is indicated with an asterisk (*). An exclamation point (!) marks the ‘fatal’ violation for any given candidate; the hand (✋) indicates the optimal candidate. Candidates are eliminated if they violate the highest ranking constraint. Surviving candidates are then considered on the basis of the next-highest constraint. Any violators of that constraint are eliminated, unless all remaining candidates violate it once. The process is repeated until there is a unique optimal candidate, or the constraints are exhausted.

An exemplar derivation is given in Table 4.1. If the speaker’s belief state is K_1 , then, the optimal utterance is H; if it is K_2 , then G. If it is K_3 , then H. Since the speaker knows that the hearer does not know her belief state, the speaker knows that from an utterance of H , the hearer will infer $K_1 \cup K_3$; likewise from an utterance of G , the hearer will infer K_2 . Since F is not optimal for any information state, an utterance of F would constitute a flouting of the maxims and would require the hearer to accommodate some

⁶Thanks to Ken Safir for enlightening discussion on this topic.

$\langle H; \sigma; \rho \rangle$	QUALITY	QUANTITY	MANNER
$\langle F, K_1 \rangle$		*!	*
$\langle G, K_1 \rangle$		*!	
$\langle H, K_1 \rangle$			
$\langle F, K_2 \rangle$		*	*!
$\langle G, K_2 \rangle$		*	
$\langle H, K_2 \rangle$	*!		
$\langle F, K_3 \rangle$		*!	*
$\langle G, K_3 \rangle$		*!	
$\langle H, K_3 \rangle$			
$\begin{aligned} \llbracket F \rrbracket^{\sigma, \rho} &= \emptyset \\ \llbracket G \rrbracket^{\sigma, \rho} &= K_2 \\ \llbracket H \rrbracket^{\sigma, \rho} &= K_1 \cup K_3 \end{aligned}$			

Table 4.1: Example of a pragmatic derivation.

additional inference or alteration to the semantic or pragmatic context.⁷

For a more concrete example, consider a scalar implicature. We have a context where there are three dogs, and it is mutually known the speaker knows of each dog whether it barked. Thus there are no possible speaker belief states containing more than one of the following worlds.

$$(33) \quad \begin{array}{ll} a = \text{Killian barked.} & c = \text{Bandit barked.} \\ b = \text{Moneypenny barked.} & \end{array}$$

$$(34) \quad \begin{array}{ll} a \& b \& c = \{w_1\} & \neg a \& b \& c = \{w_4\} \\ a \& b \& \neg c = \{w_2\} & \neg a \& \neg b \& \neg c = \{w_5\} \\ a \& \neg b \& c = \{w_3\} & \end{array}$$

The implicature for the utterance in Table 4.2 is computed as below. This is a case of a typical implicature, where the pragmatic inference generated is stronger than the semantic content.

⁷This is static. A dynamic version would be more complicated, but not different in any important way.

	$\langle \text{some of the dogs barked}; \sigma; \rho \rangle$	QUALITY	QUANTITY	MANNER
☞	$\langle \text{all of the dogs barked}, \{w_1\} \rangle$			
	$\langle \text{some of the dogs barked}, \{w_1\} \rangle$!*	
	$\langle \text{none of the dogs barked}, \{w_1\} \rangle$!*		
	$\langle \text{all of the dogs barked}, \{w_2\} \rangle$!*		
☞	$\langle \text{some of the dogs barked}, \{w_2\} \rangle$		*	
	$\langle \text{none of the dogs barked}, \{w_2\} \rangle$!*		
	$\langle \text{all of the dogs barked}, \{w_3\} \rangle$!*		
☞	$\langle \text{some of the dogs barked}, \{w_3\} \rangle$		*	
	$\langle \text{none of the dogs barked}, \{w_3\} \rangle$!*		
	$\langle \text{all of the dogs barked}, \{w_4\} \rangle$!*		
☞	$\langle \text{some of the dogs barked}, \{w_4\} \rangle$		*	
	$\langle \text{none of the dogs barked}, \{w_4\} \rangle$!*		
	$\langle \text{all of the dogs barked}, \{w_5\} \rangle$!*		
	$\langle \text{some of the dogs barked}, \{w_5\} \rangle$!*	*	
☞	$\langle \text{none of the dogs barked}, \{w_5\} \rangle$			
$\begin{aligned} & \llbracket \text{all of the dogs barked} \rrbracket^{\sigma, \rho} = \{w_1\} \\ & \llbracket \text{some of the dogs barked} \rrbracket^{\sigma, \rho} = \{w_2, w_3, w_4\} \\ & \llbracket \text{none of the dogs barked} \rrbracket^{\sigma, \rho} = \{w_5\} \end{aligned}$				

Table 4.2: Derivation of a scalar implicature with quantifiers.

The pragmatic inference (35b) is stronger than the semantic content (35a) because while *some of the dogs barked* satisfies QUALITY in knowledge state $\{w_1\}$, the hearer knows that the speaker would have been able to utter something stronger in said knowledge state. And because ignorance does not factor in, since it is mutually known that the speaker knows the status of each dog, no possible knowledge state of the speaker contains w_1 .

- (35) a. $\llbracket \text{some of the dogs barked} \rrbracket^\sigma = \{w_1, w_2, w_3, w_4\}$
 b. $\llbracket \text{some of the dogs barked} \rrbracket^{\sigma, \rho} = \{w_2, w_3, w_4\}$

In the usual terminology, we might say that the utterance gives rise to a conventional inference (some of the dogs barked), with an additional Gricean inference on top of it (not all of the dogs barked). However, in this set-up, there is just a single inference.⁸ If *all of the dogs barked* had been uttered it would have generated the inference indicated in gray in the tableau. Thus nothing about this set-up requires that the pragmatically generated inference is stronger than conventional meaning; they could in fact be weaker. This is what allows for imprecision, wherein the pragmatically generated meaning is weaker than the conventional, as pointed out by [Lauer \(2012\)](#).

One important point to make is that the tableau in Table 4.2 represents only some of the candidate utterances and information states. In principle, it should compare all possible utterances and information states – both of which are infinite candidate sets. Of course, this is limited by the natural constraints on human cognition. The utterance candidates could be determined by starting with the actual utterance⁹ and taking others which are reasonably similar; after all, anything that moves too far afield is likely to violate one of the constraints (perhaps especially RELEVANCE) far too heavily to have a chance at winning. Likewise the hypothesized information states should begin with the common ground (what the speaker is already publicly committed to believing) and whatever information about the speaker’s knowledge state is contained in it. But beyond that the range of possibilities considered might be limited only by human cognition.

Turning now to imprecision, given the definition of QUALITY, it should

⁸Competent speakers may of course still be able to track and distinguish between the ‘conventional inference’ and the Gricean one. But only the Gricean inference is relevant to normal communication.

⁹In fact this pragmatic computation could be largely done before the relevant utterance is even made, based on the expectations by the hearer of the speaker’s utterance.

be clear how approximate inferences can arise in this system. Crucially, however, QUALITY is sensitive to the pragmatic context in a way that MANNER is not. MANNER can be computed just with the knowledge of the forms and their alternatives, while QUANTITY can be computed just with the semantic interpretation of the forms and the semantic context. Whether a given utterance satisfies QUALITY, however, depends crucially on what distinctions matter in the pragmatic context.

4.2.4 Implicision

Consider (36).

(36) There were fifty people.

The interpretation of *fifty* as precise or imprecise depends on the pragmatic context. Consider two such contexts, ρ_{hi} , where the difference between fifty and forty-nine matters, and ρ_{low} , where it does not. Consider also the relevant utterances and information states.

(37)	<i>Possible utterances</i>	(38)	<i>Possible speaker belief states</i>
	$49 \rightarrow \text{There were } 49 \text{ people.}$		$=49 \rightarrow \lambda w[\text{exactly } 49 \text{ people in } w]$
	$50 \rightarrow \text{There were } 50 \text{ people.}$		$=50 \rightarrow \lambda w[\text{exactly } 50 \text{ people in } w]$
	$51 \rightarrow \text{There were } 51 \text{ people.}$		$=51 \rightarrow \lambda w[\text{exactly } 51 \text{ people in } w]$
	$\approx 49 \rightarrow \text{There were about } 49$ <i>people.</i>		

Among the candidate utterances is included versions of the utterance with approximators, which unlike slack regulators affect the semantic truth conditions of the expressions they combine with [Lauer \(2012\)](#). Also included are information states where the speaker does not know the exact actual value. Thus cases of both speaker certainty and uncertainty are included. Lastly, the tableaux below do not include the ‘at least’ readings of any of the relevant utterances for brevity; in any case these readings would be ruled out by scalar implicature along the lines of [Table 4.2](#).

Very much along the lines of [Krifka \(2007\)](#), non-round numbers incur a penalty to MANNER. This could be for both articulatory and cognitive reasons. Under high standards of precision, QUALITY is enough to eliminate

	$\langle 50; \sigma; \rho_{hi} \rangle$	QL	QN	M
	$\langle 49, = 49 \rangle$			*
	$\langle \approx 49, = 49 \rangle$!*	**
	$\langle 50, = 49 \rangle$!*		
	$\langle \approx 50, = 49 \rangle$!*	*
	$\langle 51, = 49 \rangle$!*		*
	$\langle \approx 51, = 49 \rangle$!*	**
	$\langle 49, = 50 \rangle$!*		*
	$\langle \approx 49, = 50 \rangle$!*	**
	$\langle 50, = 50 \rangle$			
	$\langle \approx 50, = 50 \rangle$!*	*
	$\langle 51, = 50 \rangle$!*		*
	$\langle \approx 51, = 50 \rangle$!*	**
	$\langle 49, = 51 \rangle$!*		*
	$\langle \approx 49, = 51 \rangle$		*	**
	$\langle 50, = 51 \rangle$!*		
	$\langle \approx 50, = 51 \rangle$		*	*
	$\langle 51, = 51 \rangle$			*
	$\langle \approx 51, = 51 \rangle$!*	**
	$\langle 49, 49 \cup 50 \cup 51 \rangle$!*		*
	$\langle \approx 49, 49 \cup 50 \cup 51 \rangle$		*	*!*
	$\langle 50, 49 \cup 50 \cup 51 \rangle$!*		
	$\langle \approx 50, 49 \cup 50 \cup 51 \rangle$		*	*
	$\langle 51, 49 \cup 50 \cup 51 \rangle$!*		*
	$\langle \approx 51, 49 \cup 50 \cup 51 \rangle$		*	*!*
$\begin{aligned} \llbracket 49 \rrbracket^{\sigma, \rho} &= [= 49] \\ \llbracket 50 \rrbracket^{\sigma, \rho} &= [= 50] \\ \llbracket 51 \rrbracket^{\sigma, \rho} &= [= 51] \\ \llbracket \approx 50 \rrbracket^{\sigma, \rho} &= [49 \cup 50 \cup 51] \end{aligned}$				

(a) High precision context.

	$\langle 50; \sigma; \rho_{low} \rangle$	QL	QN	M
	$\langle 49, = 49 \rangle$!*
	$\langle \approx 49, = 49 \rangle$!*	**
	$\langle 50, = 49 \rangle$			
	$\langle \approx 50, = 49 \rangle$!*	*
	$\langle 51, = 49 \rangle$!*
	$\langle \approx 51, = 49 \rangle$!*	**
	$\langle 49, = 50 \rangle$!*
	$\langle \approx 49, = 50 \rangle$!*	**
	$\langle 50, = 50 \rangle$			
	$\langle \approx 50, = 50 \rangle$!*	*
	$\langle 51, = 50 \rangle$!*
	$\langle \approx 51, = 50 \rangle$!*	**
	$\langle 49, = 51 \rangle$!*
	$\langle \approx 49, = 51 \rangle$!*	**
	$\langle 50, = 51 \rangle$			
	$\langle \approx 50, = 51 \rangle$!*	*
	$\langle 51, = 51 \rangle$!*
	$\langle \approx 51, = 51 \rangle$!*	**
	$\langle 49, 49 \cup 50 \cup 51 \rangle$!*
	$\langle \approx 49, 49 \cup 50 \cup 51 \rangle$!*	**
	$\langle 50, 49 \cup 50 \cup 51 \rangle$			
	$\langle \approx 50, 49 \cup 50 \cup 51 \rangle$!*	*
	$\langle 51, 49 \cup 50 \cup 51 \rangle$!*
	$\langle \approx 51, 49 \cup 50 \cup 51 \rangle$!*	**
$\begin{aligned} \llbracket 49 \rrbracket^{\sigma, \rho} &= \emptyset \\ \llbracket 50 \rrbracket^{\sigma, \rho} &= [49 \cup 50 \cup 51] \\ \llbracket 51 \rrbracket^{\sigma, \rho} &= \emptyset \\ \llbracket \approx 50 \rrbracket^{\sigma, \rho} &= \emptyset \end{aligned}$				

(b) Low precision context.

Table 4.3: Pragmatic derivation of (im)precise inferences with numerals.

competing unmodified numerals, while approximated numerals are defeated on QUANTITY (they would lose out on MANNER too, if it got that far), with the one exception being the ignorance case, where the approximated round

number wins over the unmodified numerals on QUALITY¹⁰ and over the other approximated numbers on MANNER. When the standard of precision is low, the situation is much like the ignorance case for the high standard; QUALITY no longer matters anywhere and MANNER becomes relevant in many cases.

When the standard of precision is high, the inference generated by *50* is identical to the semantic content. When the standard of precision is low, however, the inference generated is weaker than the semantic content. This is true even if ignorance is not an option; thus even if the semantic context entails that the speaker knows the exact value, and the last tableau in the low precision context is not included, the inference generated by *fifty* is the union of the remaining optimal information states, which adds up to the same thing.

This explains the static cases where the standard of precision is stable. Still to be explained, however, are cases where the standard of precision shifts.

4.2.5 Precisification

Return now to the cases of precisification. Consider (18) again, repeated below.

- (39) *The facts: Julian arrived at 2:59.*
- a. *Helena:* Julian arrived at 3.
 - b. *Itamar:* No, he arrived 2:59.
 - c. *Helena:* Well, okay, whatever.

Here we have what appears to be a shift in the standard of precision. I argue, however, that this is not really a seamless ‘shift’ the same way the update of the semantic context may constitute one. Since the standard of precision is determined by the conversational goals of the participants, which generally do not shift in a discourse¹¹ the standard of precision should not shift either. Rather, I argue that cases like (39) are simply *disagreements* about what the standard of precision governing the whole discourse is.

¹⁰Note that QUALITY as defined captures both of Grice’s (1989) clauses of QUALITY; thus if a speaker utters something which rules worlds out of the semantic context which are still live possibilities in her knowledge state, it incurs a violation of QUALITY, as with unmodified numerals in the ignorance case.

¹¹They may shift in some discourses, but not obviously in any of the cases of precisification examined here.

This has to be the case given the fact the shifts obviously constitute disagreements of a certain kind. The apparent ‘retroactive shift’, where (39a) becomes false only after Helena utters it, is not really retroactive at all. Rather, for Itamar’s purposes, it was false all along. Helena therefore has two choices: She can acknowledge the lack of common ground (or common domain goals) and reject the discourse, or she can adopt Itamar’s version of the common ground, as in (39), which involves admitting to having uttered something technically false in the past, even if it was true given what Helena thought the common ground looked like at the time. But she cannot maintain her own version of the standard once she knows that Itamar’s is different; the pragmatic context must be uniform.

(40) **Uniform Pragmatic Context (UPC)**

Speakers must agree upon a single uniform pragmatic context, and the pragmatic context does not change, unless with explicit metalinguistic negotiation.

Returning to (39), Helena and Itamar have different standards of precision in mind, and this is only revealed at the point of the discourse where this difference becomes relevant, i.e., when Itamar utters (39b), which under the low standard of precision that Helena is operating under, fails to generate an inference. In other words, it is not optimal under any possible information state given the parameters Helena believes to be operative. This disagreement about the context is then corrected as the interlocutors converge on the same standard.

This contrasts with a view where Helena and Itamar agree on one standard, then shift together to a new standard. This seems especially unlikely in light of the fact that each participant seems to be pragmatically committed to having a consistent personal standard of precision.

(41) *The facts: Julian arrived at 2:59; Gallagher arrived at 2:58.*

- a. *Itamar:* Gallagher arrived at 3.
- b. *Helena:* Right, Julian also arrived at 3.
- c. *Itamar:* #No, he arrived 2:59.

It goes beyond simple pedantry for Itamar to speak imprecisely about Gallagher but not about Julian. Even though the standard is being raised here (in the right direction for Lewis’s observation of unidirectionality) Itamar cannot felicitously make such a move. Or, put more carefully, such a move is

much more infelicitous than simple pedantry displayed in (39). Even though his use of the non-round time expression in (41c) should trigger Helena to move to the higher standard with him, this is illicit, due to the UPC constraint. Thus, these ‘shifts’ are not really shifts, but rather resolutions of underlying disagreement about the pragmatic parameters of discourse.

A question we can still ask here is: if there is a disagreement about the standard of precision, why should it so ‘automatically’ resolve in favor of the higher standard? The first part of the explanation is that disagreements about the pragmatic parameters of the discourse are much more intolerable than disagreements about what is true, or what the interlocutors ought to be mutually committed to. Interlocutors can and do disagree about a range of things, but these simple disagreements about the facts are not necessarily infelicitous. A disagreement about the conversational goals of the discourse, though, is much more problematic; it undermines the Cooperative Principle. Common goals are crucial to computing Gricean maxims, and thus, communication in general. So there is a strong pressure to resolve such disagreements right away.

But why should they be resolved in the direction of higher precision rather than lower precision? Because it cannot be any other way. If a discourse contains both a (strictly false) use of a round number as well as a (strictly true) use of a non-round number, then resolving to a higher standard of precision will make the former false and the latter true. We could imagine this comes with some kind of cost since we would prefer not to say false things. But if, given the same choice, we resolve to a low standard of precision, instead of having one false utterance and one true utterance we have one true utterance and one utterance that does not even generate an inference. Moreover the pedantic rebuttals in these discourses include explicit overt denials of the previous loose claim – thus regardless of the resolution of the standard of precision, someone has made a false utterance. But only one resolution of the standard allows for the discourse to proceed.

This can be seen in Table 4.3. Suppose two speakers disagree about the standard of precision; then the expression *forty-nine* is used. The low-standard participant generates no inference from this, as indicated in Table 4.3b. That speaker is then forced to reconsider something – either the utterance (did the hearer misunderstand? did the speaker misspeak?), the semantic context, the pragmatic context, or the candidates considered. In this case, neither changing the candidates nor semantic context will have any positive effect, since a non-round expressions is guaranteed to be bad on

low standard of precision. The only way to save the utterance, assuming no misunderstanding, is to shift to a high standard of precision, in which case the interlocutors converge on the same standard.

On the other hand, suppose in the same situation, the expression *fifty* is used. This expression generates an inference for both the high-standard and low-standard participant. So the disagreement in standard persists, undetected.

In other words: round numbers are admissible in both high and low precision contexts, even if their use in high precision contexts might sometimes be false. But non-round numbers are admissible only in high precision contexts. Thus utterance of a non-round number rules out the possibility of a low-standard of precision; round numbers can never have this effect. So the individual with the low-standard receives a cue that the standard of their fellow interlocutor is higher; but the pedant never receives a signal that their fellow interlocutor has a lower standard.

Moreover, this is not just true of numbers; this asymmetry holds for all kinds of predicates. Simple absolute adjectives like *full* are admissible at both high and low standards of precision, while *very nearly full* is only acceptable at high standards. In principle, no expression can, just by its utterance, force the pragmatic context to resolve to a lower standard of precision, because any expression admissible on a low standard of precision must also be admissible on a high standard. It is this asymmetry which accounts for unidirectionality.

4.2.6 Previous Discussion of Imprecision

Before applying the above theory of imprecision to cases of modality, I review previous discussion of imprecision in non-modal domains. First I discuss previous discussions of unidirectionality and precision, before turning to discussions of imprecision as being due to semantic imposition of granularity.

4.2.6.1 Motivating Unidirectionality

The asymmetry in how the standard of precision can be ‘shifted’ is related to the asymmetry pointed out by Lasersohn (1999) and Lauer (2012) regarding slack regulators and approximators. Slack regulators are expressions which raise the standard of precision, like *exactly* discussed above, while approximators alter the truth conditions to weaken them, like *about* discussed above. As Lauer points out, there are no slack regulators that lower the standard of

precision, nor are there truth-conditional antonyms of approximators. The absence of the latter is easy to explain since the strict truth-conditions of various expressions subject to imprecision *are* the most precise meanings possible.

Lauer explains the former as well. Any overt expression which is intended to lower the standard of precision must increase the complexity of the utterance, i.e., violate MANNER. Thus Lauer argues that a rational speaker would only use such expressions if they cared about the difference between the strict truth and the looser truth; otherwise why bother to incur the violation to MANNER? But if such a speaker does in fact care about the difference, by definition they are holding to a higher standard of precision.

My system predicts this as well – looking simply at the tableaux above, on a low standard of precision, MANNER violations are enough to rule out a given expression – modified and non-round numerals both. Thus as soon as the pragmatic context were accommodated to fit whatever constraints were imposed by such a lexical item, that use of the item would immediately be infelicitous since a simpler expression would be superior. Given the retroactive effect of such shifts in the pragmatic context as discussed above, such an expression can never be licensed.

Lauer’s explanation of the unidirectionality of shifting, however, is somewhat different from the one proposed here. He proposes the following Declarative Convention.

(42) **Declarative Convention** (Lauer, 2012)

When a speaker utters a declarative S with the denotation $\llbracket S \rrbracket^\sigma$ in context σ she thereby commits herself to act as though $\llbracket S \rrbracket^\sigma$ is true.

Thus if someone utters (43) they are committed to its strict semantic truth, regardless of the context.

(43) Mary arrived at three.

It’s just that in a low-standard context, acting like (43) is true is not distinguishable from acting like a slight variant of it is true; it gives rise to all the same desired outcomes. But should a distinction be made relevant that was not relevant before, and the difference between what was literally uttered and what is true starts to matter, the speaker is still committed to the literal truth so must retract it. Thus sentences like (44) can never be felicitous, since a speaker can never be committed to a contradiction.

- (44) #Mary arrived at three, but she did not arrive until a few minutes after three.

Some version of this convention must be true – whatever we utter has some kind of effect on the common ground and we are in some sense committed to the common ground; we have to retract things we say if they are shown to be false (and it matters); we have to reject things others say if we know them to be false (and it matters). But Lauer’s convention is a bit too weak. After all, when we utter things whose semantic content underdefines their pragmatic content, we are still “on the hook” for that pragmatic content. Although implicatures are cancelable, they have to be explicitly canceled for the speaker to not be committed to them. To act otherwise is uncooperative. But if we extend Lauer’s convention to include pragmatically generated inferences, we are back where we started – if all speakers have to do is act as if the pragmatic inferences of what they say are true, then (44) should be acceptable at a low standard of precision.

Of course, Lauer can still appeal to the notion that an expression like *a few minutes after three* requires a high standard of precision due to MANNER, which on my model makes (44) infelicitous on any pragmatic grounds. He could moreover clarify his Declarative Convention to state that a speaker is committed *both* to the pragmatic inferences she generates with a speech act as well as with its strict conventionally associated meaning.¹² But even this is not quite enough to rule out (41), since (41a) and (41c) are not contradictory.

On my view what rules this out is a meta-pragmatic constraint which says that the pragmatic context is fixed for an entire discourse and, even if it is only resolved after the discourse has begun, it governs the pragmatic inferences of every utterance within the discourse. Thus a shift in the pragmatic context is possible, but not without some acrimony – an intuition of pedantry in the case of disagreement about the standard of precision, or in the case of a single individual shifting the pragmatic context, perhaps because she realizes a different pragmatic context is required, re-evaluation and possible retraction of claims made under the previous pragmatic context.

- (45) *The facts: Julian arrived at 2:59; Gallagher arrived at 2:58.*
 a. *Itamar:* Gallagher arrived at 3.
 b. *Helena:* Right, Julian also arrived at 3.

¹²On this view, universal commitment to strict meaning only becomes relevant in the rare cases where strict meaning is not entailed by pragmatic inference, i.e., imprecision.

- c. *Itamar*: Oh, no, he arrived at 2:59. And actually Gallagher arrived at 2:58.
- d. *Helena*: Uh...
- e. *Itamar*: Sorry, I forgot that it matters when they showed up down to the minute.

Thus in this context while Itamar changes the pragmatic context, the new pragmatic context that has been settled upon now governs the past utterances made under the previous pragmatic context, even though he and Helena had basically agreed on that context. And though Itamar violates the UPC, he at least indicates that he is doing so, and why, so Helena can adjust with him.

So what makes (41) bad is the failure of Itamar to retract his claim under the new pragmatic context (as well as a failure to indicate why he suddenly shifted contexts, which otherwise seems very uncooperative). Although Lauer argues explicitly against a version of this claim, namely Lasersohn's explanation of the badness of (44), that it is simply infelicitous to shift the standard of precision in the middle of a sentence, something like my Uniform Pragmatic Context constraint is not incompatible with his framework. However, it does not leave much room for his Declarative Convention to have a role in explaining the facts surrounding imprecision.¹³

Finally, while I argue that round numbers (or other likewise MANNER-friendly expressions) are equally useable on both low and high standard contexts, Krifka (2007) makes a point of arguing that simple expressions make a low standard context *more likely*. Consider a discourse like (46).

- (46) a. *Steve*: How many students were there?
- b. *Karen*: Twenty.

Imagine a context in which Steve's expectations here make *seventeen*, *eighteen*, *nineteen*, *twenty-one*, *twenty-two*, and *twenty-three* all just as likely responses as *twenty*. Thus the odds of *twenty* being the exact truth are relatively low. However, if Steve assumes a low standard of precision, the expectedness of the answer is considerably more. Especially in natural continuous domains where exact round numbers are exceedingly rare, the mere presence of such a number may indicate a low standard precision.

Although Krifka's proposal does not take into account the possibility of

¹³It may of course be very useful for other purposes.

the standard of precision having already been set, I basically agree with this for some cases, namely, those very early in a discourse where the standard of precision is not yet made explicit, *and* where the likelihood of an exact round number is exceptionally low. This requires enriching the pragmatic model to allow for the hearer’s expectations about what is true to factor in.¹⁴ Moreover it provides a principled counterexample to the unidirectionality of standard shifting. However, this type of backwards shifting will not be possible in the domain we are concerned with, namely modality.

4.2.6.2 Granularity

Another way to view imprecision is in terms of granularity. Krifka (2007) does this, and Sauerland and Stateva (2011) develop a mostly semantic theory of granularity on this basis. This can be seen most clearly with numerals. Depending on what the pragmatic context says is appropriate, different granularities or resolutions can be used.

- (47)
- a. very low precision: [0 ... 50 ... 100 ...]
 - b. low precision: [... 30 ... 40 ... 50 ... 60 ... 70 ...]
 - c. medium precision: [... 40 ... 45 ... 50 ... 55 ... 60 ...]
 - d. high precision: [...48 ... 49 ... 50 ... 51 ... 52 ...]
 - e. very high precision: [... 49.8 ... 49.9 ... 50.0 ... 50.1 ... 50.2 ...]

On this view the expressions *fifty* can be taken to denote the interval from 25 to 74, as in context (47a), the interval from 45 to 54, as in context (47b), and so on, all the way to denote simply the point ‘50’ on the number line, in a mathematically maximally precise context. Sauerland and Stateva (2011) cache this out by taking numerals to combine with granularity operators which are functions from numerals to intervals. Thus the operator *gran_{coarse}* maps 50 to the interval [45,54], for example. Approximators and slack regulators like *approximately* and *exactly* are treated as overt granularity operators. In this way, the covert *gran* can be thought of as the analogous to the silent degree modifier *pos*, while approximators and slack regulators can be thought of as overt degree modifiers.

¹⁴There are several much richer and more complex pragmatic models on the market; I do not claim the one presented here to be superior. Rather it is sufficient to illustrate my point that imprecision can be modeled as a simple pragmatic effect, and that it can be extended to modals.

The analysis given here is entirely compatible with the spirit of this way of viewing imprecision. On my analysis, the similarity operator \approx_ρ partitions the common ground into equivalence classes, sets of worlds which are indistinguishable for the purposes of the pragmatic context. The size of these partition cells is entirely analogous to granularity. Thus, if we are counting guests at a party, and we are in a low precision context where 47 and 52 fall into the same partition-cell on the number line, then that means a world where there were 47 guests at the party is in the same partition cell as a world where 52 guests were at the party, i.e., $w_{47} \approx_\rho w_{52}$.

w_{45}	w_{46}
w_{49}	w_{50}
w_{95}	w_{96}
w_{99}	w_{100}

(a) Low precision.

w_{44}	w_{45}
w_{49}	w_{50}
w_{94}	w_{95}
w_{99}	w_{100}

(b) Medium precision.

w_{44}	w_{45}
w_{49}	w_{50}
w_{94}	w_{95}
w_{99}	w_{100}

(c) High precision.

Table 4.4: Precision as granularity.

However, there are significant challenges for the theory of imprecision espoused by [Sauerland and Stateva \(2011\)](#). The first is that the theory is built into the semantics, which fails to capture the pragmatic motivations of imprecision. The special properties of precisification cannot be captured in a semantic framework, as has been discussed above. Moreover, Sauerland and Stateva still have to appeal to pragmatics to explain what silent granularity operator is used, as well as to explain why, e.g., non-round numbers and round numbers have different granularity possibilities. What’s more, their view of approximators and slack regulators as granularity operators collapses these two distinct categories into one, which ignores the significant differences between the two categories; see [Lauer \(2012\)](#).

4.2.7 Interim Summary

This section provides a formal pragmatics which derives imprecision. On this view imprecision is not a special facet of interpretation but rather falls out from (a slight variation on) the classic Gricean maxims. My view is largely in the spirit of [Lasersohn \(1999\)](#), [Krifka \(2007\)](#), and [Lauer \(2012\)](#), though it diverges from each in some aspects. My maxim of QUALITY embeds Lasersohn’s view that for an utterance to be pragmatically true in a discourse

it must be close enough to the truth for the conversational purposes at hand, but I eschew his word-level computation of pragmatic halos in favor of a much more general pragmatic view where halos are simply pragmatic inferences that are generated like all other inferences. This however, does not preclude a pragmatic consideration of word-by-word composition especially if such a thing becomes necessary for a proper semantics of slack regulators. My faux-Optimality-Theoretic view of Gricean implicature is inspired by Krifka, but my framework diverges in how interpretations are generated and includes a more robust consideration for the surrounding discourse. Finally, I broadly agree with Lauer’s view that imprecision is driven by the distinctions made relevant or irrelevant by the goals of the participants and a general tendency against complex expressions, but I reject his Declarative Convention in favor of my own Uniform Pragmatic Context constraint.

This view derives three crucial properties of imprecision, or rather of precisification: pedantry, faultless disagreement, and unidirectionality. Pedantry comes about when speakers disagree about the standard of precision, and one interlocutor is forced from a more relaxed standard to an apparently unnecessarily strict standard by the conventions of communication. Inessential disagreement comes about when an utterance is evaluated against two different pragmatic contexts, and thus receives two different truth judgments, despite the speakers potentially having the same belief states. Finally, unidirectionality comes about because of a necessary truth about the lexicon: if an expression is acceptable under one standard of precision it is acceptable under any looser standards of precision, but the reverse is not true. Thus, while some expression may, by their very utterance, require the standard of precision to rise, no expression can have the opposite effect.

In the next section I explain how this theory can apply to modals and explain the kind of modal strength variability witnessed in Sections 1 and 2.

4.3 Imprecise Modality

In this section I show how the pragmatic framework developed above accounts for strength variability in modals. The crucial premise is this: Raising very unlikely or nonstereotypical possibilities is cognitively difficult for both speaker and hearer, and therefore dispreferred by MANNER. Additionally, acting as if those possibilities are *impossible* rather than unlikely is not a large diversion from the truth (the less unlikely, the larger the diversion). Thus,

discussing such possibilities, either as the prejacent of a possibility modal or in the antecedent of a conditional, is like using non-round numbers – they are unnecessary in low-precision contexts, and therefore cause shifts to a high standard of precision.

First I show how the model given above can account for stronger and weaker interpretation of necessity modals in varying pragmatic contexts. Then I show how possibility modals may serve to force a higher standard of precision. Finally I discuss how (Equivocal) Sobel Sequences are accounted for, and how this account is superior to prior analyses.

4.3.1 Strength of Necessity in Context

Recall the basic example of a Lewis sequence in (7) repeated below.

- (48) *It is raining outside.*
- a. *Alice:* I want to go outside, but I don't want to get wet.
 - b. *Bryan:* You have to wear a raincoat.
 - c. *Alice:* No, I don't have to. I could cover every inch of my skin in duct tape.
 - d. *Bryan:* I suppose that's technically true, but you get my drift.

We can now see that this discourse has all the defining properties of precisification discussed above. First, Alice comes across as pedantic in uttering (48c). Second, this appears to be a case of inessential disagreement, since (48d) indicates that Bryan is not factually mistaken – what he said in (48b) was true in light of the parameters of discourse he was operating under at the time, even if he has consented to go along with the new parameters insisted upon by Alice. Finally, unidirectionality; Bryan's concession is forced – he could not easily proceed any other way, except to utterly abandon the discourse.

The treatment of this phenomenon as a case of imprecision is further motivated by the appearance of the same slack regulators that are used in other domains.

- (49) a. Technically not *everyone* was there.
- b. It's technically not flat.
- c. Technically she doesn't *have to* wear a raincoat.
- (50) a. Strictly speaking not *everyone* was there.

- b. It's not strictly speaking flat.
- c. Strictly speaking she doesn't *have to* wear a raincoat.

If these expressions have to do with modulating the standard of precision, it naturally follows that modals like *have to* are sensitive to it. Truth-conditional approximators also can be used for each of these expression types.

- (51) a. Basically everyone was there.
- b. It's basically flat.
- c. You basically have to wear a raincoat.
- (52) a. Practically everyone was there.
- b. It's practically flat.
- c. You practically have to wear a raincoat.

As with numerals in the previous section, what drives imprecision is a combination of low standards for QUALITY and an adherence to MANNER in light of more longwinded possible utterances. In this case, round numbers correspond to simple modal expressions like (48b), where non-round numbers correspond to their more longwinded and qualified counterparts, like (51c), (52c), or (53).

- (53) Assuming you're not going to do anything crazy, you have to wear a raincoat.

The tableaux in Tables 4.6a and 4.6b illustrate this pragmatic inference. Let the following semantic formula be represented as below.

- (54) a. $\phi = \lambda w[\text{alice wears a raincoat in } w]$
- b. $\Box\phi = \lambda w[\forall v \in \text{BEST}_{g(w)}(\cap m(w))[v \in \phi]]$
where m is a circumstantial modal base and g is a teleological ordering
- c. $\Delta\phi = \lambda w[\forall v \in \text{BEST}_{g(w),h(w)}(\cap m(w))[v \in \phi]]$
where m is a circumstantial modal base, g is a teleological ordering, and h is a stereotypical ordering
- d. $\Diamond\phi = \lambda w[\exists v \in \text{BEST}_{g(w)}(\cap m(w))[v \in \phi]]$
where m is a circumstantial modal base and g is a teleological ordering

Based on these building blocks, consider four possible propositions, and their shorthands, which I will use to represent both linguistic expressions and

information states.

- (55)
- a. $\Box\phi$: The speaker believes that Alice wears a raincoat in all the worlds where she achieves her goals, consistent with what circumstances already hold; i.e., the strict semantic interpretation of *Alice has to wear a raincoat*.
 - b. $\Delta\phi$: The speaker believes that Alice wears a raincoat in all the reasonable worlds where she achieves her goals, consistent with the circumstances; i.e., the strict semantic interpretation of *Alice basically has to wear a raincoat*.¹⁵
 - c. $\neg\Delta\phi\&\Diamond\phi$: The speaker believes that Alice doesn't wear a raincoat in *all* the reasonable worlds where she achieves her goals, just some of them; i.e., the strict semantic interpretation of *Alice could wear a raincoat but she doesn't have to*.
 - d. $\Box\neg\phi$: The speaker believes that Alice wears a raincoat in none of the worlds where she achieves her goals, consistent with what circumstances already hold; i.e., the strict semantic interpretation of *Alice could not wear a raincoat*.

Now consider the tableau in Table 4.6a for a context with a low standard of precision, and compare to Table 4.6b, where the standard is high. The necessity claim *you have to wear a raincoat* is optimal if the speaker knows it to be literally true, but also if she only knows a more limited claim like *you basically have to wear a raincoat*. On both information states, the latter claim does not lose out on QUANTITY, since at a low standard of precision the difference in strength between the two is negligible. But on both information states the modified version loses out on MANNER.

On a higher standard of precision (Table 4.6b), under information state $\Delta\phi$, the unmodified necessity assertion is filtered out by QUALITY, since it diverges too greatly from the information state $\Delta\phi$. Thus, on a higher standard of precision, the hearer is able to infer $K = \Box\phi$ from $F = \Delta\phi$. On a low standard, an unmodified necessity assertion does not violate QUALITY in the information state $\Delta\phi$ because worlds where $\Delta\phi$ are true are not distinguishable from worlds where $\Box\phi$ are true. This is because the worlds excluded from the domain of Δ but not \Box are precisely the worlds that are most easily ignorable,¹⁶ i.e., the non-stereotypical (or unlikely or unexpected,

¹⁵Or something similar. See Chapter 5 for discussion of *basically*.

¹⁶That is, accessibility to these worlds is most easily ignorable.

	$\langle F = \Box\phi; \sigma; \rho_{low} \rangle$	QL	QN	M
\mathbb{E}	$\langle F = \Box\phi, K = \Box\phi \rangle$			
	$\langle F = \Delta\phi, K = \Box\phi \rangle$!*
	$\langle F = \Diamond\phi, K = \Box\phi \rangle$!*	
	$\langle F = \Box\neg\phi, K = \Box\phi \rangle$!*		
\mathbb{E}	$\langle F = \Box\phi, K = \Delta\phi \rangle$			
	$\langle F = \Delta\phi, K = \Delta\phi \rangle$!*
	$\langle F = \Diamond\phi, K = \Delta\phi \rangle$!*	
	$\langle F = \Box\neg\phi, K = \Delta\phi \rangle$!*		
	$\langle F = \Box\phi, K = \neg\Delta\phi \& \Diamond\phi \rangle$!*		
	$\langle F = \Delta\phi, K = \neg\Delta\phi \& \Diamond\phi \rangle$!*		*
\mathbb{E}	$\langle F = \Diamond\phi, K = \neg\Delta\phi \& \Diamond\phi \rangle$		*	
	$\langle F = \Box\neg\phi, K = \neg\Delta\phi \& \Diamond\phi \rangle$!*		
	$\langle F = \Box\phi, K = \Box\neg\phi \rangle$!*		
	$\langle F = \Delta\phi, K = \Box\neg\phi \rangle$!*		*
	$\langle F = \Diamond\phi, K = \Box\neg\phi \rangle$!*	*	
\mathbb{E}	$\langle F = \Box\neg\phi, K = \Box\neg\phi \rangle$			
$\llbracket \text{You have to wear a raincoat} \rrbracket^{\sigma, \rho} = \Box\phi \vee \Delta\phi = \Delta\phi$ $\llbracket \text{You basically have to wear a raincoat} \rrbracket^{\sigma, \rho} = \emptyset$ $\llbracket \text{You could wear a raincoat} \rrbracket^{\sigma, \rho} = \neg\Delta\phi \& \Diamond\phi$ $\llbracket \text{You could not wear a raincoat} \rrbracket^{\sigma, \rho} = \Box\neg\phi$				

(a) Low precision context.

	$\langle F = \Box\phi; \sigma; \rho_{hi} \rangle$	QL	QN	M
\mathbb{E}	$\langle F = \Box\phi, K = \Box\phi \rangle$			
	$\langle F = \Delta\phi, K = \Box\phi \rangle$!*	*
	$\langle F = \Diamond\phi, K = \Box\phi \rangle$!**	
	$\langle F = \Box\neg\phi, K = \Box\phi \rangle$!*		
	$\langle F = \Box\phi, K = \Delta\phi \rangle$!*		
\mathbb{E}	$\langle F = \Delta\phi, K = \Delta\phi \rangle$		*	*
	$\langle F = \Diamond\phi, K = \Delta\phi \rangle$!**	
	$\langle F = \Box\neg\phi, K = \Delta\phi \rangle$!*		
	$\langle F = \Box\phi, K = \neg\Delta\phi \& \Diamond\phi \rangle$!*		
	$\langle F = \Delta\phi, K = \neg\Delta\phi \& \Diamond\phi \rangle$!*	*	*
\mathbb{E}	$\langle F = \Diamond\phi, K = \neg\Delta\phi \& \Diamond\phi \rangle$		**	
	$\langle F = \Box\neg\phi, K = \neg\Delta\phi \& \Diamond\phi \rangle$!*		
	$\langle F = \Box\phi, K = \Box\neg\phi \rangle$!*		
	$\langle F = \Delta\phi, K = \Box\neg\phi \rangle$!*	*	*
	$\langle F = \Diamond\phi, K = \Box\neg\phi \rangle$!*	**	
\mathbb{E}	$\langle F = \Box\neg\phi, K = \Box\neg\phi \rangle$			
$\llbracket \text{You have to wear a raincoat} \rrbracket^{\sigma, \rho} = \Box\phi$ $\llbracket \text{You basically have to wear a raincoat} \rrbracket^{\sigma, \rho} = \Delta\phi$ $\llbracket \text{You could wear a raincoat} \rrbracket^{\sigma, \rho} = \neg\Delta\phi \& \Diamond\phi$ $\llbracket \text{You could not wear a raincoat} \rrbracket^{\sigma, \rho} = \Box\neg\phi$				

(b) High precision context.

Table 4.5: Pragmatic derivation of (im)precise inferences with modals.

etc.) worlds.

Since the unmodified necessity assertion wins out under both $K = \Delta\phi$ and $K = \Box\phi$, a hearer cannot infer $\Box\phi$, but rather must infer the disjunction of the two, which is logically equivalent to inferring the weaker knowledge state, $\Delta\phi$. Thus, the pragmatic inference generated by an unmodified necessity assertion on a low standard of precision is equivalent to the strict semantic meaning of the modified necessity assertion. So imprecision achieves an effect not unlike a Kratzer-style stereotypical ordering source, but without explicitly including one in the semantics.

I leave a detailed discussion of the meaning of the modifiers like *basically* for Chapter 5.

4.3.2 Possibility and Precisification

As seen in (48), some uses of possibility modals can have the effect of forcing a high standard of precision. In this way they are like non-round numerals. In (48), the possibility modal in question is used to raise a relatively outlandish possibility to prominence. I argue that like non-round numerals, outlandish possibility assertions are suppressed in low precision contexts by MANNER, which penalizes the cognitive difficulty of processing unexpected/non-stereotypical propositions.

To exemplify this, consider now a slightly different derivation, with the additional formula in (56), the candidate knowledge states in (57), and the additional information (in the semantic context) that ψ -worlds are extremely nonstereotypical in $\cap m(w)$.

$$(56) \quad \psi = \lambda w[\text{alice covers herself in duct tape in } w]$$

- $$(57) \quad \begin{array}{ll} \text{a.} & \Box\phi \text{ } (\rightsquigarrow \text{You have to wear a raincoat.}) \\ \text{b.} & \Delta\phi \& \Diamond\psi \text{ } (\rightsquigarrow \text{You basically have to wear a raincoat but you could cover yourself in duct tape.}) \\ \text{c.} & \neg\Delta(\phi \vee \psi) \end{array}$$

Consider first the high precision context (Table 4.6b). If the speaker knows that $\Box\phi$, she will utter *you have to wear a raincoat*, and it will give rise to exactly that inference. If the speaker knows that there is a possibility (albeit a weak one) besides ϕ , there is a tie and thus two winners.¹⁷ One optimal utterance for the speaker is to utter a weak necessity claim, $F = \Delta\phi$. In Table 4.6b below, ψ stands in for the whole range of possible alternatives that the speaker might believe are possible; thus the inference drawn from an utterance of *you basically have to wear a raincoat* (or something similar) is $\Delta\phi$, but for at least one other ψ , $\Diamond\psi$; or in other words, $\Delta\phi \& \neg\Box\phi$. The

¹⁷This is actually very unlikely in the stated context, since the speaker is also very likely to have beliefs about which possibilities are worth considering and which aren't, and the hearer is likely to know or assume this. Thus, on a knowledge state where the speaker believes $\Delta\phi \& \Diamond\psi$ but also believes that ψ is not a possibility worth raising, $F = \Diamond\psi$ will lose because it will violate QUALITY (since raising a possibility that you believe is not worth raising is a contradiction). But on a knowledge state where the speaker believes that ψ is worth raising, saying just $\Delta\phi$ will violate QUALITY for the same reason; on this knowledge state $F = \Delta\phi \& \Diamond\psi$ will win, as it would on a knowledge state where the speaker does not have any attitudes about the noteworthiness of each possibility. In this competition it would defeat both $\Delta\phi$ and $\Diamond\psi$ on QUANTITY despite taking a penalty on MANNER.

	$\langle F = \Box\phi; \sigma; \rho_{low} \rangle$	QL	QN	M
\mathbb{E}	$\langle F = \Box\phi, K = \Box\phi \rangle$			
	$\langle F = \Delta\phi, K = \Box\phi \rangle$!*
	$\langle F = \Diamond\phi, K = \Box\phi \rangle$!*	
	$\langle F = \Diamond\psi, K = \Box\phi \rangle$!*
\mathbb{E}	$\langle F = \Box\phi, K = \Delta\phi \& \Diamond\psi \rangle$			
	$\langle F = \Delta\phi, K = \Delta\phi \& \Diamond\psi \rangle$!*
	$\langle F = \Diamond\phi, K = \Delta\phi \& \Diamond\psi \rangle$!*	
	$\langle F = \Diamond\psi, K = \Delta\phi \& \Diamond\psi \rangle$!*
	$\langle F = \Box\phi, K = \neg\Delta(\phi \vee \psi) \rangle$!*		
	$\langle F = \Delta\phi, K = \neg\Delta(\phi \vee \psi) \rangle$!*		*
\mathbb{E}	$\langle F = \Diamond\phi, K = \neg\Delta(\phi \vee \psi) \rangle$			
	$\langle F = \Diamond\psi, K = \neg\Delta(\phi \vee \psi) \rangle$!*
$\llbracket \text{You have to wear a raincoat} \rrbracket^{\sigma, \rho} = \Box\phi \vee \Delta\phi = \Delta\phi$ $\llbracket \text{You basically have to wear a raincoat} \rrbracket^{\sigma, \rho} = \emptyset$ $\llbracket \text{You could wear a raincoat} \rrbracket^{\sigma, \rho} = \neg\Box\phi \& \Diamond\phi$ $\llbracket \text{You could cover yourself in duct tape} \rrbracket^{\sigma, \rho} = \emptyset$				

(a) Low precision context.

	$\langle F = \Box\phi; \sigma; \rho_{hi} \rangle$	QL	QN	M
\mathbb{E}	$\langle F = \Box\phi, K = \Box\phi \rangle$			
	$\langle F = \Delta\phi, K = \Box\phi \rangle$!*	*
	$\langle F = \Diamond\phi, K = \Box\phi \rangle$!**	
	$\langle F = \Diamond\psi, K = \Box\phi \rangle$!*		
	$\langle F = \Box\phi, K = \Delta\phi \& \Diamond\psi \rangle$!*		
\mathbb{E}	$\langle F = \Delta\phi, K = \Delta\phi \& \Diamond\psi \rangle$			*
	$\langle F = \Diamond\phi, K = \Delta\phi \& \Diamond\psi \rangle$!**	
\mathbb{E}	$\langle F = \Diamond\psi, K = \Delta\phi \& \Diamond\psi \rangle$			*
	$\langle F = \Box\phi, K = \neg\Delta(\phi \vee \psi) \rangle$!*		
	$\langle F = \Delta\phi, K = \neg\Delta(\phi \vee \psi) \rangle$!*	*	*
\mathbb{E}	$\langle F = \Diamond\phi, K = \neg\Delta(\phi \vee \psi) \rangle$			
	$\langle F = \Diamond\psi, K = \neg\Delta(\phi \vee \psi) \rangle$!**
$\llbracket \text{You have to wear a raincoat} \rrbracket^{\sigma, \rho} = \Box\phi$ $\llbracket \text{You basically have to wear a raincoat} \rrbracket^{\sigma, \rho} = \Delta\phi \& \neg\Box\phi$ $\llbracket \text{You could wear a raincoat} \rrbracket^{\sigma, \rho} = \neg\Delta\phi \& \Diamond\phi$ $\llbracket \text{You could cover yourself in duct tape} \rrbracket^{\sigma, \rho} = \Diamond\psi \& \neg\Delta\psi$				

(b) High precision context.

Table 4.6: Pragmatic derivation of (im)precise inferences with modals.

other optimal utterance would be for the speaker to utter *you could cover yourself in duct tape*, which has the usual kind of scalar inference, $\Diamond\psi \& \neg\Delta\psi$. Lastly, in a context where the speaker does not believe of any proposition that it is weakly necessary, she can only utter possibility claims.

To generalize, on a high precision context, these expressions act just like the usual scalar implicatures tell us they act: $O\phi$ means that the speaker believes $O\phi$ and doesn't believe anything stronger than $O\phi$, because of the interplay of QUALITY and QUANTITY; MANNER doesn't factor in.

On a low precision context, however, MANNER becomes ascendant. In a context where the speaker believes that ϕ is weakly necessary, but that there are other very unlikely possibilities, the only winner is $\Box\phi$. Thus, uttering *you have to wear a raincoat* conveys the disjunction of the information states on which such an utterance is optimal, which again simply adds up to $\Delta\phi$. As before, uttering *you basically have to wear a raincoat* does not generate an inference. Neither does the utterance *you could cover yourself in duct tape*.

Returning to (48), we see that the analysis of a disagreement about the standard of precision again captures the data. Even though Bryan and Alice

share the same knowledge state, because they have different standards of precision they believe different utterances are optimal. Bryan utters (48b) in accord with his standard of precision. Alice, however, evaluates it against her standard of precision, and sees that it is false. Since Alice cannot allow the common ground to be updated with information that crashes it, she utters the negation of the same sentence, and follows up with another sentence which provides support for her denial. I will not provide a pragmatic analysis of the reasoning which leads to these two utterances, but it should be fairly clear that these are broadly pragmatically motivated.

Crucially, however, Alice's utterance creates an even more problematic situation for Bryan. Whereas his utterance crashed Alice's common ground because it was false (a situation that can be rectified by denial), Alice's utterance created a total communicative failure for Bryan, because on the standard of precision he believed to be in effect, Alice's utterance simply does not generate an inference. In order to rectify the situation he has no choice but to adopt a higher standard of precision, thereby allowing for an inference to be generated and bringing the pragmatic (and semantic) context into accord.¹⁸

This analysis thus captures unidirectionality, since as above, no utterances can force a shift to a lower standard of precision, due to the natural and necessary asymmetry in the lexicon. Inessential disagreement is captured since we can posit Bryan and Alice as starting with the same knowledge states, but still deciding on different optimal utterances.

Pedantry is captured simply by the fact that Alice has a higher standard of precision than Bryan believes appropriate. Similarly, Alice might have an intuition of ignorance or carelessness on the part of Bryan due to his having a lower standard of precision than she believed appropriate.

¹⁸Note that Bryan could just as well see that Alice intended a higher standard of precision (since it is the only conversational configuration that allows for an inference to be generated) but reject the move to a higher standard. This would involve either rejecting the conversation entirely, or muddling along with a kind of defective conversation.

4.4 Conditionals



Figure 4.1: Munroe (2013)

As Kratzer (1986) argues, conditionals are just cases of modals¹⁹ whose domains are restricted by explicit *if*-clauses, in addition to the usual kind of contextual domain restriction. So it should come as no surprise that modals with *if*-clauses show the same behavior as those without. One difference that conditionals provide is that they make the domain, and thus the question of which possibilities are being ignored and which aren't, slightly more explicit.

The canonical illustration of this is the Sobel Sequence (Lewis, 1973, Sobel, 1970), which has been used to argue that (counterfactual) conditionals are not monotonic on their domains. Sobel Sequences are typically taken to be sequences of counterfactuals of the type given below, where $\Box \rightarrow$ is the counterfactual operator. An actual example commonly pointed to is in (59).

- (58) *Sobel Sequence schematic*
 $p \Box \rightarrow q$, [but] $((p \ \& \ r) \Box \rightarrow \neg q)$
- (59) If the U.S. **threw** all its nuclear weapons into the sea, there **would be** war; but if all nations with nuclear weapons **threw** them into the sea, there **would be** peace.

It is worth pointing out, however, that (59) is arguably not a counterfactual conditional, but rather, a future-less-vivid conditional (Iatridou, 2000), indicated by the lack of *have* and its attendant verbal morphology in the antecedent clause. This illustrates a broader point, which is that Sobel

¹⁹Or other sentential operators.

Sequences may occur with all manner of non-counterfactual conditionals (Williams, 2008; Moss, 2012; Willer, 2013b). (Despite this, nearly all discussion of Sobel Sequences has centered on counterfactuals.) A true counterfactual version is given in (60), and an indicative/predictive version given in (61).

- (60) If the U.S. **had thrown** all its nuclear weapons into the sea, there **would have been** war; but if all nations with nuclear weapons **had thrown** them into the sea, there **would have been** peace.
- (61) If the U.S. **throws** all its nuclear weapons into the sea, there **will be** war; but if all nations with nuclear weapons **throw** them into the sea, there **will be** peace.

As with Lewis Sequences, Sobel Sequences show that certain possibilities may be ignored in some contexts and not in others: In the first sentence in (60), the possibility that all the nations besides the United States would throw their nuclear weapons into the ocean can be ignored while in the second it cannot. And as with Lewis Sequences, mentioning such ignorable possibilities has the effect of (in descriptive terms) unidirectionally shifting the domain of admissibility to include them. This should be enough to show that Sobel Sequences and Lewis Sequences are related phenomena. However, conditionals, and especially counterfactuals, introduce many complications.

One important difference between the two is that in Lewis Sequences, the utterance that expands the domain of admissibility is a possibility claim which is also a denial of the previous necessity claim. In Sobel Sequences, however, there is no explicit denial; in fact, in the classic case, the two sentences are spoken in sequence. Edgington (1995) goes as far as to argue that counterfactual Sobel Sequences constitute “single pointful pieces of discourse” and not, e.g., a case of a speaker changing her mind and retracting a previous statement. I refer to this property as **unequivocality**. Edgington’s judgment of unequivocality is echoed by von Fintel (2001) and Gillies (2007).

However, Sobel Sequences can be modified into the style of Lewis Sequences.

- (62) a. *Bryan*: If the US threw all of its nuclear weapons in the sea, there would be war.
- b. *Alice*: Not so; if the US threw all its weapons in the sea, and

then everyone else did, there would not be war.

One thing that makes this still not quite like the Lewis Sequences is that counterfactual conditionals obey the Principle of the Excluded Middle (Stalnaker, 1984, von Fintel, 1997); though there is some debate on this point (see e.g., Lewis (1973)). If this is true, Alice’s utterance in (62b) cannot be taken as a true *denial* of (62a) the way that her rebuttal in the duct tape case is a true denial. If in some of the relevant worlds where the US throws its nuclear weapons in the sea, there is war, but in some there is not, then (62a) is seemingly neither true nor false. This may contribute to the notion that Sobel Sequences are not cases of equivocation, because the second sentence doesn’t make the first one false; just odd.

In fact, I argue that *some*, but not all, Sobel Sequences are actually infelicitous if uttered as a “single pointful piece of discourse”, and that to be felicitous they must represent some kind of change of heart on the part of the speaker, or disagreement between speakers in the case of the two-speaker variant like (62). I call these Equivocal Sobel Sequences. Returning to (4), repeated as (63a-b), consider that after (63b) is uttered, there is a feeling that looking back on (63a) in retrospect, it is odd.

- (63) a. *Julia*: If Karlos had come to the party, it would have been a good time.
 b. *Jon*: If Karlos had come to the party and jumped off the balcony, it would not have been a good time.
 c. *Jackson*: So what Julia said was a bit too strong.

von Fintel argues that if (59) is uttered “someone else can then rejoin that the initial conditional is ‘no longer’ true” (von Fintel, 2001, p. 131). But he doesn’t intend to say that the initial conditional is actually false in retrospect, rather that “the parameters of the discourse have changed so that the proposition expressed by the first counterfactual in the initial context can no longer be expressed by the same linguistic expression in the new context”. In other words, if (63a) is uttered again after (63b) has been, it simply does not denote the same proposition any more.²⁰ This is another way of saying that uttering (63b) involves ‘changing the conversational score’. As with

²⁰von Fintel’s theory is a dynamic one, so he is committed to the notion that the expression denotes the same context change potential. But since the context has changed in a way that the CCP denoted by the expression is sensitive to, the update that it effects is in some sense different.

Lewis Sequences, I argue that this is not really what is happening; rather the conversational participants are converging on an aspect of the score which they had previously, implicitly, disagreed about; namely the standard of precision. However, there is a significant empirical muddle to be resolved before this point can be made. In this section I argue that Sobel Sequences can in fact be divided into two distinct phenomena with distinct explanations; Unequivocal Sobel Sequences, which can be handled by a version of Lewis’s classical semantics, and Equivocal Sobel Sequences, which are special cases of Lewis Sequences and require a pragmatic approach.

4.4.1 Two Kinds of Sobel Sequence

Prior discussions of Sobel Sequences (Lewis, 1973, Warmbrod, 1981a,b, 1983, Edgington, 1995, von Fintel, 2001, Gillies, 2007, Moss, 2012) and Reverse Sobel Sequences (von Fintel, 2001, Gillies, 2007, Moss, 2012) have generally presented the following basic facts:

1. Unequivocality: A basic Sobel Sequences is a “single pointful piece of discourse”
2. Unidirectionality: Reverse Sobel Sequences are infelicitous

Unequivocality is not predicted by the most simple of counterfactual semantic accounts, but is famously handled by Lewis (1973), Kratzer (1981b) and Stalnaker (1984), who add an ordering component to the semantics. This analysis, called the Variably Strict Conditional analysis, says that a counterfactual $p \Box \rightarrow q$ is true if q is true in all the worlds in which p is true *which are closest to the evaluation world*. A formal representation can be given for Lewis’s analysis, adapted to the Kratzerian view on modality, below.²¹

$$(64) \quad \llbracket \text{would} \rrbracket = \lambda p \lambda m \lambda g \lambda w [\forall v \in \max_{g(w)}(\cap m(w))[p(v)]]$$

iff g is a closeness ordering and m is an empty modal base

This accounts for unequivocality because the two conditionals in a Sobel Sequence will be quantifying over two wholly disjoint sets of worlds. However, neither account can handle unidirectionality, predicting that Reverse Sobel

²¹With the added assumption that the antecedent will be added to the modal base through compositional means.

Sequences should be just as good as Sobel Sequences, since the two domains will be disjoint regardless of the order of utterance.

Thus recent accounts have tried to accommodate both properties of Sobel Sequences. However, this is in part misguided because prior accounts have failed to distinguish two distinct varieties of Sobel Sequences: what I call Equivocal Sobel Sequences (ESSs) and Unequivocal Sobel Sequences (USSs). USSs have the property of unequivocality, but not unidirectionality, while ESSs have the reverse. Thus, there does not need to be any attempt accommodate both properties in a single analysis. An analysis along the lines of Lewis or Stalnaker is indeed sufficient to account for USSs, while the pragmatic account needed for non-conditional sentences above can also account for ESSs.

4.4.1.1 Unequivocal Sobel Sequences

USSs behave as “single pointful pieces of discourse” as [Edgington \(1995\)](#) claims all Sobel Sequences do. Thus, uttering the second conditional in a USS does not invalidate the first – it neither requires the speaker to retract or qualify the first conditional, nor do they give rise to a retrospective intuition of falsity with regard to the first conditional.

- (65) *Construction workers Daryl, Aaron, and Ida, stand around a construction site. Daryl is not wearing a helmet. A large beam falls from above them and lands where no one was standing, but near to Daryl.*
- a. *Aaron:* Daryl, if you had been standing there, you would have been killed.
 - b. *Ida:* And if you had been standing there and wearing a helmet, you would not have.
 - c. *Aaron:* Exactly.

In this example, Ida is not seen as contradicting Aaron at all – in fact she is supporting his implicit argument (that Daryl should wear a helmet). Aaron is in no way compelled to retract his assertion, nor does his assertion seem false in retrospect. Its interpretation is the same in retrospect as it was in the moment: If Daryl had been standing there and *everything else was like it was in reality*, he would have been killed.

Another example:

- (66) *Karlos is known for being fun at parties. But his house is small and smelly.*
- a. *Martina:* If Karlos had come to the party, it would have been a good time.
 - b. *Ben:* But if Karlos had hosted the party, it would not have been a good time.
 - c. *Martina:* ...Sure.

Since *Karlos hosts the party* entails *Karlos comes to the party*, this is another case of strengthening the antecedent. But again Ben's comment does not contradict Martina's or cause her to retract or qualify her assertion. Once again, her assertion seems to have the same meaning before and after Ben's utterance, namely if Karlos had come to the party and everything else was the same, it would have been a good time. This of course requires knowing what the party was like in reality.

It is worth pointing out that Ben's comment may seem like a non-sequitur, depending on what the larger question under discussion is. For example, if the party was in reality not enjoyable, and the party's organizers (who all live together) are contemplating after the fact what could have made the party better, Martina might utter (66a), its relevance being the implicit inference that Karlos should have been invited to the party. But if Ben utters (66b), while true and not contradictory, it may seem irrelevant if the possibility of changing the location of the party was not really a 'live option'. Moreover, it does not relate to the previous utterance – Karlos's coming to the party is causally independent of his hosting the party.

Thus, USSs, as advertised, display unequivocality. Unidirectionality is slightly harder to diagnose. On its face, it may seem that USSs display unidirectionality as well.

- (67) a. *Ida:* If you had been standing there and wearing a helmet, you wouldn't have been killed.
- b. *Aaron:* # But if you had been standing there, you would have been killed.

In this case, reversing the order causes Aaron's assertion to be not just false, but infelicitous. One possible explanation for the infelicity is that Aaron is contradicting Ida's assertion. This could be because *would* is sensitive to modal subordination, and thus, (67b) could be interpreted as meaning that

if Daryl had been standing there *and wearing a helmet* he would have been killed, i.e., a direct contradiction of what Ida asserts in (67a).

Another possibility is that (67b) is not false at all, but simply infelicitous, because there is a pressure to stress the antecedent of (67b) contrastively with (67a). Since the antecedent of (67b) is syntactically a subset of the antecedent of (67a), this is impossible just with prosody. The use of an exhaustive operator could suffice, as in (68).

- (67) b'. *Aaron*: But if you had *just* been standing there, you would have been killed.

This arguably defeats the purpose since the exclusion of worlds where Daryl wears a helmet is now achieved just by the semantics of the antecedent itself.

Consider the reverse of (66).

- (68) a. *Ben*: If Karlos had hosted the party, it would not have been a good time.
 b. *Martina*: But if Karlos had *come* to the party, it would have been a good time.

Here simple contrastive stress is possible, and the reverse sequence is acceptable. However, it could be argued that (68b) also carries an exhaustive interpretation as well, where the antecedent alone means *Karlos comes to the party but does not host it*. This can be remedied if the Reverse Sobel Sequence overlaps a proper Sobel Sequence.

- (69) *Karlos is known for being fun at parties. But his house is small and smelly.*
 a. *Martina*: If Karlos had come to the party, it would have been a good time.
 b. *Ben*: But if Karlos had *hosted* the party, it would not have been a good time.
 c. *Martina*: Sure, but what I said is still right: If Karlos had come to the party, it would have been a good time.

This strategy works for reversing (65) too.

- (70) *Construction workers Daryl, Aaron, and Ida, stand around a construction site. Daryl is not wearing a helmet. A large beam falls from above them and lands where no one was standing, but near to*

Daryl.

- a. *Aaron:* Daryl, if you had been standing there, you would have been killed.
- b. *Ida:* And if you had been standing there and wearing a helmet, you would not have.
- c. *Aaron:* Exactly. But what I said is still right: If you had been standing there, you would have been killed.

Similar sentences are pointed out by Moss (2012), though her explanation is different (see below).

It is still certainly true that in the simplest cases, reverse USSs are degraded, and while I have speculated about a possible explanation for these judgments, my theory ultimately has nothing to say about them. The most important point I would like to make is that reverse USSs are a totally different animal from reverse ESSs and require a different explanation, where the explanation of the latter is my pragmatic account. This says nothing about the debate about, e.g., Strict versus Variably Strict theories of the semantics of counterfactuals. We may regard USSs as the phenomenon that most authors have been attempting to explain; after all, the unequivocality of USSs is what weighs heaviest on the debate over Strict and Variably Strict theories of counterfactuals. However, the existence of ESSs and their similarity to USSs have muddled the debate about the latter. Presently, I attempt to unmuddle.

4.4.1.2 Equivocal Sobel Sequences

ESSs do not behave as “single pointful pieces of discourse”. They give rise to a feeling of contradiction, or at least equivocation, and require a retraction or qualification on the part of the original asserter.

(71) *Construction workers Daryl, Aaron, and Ida, stand around a construction site. Daryl is not wearing a helmet. A large beam falls from above them and lands where no one was standing, but near to Daryl.*

- a. *Aaron:* Daryl, if you had been standing there, you would have been killed.
- b. *Ida:* But if he had been standing there and he saw a the shadow of the falling beam and managed to jump out of the way in time,

- he would not have.
- c. *Aaron*: #Exactly.

Even if Ida's second premise, that Daryl detects the falling beam and evades it, is very unlikely given the first, this is enough to force Aaron to retract his prior statement; he cannot respond as he did in (65), because there is a retrospective intuition of infelicity with regard to (71a).

- (71) c'. *Aaron*: Well, okay, I guess you're right. But he should still wear a hardhat.

Another example:

- (72) *Karlos is known for being fun at parties. But his house is small and smelly.*
- a. *Martina*: If Karlos had come to the party, it would have been a good time.
- b. *Ben*: But if Karlos had come to the party and fallen off the balcony, it would not have been a good time.
- c. *Martina*: Well, alright, but he probably wouldn't have done that – my point is, we should have invited him.

As mentioned above, neither conditional in (72) is strictly false since conditionals obey the Principle of the Excluded Middle. However, crucially, they are too strong and must be retracted; thus ESSs do not display unequivocality.

But they do display unidirectionality.

- (73) *Karlos is known for being fun at parties. But his house is small and smelly.*
- a. *Martina*: If Karlos had come to the party, it would have been a good time.
- b. *Ben*: But if Karlos had come to the party and fallen off the balcony, it would not have been a good time.
- c. *Martina*: #Well, alright, but what I said is still right: If Karlos had come to the party, it would have been a good time.

Since (73b) is a refutation of (73a), clearly (73c) cannot be felicitously uttered in response. Thus even if “well, alright, but what I said is still right” serves

to reset the context in some way, the resetting of the context is not enough to save ESSs.

Note that while (66b) sounds like a non-sequitur, the second part of ESSs never do – this is supporting evidence for the claim that ESSs always involve a kind of equivocation. Since the second part of ESSs are always contrary to the first, they are also always relevant, since they are (implicit) denials of the (presumably relevant) first part.

Next I explain why USSs and ESSs have the distinct behavior that they do.

4.4.2 Similarity and Causality

The single feature which determines whether a Sobel Sequence will be a USS or an ESS is whether or not there is a causal relationship between the two antecedents in the second conditional. In USSs there is no causal relationship, while in ESSs there is. This is predicted by a very conservative semantics for counterfactuals.

The semantics for counterfactuals (i.e., for *would*) I adopt is indeed a conservative one, and I so do not probe issues relating to their dynamic properties relating to modal subordination (see e.g., Asher and McCready, 2007, Klecha (2011)), or advanced issues relating to their composition, particularly their relationship to past tense (see, e.g., Arregui (2009)). Rather I focus on the question of how their domain is determined, the central question bearing on this study.

An analysis following from Lewis (1973), and Kratzer (1981b, 1981a, 2012) gives *would* the following semantics.

$$(74) \quad \llbracket \text{would} \rrbracket = \lambda p \lambda m \lambda g \lambda w [\forall v \in \max_{g(w)}(\cap m(w)) [p(v)]]$$

iff g is a closeness ordering and m is an empty modal base

I assume that *would* is obligatorily conditional, though it does not always need to take an overt *if*-clause. See Klecha (2011). I also assume that the presupposition that m be an empty modal base hold before conditionalization, thus, $\cap m(w)$ in the above formula will always simply denote the antecedent.

The biggest question, then, is: What is a closeness ordering? A simple implementation of Lewis’s (1973) analysis would be to take “closeness” as simply similarity. Consider (75).

$$(75) \quad \text{If Gore had won in 2000, the U.S. would not have invaded Iraq.}$$

On this view, sentences like (75) are supposedly accounted for in the following way. Begin by taking the set of worlds in which the antecedent is true (as determined by the modal base). Then take the set of worlds most similar to the evaluation world which are also in the modal base (the effect of the ordering source). Since no world can be more similar to the evaluation world than itself, if the antecedent is true, this will return only the evaluation world and the truth value of the whole counterfactual will simply be the truth value of the consequent. But if it is false (the more typical case), the counterfactual will be true in the set of worlds which diverge from the evaluation world in terms of the antecedent, but otherwise as little as possible.

However, this does not quite suffice, as [Bennett \(2003\)](#) explains. This is because a world where Gore wins in 2000 and the United States does not go to war with Iraq is probably more different from the evaluation world than one where Gore is actually a Republican and behaves just like George W. Bush did in reality, and thus the Iraq War, with all its geopolitical consequences, happens just as it did in reality. In other words, any counterfactual with a consequent which entails a big difference with the evaluation world is predicted to be false, which is of course a bad result.

[Bennett's \(2003\)](#) solution (see also [Arregui \(2009\)](#) for a compositional account) is to argue that our closeness measure must be indifferent to matters which follow in a causal chain from the antecedent eventuality. A simplified version of Bennett's proposal says that the closeness of two worlds is just their similarity in all matters except those which pertain to the antecedent and what follows causally from the antecedent. In other words, if two worlds differ in a way that has nothing to do with the antecedent p , then they will count as distant for that reason. But if they differ in terms of q , which is due to a causal chain begun by p , and which is true in only one of those worlds, this has no effect on similarity.

So worlds where Al Gore is actually a Republican count as distant because his being a Republican is not causally associated with his winning in 2000 (or in any case, we can imagine worlds where he wins despite being a Democrat). But the non-occurrence of the Iraq War, which follows causally from Al Gore's winning the election, does not cause those worlds to count as distant from the evaluation world. Thus, (75) is (plausibly) true, because worlds where Al Gore is a Republican are successfully kept out of the domain.

The unequivocality of USSs is neatly accounted for by this analysis, as [Lewis \(1973\)](#) originally argued. Consider again a typical USS.

- (76) *Construction workers Daryl, Aaron, and Ida, stand around a construction site. Daryl is not wearing a helmet. A large beam falls from above them and lands where no one was standing, but near to Daryl.*
- a. *Aaron:* Daryl, if you had been standing there, you would have been killed.
 - b. *Ida:* And if you had been standing there and wearing a helmet, you would not have.
 - c. *Aaron:* Exactly.

When we evaluate the truth of (76a), we consider worlds where Daryl was standing under the falling beam, but which are otherwise just like the evaluation world *except* in all ways that proceed causally from the counterfactual premise, Daryl's standing under the beam. Thus, the fact that he is killed in such worlds (and that his mother grieves him, and that his employer replaces him, etc.) does not count against including them in the modal domain. But we *do* exclude worlds where Daryl is standing under the beam and is wearing a helmet, because wearing a helmet is causally independent of the standing under the beam. Such a divergence from the facts of the evaluation world *does* count as being unlike the evaluation world.

On the other hand, in (76b), his wearing a helmet is explicitly included in the antecedent, and so the set of worlds considered in the domain is different (and in fact entirely non-overlapping). Since these two expressions quantify over different worlds, there is no reason to expect any kind of entailment relations between them; thus, unequivocality of USSs is derived. Moreover, nothing about this theory predicts unidirectionality, and apparent unidirectionality effects can be explained, as in the previous section.

Crucially, however, this theory does *not* predict that ESSs should be unequivocal, because ESSs differ from USSs precisely in terms of the causal relationship between the premises. Whereas Daryl's wearing the helmet does not causally proceed from his standing beneath the beam, and Karlos's hosting the party does not causally proceed from his coming to it, Daryl's jumping out of the way of the beam, and Karlos's falling off the balcony do, respectively.

Note here that *causally proceed from* does not mean 'caused by'. Daryl's standing under the beam caused the conditions under which he could have jumped out of the way, or not. Either outcome, avoiding or failing to avoid, would count as proceeding causally from his standing under the beam.

So the relatively conservative semantics adopted here captures USSs, but not ESSs. Both outcomes are desirable, since the behavior seen in ESSs is more like the behavior of Lewis Sequences discussed in the previous section. In fact, this account allows us to neatly explain the difference between ESSs and USSs: ESSs result from strengthening the antecedent with a causally related proposition, while USSs result from strengthening with an unrelated proposition.

Further support for this distinction comes from non-counterfactual conditionals. Sobel Sequences do not only afflict counterfactuals, as was already seen in (59), a future-less vivid conditional (Iatridou, 2000). Williams (2008) points out that they occur in indicative conditionals as well, and Moss (2012) does the same for predictive conditionals; see also Willer (2013b). And while future-less vivid conditionals may have a lot in common with counterfactuals (especially the lexical item *would*), indicative and predictive conditionals presumably have less.

- (77) *Neither Ryan nor Eva know if Karlos went to the party, or if the party was fun.*
- a. *Ryan:* If Karlos went to the party, it was a good time.
 - b. *Eva:* But if Karlos went to the party and fell off the balcony, it was not a good time.
 - c. *Ryan:* #Right, but what I said is still true.
 - c'. *Ryan:* Well, okay, sure, but assuming nothing crazy happened, if Karlos went to the party, it was a good time.
- (78) a. *Ryan:* If Karlos goes to the party, it'll be a good time.
- b. *Eva:* But if Karlos goes to the party and falls off the balcony, it won't be a good time.
 - c. *Ryan:* #Right, but what I said is still true.
 - c'. *Ryan:* Well, okay, sure, but assuming nothing crazy happens, if Karlos goes to the party, it'll be a good time.

Both cases are ESSs; there is a distinct intuition of equivocation, thus unidirectionality.

These non-counterfactual conditionals do not appear in USSs, however.

- (79) a. *Ryan:* If Karlos goes to the party, it'll be a good time.
- b. *Eva:* But if Karlos hosts the party, it won't be a good time.
 - c. *Ryan:* #Right, but what I said is still true.

- c'. *Ryan*: Well, okay, sure, but assuming the party is hosted by Graham, if Karlos goes to the party, it'll be a good time.

Though *Karlos goes to the party* and *Karlos hosts the party* are causally independent from each other, this is still an ESS, as long as the second conjunct is sufficiently outlandish. If it is *so* outlandish as to actually be ruled out as a future possibility, (79b) is simply infelicitous. This provides significant support for the view that there are two varieties of Sobel Sequences; one due to the special semantics of counterfactuals, and one due to very general pragmatic principles.

USSs are derived from the similarity ordering ascribed to the semantics of the counterfactual, particularly the lexical item *would*, which does not occur in predictive or indicative conditionals. ESSs meanwhile are derived from pragmatic principles which should affect all conditionals. The fact that indicative and predictive conditionals undergo ESSs but not USSs is therefore exactly what we should expect.

Another way to frame this distinction is that a Sobel Sequence $p \Box \rightarrow q$ & $p \& r \Box \rightarrow \neg q$ is consistent as long as $p \Box \rightarrow \neg r$ is true, i.e., there are no r worlds in the domain of the simple counterfactual. This is the case for USSs, but not ESSs.²²

- (80) a. If Daryl had been standing there, he would have been killed.
 b. But if he had been standing there and saw his shadow and jumped out of the way, he would not have.
- (81) If Daryl had been standing there, he would not have seen his shadow and jumped out of the way. (false)
- (82) a. If Daryl had been standing there, he would have been killed.
 b. But if he had been standing there and wearing a helmet, he would not have.

²²It's worth pointing out that (79) is questionable, but can be ameliorated by a better contextual set up. Suppose the conversants continue to discuss Daryl's near miss. Aaron mentions that the contracting firm that employs them all is being fined by a government agency for Daryl's forgetfulness, since not wearing a helmet is against regulations. Ida then says she wishes Daryl *had* been standing under the beam. Aaron may then respond with (i).

- (i) Why? If Daryl had been standing there, he (still) would not have been wearing a helmet.

- (83) ?If Daryl had been standing there, he would not have been wearing
a helmet. (true)

So the cut between ESSs and USSs may seem unsurprising. After all, any theory of conditionals as sophisticated as Lewis’s will validate this basic point – the Sobel Sequence is truly consistent only if the corresponding counterfactual $p \Box \rightarrow \neg r$ is true. But a pragmatic theory is still needed to account for *apparent* consistency of Sobel Sequences when this corresponding counterfactual is not true. This pragmatic theory is discussed below.

4.4.3 Applying Imprecision to Conditionals

The purpose of the previous subsection has been to establish the empirical point that there are two kinds of Sobel Sequences: Those which are unequivocal and can be given a fully semantic account, and those which cannot. This latter variety, ESSs, have several characteristics in common with Lewis Sequences discussed in the previous section.

Both also involve the expansion of the domain of a modal to include worlds yet unconsidered by the speaker of the original sentence. Both involve felicitous equivocation, unlike USSs, which are not equivocal, or cases of misunderstanding like (9), which are not felicitous. And perhaps most compellingly, both are subject to unidirectionality.

ESSs also give rise to intuitions of pedantry, intuitively for the same reason that Lewis Sequences feature pedantry: Because a relatively unlikely eventuality is mentioned, and thus a violation of MANNER incurred, in a context where doing so was not pragmatically necessary. Moreover ESSs feature the type of partial concessions that also occur in Lewis Sequences, seen in (71c) and (72c), which diagnose inessential disagreement. (USSs, on the other hand, do not feature concessions of any kind.)

Thus ESSs behave unlike USSs, and just like LSs, in featuring unidirectionality (shown in Section 2), pedantry, and inessential disagreement, while not featuring unequivocality. This more fine-grained empirical view makes an analysis of SSs generally much easier. As argued above, USSs can be accounted for semantically, while ESSs can be accounted for with the pragmatic account developed above, which says that they are, like many other expressions, subject to (im)precision. Hájek (Ms) argues for a very similar view, arguing that false counterfactuals can be treated as true if they are close enough to true (but practically unutterable) counterfactuals. Hájek does not

formalize his view of imprecision, or discuss Sobel Sequences, however.

ESSs can then be accounted for very straightforwardly in the same manner as Lewis Sequences. Consider a simple example again.

- (84) *Karlos is known for being fun at parties. But his house is small and smelly.*
- a. *Martina:* If Karlos had come to the party, it would have been a good time.
 - b. *Ben:* But if Karlos had come to the party and fallen off the balcony, it would not have been a good time.
 - c. *Martina:* Well, alright, but he probably wouldn't have done that – my point is, we should have invited him.

Before Martina's first utterance is made, there is a standing disagreement between her and Ben about what the standard of precision is in this discourse. Both are unaware of it, however. Martina utters (84a) and (justifiedly) believes it to be true because of the standard of precision she is operating under. Her actual belief state is more precisely captured by (85), but she did not utter this, because at the low standard of precision she was assuming, (85) and (84a) are interchangeable as far as QUALITY is concerned, whereas (85) incurs a violation of MANNER.

- (85) If Karlos comes to the party, it almost certainly would have been a good time.

Thus, on Martina's standard, (84a) implicates the disjunction of the strict semantic values of (84a) and (85), which is equivalent to the semantic value of (85).

Ben, however, is operating under a higher standard of precision, so for him the (84a) simply implicates the strict semantic value of (84a); because it is not interchangeable with (85). Ben, therefore, objects in the expected way, falsifying Martina's claim by uttering (84b). For Ben, this is an ordinary disagreement.

For Martina, however, it is extraordinary in the sense that Ben's utterance does not, on the standard of precision under which she operates, even generate an implicature. It fatally violates MANNER because of the extreme unlikelihood or unexpectedness of the second conjunct given the first.²³ Mar-

²³Of course, if it were not very unlikely it would make the first counterfactual of the

tina is forced to re-evaluate her pragmatic parameters. Only by raising to meet Ben’s standard of precision can she (easily) salvage the discourse. In doing so, she, by the UPC, renders her own utterance false, and admits the truth of his. But she of course still recognizes the truth of her original *belief* and the intent behind her original utterance, thus the partial concession and the impression of pedantry on Ben’s part.

Now that this has occurred, the standard of precision has been settled upon at the value Ben initially chose. Neither party can felicitously ignore the possibility that Ben raised, at least not without flagging it in some way. And outside the realm of metalinguistic negotiation (*Hey, let’s stop talking about such ridiculous possibilities*) there is simply no method offered to Martina to lower the standard in the way that Ben raised it.

The analysis of LSs therefore extends naturally to ESSs, but while previous literature addressing a formal analysis of Lewis Sequences is almost non-existent, discussion of Sobel Sequences abounds, and must be addressed.

4.4.4 Previous Discussion

4.4.4.1 Semantic Theories of Sobel Sequences

The first attempt to reconcile the apparent acceptability of Sobel Sequences with the unacceptability of reverse Sobel Sequences was von Fintel (2001), who provided a semi-dynamic²⁴ account of the semantics of counterfactuals. On his account, counterfactuals are interpreted with respect to a particular modal domain determined by the accessibility relation f_σ as determined by the context σ . However, the particular semantics of counterfactuals are such that using the counterfactual updates σ , and thus f_σ , so that the domain it determines is not empty after conditionalization.

$$(86) \quad \text{a. } \textit{Context Change Potential} \\ f_\sigma + \llbracket \text{would} \rrbracket_{KvF}(p)(q)(w) = f_\sigma^q = \lambda w [f_\sigma(w) \cup \{w' : \forall w'' \in q[w' \leq_w w'']\}]$$

sequence more straightforwardly false – there would be a real concession rather than a partial one.

²⁴Von Fintel’s account is semi-dynamic in the sense that he states the semantics of the counterfactual as having two components: a dynamic component which updates the context, and a static component which is interpreted against the new context. It is dynamic in that the new updated context then survives for future utterances within the discourse, but static in that it still is stated in terms of truth-conditions.

- b. *Truth Conditions*

$$\llbracket \text{would} \rrbracket_{KvF}^{\sigma} = \lambda p[\lambda q[\lambda w[\forall v \in f_{\sigma}^q(w) \cap q[p(v)]]]]$$

Thus a discourse-initial counterfactual will be evaluated as a conditional with the domain being the set of worlds closest to the evaluation world in which q is true. The next conditional's domain will start from the previous conditional's domain, the closest q -worlds, as well as with all non- q -worlds which are at least as close, and then iterate the process.

Thus for a Sobel Sequence schematized in (87), the first counterfactual is evaluated against the domain including the closest worlds p -worlds. The second counterfactual is evaluated against the domain used for (87a), expanded to include the nearest r -worlds, and restricted just to r -worlds, thus a totally different domain.

- (87) *Sobel Sequence Schematic*
 a. if had been p , would have been q
 b. if had been p and r , would have been q
- (88) *Reverse Sobel Sequence Schematic*
 a. if had been p and r , would have been q
 b. #if had been p , would have been q

In Reverse Sobel Sequences, the domain is widened out to include p and r worlds right away. Thus when the second sentence is uttered, expansion is trivial and the domain is the same, guaranteeing incompatibility.

Gillies's (2007) fully dynamic framework works similarly, with the context change potential of a counterfactual conditional working to minimally expand the modal domain to include worlds in which the antecedent is true, with his account going further to explain the fact that possibility modals likewise expand the domain.

- (89) *Hegel Sequence* (Gillies, 2007)
 a. If Karlos had come to the party, it would have been a good time.
 b. But if Karlos had come to the party, he might have jumped off the balcony. Then it would not have been a good time.

Willer's (2013b) account, which focuses on indicative conditionals, likewise builds modulation of the modal domain into the semantics of possibility modals, as well as conditionals, which, like on Gillies's analysis, presuppose

the possibility of their antecedents.

I would first like to point out that these accounts, which are properly accounts of USSs alone, cannot be extended to handle ESSs. First, these accounts do not connect to the phenomenon of imprecision. Though the analysis might be expanded to indicative or predictive conditionals, and might with some more wrangling be extended to modals in general, there is little hope that imprecision in quantifiers or definites or numerals can be given a unified account.

There are problems with the narrow account of counterfactuals as well. One is that the domain can be expanded by entirely non-conditional assertions, as pointed out [Moss \(2012\)](#).

- (90) a. *Ida*: Remember when Julia was standing under a falling beam and managed to dive out of the way?
b. *Aaron*: (#)Yes, but if Daryl had been standing beneath that falling beam he would have been killed.

(90b) is felicitous, but only if we derive the inference that Aaron believes that Daryl would not have been able to dive out of the way; in other words, (90a) has the effect of expanding the domain for (90b) to include worlds where people dive out of the way of falling beams. If such worlds were still excluded semantically from the domain, no inference would be generated.

Since these authors put the work of domain expansion in the semantics of counterfactuals, an extension of their analyses cannot account for the fact that here, the domain is expanded by a simple indicative assertion. What's more, discourses just with plain assertions can have the similar properties to these conditionals, also pointed out by [Moss \(2012\)](#).

- (91) a. *Sarah*: That zebra was born with stripes.
b. *Ben*: But cleverly disguised mules are not born with stripes.
(92) a. *Ben*: Cleverly disguised mules are not born with stripes.
b. *Sarah*: #But that zebra was born with stripes.

On my proposal, just mentioning an unlikely proposition is sufficient to generate the inference that the interlocutors are/should be at a higher standard of precision, and thus, alter the domain of worlds that those interlocutors are willing to consider. ([de Jager \(2009\)](#) similarly argues that the raising of possibilities should not be limited to the semantics of any given expression, but rather occurs as a side effect of mentioning the proposition.) So far I

have only examined modal cases, where the worlds that are being ignored are worlds in a particular modal domain, and thus their exclusion affects the interpretation of modal expressions. But the same logic can apply to the common ground itself; we can exclude certain worlds from the common ground because they are so outlandish as to be ignorable. But by mentioning them, we are forced to include them. Here, ‘cleverly disguised mules’, which are of a class so rare as to be generally ignorable, are mentioned, triggering a violation of manner, which forces us to accommodate a higher standard under which we consider worlds where there are cleverly disguised mules, which of course means that the zebra under discussion could also be such a mule.

However, none of the above is a criticism of von Fintel, Gillies, or Willer. As I have argued at length, USSs and ESSs are distinct phenomena which deserve distinct treatments. It can be said that these authors are all truly attempting to account for what I call USSs, and *not* ESSs, since ESSs are a different phenomenon and display different behavior (namely, equivocality) which means that they do not bear on the issue these authors are really after: The debate over strict versus variably strict accounts of conditionals. What’s more, USSs do show a limited kind of unidirectionality, being certainly *easier* in one direction than the other; so von Fintel, Gillies, and Willer are right to pursue an account of USSs as, at least weakly, unidirectional, which is distinct from the account of ESSs presented here. But this is not to say that I have nothing to say about the debate on USSs.

First of all, all of the sequences considered by Gillies (2007) are ESSs by my reckoning; thus, he does not account for (or point out) the fact that the first sentence in an ESS is false in retrospect, which the data in (71), (72), and (73) show. It is a very common point, going back to Lewis (1973), that Sobel Sequences are consistent and can be uttered as a single felicitous piece of discourse. As I have argued this may be true of USSs but not ESSs.

There are two notions of ‘felicitous’ or ‘consistent’ discourse that I believe are confounded in much of the discussion on Sobel Sequences. A discourse like (93) is a felicitous discourse in a sense, namely that both individuals end up on the same page.

- (93) a. *Katie*: Gallagher isn’t at school.
- b. *Asia*: Yes, he is, he just texted me to say he was at his meeting with Chris.
- c. *Katie*: No, he isn’t.

- d. *Katie*: Chris broke his leg so they're meeting at Chris's house.
- e. *Asia*: Oh, okay, I guess he isn't at school.

But there is a sense in which we could say certain components of the discourse are infelicitous. (93a) seems infelicitous to Asia because it is contrary to what she believes. (93b) seems infelicitous to Katie because it is contrary to what she believes, at least the part about Gallagher being at school.

Moreover, if we cut off the conversation after (93c), the discourse would seem as a whole infelicitous to any outside observer (as well as Asia) because Katie rejects Asia's assertion without providing any kind of support. Cutting off after (93b) or (93d) would seem better, but still slightly odd because there is no concession. But taking the whole discourse, we find nothing wrong. Likewise, a discourse like (94) is in some sense felicitous as well.

- (94) a. *Alma*: #The King of France is bald.
- b. *Na'ama*: There is no King of France.
- c. *Alma*: Oh, that guy must have been the president of France.

ESSs are likewise felicitous. Not because their first and second parts are not contradictory but because they involve the implicit retraction of the false first sentence and adoption of the true second sentence. Given all the information we have by the end of the discourse, the speaker seems to have settled on a good information state. What makes ESSs different from the discourses above is that they are compatible with partial concessions, and they give rise to intuitions of pedantry. Both of these things require the SS to be distributed across two speakers in a discourse, unlike the way that Gillies presents them. But single-speaker ESSs still give rise to a feeling of equivocation, especially in contrast to USSs, as shown below.

- (95) a. *Martina*: If Karlos had come to the party, it would have been a good time; but if Karlos had come to the party and fallen off the balcony, it would not have been a good time.
- b. *Martina*: #But what I said before is still right: if Karlos had come to the party, it would have been a good time.
- (96) a. *Martina*: If Karlos had come to the party, it would have been a good time; but if Karlos had hosted the party, it would not have been a good time.
- b. *Martina*: But what I said before is still right: if Karlos had come to the party, it would have been a good time.

Thus the debate on USSs is muddled if what are truly ESSs are considered as part of the empirical basis. [Willer \(2013b\)](#) likewise argues for a semantic account of SSs in indicative conditionals, but I have argued here that indicative conditionals only ever participate in ESSs, never in USSs. So even if I have nothing to say about, for example, the weak unidirectionality of USSs, I do argue that the view of unidirectionality in (U)SSs is distorted by consideration of unidirectionality in ESSs, which is much more robust.

4.4.4.2 Pragmatic Theories

[Moss's \(2012\)](#) account, on the other hand, does attempt a pragmatic approach, and gets much of the generality that von Fintel and Gillies do not. However, there are several problems with her account as well. The first, shared with von Fintel and Gillies, is the lack of a connection between these modal phenomena and precisification generally, particularly the way both give rise to intuitions of pedantry, inessential disagreement, and unidirectionality.

Moss's account is to maintain a Lewis-style approach with an additional pragmatic constraint, much like my theory. Her constraint, EI, says the following:

- (97) It is epistemically irresponsible (and thus infelicitous) to utter sentence S in context C if there is some proposition ϕ and possibility μ such that when a speaker utters S :
- a. S expresses ϕ in C
 - b. ϕ is incompatible with μ
 - c. μ is a salient possibility
 - d. the speaker of S cannot rule out μ

This principle is largely just the classical definition of the Maxim of Quality, which generally says that you should not assert anything which is incompatible with something you cannot rule out. What makes it different is the condition that μ be salient. However, Moss crucially never gives a definition for salience. Moreover, she argues that clauses (c) and (d) of (97) can sometimes be ignored without explaining under what conditions.

My theory, on the other hand, makes much more explicit what the condition on μ is for EI to hold, namely, if μ is not so unlikely (in the relevant modal domain) that it can be ignored in the current pragmatic context. A

proposition can thus be made ‘salient’ if any expression which denotes it is uttered, regardless of what that expression is embedded under. Uttering such a proposition automatically gives rise to the inference that it should be considered in the present context.

The cases where Moss claims the EI does not hold, like (98), are simply USSs by my reckoning.

- (98) a. If Sophie had gone to the parade and been shorter than she actually is, she would have seen Pedro.
 b. But if Sophie had gone to the parade, she would have seen Pedro.

Thus while Moss is right to point to a pragmatic explanation for (E)SSs, she does not really tackle the bigger question: What does it take for a possibility to be ignored for the purposes of the Maxim of Quality? The present study provides an answer to that question, and moreover connects SSs to imprecision generally, which Moss does not attempt.

4.4.4.3 Assumption

Another approach which might be called pragmatic is [de Jager’s \(2009\)](#), which is in many regards the closest to the one presented here. His theory of Sobel Sequences is couched in a larger proposal on awareness and assumption and their interplay with natural language. On de Jager’s view, our typical model of belief and common ground must be revised to be sensitive to the distinction between belief and assumption. A world may be excluded from a belief model because an agent actively believes it does not represent reality, or a world may be excluded from the model simply by assumption. In the latter case, merely mentioning or drawing attention to the proposition will cause the agent to revise the assumption, which may lead to its re-inclusion in the belief model.

On de Jager’s view, assumption can straightforwardly derive Sobel Sequences. We exclude certain worlds from the domain because we assume they are not possible; once they are mentioned they are brought back into play. What’s more, this process is very naturally unidirectional. De Jager’s view very naturally explains non-counterfactual conditionals, non-conditional modals, and non-modal sentences like the ones [Moss \(2012\)](#) brings up. What’s more, he has an extremely well-formalized notion of salience, awareness, and assumption and the dynamics thereof.

However, there are a few drawbacks to [de Jager’s \(2009\)](#) approach. Admirably, de Jager does seem to make the connection between modal cases like Sobel Sequences and precision, mentioning that the account he presents could be extended to cover imprecision generally. However, such an extension is not given explicitly and it is not clear how it could be made. For example, when an expression like *fifty* is uttered imprecisely, this does not necessarily mean that the possibility that 49 is *assumed to be false* or that the conversants are unaware that it is a possibility. In fact, this case is very hard to imagine.

This gets back to the modal cases as well. Clearly assumption and unawareness are very good reasons to ignore a possibility, and imprecision is a very useful mechanism to that end. But they are not the only reasons, as [Willer \(2013a\)](#) points out. We may simply *choose* to ignore a possibility because it is so outlandish that is simply not worth attending to it.

Likewise, a mathematically or scientifically naïve person may actually be unaware that there is no upper limit to the fine-grainedness of continuous quantities, and for example believe that there is no height which is taller than 5.5438 inches and yet shorter than 5.5439 inches. But this is by no means the only use for imprecision, however. We may also be aware of the possibility of finer-grained distinctions, but be unable or unwilling to distinguish them. What’s more, we may even know precise values but find it laborious to talk about them and so simply ignore them.

All of these possibilities are captured under a general theory of imprecision like the one presented here.

4.5 Conclusion

Imprecision plays a crucial role with modal expressions. Though there is debate on this point, many take a typical possible worlds model to involve an infinitude of possible worlds, and even those assume finite sets of worlds will admit that the number of worlds in any linguistically useful set is too large to count.

Imprecision allows us to reconcile this fact with the obvious limitations of human cognition. Lewis and Sobel Sequences serve to illustrate an already obvious fact: we cannot keep track of all possibilities at any given time. The more outlandish a possibility the less likely we are to (be able to) mentally attend to it. Imprecision allows speakers to deal with this fact while still

remaining cooperative and communicative. In most contexts we are licensed to ignore most possibilities and focus on the ones most likely to affect our reasoning.

In this way, the effect of imprecision on modality is identical to its effect in the case of, say, numerals. Humans simply cannot know natural quantities in continuous domains, as illustrated in (28), and imprecision safeguards us from this fact.

This chapter presents aspects of modal strength variability that has been known since at least Lewis (1973) and Lewis (1979). I account for these by proposing an utterly pragmatic account, which complements the usual Lewis-Kratzer semantics for modals. On this view, truly non-gradable, categorical necessity and possibility modals can come to have gradient meanings, where the grade is determined by the pragmatic context. Thus necessity modals can behave like weak necessity modals. On this view imprecision serves a role that might otherwise be filled by a Kratzer-style stereotypical ordering source. This achieves a classical goal of semantic analysis: simplifying the semantics by explaining phenomena with well-motivated pragmatic principles.

Chapter 5

Modification and Modality

The purpose of this chapter is to lay out the empirical landscape of expressions which modify modals (especially modal adjectives and modal verbs and auxiliaries) and explore how these expressions can induce variability in the strength of a modal in a scalar way, even when interacting with modal expressions which have a non-gradable meaning. The essential point here is that modifiers can manipulate the domain of the modal and modulate its strength, albeit through varied means: domain restriction, manipulation of a degree argument, quantification over contexts (and thus domains), and manipulation of the standard of precision through evocation of focus alternatives.

The term *modify* is not terribly well-defined. [Morzycki \(2015\)](#) discusses this at length, pointing several proposals: A modifier is any expression of type $\langle\alpha, \alpha\rangle$, or any expression which combines with something that it is not an argument to, or simply is defined in syntactic terms, as adjectives and adverbs. I will tentatively adopt the first definition here, but nothing much hinges on it; the central of topic of this chapter is how certain expressions (call them ‘modifiers’) combine with non-gradable modals to lend some gradience to their interpretation.

Depending on how exactly this definition goes, many expressions can be called modifiers of modals. [Huitink \(2014\)](#) gives an overview of various expressions that have been discussed in the context of analyzing modal concord, the phenomenon by which two seemingly independent modals co-exist in a sentence with only one apparent semantic exponent.

- (1) We **can legitimately** deny your request.

Both *can* and *legitimately* can appear on their own and provide their own kind of modal meaning, but appear to collapse into a single modal in (1). Huitink terms cases like these ‘true concord’, as opposed to other cases which I discuss below.

But most analyses of these do not actually ascribe any ‘modification’ to expressions like *legitimately*. Geurts and Huitink (2006) argue for the existence of a ‘concord operator’ which takes both expressions, and returns their common meaning, presupposing that they are synonymous. Anand and Brasoveanu (2010), on the other hand, argue that both meanings occur in parallel, i.e., that (1) can be paraphrased as in (2).

- (2) We can deny your request and it is legitimate to deny your request.

In any case modal concord involving adverbs like *legitimately* does not appear to be ‘modal modification’ in the sense that I am interested in – namely, cases where the adverbial expression either denotes a function from a modal meaning to modal meaning, or serve as an input to such a meaning, ultimately strengthening or weakening the original modal meaning; so I will not discuss them any further.

Another set of expressions discussed by Huitink (2014) are what can be called flavor-specifiers, like the adverbs below.

- (3) You are **legally required to** submit your taxes by April 15th.

These expressions, discussed by Huitink (2012) among others, can simply be taken to specify the modal base and/or ordering source, given Kratzer’s (1981a) classic analysis of modals. These may be cases of ‘modal modification’ by my definition above, but do not involve a change to the strength of the modal given its baseline modal domain. As Huitink discusses, cases like (3) may be thought of as simply disambiguators, since context alone is often sufficient to specify the flavor of a modal. So I will leave these aside as well.

It is rather a third class of expression, which Huitink (2014) terms ‘intensifiers’, which are the object of present study. Below are some examples of what could be termed modal intensifiers.

- (4) a. It’s quite possible Katie will win.
b. The vase could easily have fallen.
c. You should really walk the dog more often.
d. I absolutely have to finish this paper tonight.

- e. The Red Wings are super likely to make the playoffs again.
- f. The bar is very unlikely to be open at this hour.

The claim made in Chapter 2 and Chapter 3 predicts that the intensifiers in (4a-d) cannot be given a unified account with the intensifiers in (4e-f), contra Grosz (2010), who attempts to unify not only these expressions,¹ but the other modifiers discussed above. The intensifiers in (4e-f) are degree modifiers and require the expression that they compose with to be gradable. Since, as argued in Chapter 2, expressions like *possible*, *should*, and *have to* are not gradable, Grosz’s account cannot be maintained.

The challenge with sentences like (4e-f) is reconciling the meaning of the modal with the degree semantics required for intensifiers like *super* or *very*, given their very broad distribution; that challenge has been taken on in Chapter 3. In the present chapter I deal with the challenge of (4a-d): Providing a meaning for the modifiers in these expressions that squares with a non-scalar interpretation for modals like *possible*, *should*, and *have to*.

The general strategy is this: I focus on a few cases of modal modification which exemplify the different ways in which scales can be constructed and integrated with the meaning of a non-scalar modal expression. There may be yet other modal modifiers which I do not discuss which do not fit readily into the schema of the terms I do analyze, but this is not problematic. Rather, the larger point here is this: The grammar provides many routes to scalarization, so modifiers of modals (and of other expressions for that matter) need to be looked at on a case-by-case basis. This is of course not to surrender the domain of modification to total idiosyncrasy – there are of course patterns and generalizations to be found within smaller classes of modifiers. The purpose of this chapter is to examine representatives of a several of these classes.

In Section 1 below I discuss modifiers which specifically target modal expressions, particularly *easily*; I term these expression Modifiers of Modal Auxiliaries. The analyses I provide for these kinds of expression revolves around the core notion that a modifier may apply a scalar, or even gradable, restriction to the domain of a necessity or possibility modal and thereby modify the strength of that modal through scalar means. The logic of this analysis has its roots in Lewis’s (1973) similarity function and Kratzer’s (1981a) ordering sources, differentiated from these prior works mostly in its compositional

¹Grosz does not discuss (4a-c).

details. For this reason, these expressions also shed light on the compositionality of modals, an issue which has generally been under-discussed in the literature on modality. In Sections 2 and 3, I discuss modifiers which are highly cross-categorial and therefore cannot be given a modal-specific treatment. These expressions, *quite*, *really*, *absolutely*, and *basically*, which have the most general distribution, are the most tempting for a degree-theorist, since degrees provide a way to achieve scalarity over arbitrary domains. I show that a non-degree theoretic account can be given for these expressions, by ranging over higher order representations (contexts in the case of *quite* and *really*, and focus-alternatives in the case of *absolutely* and *basically*.) In Section 4 I conclude.

5.1 Modifiers of Modal Auxiliaries

Consider the use of *easily* as a modifier of a modal auxiliary.

- (5) The vase could **easily** fall.

This expression ‘intensifies’ the meaning of *could*, giving rise to the following asymmetric entailment pattern.

- (6) a. The vase could easily fall. \rightarrow The vase could fall.
 b. The vase could fall. \nrightarrow The vase could easily fall.

This expression requires the presence of a modal auxiliary and is thus distinct from modal adverbials like *probably* (see, e.g., [Giannakidou and Mari \(to appear\)](#)).

- (7) *The vase easily fell.

The distribution of *easily* is quite limited. First, it cannot appear with any necessity modals.

- (8) a. *The vase must easily have fallen.
 b. *The vase should easily have fallen.
 c. *The vase would easily have fallen.

Second, it can only appear with possibility modals on epistemic, metaphysical, or counterfactual readings (not, e.g., deontic or teleological readings).

- (9) a. The vase could easily have fallen.
 b. The vase might easily have fallen.
 c. ?The vase may easily have fallen.
 d. #The boys can easily go to bed late tonight. (deontic reading intended)
 e. To get down town, you can easily take the six. (only has reading of ‘with ease’)

If the modal that *easily* modifies does not appear together with the perfect marker *have*, *easily* must appear immediately after the modal.

- (10) a. *The vase easily could fall.
 b. The vase could easily fall.

However, if the auxiliary does appear together with the perfect marker, the position of *easily* becomes much more flexible; it can appear either before the modal, immediately after it, or after *have*.

- (11) a. The vase easily could have fallen.
 b. The vase could easily have fallen.
 c. The vase could have easily fallen.

With or without *have*, it is only marginally acceptable in sentence final position.

- (12) a. ?The vase could fall easily.
 b. ?The vase could have fallen easily.

Finally, *easily* is gradable.

- (13) a. The vase could have very easily fallen.
 b. The painting fell, but it just as easily could have been the vase.

(13b) is true if, for some point *t* prior to the painting’s falling, the ‘strength’ of the possibility of the vase falling in the future of *t* was just as good as the strength of the possibility of the painting falling at *t*. The question of what ‘strength of possibility’ means here (likelihood? stereotypicality or expectedness?) is a very subtle one which I will not focus on here. Rather, the purpose of this study is to show how compositionally a gradable expression like *easily* may induce this strengthening effect on a non-gradable, Kratzerian possibility modal. Thus I will assume that the scale that *easily* relates worlds

to is the stereotypicality scale, but further analysis of the expression may show otherwise.

Thus in this section I argue that *easily* is a gradable property of worlds which narrows the domain of the modal it attaches to. Since it only combines with possibility modals, this always results in a strengthening effect. In particular, *easily* narrows the modal domain to just the worlds in it which meet a standard for stereotypicality. This analysis is provided in detail below.

5.1.1 What *easily* Is Not

At first glance this expression may seem to simply be an instance of the manner adverbial *easily*_{MA}, seen in (14).

(14) He lifted the vase easily.

But two tests show that this is not the case. First of all, *easily*_{MA} is an agent-oriented adverbial, and so cannot appear with non-agentive verbs.

(15) *The vase fell easily.

The acceptability of (5) can only be explained if the *easily* in that sentence is a distinct expression without a restriction to agentive verbs.

Second, *easily*_{MA} can be paraphrased with the PP *with ease*, while the MMA *easily* cannot.

- (16) a. He could easily lift it over his head. ↔ He could lift it over his head with ease.
 b. The vase could easily have fallen. ↔ #The vase could have fallen with ease.

A sentence with *could*, *easily*, and an agentive verb, like in (16a), may be ambiguous between the two readings, though the manner adverbial reading is generally favored. But the distinct MMA reading is also present, and is especially salient when the agent's intentions are unknown and of present concern. Moreover, the reading of the modal *could* is always different, at least in future-oriented cases; *easily*_{MA} co-occurs with *could* on its ability readings (which can be paraphrased by *able to*, but not *might*), while the MMA *easily* co-occurs with *could* on its metaphysical/historical and counterfactual readings.

Yet a third expression, which I will call *easily_{ep}* is homophonous with these two adverbials, and provides an inference of epistemic certainty (something like “it’s obvious that...”, or “it’s easy to discern that...”) and patterns with the MMA *easily* in terms of both of the diagnostics above.

- (17) a. That book is easily 300 pages long.
b. #That book is 300 pages long with ease.

But this expression has its own paraphrase, the sentence-final particle *easy*, which distinguishes it from the MMA *easily*.

- (18) a. That book is 300 pages long, easy.
b. #That vase could fall, easy.

Ideally, some account would be given of what links these three expressions, whether it is synchronic, or perhaps more likely, diachronic; but here I will focus only on an analysis of the MMA usage of *easily*.²

5.1.2 Degree Semantics

I propose that *easily* denotes a gradable property of worlds, relating worlds to degree concepts on a scale of stereotypicality (ST below).

$$(19) \quad \llbracket \text{easily} \rrbracket = \lambda v[\lambda w[\text{ST}(v)(w)]]$$

Thus *easily* has the type $\langle s, \langle s, d \rangle \rangle$. It takes two worlds arguments, v and w and returns the degree of stereotypicality or expectedness of v given circumstances in w , the evaluation world.

Since *easily* is of type $\langle s, \langle s, d \rangle \rangle$ it is an intensional measure function and therefore gradable. Given the flexible type denotation for degree modifiers suggested in Chapter 2, *easily* is combinable with degree modifiers. In (19) below a denotation is given for the degree modifier *just as*.³

$$(20) \quad \llbracket \text{just as}_7 \rrbracket^g = \lambda G_{\langle \alpha, \langle s, d \rangle \rangle}[\lambda x_\alpha[\lambda w[\max(G(x)(w)) = g(7)]]]$$

²I will speculate that there is a diachronic link between these three expressions which begins with the manner adverb/adjective *easy(ly)*, the antonym of *difficult*. As noted above, *easily_{ep}* can be paraphrased “it is easy to discern that ϕ ”, while *could easily* ϕ might be paraphrased “it is easy to imagine a world in which ϕ ”. However I will not speculate beyond this.

³Here I am simplifying a considerable typology of degree modifiers; see, e.g., McNabb (2012b).

The degree modifier *just as* takes a gradable property of objects of type α and returns the non-gradable property of objects of type α , which is true of such objects iff they possess the gradable property to exactly the degree specified by a contextual parameter specified by the assignment g . The flexible of type of such an expression is independently motivated by the fact that it can combine with gradable properties of, e.g., events, as well as entities.

$$\begin{aligned}
(21) \quad & \llbracket \text{tall} \rrbracket^g = \lambda x_e \lambda w [\mathbf{height}(x)(w)] \\
& \llbracket \text{just as}_7 \rrbracket^g = \lambda G_{\langle \alpha, \langle s, d \rangle \rangle} \lambda x_\alpha \lambda w [G(x)(w) = g(7)] \\
& \llbracket \text{just as}_7 \text{ tall} \rrbracket^g = \lambda x_e \lambda w [\mathbf{height}(x)(w) = g(7)] \\
(22) \quad & \llbracket \text{early} \rrbracket^g = \lambda x_e \lambda w [\mathbf{earliness}(x)(w)] \\
& \llbracket \text{just as}_7 \rrbracket^g = \lambda G_{\langle \alpha, \langle s, d \rangle \rangle} \lambda x_\alpha \lambda w [G(x)(w) = g(7)] \\
& \llbracket \text{just as}_7 \text{ early} \rrbracket^g = \lambda x_e \lambda w [\mathbf{earliness}(x)(w) = g(7)]
\end{aligned}$$

Thus, such an expression can also combine with *easily*.

$$\begin{aligned}
(23) \quad & \llbracket \text{easily} \rrbracket^g = \lambda v \lambda w [\mathbf{ST}(v)(w)] \\
& \llbracket \text{just as}_7 \rrbracket^g = \lambda G_{\langle \alpha, \langle s, d \rangle \rangle} \lambda x_\alpha \lambda w [G(x)(w) = g(7)] \\
& \llbracket \text{just as}_7 \text{ easily} \rrbracket^g = \lambda v \lambda w [\mathbf{ST}(v)(w) = g(7)]
\end{aligned}$$

Since *easily* is analyzed as a measure function, any instance of it without an overt degree modifier is analyzed as combining with the silent *pos* morpheme, which relates it to a contextual standard.

$$\begin{aligned}
(24) \quad & \llbracket \text{easily} \rrbracket^g = \lambda v \lambda w [\mathbf{ST}(v)(w)] \\
& \llbracket \text{pos}_8 \rrbracket^g = \lambda G_{\langle \alpha, \langle s, d \rangle \rangle} \lambda x_\alpha \lambda w [G(x)(w) \succeq s(G)(g(8))(w)] \\
& \llbracket \text{pos}_8 \text{ easily} \rrbracket^g = \lambda v \lambda w [\mathbf{ST}(v)(w) \succeq s(\mathbf{ST})(g(8))(w)]
\end{aligned}$$

5.1.3 Composition

Once *easily* has combined with its degree modifier, it denotes an accessibility relation. This set can be intersected with the modal domain of *could* or other possibility modals to achieve domain restriction. In order for *easily* to access the modal domain itself, however, the modal domain must be represented in the syntax.

Thus I adopt one tweak to the standard model. Following some authors, I posit that modals combine with **accessibility relations**, functions from worlds directly into (preliminary) modal domains, type $\langle s, \langle s, t \rangle \rangle$, rather than modal base intensions. This is necessary for the present account of MMAs

like *easily*.

However I do not sacrifice the Kratzerian notion of relativizing modals to sets of propositions rather than sets of worlds which may be gotten from them. There is data which shows that simple accessibility relations are not rich enough to capture the variable behavior of modals' domains. Consider (25) in a context in which a vase is locked in a very secure safe.

- (25) a. The vase can be broken.
 b. The vase can't be broken.

Both sentences could be true depending on finer aspects of the context. (25a) can be true if the modal is relativized to the proposition that the vase is fragile (and *not* the proposition that the vase is in the safe) while (25b) is true if we include the proposition that the vase is in the safe. Note that (25a) could not be accounted for by any kind of implicit modal domain restriction – here the domain of *worlds* is larger than it is in (25b). But it can be accounted for by implicit modal *base* restriction.

Thus I argue that while modals combine with a simple accessibility relation, this accessibility relation is determined by a hidden indexical in the modal's specifier which is anaphorically related to a modal base intension. I call this modal base pronoun ***mbro***. In (26) I provide its denotation relative to an assignment function g which maps indices to semantic objects; I assume *mbro* bears such an index.

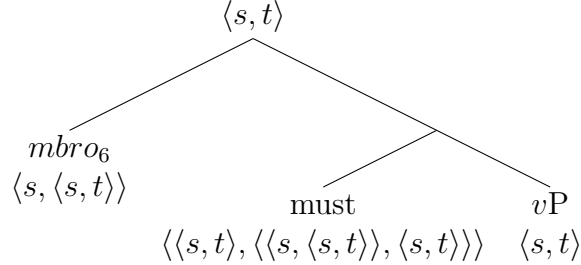
$$(26) \quad \llbracket mbro_6 \rrbracket^g = \lambda v \lambda w [v \in \cap g(6)(w)]$$

The index *mbro* bears is mapped by g to a modal base intension, but the semantics of *mbro* shifts this into an accessibility relation.

More along traditional lines, however, I assume that the ordering source is still anaphorically keyed to the modal itself. The denotation of a modal like *must* is given below, with a schematic LF showing its derivation.

$$(27) \quad \llbracket must_7 \rrbracket^g = \lambda p \lambda R \lambda w [\forall v \in \text{BEST}_{g(7)(w)} R(w)[p(v)]]$$

(28)



Distinct versions of *mbro* may carry distinct presuppositions determining their flavor. Each lexical modal may then impose distinct *syntactic* selectional restrictions on which kinds of accessibility relations they may combine with. Below ∂ introduces presuppositions.

- (29) a. $\llbracket mbro[EPI]_6 \rrbracket^g(v)(w)$ is defined iff $g(6)(w) \subseteq \{p : p \text{ is known in } w\}$
 $\llbracket mbro[EPI]_6 \rrbracket^g = \lambda v \lambda w [v \in \cap g(6)(w)]$
 b. $\llbracket mbro[CIR]_6 \rrbracket^g(v)(w)$ is defined iff $g(6)(w) \subseteq \{p : w \in p\}$
 $\llbracket mbro[CIR]_6 \rrbracket^g = \lambda v \lambda w [v \in \cap g(6)(w)]$

Several factors complicate the issue of modal domain determination significantly. The first is the role of time in determining the modal domain. The second is the role of individual anchors in determining the modal domain. For example, an epistemic modal base is a set of facts known by a given individual or individuals. It could be that the precise subset of information is contextually determined, or the individual anchor is as well. Two possible ways of representing epistemic-*mbro* are given below.

- (30) a. $\llbracket mbro[EPI]_{6,8} \rrbracket^g(v)(w)$ defined iff $g(6)(w) \subseteq \{p : p \text{ is known by } g(8) \text{ in } w\}$
 $\llbracket mbro[EPI]_{6,8} \rrbracket^g = \lambda v \lambda w [v \in \cap g(6)(w)]$
 b. $\llbracket mbro[EPI]_6 \rrbracket^g(v)(w)$ defined iff $g(6)(w) \subseteq \{p : \exists x [p \text{ is known by } x \text{ in } w]\}$
 $\llbracket mbro[EPI]_6 \rrbracket^g = \lambda v \lambda w [v \in \cap g(6)(w)]$

However, I will ignore these concerns for present purposes and thus not include any temporal or individual anchors in modal domain representation.

The basic schematic syntax for modals in place, it is possible to provide a compositional analysis of *easily*. After combining with its degree modifier, *easily* adjoins directly to *mbro*, Nx a Generalized Predicate Modification rule applies, which takes two properties of type $\langle \alpha, st \rangle$ and returns their intersection. This allows us to maintain a simple type for *easily* of $\langle s, \langle s, d \rangle \rangle$, i.e., the type of a gradable property – this in turn allows us to keep a simple

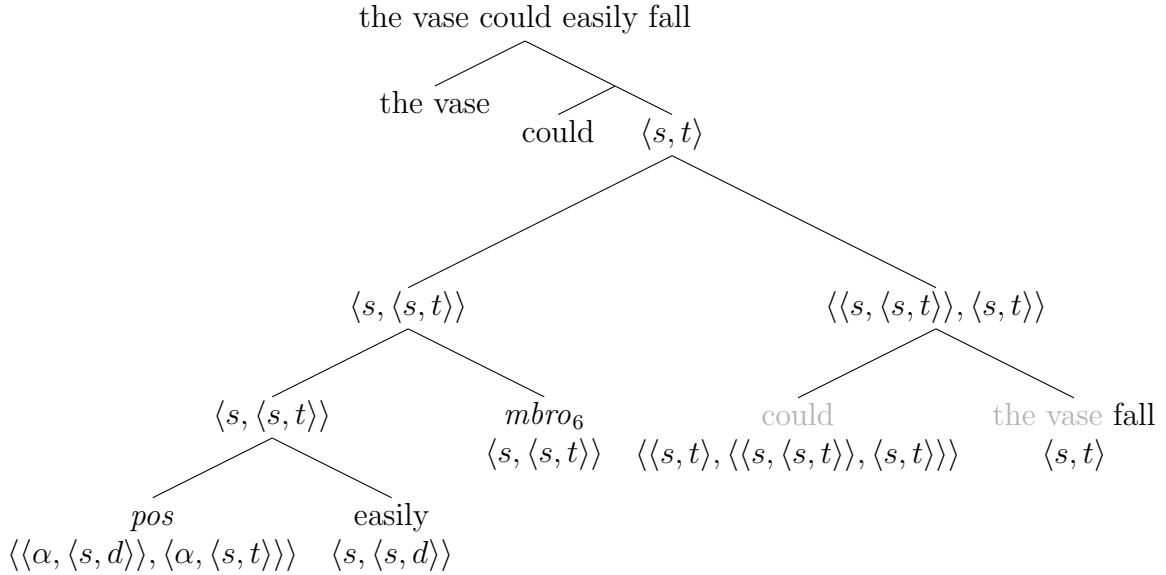
compositional semantics for degree modifiers.

Generalized Predicate Modification If a node α has two daughters, β and γ , both of type $\langle \tau, \langle s, t \rangle \rangle$, then let $\llbracket \alpha \rrbracket = \lambda x_\tau [\lambda w [\llbracket \beta \rrbracket (x)(w) \& \llbracket \gamma \rrbracket (x)(w)]]$

The tree in (32) illustrates the compositional analysis of (31), which is spelled out in detail in (33).

(31) The vase could easily fall.

(32) *Structure of (31)*



(33) *Derivation of (31)*

1. $\llbracket \text{the vase fall} \rrbracket^g = \lambda w [\mathbf{tvf}(w)]$ LEX
2. $\llbracket \text{could} \rrbracket^g = \lambda \phi_{\langle s, t \rangle} \lambda m_{\langle s, \langle s, t \rangle \rangle} \lambda w [\exists v \in m(w) [\phi(v)]]$ LEX
3. $\llbracket \text{the vase could fall} \rrbracket^g = \lambda m_{\langle s, \langle s, t \rangle \rangle} \lambda w [\exists v \in m(w) [\mathbf{tvf}(v)]]$ FA
4. $\llbracket \text{mbro}_6 \rrbracket^g = \lambda v \lambda w [v \in \cap g(6)(w)]$ LEX
5. $\llbracket \text{pos}_8 \text{ easily} \rrbracket^g = \lambda v \lambda w [\text{SIM}(v)(w) \succeq s(\text{SIM})(g(8))(w)]$ (24)
6. $\llbracket \text{pos}_8 \text{ easily } \text{mbro}_6 \rrbracket^g =$
 $\lambda v \lambda w [\text{SIM}(v)(w) \succeq s(\text{SIM})(g(8))(w) \& v \in \cap g(6)(w)]$ GPM

7. $\llbracket \text{the vase could } pos_8 \text{ easily } mbro_6 \text{ fall} \rrbracket^g =$
 $\lambda w [\exists v [\text{SIM}(v)(w) \succeq s(\text{SIM})(g(8))(w) \ \& \ v \in \cap g(6)(w) \ \& \ \mathbf{tf}(v)]]$
 FA

This gives the desired domain restriction and corresponding stronger interpretation. Intensification of *easily* will correspondingly give greater domain restriction, and thus further strengthening.

One important addition is the constraint that *easily* only can be used when it leads to a stronger interpretation; this would mean that it can only be used with possibility modals. Similar proposals have been made for polarity sensitive items (Kadmon and Landman, 1993; Chierchia, 2006), though see Giannakidou (1997, 1998, 2006) for compelling arguments against this view. In fact, as Giannakidou argues for NPIs, a general pragmatic approach based on strength of interpretation cannot work to explain the distribution of this kind of modal modification. See the discussion below of *basically*, an expression which functions to *weaken* necessity modals; this shows that a general pragmatic constraint in favor of strengthening fails to predict the possibility of such an expression. Rather, the distribution of *easily* must be due to semantic properties of the relevant expressions; but this could still be stated in terms of ‘strengthening’; i.e., *easily* is only felicitous when it combines with an expression to result in a strengthened meaning. Note that, depending on how this constraint is realized, this might also predict that necessity modals under the scope of negation can combine with *easily*; however, necessity auxiliaries with epistemic, metaphysical, or counterfactual domains do not scope under negation, so this is difficult to test.

- (34) a. Helena must not be at home. ($\Box > \neg$)
b. If I came in late, Gallagher would not notice. ($\Box > \neg$)

Notice also that the modal *could* is represented without an ordering source – i.e., it quantifies directly over the worlds in its modal base with further restriction. This is a highly non-trivial move because *easily* fills a role potentially very similar to that of an ordering source. If *could* has a stereotypical ordering source already, the effect of *easily* should be redundant. However, in Chapter 4, I argue that the apparent restriction to ‘reasonable’ or ‘stereotypical’ worlds seen in many modals is due not to a semantically specified ordering source but rather the pragmatic phenomenon of imprecision.

Consider the following Lewis Sequence:

- (35) *A man walks along a tightrope between two buildings, secured by a safety line*
- a. A: You could have fallen to your death!
 - b. B: No, I couldn't have – I had a safety line which was tested right beforehand.
 - c. A: Yes, but the safety line could have broken in some unforeseen way!

Compare this to a similar example with *easily*.

- (36) *A man walks along a tightrope between two buildings, secured by a safety line*
- a. You could easily have fallen to your death!
 - b. No, I couldn't have – I had a safety line which was tested right beforehand.
 - c. #Yes, but the safety line could have broken in some unforeseen way!

While in (35), speaker B is forced to admit the strict truth of (36a) (and thus the presence of unstereotypical worlds in the domain of *could*), in (36), speaker B is not compelled to do the same. The use of *easily* really does make the modal too strong for this context, no matter what kind of precisification goes on. Thus, I argue there is no ordering source in *could*; *easily* therefore is not redundant.

5.1.4 Context Sensitivity

One important flaw that this theory may seem to be susceptible to concerns monotonicity. Since the modal domain is determined intersectively, it appears that with *easily*, this domain should be monotonic. To make this point clear, consider the following conditional example.

- (37) If a strong gust of wind had come along, you could easily have fallen.

If we suppose that in this context, all worlds in which a gust of wind comes along are below the salient standard of stereotypicality, this should result in an empty domain (making (37) trivially true). This is contrary to intuitions about (37), which suggest that (37) should always range over some non-empty set of worlds.

The original semantics for ordering sources given by Kratzer (1981a), with its lexicalized superlative meaning, was formulated to avoid this very problem. As discussed in Chapter 1, this semantics avoids ever getting an empty domain with conditionals (except where the antecedent is contradictory) because the modal always takes the *best* worlds of the domain, rather than the worlds which exceed some externally determined standard. Thus the standard is always sensitive to what is in the domain.

However, the present theory of *easily* actually does not make use of an ‘externally determined standard’ either. Recall that when *easily* does not appear with a degree modifier, its standard is fixed by the silent positive morpheme. This morpheme determines a standard from a contextually given comparison class (Kennedy, 2007). However, when a positive adjective is in attributive position, this comparison class can be, at least partially, determined by its sister.

- (38) a. That is a tall boy.
 b. That is a tall basketball player.
 c. That is a tall skyscraper.

Correspondingly, we should expect that the standard in play for *easily* is also determined in part by its sister – i.e., the modal domain. Thus, the standard for stereotypicality will be relativized to what worlds are already in the domain, and therefore avoid the problem of creating an empty domain.

Thus *easily* in a sense replicates the semantics of the ordering source, but relativizes to the modal base via positive rather than superlative semantics.

5.1.5 Interim Conclusion

Most modifiers of modal auxiliaries seem to be of the flavor specifying variety discussed by Huitink (2014), but a few have intensifying behavior, including *easily*, discussed above, and *well*, seen below, which could be given a comparable analysis.⁴

⁴These expressions often display a high level of idiosyncrasy – notice that *well* can combine with *very*, but is marginal in the positive form, and unacceptable with seemingly any other degree modifier.

- (i) a. #They better could be there.
 b. #They well enough could be there.

(39) They [?](very) well could be there.

Giannakidou and Mari (to appear) consider expressions for Greek and Italian which they translate into English expressions like *maybe*, *probably*, and *definitely*. Particularly, they consider how these expressions interact with epistemic modals and with predictive expressions (corresponding to *will* in English) which they take to be modals (see also Klecha (2014) for arguments for English).

(40) Probabilmente Giacomo sarà un dottore.
 probably G. be.FUT a doctor
 “Giacomo must probably be a doctor.” (Giannakidou and Mari’s translation)

On Giannakidou and Mari’s view, these expressions are ‘modal modulators’, which contribute, in addition to the modal’s at-issue semantics, a presupposition about the likelihood that the real world is in the domain of the modal. This is another route to intensification (or de-intensification) for modifiers of modal auxiliaries, since the function they ascribe to these expressions is specially crafted to modality.

There are many other expressions, however, which are not special to modality, but which nonetheless have the effect of modulating the strength of modal expressions, and so must be explained. Given a null hypothesis that these expressions are in fact degree modifiers, it must be shown that a non-degree interpretation can be given for these expressions which preserves an essentially Kratzerian account of the modals they modify.

5.2 Non-Degree Intensification

While MMAs like *easily* have a semantics specialized explicitly for modals, other expressions which modify modals are cross-categorical and therefore demand a general analysis. The first variety I examine are non-degree intensifiers, which give a strengthened meaning to expressions both gradable and non-gradable, below I examine two of these expressions in English, *quite* and *really*, providing each an analysis along the lines of the non-degree intensification discussed by Beltrama and Bochnak (to appear) and McNabb (2012a),

I will refrain from speculating as to why this is not a more robustly attested class.

which have an effect on modals much like the one proposed by [Anand and Brasoveanu \(2010\)](#) for *absolutely*. As I will discuss in the next section below, however, I do not consider *absolutely* to belong to this class.

5.2.1 Adjectival Modification

One such expression already examined is *quite*, which has an intensifying effect when paired with the nongradable modal adjective *possible*, discussed in Chapter 2.

- (41) a. It's possible he has a torn ligament.
- b. It's quite possible he has a torn ligament.

Expressions like these provide the strongest motivation for treating expressions like *possible* as simply gradable. However, in the particular case of *quite*, an analysis of it as a degree modifier fails to capture its propensity to combine with expressions which clearly otherwise not gradable (see Chapter 2).

- (42) a. It's quite possible he has a torn ligament.
- b. *It's too possible he has a torn ligament.
- (43) a. Your response is quite correct.
- b. *Your response is too correct.
- (44) a. He is quite unable to get out of bed.
- b. *He is too unable to get out of bed.
- (45) a. She is quite capable of breaking the law.
- b. *She is too capable of breaking the law.

Ideally, an analysis for expressions like should be given which does not require the expressions they 'intensify' to be gradable. In fact, just such an analysis has already been given for expressions very similar to *quite*. Such analyses argue that expressions which combine with both gradable and non-gradable adjectives to provide an intensifying-type meaning do so by introducing universal quantification over contexts. [McNabb \(2012a\)](#) argues this for Hebrew *mamaš* and English *really*, while [Beltrama and Bochnak \(to appear\)](#) argue for a similar approach for Italian *-issimo* and Washo *šemu*. Below is a semantics for *quite* which closely mirrors [Beltrama and Bochnak's \(to appear\)](#) for *-issimo* and *šemu* (somewhat simplified).

$$(46) \quad \llbracket \text{quite} \rrbracket^g = \lambda P_{\langle \alpha, st \rangle} \lambda x_\alpha \lambda w [\forall h [gRh][P(x)(w, h)]]$$

On this account, expressions of type $\langle s, t \rangle$ are actually functions from world-assignment pairs to truth values, which has been an assumption made (without consequence) throughout this dissertation, and in much semantic work in the tradition of Heim and Kratzer (1998). To put it more transparently, where $\langle c \rangle$ is the type of world-assignment pairs:

$$(47) \quad \llbracket \text{quite} \rrbracket = \lambda P_{\langle \alpha, ct \rangle} \lambda x_\alpha \lambda w, g [\forall h [gRh][P(x)(w, h)]]$$

In the above formulas, R represents an accessibility relation which selects assignment functions which are similar to the assignment g determined by the context. The essential idea here is that when such an expression combine with an adjective, it is combining with positive form rather than the bare adjective. Thus in the case of a positive gradable adjective, the gradable property P must be held by the target to a degree exceeding not just the standard determined by context, but every standard determined by any relevant context. This sets a much higher bar for truth and gives rise to the intensified reading.

Note that Beltrama and Bochnak (to appear) introduce quantification over ‘contexts’, explicitly eschewing quantification over assignment functions. This is done to avoid the possibility of introducing variance in contextual parameters which clearly do not vary under the scope of of such intensifiers, e.g., pronouns.

$$(48) \quad \text{John really went after him.}$$

Clearly (48) cannot mean that John went after every individual who the index associated with *him* could be mapped to. However, it’s not clear what a ‘context’ might be, if not an assignment function – on dynamic semantic models where the notion of context is made very explicit, e.g., File Change Semantics (Heim, 1982; Condoravdi and Gawron, 1996), the assignment function *is* the representation of context. Moreover, this concern can be alleviated precisely by defining the accessibility relation R appropriately. Thus, and assignment g is related to h by R iff all mappings from indices to the domain of individuals are preserved exactly. Things like standards of comparison, however, may vary.

In the case of the non-gradable modal *possible*, however, this context manipulation may affect the modal domain rather than a contextual standard.

Thus, the possibility claim will be evaluated against not only the contextually determined modal base, but variants on it as well, including smaller and larger domains. Since possibility will be evaluated at several subdomains of the contextually determined domain, the assertion is much stronger. This can also explain why *quite* also strengthens necessity modals, which would be unexpected if it were strictly a domain restrictor (a la *easily*).

(49) This course of action is quite necessary.

Since the necessity claim ranges over all contexts, both larger and smaller domains will be included. Thus any kind of modal will be given a stronger interpretation.

This is almost precisely the analysis that is given by [Anand and Brasoveanu \(2010\)](#) for *absolutely*. In the next section I will argue explicitly against their treatment of that expression, but only because I disagree with them about the empirical distribution of *absolutely*; the logic of their argument applies perfectly to expressions like *quite* which modify both necessity and possibility modals. [Anand and Brasoveanu \(2010\)](#) do state their analysis as involving explicit quantification over modal bases, rather than over contexts as a whole, which in this case does not capture the cross-categoricity of *quite*. However, following [Beltrama and Bochnak \(to appear\)](#) and [McNabb \(2012a\)](#), we are able to generalize appropriately.

5.2.2 Auxiliary Modification

Turn now to the contribution of the expression *really*, which [McNabb \(2012a\)](#) also analyzes as a non-degree intensifier, when paired with a weak necessity modal like *should* or *ought*.

- (50) a. To get downtown, you should take the 6.
b. To get downtown, you should really take the 6.
- (51) a. To get downtown, you ought to take the 6.
b. To get downtown, you really ought to take the 6.

As discussed by [McNabb \(2012b\)](#), [Kennedy and McNally \(2005\)](#), and many others, there are (at least) two uses of *really*; what I will term *really*_{PM} and *really*_{DM}. *really*_{PM} functions as a propositional modifier distinct from *really*_{DM}. It is used to provide mirativity or verum focus to a sentence, often to contrast

a preceding utterance; see e.g., Paradis (2003). It does not depend on the presence of a modal, or even a gradable expression.

- (52) a. Clark Kent is not Superman.
b. No, Clark Kent really [is]_F Superman.

When combining with *should*, it can only appear before the modal auxiliary, whereas *really*_{DM} can occur either before or after the auxiliary. (53a), but not (53b), can paraphrase (53c).

- (53) a. You really should take the 6.
b. You should really take the 6.
c. It really is the case that you should take the 6.

It is often distinguished by prosodic stress, but this stress pattern is not required in the presence of *really*_{PM}. However, only *really*_{PM} is compatible with such a stress pattern.

- (54) a. You really [should]_F take the 6.
b. #You [should]_F really take the 6.

Thus I will set *really*_{PM} aside.

Intuitively, if (55a) is a paraphrase of (54a) (as discussed in Chapter 3), then (55b) is a paraphrase of (54b).

- (55) a. Taking the 6 is the best option for getting downtown.
b. Taking the 6 is *by far* the best option for getting downtown.

This is not obviously a case of domain modulation at all. If anything, this seems like degree modification, or more specifically, differential degree modification. However, as discussed above, the original analyses of non-degree intensifiers was intended to capture their behavior which seemed very much like degree modification. Thus Washo *šemu* can be used with expressions like *tall* which are gradable in English, but not in Washo. Likewise, weak necessity expressions like *should* are built on a scalar notion, as argued in Chapter 3, and thus could be manipulated in manner similar to degree modification, despite not exhibiting a true degree argument.

Returning to the denotation given for *should* in Chapter 3, consider that we could add a degree parameter which specifies the differential between the highest ranked possibility (over which *should* quantifies) and the next

highest-ranked one. This is not unlike the analysis sketched by [Morzycki \(2015\)](#) for differential comparatives with adjectives, where a degree argument is added to accommodate the additional argument in those expressions.

(56) John is three feet taller than Bill.

On this analysis, the meaning of the comparative is given in (57b), as opposed to (57a), the more conservative denotation for the comparative given in Chapter 2.

- (57) a. $\llbracket \text{more than} \rrbracket = \lambda g_{\langle e, d \rangle} \lambda d \lambda x [\max\{d' \in g(x)\} > d]$
b. $\llbracket \text{more than} \rrbracket = \lambda g_{\langle e, d \rangle} \lambda d \lambda d' \lambda x [\max\{d' \in g(x)\} > d + d']$

The additional degree argument in (57b) is saturated by the measure phrase, like *three feet* in (56); if no measure phrase is present, it can be taken to be existentially closed.

In Chapter 3 the following semantics is given for *should*, following the intuition that it has an essentially superlative meaning.

- (58) $\llbracket \text{should} \rrbracket = \lambda P \lambda m \lambda o \lambda x \lambda w [\forall v \in \mathbf{best-fit}(o, x, m, \Gamma, w) [v \in P(x)]]$
iff m is metaphysical and o is a teleological ordering source

This says that all the worlds in the partition-cell of the decision problem which best fits the agent's goals are worlds where P holds of the agent. Below is a list of the definitions of the terms referenced in the denotation above, which are discussed in detail in Chapter 3.

- (59) a. $\mathbf{best-fit}(o, x, m, \Gamma, w) = \bigcup \{p : \langle \bigcap m(w), p \rangle \in \Delta(\mathbf{fit}_{o(x)(w)})(\Gamma_x^*(\bigcap m(w)))\}$
b. $\Gamma_x^*(\bigcap m(w)) := \{\langle \bigcap m(w), p \rangle : p \in \Gamma_x(\bigcap m(w))\}$
c. $\Delta(G)(C) = \{x \in C : \neg \exists y \in C [G(y) > G(x)]\}$

The relevant ingredient here is (59c), the superlative operator. (60a) shows it altered to include a degree argument along the lines of (57b), with an altered definition for **best-fit** in (60b), and a correspondingly altered denotation for *should* in (60c).

- (60) a. $\Delta(d)(G)(C) = \{x \in C : \neg \exists y \in C [G(y) + d > G(x)]\}$
b. $\mathbf{best-fit}(d, o, x, m, \Gamma, w) = \bigcup \{p : \langle \bigcap m(w), p \rangle \in \Delta(d)(\mathbf{fit}_{o(x)(w)})(\Gamma_x^*(\bigcap m(w)))\}$
c. $\llbracket \text{should}_5 \rrbracket^g = \lambda P \lambda m \lambda o \lambda x \lambda w [\forall v \in \mathbf{best-fit}(g(5), o, x, m, \Gamma, w) [v \in P(x)]]$
iff m is metaphysical and o is a teleological ordering source

If *really* is given a semantics like *quite*, it could likewise intensify *should* by manipulating the degree term that appears in (60c). Note that this doesn't contract the domain, it only increases the differential between the domain and the next best domain in the ranking of possible domains induced by the gradable property **fit**. Thus non-degree intensifiers can effect intensification through both domain modulation and degree manipulation.

Note that it is even possible that an expression could perform a similar function more straightforwardly by interacting with a degree argument by function application. What suggests that this is not what is happening in the case of *should really* is that we do not see a range of degree modifying-like expressions interacting with the supposed degree argument of *should*, only *really*.

The reader is referred to [Beltrama and Bochnak \(to appear\)](#) for a more detailed discussion of this approach to non-degree intensification; the point here is only to argue that an analysis of *quite* and *really* can be given which preserves an account of *possible*, as well as others like *unable* and *necessary*, as non-gradable modals. Thus a traditional Kratzerian analysis of both necessity and possibility modals can be maintained.

5.3 Alternative Manipulators

[Horn \(1972, 2005\)](#) coins the term 'A-Adverb' for expressions like *absolutely*, which seem to attach only to expressions which constitute 'scalar-endpoints', in the sense of Horn scales.

- (61)
- a. Absolutely {everyone/noone/# someone} was there.
 - b. It's absolutely {necessary/certain/impossible/#possible/#likely}.
 - c. You absolutely {have to/cannot/#may/#can} go.
 - d. He absolutely {always/never/#sometimes/#often/#rarely} eats meat.
 - e. I absolutely {adore/love/despise/#like/#dislike} you.

This behavior is reminiscent of that of degree maximizers like *completely*, which take a scale-denoting expression and pick out the property associated with its endpoint. However this clearly cannot be the analysis of *absolutely* since a) none of the expressions in (61a-d) are gradable, and b) while some expressions in (61e) are gradable, they do not denote upper-closed scales.

- (62) I love Fred, and I love Bill even more.
- (63) It's absolutely freezing outside. And it's going to be even colder tomorrow.

As Horn points out, among the expressions that *absolutely* combines with are necessity modals, or negated-possibility modals, including verbs and auxiliaries, as well as adjectives. The same can be said for *basically*.

- (64) a. I absolutely have to finish this today.
b. You basically have to wear a raincoat.

In this section I analyze these expressions as introducing focus alternatives requiring either the strongest interpretation (in the case of (64a)) or a close-to-strongest interpretation (in the case of (64b)) among the alternatives, which has the secondary effect of precisification, and thus strengthening. Below I first consider prior accounts of *absolutely* before presenting the analysis.

5.3.1 Precisifiers

Anand and Brasoveanu (2010) contradict Horn's claim about the distribution of *absolutely* by arguing that it can combine with (non-negated) possibility modals as well.

- (65) a. May I have another?
b. Yes sir, you absolutely may. ((65b) from Anand and Brasoveanu (2010))

However, this conflates two distinct uses of the expression *absolutely*.⁵ One is the A-Adverb discussed by Horn (2005), (henceforth *absolutely*_A) which, as Horn discusses, involves focus on the modified expression, and is restricted only to high-scale expressions (thus necessity and not possibility modals).

- (66) a. I gave absolutely [every book I own]_F to Roger.
b. I absolutely [have to]_F finish this by tomorrow.
c. He absolutely [never]_F eats meat.
d. Norm absolutely [despises]_F you.

⁵It is worth pointing out that Anand and Brasoveanu's (2010) finding are based a corpus search, which makes it very difficult to tease apart cases of homophony.

The other homophone, *absolutely_B*, has only contrastive or question-answering usages which, in non-elliptical cases, requires stress on *absolutely* or stress consistent with verum focus, i.e., on the highest auxiliary.

- (67) a. Am I allowed go to the store?
 b. You are [absolutely]_F allowed to go to the store.
 b'. ?You [absolutely]_F are allowed to go to the store.
 b''. You absolutely [are]_F allowed to go to the store.
 b''#You [are]_F absolutely allowed to go to the store.

This variant imposes no constraint on what kinds of expressions it may combine with, which is best explained by assuming that it is a propositional modifier.

- (68) a. Was John at fault in the accident?
 b. He absolutely [was]_F at fault.
 b'. ?He [absolutely]_F was at fault.
 b''. He was [absolutely]_F at fault.

Since *absolutely_B* is consistent with verum focus if it precedes the auxiliary, and *absolutely_A* requires focus on the expression it modifies, modal auxiliaries conflate the prosodic patterns of these two expressions. But consider *absolutely* preceding a non-auxiliary modal.

- (69) #You are absolutely [allowed]_F to go to the store.

No context makes (69) acceptable.

This is an important point, because Anand and Brasoveanu (2010) hold up *absolutely* as an exemplar for a novel category of modal-modifying-adverb, namely the category of adverbs which impose constraints on neither the force nor the flavor of the modal. As discussed above, *easily* combines with only possibility modals.

The account given here for *quite* and *really*, derived from accounts by Beltrama and Bochnak (to appear) and McNabb (2012a), however, does have a universally intensifying effect, due to a generalized version of the same semantics as that given by Anand and Brasoveanu. But the crucial point here is that *absolutely_A* fits into a category which Anand and Brasoveanu (2010) claim has no exemplars, namely expressions which constrain the force of the

modal but not the flavor.⁶

5.3.1.1 Extremeness and Domain Widening

Morzycki (2012) argues that *absolutely* is an extreme degree modifier (EDM), a class of degree modifiers which combine only with extreme adjectives, like *gigantic*.

(70) That thing is absolutely [gigantic]_F.

Morzycki's account of adjectival extremeness is to say that typical gradable adjectives are, like quantifiers, constrained to a salient domain, in this case a domain of degrees. Thus a non-extreme gradable adjective, like *big*, has a contextually determined upper-bound, even though it lexically is upper-open. On Morzycki's account, extreme adjectives and extreme degree modifiers involve widening that domain to include yet higher degrees, account for the intuition of extremeness.

Morzycki notes Horn's (1972) observation about *absolutely*'s modification of quantifiers, and points out that a domain widening account may intuitively explain the full distribution of *absolutely*. Morzycki's account, however, is not actually general, since it is stated in terms of degrees and thus accounts only for *absolutely*'s behavior with gradable extreme adjectives. Morzycki suggests that the use of *absolutely* which Horn attests, and its use as an EDM, are distinct but related. This proposal is in that spirit as well, providing an account which aligns with Morzycki's for EDMs (in spirit if not in details) in terms of domain 'widening', but not in terms of degree modification.

Strictly speaking, however, it cannot truly be about domain widening; consider its use with *everyone* below.

(71) I went to the faculty meeting, and to my great shock, absolutely [everyone]_F was there.

Naturally, with or without *absolutely*, we would say that the domain here is 'the set of all faculty members in the relevant department'. It would be a mistake to say that *absolutely* truly widens the domain from this baseline, since we do not take (71) to mean that anyone other than faculty members were present. Rather, I argue that the apparent effect of 'domain widening'

⁶As noted above, *easily*_{MMA} constrains both force and flavor, a variety which Anand and Brasoveanu (2010) do attest.

in (71) is actually a case of ensuring a high standard of precision to *prevent* the domain-contracting effect of imprecision.

Thus there are two facts about *absolutely* which the analysis must capture: First, it can only appear with expressions which are endpoints of Horn scales. Second, it gives rise to precisification, which in turns smacks of intensification. What unites these two facts is that they involve the consideration of alternatives; Horn scales are simply sets of alternative expressions which can be ordered by entailment relations, while (im)precision and precisification is driven by comparison between alternatives, as discussed in Chapter 4.

To make the latter point clear, recall the analysis of (im)precision in numerals. The round numeral *fifty* can be uttered when the true precise value is 49 because *fifty* is, in comparison to *forty-nine*, an easy expression to use. Likewise, the use of *forty-nine* signals that the speaker cares to have a high standard of precision, because if the speaker did not, she would have used the alternative *fifty*.

Thus I argue that *absolutely* makes use of alternatives to both restrict its use to Horn-scale endpoints, and to induce a comparison between alternatives that gives rise to precisification. The logic of the latter process is as follows: By evoking alternatives to the strict semantic content, the pragmatic inference is generated the he difference between those alternatives and the strict semantic content matters. If this evocation occurs in a context where it is (possible that it is) assumed by some of the participants that these differences do *not* matter, this will cause a crash and a shift to a higher standard where, indeed, such differences matter.

The logic of this account is very similar to that of the account given for non-round numerals like *forty-seven*. Nothing about the conventional, semantic meaning of *forty-seven* says anything about a standard of precision or what should happen to it. Rather, it is a pragmatic side-effect of the expression that it induces precisification. And while surely there are expressions that do, as a direct semantic consequence, involve a shift in the standard of precision, these are very general expressions which are not distributionally limited, like *absolutely*.

- (72) a. Strictly speaking, everyone was there.
b. Strictly speaking, there were fifty people there.

Ideally, the distribution of *absolutely* and its precisifying effect fall out from a single fact about its conventional meaning. Thus, I will pursue the ‘side-

effect’ analysis of precisification for *absolutely* presently.

5.3.1.2 Analysis

I argue that *absolutely* attaches to a quantifier Q with an unsaturated domain argument. It then constructs a set of alternatives to Q , with the felicity condition that the interpretation of Q is stronger than any of the alternatives. The set of alternatives is limited just to alternative quantifiers which are conservative, where $\mathbf{cons}(Q)$ below is true iff Q is conservative.

- (73) For any type α , s.t. the type of $\llbracket \text{XP} \rrbracket$ is $\langle \langle \alpha, t \rangle, \langle \langle \alpha, t \rangle, t \rangle \rangle$,
 $\llbracket \text{absolutely XP} \rrbracket$ is defined iff $\forall P_{\langle \alpha, t \rangle} [\llbracket \text{XP} \rrbracket(P) \subseteq \bigcap \{Q_{\langle \langle \alpha, t \rangle, \langle \langle \alpha, t \rangle, t \rangle \rangle}(P) : \mathbf{cons}(Q)\}]$
 $\llbracket \text{absolutely} \rrbracket = \lambda x[x]$

It is this felicity condition which naturally derives the constraint that *absolutely* only combine with high-Horn scale items. For each Horn scale of quantifiers (the nominal ones, the temporal ones, the modal ones, and their negative inversions) only the strongest survive the felicity condition, thus the (strong) universal quantifiers and negated existential quantifiers.

Since these weaker quantifier meanings are evoked and contrasted with the domain in the description of the felicity condition, the standard of precision is forced to raise so that, e.g., the distinction between $\llbracket \text{everyone} \rrbracket$ and $\llbracket \text{nearly everyone} \rrbracket$ matters in this context. The effect is thus derived that *absolutely* is a slack regulator, rather than an intensifier.

This approach is evocative of Krifka’s (1999) analysis of modified and unmodified numerals, and the scalar implicatures which the latter, but not the former, give rise to. On Krifka’s account, numerals, and any other Horn-scale-inducing expressions, are lexically specified as having focus-semantic values, i.e., being associated with alternative sets that ultimately feed into the pragmatic calculation of scalar implicature. Numeral modifiers like *at least* suppress those focus alternatives, essentially introducing disjunction over the alternatives and making that the ordinary semantic value of the expression. On my account, however, *absolutely* does not depend on pre-existing focus, since it constructs the alternatives itself.

One concern, given Krifka’s analysis, is that *any* expression which bears focus, and thus evokes alternatives, might have a precisifying effect. But this is not the case. On Krifka’s account, all expressions which evoke Horn-scales have inherent focus. Thus, quantifiers like *everyone* or *have to* evoke

their Horn-scale-mates as focus alternatives lexically. But this alone will not give rise to precisification, because there are no alternative expressions which do not give rise to these alternatives. Thus, a rational hearer will not assume that the speaker cares to distinguish between what is uttered the focal alternatives, because the hearer knows the speaker had no choice. It is only when using *absolutely*, which contributes no truth-conditional content, but relies crucially on these alternatives for determining the felicity condition that it establishes, that a hearer would be licensed to infer that the speaker cares to distinguish.

Thus *absolutely* gives rise to precisification, which results in a stronger interpretation, despite a non-gradient meaning for the modal it attaches to.

5.3.2 Approximators

Every expression examined so far can be thought of, at least loosely, as intensifiers – in the sense that they strengthen the expression the combine with, either semantically or pragmatically. But not all of Horn’s A-Adverbs have this effect. Expressions like *almost* have the opposite:

- (74) a. Almost {everyone/noone/# someone} was there.
- b. He almost {always/never/#sometimes/#often/#rarely} eats meat.

There is an extensive literature on *almost* – see [Morzycki \(2015\)](#) for a summary and citations. But it should be pointed out that, with regard to modality, the distribution of *almost* is curious and not easily relatable to *absolutely*; it can combine with some (but not all) non-gradable modals. It can also modify several non-gradable expressions that *absolutely* cannot combine with.

- (75) a. It’s almost {impossible/??mandatory}.
- b. #To get downtown, you almost have to take the number 6.
- (76) a. He’s almost dead.
- b. #He’s absolutely [dead]_F.
- (77) a. There were almost fifty people there.
- b. #There were absolutely [fifty]_F people there.
- (78) a. He’s almost able to walk without support.
- b. #He’s absolutely [able]_F to walk without support.

While this is puzzling, I will not attempt to explain it at present, though one possible avenue of explanation is *almost*'s requirement that it combine with expressions with “exact values” in the words of Horn (2005); this may mean non-vague expressions, or expressions dealing with objective or quantifiable properties. But another expression has a very similar meaning but does combine freely with necessity modals.

- (79) a. Basically {everyone/noone/# someone} was there.
 b. He basically {always/never/#sometimes/#often/#rarely} eats meat.
- (80) a. It's basically {impossible/mandatory}.
 b. To get downtown, you basically have to take the number 6.

Moreover, while its meaning is similar to *almost*, it does contrast subtly:⁷

- (81) a. I'm almost done, so, hold on a second.
 b. I'm basically done, so, we can go.
- (82) a. You almost called him a liar. But you didn't, so, phew!
 b. You basically called him a liar. #But you didn't, so, phew!

While *almost* and *basically* both give rise to the inference that the salient endpoint was approximated but not reached, *basically* additionally conveys that the salient endpoint *may as well* have been reached. In other words, *basically* suggests that the unmodified proposition would be true on a lower standard of precision; that the difference between the unmodified proposition and what is true does not matter. Unlike *absolutely*, *basically* does not appear to associate only with domain-bearing expressions; e.g., (82b). However, the focus-alternative semantics given for *absolutely* does appear to be on the right track for *basically* as well, given that it is itself focus-sensitive:

- (83) a. You basically [slandered]_F every person in this room.
 b. You basically slandered [every person]_F in this room.

(83a) suggests that the addressee insinuated negative things about every person in the indicated room, not reaching the technical definition of slander in every case, but always coming close enough that the distinction does not matter. (83b) on the other hand suggests that the addressee did, strictly

⁷In this regard, *nearly* behaves like *almost*, while the following expressions behave like *basically*: *essentially*, *practically*, *more or less*, and *pretty much*.

speaking, slander almost every person in the room. This focus-sensitivity may be important to understanding the cross-categorical nature of approximators that Morzycki (2001) wrestles with for *almost*.

Consistent with Lauer’s (2012) point that approximators, but not slack regulators, operate truth conditionally, the approximating effect is written into the semantics for *basically*, rather than left as a side-effect, as with *absolutely*. Correspondingly, *basically* does not give rise to focus-alternatives as *absolutely* does in service of creating its slack regulating effect, rather, it is fed by such alternatives. Thus I characterize it as a propositional operator, taking a pair of an ordinary semantic value (a proposition) and a focus semantic value (a set of propositions), established previously by focus and the \approx operator a la Rooth (1992).

$$(84) \quad \llbracket \text{basically} \rrbracket = \lambda \langle p_o, p_f \rangle \lambda w [\neg p_o \ \& \ \exists p \in p_f [p(w) \ \& \ \forall v \in p [\exists u \in p_o [v \approx_x u]]]] \text{ (v1)}$$

This expression first establishes that the ordinary semantic value of the sentence is false; this point has been argued extensively for *almost* (see Sadock, 1981; Rapp and von Stechow, 1999) but does not bear on the main point. It then says that some proposition out of the focal alternatives, which is similar to ordinary semantic value, is true. Here ‘similar’ means that every world in the proposition is similar to a world in the ordinary semantic value.

Note that the similarity operator \approx is not anchored to the pragmatic context ρ – if it was, it would simply duplicate the pragmatic function of imprecision and *basically* would not contribute anything. So what does determine similarity? I argue it is some subset of the goals or concerns of the discourse, which are pragmatically necessarily distinct from the larger goals the conversation is operating under. (If they weren’t, there would be no reason to utter *basically*.) For example, (81b) might uttered when, for the purpose of *some* goals, the distinction between done and nearly done matters (thus the importance of conveying “not done”), while for the purpose of others, the distinction does not matter. The final version of *basically* is therefore given below:

$$(85) \quad \llbracket \text{basically} \rrbracket^\rho = \lambda \langle p_o, p_f \rangle \lambda w [\neg p_o \ \& \ \exists p \in p_f [p(w) \ \& \ \exists x \in \rho [\forall v \in p [\exists u \in p_o [v \approx_x u]]]]]$$

I will not attempt to show that this same approach can be carried over to *almost*, but will suggest that the distinction between *basically* and *almost*

seen in (81) can be accounted for by removing all reference to ρ .

What is crucial is that *basically* highlights another route through which modals may be modified – in this case de-intensified; by existentially quantifying over focus-alternatives under a similarity ordering and mimicking the effect of imprecision within the semantics.

5.4 Conclusion

This chapter has shown many methods for modal modification that does not rely on a degree semantics. These include gradable domain restriction (*easily*), manipulation of a degree argument on the domain (*really*), quantification over contexts, and therefore domains (*quite*), manipulation of the standard of precision through evocation of focus alternatives (*absolutely*), and truth-conditional mimicry of imprecision through sensitivity to focus alternatives (*basically*).

A criticism that might be brought against this kitchen-sink approach may be that it suggests that the clever semanticist may be able to simply wriggle out of any empirical situation in defense of a non-de-scalar view of modal auxiliaries and verbs. But this is not the case – with the possible exception of *quite*, which admittedly has a very flexible semantics, each of these expressions is constrained significantly in terms of the force that modals which combine with it may show. The MMA *easily* combines only with possibility modals, while *really* modifies only weak necessity. Meanwhile *absolutely* and *basically* only combine with high-Horn-scale expressions, thus, necessity modals. This is an important point.

If e.g., *easily* combined with necessity modals, it should, on my account, have a *weakening* effect, since reversing the force reverses the effect of domain narrowing/widening. A purely scalar account predicts that no such expressions should exist – if there is a single scale, and it has a single polarity, an intensifier should always strengthen and never weaken such an expression. It might be taken as a kind of weak support for my claim that *easily* does not appear with necessity modals, since an informativity-type account can be given for this restriction, whereas a purely scalar account of modal auxiliaries and verbs would be incompatible with such an account. But recall that Lassiter (2011) take the possibility/necessity distinction to be one of minimality/maximality, which on his account, degree modifier distribution is indeed sensitive to. So with the present data, it appears to be a stalemate.

What’s worse, in light of modifiers like *quite*, which do in fact strengthen both possibility and necessity modals without reference to degrees, any claim that intensifiers show evidence for non-scalarity is not falsifiable; any intensifier which *does* strengthen both possibility and necessity modals can simply be claimed to have a semantics like *quite*.

However, all is not lost: We are in a position to make a strong prediction about modification of modals. If a particular modal expression is not scalar then there should *not* be any expressions which combine with it which compare to an explicit standard (a la *more*, *too*, or *enough*) and which is insensitive to either force or flavor. The degree modifiers which have this effect do not constraint scale structure or positive meaning at all, and so we should not expect the same out of such MMAs. What’s more, expressions like *easily*, which can display such meanings (by combining with *more*, *too*, or *enough*) specify a particular measure function to manipulate (stereotypicality in the case of *easily*) and therefore limit themselves to epistemic or metaphysical modals. We should not expect such limitations out of general degree modifiers.

This is not to say that a de-scalar modal auxiliary or verb should have to be compatible with the familiar degree modifiers, since, as Lassiter points out, there are real syntactic reasons to suspect ad-adjectival degree modifiers cannot combine with such expressions. The set of MMAs seems to be a highly idiosyncratic one. But we should see expressions like ‘really’ which instead have a comparative or excessive meaning, and can combine with modal auxiliaries of any force or flavor. An approach along the lines of [Beltrama and Bochnak \(to appear\)](#) and [McNabb \(2012b\)](#) would not be able to account for such expressions, nor would any of the approaches proposed here. But I know of no such expressions in English.

These proposals of course only the first steps in a full theory of modal modification. Much work can and should be done to test the predictions made here, both with regard to the expressions discussed here and with regard to the larger patterns observed. This proposal serves to a) provide analyses for a wide range of intensifying-type expressions without appeal to a purely scalar semantics for modality, and b) predicts expressions that would provide evidence for widespread scalarity in the modal auxiliary domain do not exist. Only further research will show whether these analyses can be refined or whether this prediction can be falsified.

Chapter 6

Conclusion

This dissertation is fundamentally about describing and analyzing the ways in which the strength of modal expressions may vary across contexts, especially in ways that can be understood as deriving from conceptual scales of some kind or another. I have referred to this very broad property of modals as **scalarity**.

In this dissertation, I have tried to show that there are many ways to derive scalarity in modals. This follows from the primary assumption about what is to be a modal: A modal is simply a relation between two pieces of information, i.e., two propositions. We distinguish between these two propositions because they tend to display consistent differences. The first sort of proposition, the modal domain, is always at least partially determined by context, sometimes entirely so, being modified overtly only by *if*-clauses and handful of other constructions. [Kratzer \(1981a\)](#) has influentially argued that the modal domain is always determined from two contextual parameters, the modal base and the ordering source, which are both sets of propositions, each employed differently in the construction of the modal domain. The second sort of proposition, the prejacent, is generally syntactically present, denoted by something like a clause, a singular foregrounded proposition. A basic assumption of this dissertation has been that there are no properties universal to modals except the ones that define the notion, the ones given above. Thus we expect significant variation.

This should not be very surprising. In non-modal domains we see this variation as well. Properties of individuals may be gradable, or not, like *tall* and *dead*. They may be subject to imprecision, or not, like *five feet tall* and *quite tall*. Even if they are not gradable may be modified by an expression

which itself is gradable, or not, like *tall man* and *man*. If they are inherently subject to imprecision, they may be precisified or approximated, like *exactly five feet tall* and *about five feet tall*.

All of these points of variation apply to modals as well, because nothing about what it is to be a modal makes scalarity necessary or impossible; modals are just relations and relations may have any of these properties. In light of the existence of these various routes to scalarity, which have all long been argued for diverse domains beyond the territory of modality, we should predict the same kind of diversity among modal expressions. This dissertation supports this prediction.

The varieties of scalarity seen in modals can be organized by two binary distinctions. Scalarity can be semantic or pragmatics, and it can be intrinsic or extrinsic. Intrinsic gradability is seen in gradable modals like *likely* or *important*, but not, say, *unable*, as indicated by compatibility with degree modifiers.

- (1)
 - a. It is not very likely that Greg will respond.
 - b. It's more important that the food arrive on time.
 - c. #How unable is Chris to play football?

Extrinsic gradability can come in several varieties. It can be seen when non-gradable modals are coerced into gradability or involve metalinguistic gradability.

- (2)
 - a. ?It's more possible now.
 - b. I would say it's more 'desireable' than 'obligatory'.

It can also be seen when modals with categorical meanings, like *could*, are modified by expressions which are themselves gradable.

- (3) The vase could (very) easily have fallen.

Intrinsic (im)precision is present in non-vague modal expressions, which allow for precisifications or approximations.

- (4)
 - a. You have to wear a raincoat.
 - b. Strictly speaking, I don't [have to]_F.
- (5)
 - a. It's likely to rain tonight.
 - b. #Strictly speaking, it's not [likely]_F.

(6) You basically have to wear a raincoat.

(7) #It's basically likely to rain tonight.

Extrinsic (im)precision is not discussed in this dissertation, because there is no reason to think that it would inform our understanding of modality in a way that the three previous cases, but it is attested.

- (8) a. There is a 70% chance of rain tonight.
b. Strictly speaking, there is only a 68.8% chance.

(9) There is approximately a 70% chance of rain tonight.

Below, I briefly review the analysis of each of these sources of scalarity.

6.1 Semantic Scalarity

Consistent with the definition of modals as relations between sets of worlds, i.e., functions of type $\langle\langle s, t \rangle, \langle\langle s, t \rangle, \alpha \rangle\rangle$, I have argued that modals may either be gradable or not. I provide empirical support for this claim in the form of a corpus study and an acceptability judgment task, which support the following basic characterization of the data.

- (10) a. It is too likely that we will encounter danger (to undertake this mission).
b. #It is too possible that we will encounter danger (to undertake this mission).

An account which essentially follows [Kratzer \(1981a\)](#) remains sufficient to account for the meanings and distributions of expressions like *possible*, but attempts by her and others to reduce gradable modals like *likely* to instances of logical universal or logical existential quantification have met with significant difficulty. The argument I make is that giving an expression like *likely* a degree based semantics, and abandoning any effort to characterize its meaning as either a (very) special case of necessity or possibility, in no way undermines the project of treating non-gradable modals as such. The crucial fact which unites modal expressions is that they relate a domain and prejacent, which allows for a unified analysis of conditionals as domain restrictors. By treating gradable modals as gradable relations between domains and prejacent, this can be maintained.

Several other facts surrounding the distribution of these expressions must be explained. One is their scalar structure as indicated by compatibility with scale-specific degree modifiers. Another is entailment relations between gradable and non-gradable modals. I have argued for two possible modes resolving the latter issue; entailment relations may be captured by simply specifying the right properties of the orderings, the other by giving an appropriate semantics to the non-gradable modal. The former strategy is exemplified by *likely*, which I argue to be Modal Additive and Scalar Conservative, and the latter is exemplified by *should*, whose domain is determined by reference to a gradable modal property, which itself forms the scalar basis of *important*.

Gradability can be introduced into the modal domain extrinsically as well, by an expression like *easily* which denotes a gradable property of worlds. This expression can be used to restrict the domain of a non-gradable modal like *could*.

6.2 Pragmatic Scalarity

Like various other expressions, modals are also subject to imprecision. This helps to explain how they can be precisified and approximated by expressions like *absolutely* and *basically*. It also explains commonalities between modal discourses and more canonical cases of precisification. This approach is helpful explaining two long standing puzzles in modality, Sobel Sequences, and what I call Lewis Sequences.

The most striking commonality between these kinds of discourses and canonical cases of precisifications involving numerals is the unidirectionality of both phenomena. In order to account for this unidirectionality, I propose a sketch of a pragmatic model of implicature which is intended to account for imprecision. This model shows why precisifications, including Lewis and Sobel Sequences, are unidirectional; namely, there are expressions, which, by virtue of their meaning, are felicitous only on higher standards of precision, but there are no expressions, and there cannot be any, which are felicitous only on lower standards of precision.

Bibliography

- Anand, P. and A. Brasoveanu (2010). Modal concord as modal modification. *Proceedings of Sinn und Bedeutung* 14, 19–36.
- Arregui, A. (2009). On similarity in counterfactuals. *Linguistics and Philosophy* 32(3), 245–278.
- Asher, N. and E. McCready (2007). Were, would, might and a compositional account of counterfactuals. *Journal of Semantics* 24, 93–129.
- Barwise, J. and R. Cooper (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy* 4, 159–219.
- Beltrama, A. and M. R. Bochnak (to appear). Intensification without degrees cross-linguistically. *Natural Language and Linguistic Theory*.
- Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford University Press.
- Bochnak, M. R. (2013). Two sources of scalarity in the verb phrase. In B. G. Boban Arsenijević and R. Marín (Eds.), *Subatomic Semantics of Event Predicates*, pp. 99–123. Dordrecht: Springer.
- Bylinina, L. (2012). “functional standards” and relative/absolute distinction. In A. A. Guevara, A. Chernilovskaya, and R. Nouwen (Eds.), *Proceedings of Sinn und Bedeutung* 16, Volume 2, pp. 141–157. MITWPL.
- Cariani, F., M. Kaufmann, and S. Kaufmann (2012). Deliberative modality under epistemic uncertainty. Manuscript.
- Carr, J. (2012). Deontic modals without decision theory. In E. Chemla, V. Homer, and G. Winterstein (Eds.), *Proceedings of Sinn und Bedeutung* 17, pp. 167–182.

- Charlow, N. (2013). What we know and what to do. *Synthese* 190, 2291–2323.
- Chierchia, G. (2006). Broaden your views: Implicatures of domain widening and the “logicality” of language. *Linguistic Inquiry* 37(4), 535–590.
- Condoravdi, C. (2002). Temporal interpretation of modals: Modals for the present and for the past. In D. Beaver, S. Kaufmann, B. Clark, and L. Casillas (Eds.), *The Construction of Meaning*, pp. 59–88. CSLI Publications.
- Condoravdi, C. and J. M. Gawron (1996). The context-dependency of implicit arguments. In *Quantifiers, Deduction, and Context*. CSLI Publications.
- Cresswell, M. J. (1976). The semantics of degree. In B. Partee (Ed.), *Montague Grammar*, pp. 261–292. Academic Press.
- Davies, M. (2008). The corpus of contemporary american english: 425 million words, 1990-present. <http://corpus.byu.edu/coca>.
- de Jager, T. (2009). ‘Now that you mention it, I wonder...’: Awareness, Attention, Assumption. Ph. D. thesis, University of Amsterdam.
- Edgington, D. (1995, April). On conditionals. *Mind* 104(414), 235–329.
- Égré, P. and M. Cozic (2011). If-clauses and probability operators. *Topoi* 30, 17–29.
- von Fintel, K. (1997). Bare plurals, bare conditionals, and *only*. *Journal of Semantics* 14(1), 1–56.
- von Fintel, K. (2001). Counterfactuals in a dynamic context. In M. Kenstowicz (Ed.), *Ken Hale: A Life in Language*, pp. 123–152. MIT Press.
- von Fintel, K. (2012). The best we can (expect to) get? challenges to the classic semantics for deontic modals. Talk at Central APA.
- von Fintel, K. and T. Gillies (2010). Must... stay... strong! *Natural Language Semantics* 18(4), 351–383.

- von Fintel, K. and S. Iatridou (2008). How to say *ought* in foreign: The composition of weak necessity modals. In J. Guéron and J. Lecarme (Eds.), *Time and Modality*, pp. 115–141. Springer.
- Geurts, B. and J. Huitink (2006). Modal concord. In *Proceedings of the ESSLLI 2006 Workshop Concord Phenomena at the Syntax Semantics Interface*, pp. 15–20.
- Giannakidou, A. (1997). *The Landscape of Polarity Items*. Ph. D. thesis, University of Groningen.
- Giannakidou, A. (1998). *Polarity Sensitivity as (Non)veridical Dependency*. Benjamins.
- Giannakidou, A. (2006). *Only*, emotive factive verbs, and the dual nature of polarity dependency. *Language* 82, 575–603.
- Giannakidou, A. and A. Mari (to appear). Future morphemes and epistemic modals: Modal adverbs measuring speaker’s confidence. In J. Blaczack (Ed.), *Tense, mood, and modality: New answers to old questions*. University of Chicago Press.
- Gibson, E., S. Piantadosi, and K. Fedorenko (2011). Using mechanical turk to obtain and analyze english acceptability judgments. *Language and Linguistics Compass* 5(8), 509–524.
- Gillies, A. S. (2007). Counterfactual scorekeeping. *Linguistics and Philosophy* 30(3), 329–360.
- Grice, H. P. (1957, July). Meaning. *The Philosophical Review* 66(3), 377–388.
- Grice, H. P. (1989). Logic and conversation. In *Studies in the Ways of Words*, pp. 22–40. Harvard University Press: Cambridge, Massachusetts.
- Grosz, P. (2010). Grading modality: A new approach to modal concord and its relatives. *Proceedings of Sinn und Bedeutung* 14, 185–201.
- Hackl, M. (2000). *Comparative Quantifiers*. Ph. D. thesis, MIT.
- Hájek, A. (2003). What conditional probability could not be. *Synthese* 137, 273–323.

- Hájek, A. (Ms.). Most counterfactuals are false. Unpublished manuscript.
- Hamblin, C. (1959). The modal “probably”. *Mind* 68(270), 234–240.
- Heim, I. (1982). *The Semantics of Definite and Indefinite NPs*. Ph. D. thesis, MIT.
- Heim, I. and A. Kratzer (1998). *Semantics in Generative Grammar*. Blackwell.
- Hintikka, J. (1969). Semantics for propositional attitudes. In J. Davis, D. Hockney, and W. Wilson (Eds.), *Philosophical Logic*, pp. 21–45. Dordrecht: Reidel.
- Holliday, W. H. and T. F. Icard (2013). Measure semantics and qualitative semantics for epistemic modals. In *Proceedings of SALT 23*, pp. 514–534.
- Horn, L. R. (1972). *On the Semantic Properties of Logical Operators in English*. Ph. D. thesis, UCLA.
- Horn, L. R. (2005). Airport ’86 revisited: Toward a unified indefinite *any*. In G. N. Carlson and F. J. Pelletier (Eds.), *Reference and Quantification: The Partee Effect*, pp. 179–205. CSLI Publications.
- Huitink, J. (2012). Modal concord: A case study of dutch. *Journal of Semantics* 29(3), 403–437.
- Huitink, J. (2014). Modal concord. ms.
- Iatridou, S. (2000). The grammatical ingredients of counterfactuality. *Linguistic Inquiry* 31(2), 231–270.
- Kadmon, N. and F. Landman (1993). Any. *Linguistics and Philosophy* 15, 353–422.
- Katz, G., P. Portner, and A. Rubinstein (2012). Ordering combination for modal comparison. In *Proceedings of SALT 22*, pp. 488–507.
- Kennedy, C. (1999). *Projecting the adjective: The syntax and semantics of gradability and comparison*. Garland Press, New York.
- Kennedy, C. (2007). Vagueness and grammar: the semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy* 30(1), 1–45.

- Kennedy, C. and L. McNally (2005). Scale structure and semantic typology of gradable predicates. *Language* 81(2), 345–81.
- Klecha, P. (2011). Optional and obligatory modal subordination. In *Proceedings of Sinn und Bedeutung* 15, pp. 365–379.
- Klecha, P. (2012). Positive and conditional semantics for gradable modals. In A. A. Guevara, A. Chernilovskaya, and R. Nouwen (Eds.), *Proceedings of Sinn und Bedeutung* 16, Volume 2, pp. 363–376. MITWPL.
- Klecha, P. (2014). Diagnosing modality in predictive expressions. *Journal of Semantics* 31(3), 443–455.
- Klecha, P. (to appear). Modal constraints on temporal reference. In H.-L. Huang, E. Poole, and A. Rysling (Eds.), *Proceedings of NELS* 43, Volume 1, pp. 239–252.
- Klein, E. (1980). A semantics for positive and comparative adjectives. *Linguistics and Philosophy* 4, 1–45.
- Kolodny, N. and J. MacFarlane (2010). Ifs and oughts. *Journal of Philosophy* 107(3), 115–143.
- Kratzer, A. (1977). What ‘must’ and ‘can’ must and can mean. *Linguistics and Philosophy* 1(3), 337–355.
- Kratzer, A. (1981a). The notional category of modality. In H.-J. Eikmeyer and H. Rieser (Eds.), *Words, Worlds, and Context*, pp. 38–74. de Gruyter.
- Kratzer, A. (1981b). Partition and revision: The semantics of counterfactuals. *Journal of Philosophical Logic* 10, 201–216.
- Kratzer, A. (1986). Conditionals. *Proceedings of CLS* 22(2), 1–15.
- Kratzer, A. (1991). Modality. In A. von Stechow and D. Wunderlich (Eds.), *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung*, pp. 639–650. de Gruyter.
- Kratzer, A. (2012). *Modality and Conditionals*. Oxford University Press.

- Krifka, M. (1999). At least some determiners aren't determiners. In K. Turner (Ed.), *The Semantics/Pragmatics Interface from Different Points of View*, Volume 1 of *Current Research in the Semantics/Pragmatics Interface*, pp. 257–291. Elsevier Science B.V.
- Krifka, M. (2007). Approximation of number words: a case for strategic communication. ms.
- Kripke, S. A. (1959). A completeness theorem in modal logic. *The Journal of Symbolic Logic* 24(1), 1–14.
- Lasnik, P. (1999). Pragmatic halos. *Language* 75(3), 522–551.
- Lassiter, D. (2010). Gradable epistemic modals, probability, and scale structure. In N. Li and D. Lutz (Eds.), *Proceedings of SALT 20*, pp. 197–215.
- Lassiter, D. (2011). *Measurement and Modality*. Ph. D. thesis, New York University.
- Lauer, S. (2012). On the pragmatics of pragmatic slack. In A. A. Guevara, A. Chernilovskaya, and R. Nouwen (Eds.), *Proceedings of Sinn und Bedeutung 16*, Volume 2, pp. 389–401. MITWPL.
- Lewis, D. (1973). *Counterfactuals*. Harvard University Press: Cambridge, Massachusetts.
- Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic* 8(3), 339–359.
- MacFarlane, J. (2008). Truth in the garden of forking paths. In M. Kölbel and M. García-Carpintero (Eds.), *Relative Truth*, pp. 81–102. Oxford University Press.
- McNabb, Y. (2012a). Cross-categorial modification of properties in hebrew and english. In A. Chereches (Ed.), *Proceedings of SALT 22*, pp. 365–382.
- McNabb, Y. (2012b). *The Syntax and Semantics of Degree Modification*. Ph. D. thesis, University of Chicago.
- Morzycki, M. (2001). *Almost* and its kin, across categories. In R. Hastings, B. Jackson, and Z. Zvolensky (Eds.), *Proceedings of SALT 11*, pp. 306–325.

- Morzycki, M. (2011). Metalinguistic comparison in an alternative semantics for imprecision. *Natural Language Semantics* 19(1), 39–86.
- Morzycki, M. (2012). Adjectival extremeness: Degree modification and contextually restricted scales. *Natural Language and Linguistic Theory* 30(2), 567–609.
- Morzycki, M. (2015). *Modification*. Cambridge University Press.
- Moss, S. (2012). On the pragmatics of counterfactuals. *Noûs* 46(3), 561–586.
- Munroe, R. (2013). Bridge. <http://xkcd.com/1170/>.
- Paradis, C. (2003). Between epistemic modality and truth conditions: The case of *really*. *Topics in English Linguistics* 44, 191–222.
- Portner, P. (2009). *Modality*. Studies in Semantics and Pragmatics. Oxford University Press.
- Prince, A. and P. Smolensky (1993/2002). *Optimality Theory: Constraint Interaction in Generative Grammar*. Rutgers Optimality Archive.
- Rapp, I. and A. von Stechow (1999). *Fast* “almost” and the visibility parameter for d-adverbs. *Journal of Semantics* 16(2), 149–204.
- Roberts, C. (2012). Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5(6), 1–69.
- Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics* 1(1), 75–116.
- Rotstein, C. and Y. Winter (2004). Total adjectives vs. partial adjectives: Scale structure and higher-order modification. *Natural Language Semantics* 12, 259–288.
- Rubinstein, A. (2012). *Roots of Modality*. Ph. D. thesis, University of Massachusetts, Amherst.
- Sadock, J. (1981). *Almost*. In P. Cole (Ed.), *Radical Pragmatics*. Academic Press.

- Sauerland, U. and P. Stateva (2011). Two types of vagueness. In P. Égré and N. Klinedinst (Eds.), *Vagueness and Language Use*, Chapter 6, pp. 121–145. Palgrave.
- Silk, A. (2013). Evidence sensitivity in weak necessity deontic modals. *Journal of Philosophical Logic*, 1–33.
- Sobel, J. H. (1970). Utilitarianisms: Simple and general. *Inquiry* 13, 394–449.
- Solt, S. (2012). Comparison to fuzzy standards. In A. A. Guevara, A. Chernilovskaya, and R. Nouwen (Eds.), *Proceedings of Sinn und Bedeutung* 16, Volume 2, pp. 557–570. MITWPL.
- Solt, S., C. Cummins, and M. Palmović (2013a). The preference for approximation. ms.
- Solt, S., C. Cummins, and M. Palmović (2013b). Preference for rounding. Talk at LSA 87.
- Sperber, D. and D. Wilson (1986a). Loose talk. *Proceedings of the Aristotelian Society* 87, 153–171.
- Sperber, D. and D. Wilson (1986b). *Relevance: Communication and cognition*. Blackwell.
- Sprouse, J. (2011). A validation of amazon mechanical turk for the collection of acceptability judgments in linguistic theory. *Behavioral Research Methods* 43(1), 155–167.
- Stalnaker, R. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in Logical Theory*, pp. 98–112. Oxford: Blackwell.
- Stalnaker, R. (1984). *Inquiry*. MIT Press.
- von Stechow, A. (1984). Comparing semantic theories of comparison. *Journal of Semantics* 3, 1–77.
- Todorova, M., K. Straub, W. Badecker, and R. Frank (2000). Aspectual coercion and the online computation of sentential aspect. In *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*, pp. 3–8.

- Warmbrod, K. (1981a). Counterfactuals and substitution of equivalent antecedents. *Journal of Philosophical Logic* 10, 267–289.
- Warmbrod, K. (1981b). An indexical theory of conditionals. *Dialogue: Canadian Philosophical Review* 20, 644–664.
- Warmbrod, K. (1983). Epistemic conditionals. *Pacific Philosophical Quarterly*, 249–265.
- Willer, M. (2012). A remark on iff-oughts. *Journal of Philosophy* 109(7), 449–461.
- Willer, M. (2013a). Dynamics of epistemic modality. *Philosophical Review* 122(1), 45–92.
- Willer, M. (2013b). Indicative scorekeeping. In M. Aloni, M. Franke, and F. Roelofsen (Eds.), *Proceedings of the 19th Amsterdam Colloquium*, pp. 249–256.
- Williams, J. R. G. (2008). Conversation and conditionals. *Philosophical Studies* 138(2), 211–223.
- Yalcin, S. (2007). Epistemic modals. *Mind* 116(464), 983–1026.
- Yalcin, S. (2010). Probability operators. *Philosophy Compass* 5(11), 916–37.