SEMANTIC UNDERSPECIFICATION AND ITS CONTEXTUAL RESOLUTION IN THE DOMAIN OF DEGREES

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF LINGUISTICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Ciyang Qing
November 2020

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Daniel Lassiter, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Cleo Condoravdi

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Judith Degen

Approved for the Stanford University Committee on Graduate Studies.

Stacey F. Bent, Vice Provost for Graduate Education

Abstract

Many natural language expressions have context-sensitive interpretations. For instance, *John is tall* can be interpreted differently depending on whether we are comparing him with the general adult population or professional basketball players. Context sensitivity can make it difficult to pin down the exact meaning of an expression. To address this challenge, it is common for semanticists to assign context-sensitive expressions underspecified meaning representations, where parts of the meanings need to be resolved in context. This naturally raises two important theoretical issues concerning their *semantics/pragmatics* interface: (i) How do we represent their conventional, context-invariant meaning (i.e., the *semantic* question)? (ii) How does context, together with the conventional meaning, determine the interpretation (i.e., the *metasemantic* question)?

In this dissertation, I focus on the domain of degrees and argue for the generality and importance of the distinction between *semantic* contextual resolution mechanisms and *pragmatic* ones in addressing these questions. The contrast is in parallel with that between the pronouns *I* (which conventionally refers to the speaker in context) and *they* (whose referent is determined via pragmatic reasoning).

Concretely, I use *directional modified numerals* (e.g., *up to 100*) and *gradable adjectives* (e.g., *tall* and *full*) as two case studies to argue that (a) the two contextual resolution mechanisms can be unified by a general principle of *informativity-applicability trade-off*, which is a quantitative generalization of the interaction between Grice's maxims of Quantity and Quality, and that (b) the way to tease them apart is by examining whether the contextual resolution of an expression is sensitive to the larger linguistic environment in which it appears. If yes, the mechanism is pragmatic, and if no, it is semantic. For directional modified numerals, the opposite inference patterns they trigger when they are unembedded and embedded under deontic modals suggest that they have a pragmatic contextual resolution mechanism. For relative and maximum gradable adjectives, I examine how they are used in definite descriptions in referential contexts and argue that, contrary to some recent proposals, their contextual resolution mechanism is in fact semantic in nature.

In addition to providing a more precise and unified semantics for *relative* (e.g., *tall*) and *maximum* (e.g., *full*) gradable adjectives, I further examine the class of *minimum* gradable adjectives (e.g., *bent*) and illustrate that many of them are in fact systematically ambiguous between a relative reading and a > 0 reading, which is based on a meaningful notion of zero rather than the minimum.

I argue that this suggests a new taxonomy of positive forms of gradable adjectives. On the one hand, some positive forms are threshold-introducing, which include maximum and relative adjectives, together with minimum adjectives with relative readings. They share a unified semantics and their differences are explained by the different contextual parameters. On the other hand, the > 0 reading of a gradable adjective is derived in a way parallel to comparative constructions, which provides a straightforward explanation of their similarities observed in the literature.

Acknowledgments

First, it goes without saying that my study at Stanford, let alone this dissertation, would have been impossible if I did not attend ESSLLI 2013 and took the courses Dan Lassiter taught there. Dan's introduction to the empirical and theoretical issues surrounding gradable adjectives and the probabilistic model he and Noah Goodman proposed to analyze them immediately fascinated me. I was deeply impressed by the elegance and the generality of the proposal, but in the meantime I couldn't help but feel that something is missing. So I tried to apply what I thought was the core insight of Dan and Noah's model, and drafted an alternative analysis motivated from a slightly different perspective. When I approached Dan to discuss this idea, even though it was so primitive and in hindsight it was probably inscrutable, Dan was very interested. He encouraged me to write down the analysis in detail and share it with him, and also to consider applying to Stanford if I would like to further work on probabilistic approaches to semantics and pragmatics. I did. Together with Michael Franke, I developed the idea into my very first SALT presentation, and later began my study at Stanford Linguistics.

This very first encounter between Dan and me is in fact quite representative of our advising relationship. At the beginning of a project, I tended to only have a very vague sense of what I was after and perhaps some very primitive ideas of how I would achieve my goal. But despite this, when I presented my ideas to Dan, no matter how fuzzy they were, he always took them seriously, and encouraged me to explore them more and to pin them down in more concrete forms. Moreover, I have always been impressed by how Dan was able to provide insightful advice about how to proceed, as well as pointers to various types of resources that would later prove highly relevant and useful. But in a way this is unsurprising, considering the highly interdisciplinary nature of his own research, which was also reflected in the courses he taught and the workshops and seminars he organized: they were all so inspiring! Finally, I really appreciate the confidence Dan had in me and how patient he was when my progress was slow (or even none), especially considering that I was aiming to show certain limitations of his own theory in part of this dissertation. Even though I felt from the get-go that something is missing in the theory and an alternative analysis is needed, the difference is quite nuanced and it took a very long time before I could properly spell out the difference and defend my position. Along the way, there were so many failed attempts and so many times where I was simply stucked, that I sometimes wondered whether this was all just a waste of time. But Dan never thought so. After pointing out that there were holes in my arguments or that they were not as clear as I had thought, he would encourage me to look for ways to improve them and suggest possible directions. Even though for a long time he was not convinced by what I was claiming, Dan did believe that I would ultimately have something real to contribute. This dissertation finally becomes what it is now thanks to this trust, and I am incredibly lucky to have Dan as my advisor. Of course, Dan's influences on me go way beyond this dissertation, but it is just impossible to enumerate them all.

I also cannot thank Cleo Condoravdi enough for her guidance throughout my time at Stanford. Cleo was a committee member for both of my qualifying papers as well as this dissertation. She was so generous with her time, and her advice was tremendously helpful to me. I especially appreciated the fact that she always encouraged (and sometimes pushed) me to get out of my comfort zone. When choosing a research project, it is very tempting to just stick with a familiar topic, and when dealing with challenging and/or nuanced issues, it is also very tempting to just provide a hand-wavy answer and be satisfied with it. Through her encouragement for exploration and insistence on rigor, Cleo helped me resist such temptations and improve the quality of my research.

Judith Degen joined the linguistics department not so long ago, but we actually knew each other before I came to Stanford. I learned a lot from her as a student in her class and a member in her lab, and I am always amazed by how organized and effective she is. But most importantly, she really showed me the importance and power of perseverance. Throughout the development of this dissertation, Judith provided invaluable comments and suggestions that helped it stand on more solid empirical grounds.

Special thanks are due to Michael Franke, without whom this dissertation would also not have been possible. Michael taught me game-theoretic pragmatics, which was in fact my very first introduction to pragmatics and greatly shaped my understanding of it. Also, this dissertation stemmed from our joint work on gradable adjectives. It was under his supervision that I first got a taste of what it is like to do research in semantics/pragmatics and fell in love with it.

This dissertation has also benefited from discussions with Heather Burnett, Noah Goodman, Nicole Gotzner, Roger Levy, Louise McNally, Chris Potts, Stephanie Solt, Benjamin Spector, Ming Xiang, and Zhuoye Zhao, among many others. I would especially like to thank Nicole Gotzner and Stephanie Solt for the very helpful discussions over the years on various occasions. At the early stage of the dissertation, I was invited by Nicole to present at ZAS. After the presentation, which was based on Chapter 3, Nicole asked how I would analyze minimum adjectives. Back then, I had not carefully considered minimum adjectives and just thought that they could be analyzed in the same way as maximum adjectives, so that was how I replied. But I guess Nicole's question planted a seed of doubt in my mind, urging me to take a closer look at the class of minimum adjectives. This ultimately led to a totally different analysis of minimum adjectives in Chapter 4. As for Stephanie, I've always admired her ability to make insightful and thought-provoking observations about the subtle differences between various expressions, both in her work and during our discussions. Her influence on this dissertation should be evident.

I am grateful to Beth Levin, Vera Gribanova, and the linguistics department in general for making the administrative aspect of the graduate program as smooth as possible. In particular, I really, really appreciate the support from the department to help me navigate through the complicated situation caused by the pandemic. It was a real lifesaver.

I want to thank my fellow semanticists/pragmaticists and regular attendants of Words With Friends for the supportive community: James Collins, Lelia Glass, Masoud Jasbi, Sunwoo Jeong, Sara Kessler, Prerna Nadathur, Brandon Waldon, and my friends in the little Linguistics-SymSys circle for all the fun activities together: Sabrina Grimberg, Jason Freeman, Ben Peloquin, Naomi Shapiro, and Robert Xu. I was very fortunate to also have some good old friends around. Special thanks to Junyang Qian for all the help and support, and all the amazing time we had together over the past six years. Also special thanks to Qiushu Chen, Jie Li, Ziqin Rong, Tailin Wu, Xinyi Yang, and Zhiyuan Yao, for the luxury of playing "Tractor" together at a place thousands of miles away from our hometown. I am also very grateful to Aunt Yu Zhu and Uncle Tong Chang for all their help, especially during my parents' visits. Finally, special thanks to Ge Du, Qian Li, Yuanhang Wang, and Siyao Xie, for being such great friends, no matter how far away we might be.

Last but not least, I want to thank my family, especially my parents, for their unconditional love and support. Any words would pale in light of what they have done and are willing to do to ensure that I am healthy, happy, and free. So I just want to say: "Thank you, Mom and Dad! I love you!"

Contents

List of Figures

Chapter 1

Introduction

1.1 Context sensitivity and the semantics/pragmatics interface

Many natural language expressions have *context-sensitive* interpretations. For example, the gradable adjective *tall*, which belongs to a class of adjectives that will be studied in detail in Chapter 3, is context sensitive (1).

(1) John is tall.

Suppose that John is $6'2''$ tall and is a professional basketball player. Then (1) would (generally) be considered true if we are comparing him with the general US male adult population, but false if we are comparing him with professional basketball players.

In general, context-sensitivity can make it difficult to spell out the meaning of an expression. In light of this difficulty, it is common for semanticists to assign *underspecified* meaning representations to context-sensitive expressions. For instance, (2a) and (2b) correspond to two major types of meaning representations for (1).

(2) a. $[\![\text{John is tall}]\!] = \mathbf{height}(\mathbf{j}) \geq \theta$, where θ is a contextually determined standard/threshold

b. $[\![\text{John is tall}]\!] = \mathbf{tall}(K)(\mathbf{j})$, where K is a contextually determined set of individuals

According to (2a), (1) is true iff John's height reaches a contextually determined standard, whereas according to (2b), (1) is true iff John is tall relative to a contextually determined set of individuals. Such semantic analyses provide a way to *represent* the context-sensitive interpretations of an expression. Suppose we are in the context where we are comparing John with the general US male adult population. Proponents of (2a) would say that in this context the standard is (presumably) around $6'$ or $6'1''$. Therefore, if John is $6'2''$ then (1) is true. Proponents of (2b) would say that the relevant set K is the set of US male adults. Presumably, if John is $6'2''$, he is tall relative to this set K, and therefore (1) is true. Now suppose we are in the context where we are comparing John with professional basketball players. Proponents of (2a) would say that in this context the standard is much higher than $6'2''$. Therefore, (1) is false when John is $6'2''$. Proponents of (2b) would say that the relevant set K is now the set of professional basketball players. Presumably, if John is $6'2''$, he is not tall relative to this set K, and therefore (1) is false.

Semantic underspecification provides us with the necessary flexibility to represent context-sensitive interpretations. This is particularly useful in that it allows us to explore how such representations are derived compositionally and how they will participate in the meaning composition of even larger expressions, which are central questions in formal compositional semantics. However, semantic underspecification per se does not provide a deep understanding of why an expression has the interpretations that it does in different contexts, because it merely represents the fact that an expression can have different interpretations in different contexts. It needs to be supplemented with a theory that spells out how the underspecified part in the semantics, e.g., the threshold θ in (2a) and the set of relevant individuals K in (2b), is determined, or *resolved* in context. In theoretical linguistics, such a theory is considered to be in the realm of *pragmatics*, in particular concerning the *semantics/pragmatics interface*. In philosophy of language, this issue of how a (possibly semantically underspecified) natural language expression gets its semantic value in context is considered to be part of the *metasemantics* (Kaplan, 1989a).[1] Following Kaplan (1989b), the literature on this metasemantic issue has traditionally been focusing on demonstratives. In recent years, however, there has been a series of discussions and debates about the metasemantic question concerning other context-sensitive expressions such as gradable adjectives (e.g., Glanzberg, 2007, 2009, 2020; King, 2014a, 2020; K. S. Lewis, 2019). In general, for each context-sensitive natural language expression, we can ask the following questions (3).

(3) a. **The semantic question**: What is the meaning of the expression in context? In particular, we are interested in which part of this meaning is due to the context-invariant, conventional meaning of the expression, and which part is underspecified and needs to be resolved in context. Also, in formal semantics, we want to know how to formally represent such a meaning in a way that allows us to build such a representation compositionally.

b. **The metasemantic question**: What determines the meaning of a context-sensitive expression in context? In other words, how is semantic underspecification resolved in context?

These two questions are interrelated. As discussed above, an analysis that only answers the semantic question is unlikely to provide deep insights into the nature of context sensitivity. Meanwhile,

[1] Kaplan (1989a) does consider the label "pragmatics," because he thinks that questions like this are related to language use, but comments that he is "not really comfortable with this nomenclature" (p. 573). Immediately after that he considers the label "metasemantics" and settles on it.

the metasemantic question presupposes an answer to the semantic question. Therefore, answers to the two questions need to be evaluated together.

Addressing the metasemantic question is important if our goal is to provide a predictive theory of how interlocutors produce and interpret context-sensitive natural language expressions. Moreover, it can provide useful considerations for evaluating answers to the semantic question. For example, the different semantic representations (2a) and (2b) have traditionally been motivated by considerations of other constructions that involve gradable adjectives, such as comparatives (e.g., *John is taller than Mary*) and measure phrase modifications (e.g., *John is* $6'$ *tall*). Each representation also has multiple variants with different auxiliary assumptions. In Chapter 3, I illustrate how considerations of the metasemantic question can inform our choice among these possibilities and lead to a new semantic analysis that has the best empirical coverage with the least stipulations.

In this dissertation, I use gradable adjectives and directional modified numerals (e.g., *up to 100*) as two case studies to show how we can address the semantic and metasemantic questions in a principled way. In particular, I argue that while they both introduce an underspecified lower bound/threshold, we need two types of contextual resolution mechanisms to answer the metasemantic question. The distinction between the two mechanisms is described and discussed in more detail below.

1.2 Contextual resolution mechanisms: Semantic vs pragmatic

In this section, I introduce the distinction between semantic and pragmatic contextual resolution mechanisms. This distinction has always been implicit in the literature, and is in line with a common understanding of the difference between semantics and pragmatics. The main purpose of the discussion below is to make this distinction explicit, and compare it with some other related distinctions in the literature. I will use the pronouns *I* and *they* as paradigmatic examples of the two contextual resolution mechanisms.

As is well known, the interpretations of the pronouns *I* and *they* are context sensitive. For example, in a context where John is the speaker of (4), (4) would be interpreted as saying that John likes coffee. In another context where Mary is the speaker, (4) would be interpreted as saying that Mary likes coffee.

(4) I like coffee.

a. Context SPEAKER-JOHN: John is the speaker.[2]

b. Context SPEAKER-MARY: Mary is the speaker.

Similarly, depending on the context, the interpretation of *they* in (5) can differ.

[2] In this dissertation, I use SMALL-CAPS to name contexts.

(5) The students$_i$ saw the firefighters$_j$ as they$_{i/j}$ left the building.

a. Context 7AM-ALARM: The speaker says the following before uttering (5)

Last Thursday, a fire alarm in our school went off at 7am. The firefighters next door were called to the scene for a close inspection.

b. Context 10AM-ALARM: The speaker says the following before uttering (5)

Last Thursday, a fire alarm in our school went off at 10am. The firefighters next door were called to the scene for a close inspection.

In 7AM-ALARM, given that the fire alarm went off early in the morning, it is plausible to assume that the firefighters came to an empty building, shut it down for a close inspection, and when they were done, they left the building, which was when the students saw them. Interpreted in this way, *they* refers to the firefighters in this context.[3] In 10AM-ALARM, however, since the fire alarm went off during school hours and the firefighters were close, it is plausible to assume that the students were still in the evacuation process when the firefighters arrived. Interpreted in this way, *they* refers to the students in this context.

These examples show that the interpretations of the pronouns *I* and *they* are context sensitive. Intuitively, their meanings in context are their referents. Therefore, an answer to the semantic question could be that both *I* and *they* have a referent that is determined in context. Formally, we can provide the following meaning representations (6). Note that such meaning representations are underspecified, in the sense that we do not spell out exactly what the referent is.

(6) a. $[\![\text{I}]\!]^c = r$, where r is a referent determined in context

b. $[\![\text{they}]\!]^c = r$, where r is a referent determined in context

However, as Kaplan (1989b) points out, the pronoun *I*, which he calls a *pure indexical*, is special in that its conventional meaning completely determines its referent in each context, i.e., the referent of *I* must be the speaker of the utterance in the context. To better analyze this conventional meaning of *I*, Kaplan distinguishes between *content*, i.e., the meaning of an expression in context, e.g., (7a), and *character*, i.e., a function that maps a context to the content in that context (7b), which he takes to be the conventional meaning of an expression.

(7) a. $[\![\text{I}]\!]^c = \textbf{speaker}(c)$

b. $[\![\text{I}]\!] = \lambda c.\, \textbf{speaker}(c)$

[3] Throughout this dissertation, I use the terms *refer* and *referent* only in an intuitive, descriptive, theory-neutral sense. Some theories use alternative notions to avoid saying that a demonstrative refers to an object. For instance, King (2001) says that a demonstrative has an object as its semantic value in context. I will also use this terminology, but again only at a descriptive level, i.e., I will use the *referent* and the *semantic value* of a demonstrative (or a definite description) interchangeably.

What is crucial for our current purpose is that the referent of *I* in context is determined by the conventional meaning (i.e., semantics) of *I*. This addresses the metasemantic question, i.e., what determines the meaning of *I* in context, by appealing to the semantics of *I*. Therefore I will also say that the referent of *I* is contextually resolved by a semantic mechanism. This has a couple of important consequences given common assumptions about conventional meanings. Under the assumption that the conventional meaning constrains possible interpretations, it follows that there is little flexibility in how the referent of *I* is supposed to be resolved in context. For instance, in SPEAKER-JOHN, *I* must refer to John. It would be infelicitous for the speaker (i.e., John) to follow (4) with "Oh, just to clarify, by that I meant Bob likes coffee." Another important consequence is that, since the mapping from a context and the referent of *I* in that context is part of the semantics of *I*, due to the principle of compositionality, the very same mapping will hold regardless of whether *I* is embedded under other operators. For instance, suppose you run into John, a friend you have not seen for a while, and you notice that he is wearing a wedding ring. If he says to you: "I am married," you would interpret *I* as referring to John. If he says to you: "I am not married," you would still interpret *I* as referring to John, even though under this interpretation what John says is surprising given that he is wearing a wedding ring (and therefore you might reasonably expect some further explanations such as "I am just bringing this ring to my friend but I lost the box").

In contrast, it is difficult to see whether the conventional meaning of the pronoun *they* plays any substantial role in determining its referent in context. Even though there is a salient interpretation of *they* in each of the two contexts in (5), it is not the only possible interpretation. For instance, in 7AM-ALARM, the most salient interpretation of *they* is the firefighters. However, it is also possible to interpret *they* as the students. The speaker can try to clarify that this is indeed the intended interpretation by saying, e.g., "Oh, just to clarify, by that I meant the students were leaving the building and these students saw the firefighters. There was a special event that day so the students came to the school very early." Therefore, it seems that the conventional meaning of the pronoun *they*, whatever it might be, imposes few constraints on and thus under-determines its contextual interpretation. One common way to formally implement this idea is to assume that the conventional meaning of *they* is just a variable (8a), whose interpretation in a context is determined by a contextual assignment function f_c (8b).[4] Quite often, (8b) is simply written as (8c), assuming that the only relevant aspect of the context for our current purpose is the assignment function.[5]

(8) a. $[\![\text{they}_i]\!] = x_i$

[4] The subscript here is used just to distinguish between multiple occurrences of the pronoun, unlike in (5), where the indices are meant to illustrate the intended referent. Also, I do not encode a plurality requirement here because there are reasons to think that it is derived pragmatically by competition with singular pronouns (see Büring, 2011 for further discussions).

[5] In some analyses of *bound* uses of pronouns, e.g., *every student$_i$ did their$_i$ homework*, the assignment function is used as a technical tool to build dependencies between the quantifier and the pronoun. Depending on the theory, the assignment function may or may not be viewed as related to context.

b. $[\![\text{they}_i]\!]^c = f_c(x_i)$, where f_c is a contextually determined assignment function

c. $[\![\text{they}_i]\!]^f = f(x_i)$, where f is a contextually determined assignment function

For the current purpose, we can think that such an analysis provides a satisfying answer to the semantic question, and we are more interested in the metasemantic question. It might seem that this analysis also provides an answer to the metasemantic question, i.e., what determines the referent of *they* is the contextual assignment function. However, there is a sense in which this answer is not quite satisfactory. To better illustrate this, I distinguish between the notions of *context* and *contextual parameter*. A context is a descriptive, intuitive notion, which can be roughly characterized as the situation in which an utterance is made. A contextual parameter is a formal, technical notion, used in our semantic analysis to represent the relevant aspect of a context. A context determines the values of a variety of contextual parameters, but should not be identified with such values (see Glanzberg, 2007 for more discussion). With this distinction in mind, we can see that the above answer to the metasemantic question makes reference to the contextual assignment function, which is a contextual parameter. As a result, unless we also provide an answer to how context determines this contextual parameter, (8c) does not provide any further insight into the metasemantic question than (6b). In general, it is important to keep in mind that the metasemantic question is not automatically solved just because we relativize our semantics to a contextual parameter. Whatever the answer to the metasemantic question is for *they*, the semantics of *they* will not provide such an answer, since we conclude from the earlier discussion that the conventional meaning of *they* imposes few constraints on and thus under-determines its referent. Therefore, I will say that the referent of *they* is contextually resolved by a *pragmatic* mechanism.

I will discuss a concrete pragmatic contextual resolution mechanism in the next section. For now, let us consider a few more examples to see how the distinction between semantic and pragmatic contextual resolution mechanisms helps us carve the space of context-sensitivity expressions with finer granularity.

Note that the two contextual resolution mechanism can be applied to *parts* of the meaning of an expression in context. Therefore, in general the meaning of an expression in context can be determined by a hybrid of semantic and pragmatic mechanisms. Consider *here*, which intuitively refers to a region determined by context. Suppose that you are visiting me at the linguistics department at Stanford and I utter (9) to you.

(9) People here love organic food.

The word *here* can be interpreted as the linguistics department at Stanford, or Stanford University, or California, and so on. Such possible interpretations are clearly related and constrained. They must be regions that include the location where the utterance takes place, and this is because of the conventional meaning of *here*. However, the conventional meaning of *here* does not fully determine its contextual interpretation. Not all interpretations that are compatible with the conventional meaning of *here* are equally salient or plausible. For example, it is unlikely that *here* would be interpreted as *the US*, and even more so for it to be interpreted as *Earth*.[6] Therefore, a pragmatic contextual resolution mechanism is also needed to pick out the salient interpretation(s) in this context and explain why certain interpretations are implausible.

In comparison, other related distinctions in the literature are used to classify *expressions*. For instance, Kaplan (1989b) distinguishes between *pure indexicals* (e.g., *I*, *now*, *here*) and *true demonstratives* (e.g., *that*, *he*). According to Kaplan, a pure indexical is a context-sensitive expression whose conventional meaning fully determines its referent in context, and a true demonstrative is one that requires demonstration to ensure that it has a referent in context. As is well known, the problem with the notion of pure indexicals is that it is questionable whether any expression other than *I* is a pure indexical. Kaplan (1989b, fn. 12) himself is aware of this problem but concludes that it does not "slur the difference between demonstratives and pure indexicals." I suggest that the distinction between semantic and pragmatic contextual resolution mechanisms can reasonably capture the difference Kaplan has in mind, without the problem caused by having to assign a category to each expression.

Similarly, King (2014a) distinguishes between pure indexicals and *supplementives*, which are simply defined as context-sensitive expressions that are not pure indexicals, and include a much wider range of expressions such as the gradable adjective *tall*. Then he proposes a single, unified answer to the metasemantic question for all supplementives. Despite the generality of his proposal, King's account ends up not providing a unified answer to the metasemantic question for gradable adjectives because some gradable adjectives such as *full* are presumed to be pure indexicals (or even not context-sensitive at all), which means that King's account does not apply to them (King, 2020). In Chapter 3, I argue that this does not provide a satisfying answer to the metasemantic question for gradable adjectives, and show how we can obtain a unified answer by considering the type of contextual resolution mechanism gradable adjectives use, i.e., whether it is semantic or pragmatic.[7]

The distinction between semantic and pragmatic contextual resolution mechanisms may seem obvious or even trivial, but this is only because we have clear intuitions about the conventional meanings of expressions such as *I*, *they*, and *here*. For many context-sensitive expressions, we do not have clear intuitions about which parts of their meanings are underspecified and whether those parts are contextually resolved semantically or pragmatically. For instance, suppose we agree that (2a) is an appropriate semantic representation of *John is tall*, repeated below in (10). It is unclear whether the underspecified threshold θ is resolved semantically or pragmatically in context, and hence the

[6] The first interpretation becomes more likely if you are visiting from outside the US, and the second one can also be improved if you were a Martian. Such observations provide further evidence for a pragmatic contextual resolution mechanism.

[7] Strictly speaking, the answer is not unified for all gradable adjectives because of the class of minimum adjectives that I will discuss in Chapter 4. However, as far as the class of relative adjectives (e.g., *tall*) and the class of maximum adjectives (e.g., *full*) are concerned, my analysis does provide a unified answer to the metasemantic question.

debates in the literature.

(10) $[\![\text{John is tall}]\!]^c = \textbf{height}(\textbf{j}) \geq \theta$, where θ is a contextually determined standard/threshold

Similarly, as I will argue in Chapter 2, there are good reasons to think that a directional modified numeral such as *up to 100* involves an underspecified lower bound in the semantic representation. For the current purpose, we can take (11) to be the semantic representation.

(11) $[\![\text{Up to 100 people will attend the meeting}]\!]^c =$

$$100 \geq n_{\text{attendants}} \geq \theta, \text{ where } \theta \text{ is a contextually determined lower bound}$$

Intuitively, it seems that the lower bound θ is often resolved to a number that is close to the upper bound 100, which is why (11) can often be interpreted as, e.g., *80 to 100 people will attend the meeting*. Now, is this lower bound resolved semantically or pragmatically in context? Again, intuition does not provide a clear answer.

Note that we are not asking whether the underspecified threshold or lower bound gets resolved semantically or pragmatically just for its own sake. As discussed earlier, the answer to this question will have empirical consequences for how the relevant expression gets interpreted when it is embedded under other linguistic environments. If the contextual resolution mechanism is semantic, then due to the principle of compositionality, we would expect the same dependency between the threshold and the relevant contextual parameter regardless of the linguistic environment the expression appears in. If the mechanism is pragmatic, we would expect more variability in how the threshold is resolved in different contexts and more sensitivity to certain contextual features such as the communicative intentions and goals of the interlocutors.

When we propose a semantic contextual resolution mechanism, we need to do three things: (i) specify the relevant contextual parameters, (ii) define the function that maps them to semantic values, and (iii) provide the necessary link from contexts to contextual parameters. For instance, for pronoun *I*, the relevant contextual parameter is the context itself, which means that there is a trivial link between the two, and the function **speaker** maps a context to its speaker. Note that this **speaker** function will always be defined, because all contexts are assumed to have a speaker. However, this need not be the case for semantic mechanisms in general. For instance, according to Kaplan (1989b), the conventional meaning (or character) of the demonstrative *that* (when accompanied by a demonstration δ) can be represented in (12a).

(12) a. $[\![\text{that}_\delta]\!] = \lambda c.\, \textbf{demonstratum}(\delta)(c)$

b. $[\![\text{that}_\delta]\!]^c = \textbf{demonstratum}(\delta)(c)$

The function **demonstratum** takes a demonstration δ and a context c and returns what is demonstrated by δ in c. Crucially, since not all demonstrations in a context successfully demonstrate something, **demonstratum** can sometimes be undefined. As a result the referent of *that* (12b) is not always defined in any context.

In contrast, when we propose a pragmatic mechanism, there is not a fixed list of things that we need to do, and in principle there are many different ways in which we can specify a pragmatic mechanism. For concreteness, in the next section, I introduce and discuss King's (2013, 2014a, 2014b) *coordination account*, which will be the starting point of the pragmatic contextual resolution mechanism that I will propose in Chapter 2.

1.3 The coordination account

King (2013, 2014a, 2014b) proposes what he calls the coordination account as a general answer to the metasemantic question. Here I summarize King's (2014a) version in (13).

(13) **The coordination account** (King, 2014a): the semantic value of an expression ϕ is o in context c (i.e., $[\![\phi]\!]^c = o$) iff the following two conditions are met

a. The speaker intends o to be the semantic value of ϕ in c.

b. A competent, attentive, reasonable hearer who knows the common ground of the conversation at the time of utterance would know that the speaker intends o to be the value of ϕ in c.[8]

Following King, I will often call the hearer as specified in (13b) an *idealized* hearer/ listener.

As a concrete example, let us see how the coordination account analyzes the demonstrative *that* (14).

(14) $[\![\text{that}]\!]^c = r$ iff the following two conditions are met

a. The speaker intends r to be the semantic value of *that* in c

b. An idealized listener would know the speaker's intention in (14a)

Crucially, (14) is different from Kaplan's (1989b) analysis (12b) in that the demonstrative *that* does not require an associated demonstration, and that its conventional meaning does not impose any constraint on its semantic value (except that it cannot be a plurality, which is omitted here). King (2013, 2014b) discusses how this coordination account can handle a variety of cases that are problematic for Kaplan's (1989b) analysis (12b). For example, suppose a speaker points in the general direction of a dog, a bike and a child and says 'Careful, that is a mean dog.' Intuitively the demonstrative *that* should have the dog as its semantic value, but in this context the demonstration

[8] King (2013) has two additional requirements. First, the hearer needs to have the properties attributed to the audience by the common ground. A relevant case is that the speaker is talking to blind people (and this fact is in the common ground). This requirement makes sure that we are considering what an idealized blind hearer would do, which means that pointing would not be a good way for such a hearer to recognize the speaker's referential intention. Second, the hearer needs to recognize the speaker's intention in the way the speaker intends the hearer to recognize it. This requirement makes sure that we rule out cases where the hearer happens to recognize the speaker's intention by a lucky accident. These additional requirements will not affect my discussions of the coordination account.

in itself is vague and therefore does not have a unique demonstratum to serve as the semantic value of *that* in (12b). In contrast, under the coordination account, if the speaker intends the dog to be the semantic value of *that* in this context, an idealized listener would know the speaker's intention, presumably because this is the only reasonable way to make what the speaker says potentially true (I will return to this assumption later). As a result, the demonstrative *that* is predicted to have the dog as its semantic value, which accounts for the intuition we have above.

Given its success in the domain of demonstratives, and its generality and uniformity, King's coordination account is a good starting point to address the metasemantic question for other context-sensitive expressions. However, it has two major limitations.

First, the account is not specific enough about what an idealized hearer would or would not know, which is entirely left to our intuition. To be sure, for most of the examples King discusses in his papers, which generally involve demonstratives, we do have clear intuitions about whether an idealized hearer would know a certain intention of the speaker. However, for other context-sensitive expressions, we often do not have clear intuitions about what intentions an idealized hearer would know. This can make the coordination account unable to state interesting generalizations about the interpretation of a context-sensitive expression, or worse, potentially unfalsifiable.

For instance, suppose that we want to apply the coordination account to (11), repeated below in (15).

(15) $[\![\text{Up to 100 people will attend the meeting}]\!]^c =$

$$100 \geq n_{\text{attendants}} \geq \theta, \text{ where } \theta \text{ is a contextually determined lower bound}$$

If the speaker intends some number, e.g., 80, to be the lower bound, would an idealized listener know this intention?[9] It is not clear what the answer should be here. To the extent that (15) seems to imply that the number of attendants is close to 100 and can therefore be interpreted as, e.g., *80 to 100 people will attend the meeting*, the answer seems to be yes. However, this way of answering the question does not seem very satisfying or interesting, and I suggest the reason is that it does not reveal *why* (15) can have such an interpretation. It is helpful to compare this with the example of vague demonstration discussed above. When we say that an idealized listener would know the speaker's intention for the dog to be the semantic value of *that*, we draw this conclusion not (just) because a competent, attentive, reasonable listener (such as ourselves) would take the speaker to intend the dog to be the semantic value of *that*, but because we have clear intuitions about *why* such a listener would do so, i.e., among the three possible semantic values of *that* suggested by pointing, only the dog would make the speaker's utterance potentially true, and we are assuming that the speaker follows Grice's (1975) Quality Maxim (or something to this effect). This suggests that the coordination account needs to be supplemented with an explicit theory of what an idealized listener

[9] Note that the speaker's intention does not need to be this specific, for instance, the speaker can intend a range of numbers to be the lower bound. King (2014a) discusses such possibilities in his analysis of gradable adjectives.

would know and why, and a good candidate is that the idealized listener assumes that the speaker is cooperative and follows Grice's Maxims. This addresses the first limitation, because we no longer need to rely on our intuitions to determine what an idealized listener would know. Rather, this is determined by the explicit theory and we can then check whether the coordination account makes predictions that track our intuitions.

Using this strategy, I argue in Chapter 2 that we can explain why (15) often seems to imply that the number of attendants is close to 100 in terms of the interaction between Grice's Maxims of Quantity and Quality. I show that this also explains why this implication disappears when *up to 100* is embedded under a permission modal and is used by a speaker with authority. For instance, *up to 100 people are allowed to attend the meeting* does not imply that only the numbers close to 100 are allowed. Intuitively, any number from 0 to 100 is allowed. Furthermore, I generalize and formalize this analysis and propose a probabilistic model that makes quantitative predictions about possible interpretations of *up to 100* in different contexts.

The second limitation of the coordination account is that, while it appears to be a pragmatic mechanism (since it does not directly make reference to the conventional meaning of the expression), it is in fact unable to distinguish between semantic and pragmatic contextual resolution mechanisms. For example, King (2014a, fn. 18) discusses how the conventional meaning of *he* can constrain what a speaker can reasonably intend (i.e., the intended referent must be male) and what an idealized hearer can reasonably recognize. In the context of this discussion, i.e., using a representation like (2a) as the semantics of gradable adjectives rather than Kennedy's (2007) formulation, King seems to imply that it is unnecessary to include the gender requirement of *he* in the semantic representation. However, this will miss an important distinction about the defeasibility of an interpretation (i.e., whether an interpretation is impossible or just implausible). Now, perhaps King's intention is that, as far as the metasemantic question is concerned, it does not matter whether the contribution of the conventional meaning is explicitly reflected in the semantic representation. While this is true, we should recall from our earlier discussion that answers to the semantic and metasemantic questions need to be evaluated together. In the case of pronouns, since their semantics are relatively uncontroversial (at the very least, it is clear that the gender requirements come from the conventional meaning), we can be more relaxed about their semantic representations and focus on the metasemantics. In contrast, in many other cases, such as gradable adjectives, the semantic question is much more complicated and subtle, and often there is no consensus on which part of the contextual interpretation should be part of the conventional meaning. Therefore we need to be more careful about the semantic representations we assume and evaluate the empirical consequences accordingly.

For this reason, I will assume that the coordination account is only used to provide a pragmatic contextual resolution mechanism, and examine whether it makes the correct predictions given a particular semantic representation. For instance, King (2014a) proposes the following coordination account for gradable adjectives such as *tall* (16).

(16) $[\![\text{tall}]\!]^c = \lambda x.\, \mathbf{height}(x) \geq \theta$ iff

a. The speaker intends θ to be the threshold

b. An idealized listener would know that the speaker's intention in (16a)

In Chapter 3, I provide arguments against this account for gradable adjectives, by considering their uses in definite descriptions in referential contexts (17).

(17) a. Give me the tall glass ($\approx$ give me the taller glass)

b. Give me the full glass ($\neq$ give me the fuller glass)

Suppose there are two glasses and one is taller than the other. When a speaker utters (17) intending some height θ between the heights of the two glasses, e.g., the height of the taller glass, to be the threshold, it seems that a competent, attentive, reasonable hearer who knows the common ground of the conversation at the time of utterance would be able to know this intention, because this is the only way to meet the presuppositions of the definite description *the tall glass* (i.e., there is a unique glass whose height is at least θ) to make it felicitous. In addition, this interpretation of *tall* in (17a) is relevant for the conversational goal at hand, which is to convey which glass the speaker is requesting. Therefore the coordination account should predict that (17) will always be felicitous and be interpreted as requesting the taller glass. However, while this is indeed generally the case (Syrett, Kennedy, & Lidz, 2010), there are exceptions. When the heights of the two glasses are very small but still noticeably different, the definite description *the tall glass* becomes infelicitous. Such cases are called *crisp judgments* by Kennedy (2007), and the coordination account is unable to explain why (17a) is infelicitous in such cases. In addition, if we assume that all gradable adjectives have a unified semantic representation similar to (16), then the coordination account will wrongly predict that (17b) is felicitous in a context where one glass is half full and the other is empty.

This suggests that the threshold of a gradable adjective does not have a purely pragmatic contextual resolution mechanism that the coordination account (16) specifies. Rather, the conventional meaning, together with various other contextual factors, must play some role in determining the threshold in context (e.g., Glanzberg, 2007, 2009; K. S. Lewis, 2019), and the challenge is how to specify such a semantic contextual resolution mechanism. In Chapter 3, I combine insights from previous approaches and propose such a formal implementation of such a semantic mechanism. Crucially, the proposal is closely related to the formal model in Chapter 2 in that functional considerations, i.e., how interlocutors should use language to best serve their communicative goals, play a major role in determining the contextual interpretation of an expression.

1.4 Semantic underspecification vs ambiguity

One common problem in semantics is to determine whether the multiple possible interpretations of an expression or a class of expressions are due to semantic underspecification or ambiguity. This can be particularly challenging when the possible interpretations are clearly related in some ways but also have some different properties.

In Chapter 3 and Chapter 4, I discuss this issue for gradable adjectives, which are commonly categorized into three main classes: *relative* (e.g., *tall, big*), *maximum* (e.g., *straight, full*) and *minimum* (e.g., *bent, dirty*). This classification is based on the relation between the standard/threshold and the underlying scale structure of a gradable adjective. For instance, *straight* is a maximum adjective because the threshold for *straight* is the maximum degree on the straightness scale, i.e., something is straight iff it is maximally straight. Similarly, *bent* is a minimum adjective because the threshold for *bent* is the minimum degree on the "bentness" scale, i.e., something is bent iff it is at least minimally bent. Finally, *tall* is a relative adjective because the threshold for *tall* is neither the maximum nor the minimum degree on the scale of height (in fact, the scale presumably does not have a maximum/minimum degree).

Given this classification, it is natural to ask the following questions (18).

(18) Questions concerning different classes of gradable adjectives

a. Is there a single, unified semantic representation for (the positive forms of) all gradable adjectives, or do different classes have different semantics?

b. What accounts for the different properties of each class of gradable adjectives and why?

c. In light of the answers to the first two questions, is there a better alternative to the traditional three-way classification?

In this dissertation, I provide the following answers to the questions in (18).

Regarding (18a), I argue that there are two kinds of semantic representations for positive forms of gradable adjectives. According to the first type of semantics, the positive form of a gradable adjective has a threshold that is contextually determined via a semantic mechanism based on a probability distribution over degrees (called a *comparison distribution*, a generalization of comparison classes), which will be spelled out in Chapter 3. This provides a unified semantic representation for maximum, relative, and (the relative interpretation of) minimum gradable adjectives. In Chapter 4, I introduce a second type of semantics, which derives the meaning of the positive form of a gradable adjective by applying an existential closure that quantifies over the positive degrees on the scale. This provides a unified treatment of minimum adjectives and comparative constructions.

Regarding (18b), in Chapter 3, I discuss how the general shape of the comparison distribution, or more precisely, its dispersion, accounts for the differences between maximum and relative adjectives. In Chapter 4, I discuss how the two kinds of semantic representations account for the differences between two types of interpretations of minimum adjectives.

Regarding (18c), the answers to (18a) and (18b) suggest a novel categorization of different readings of gradable adjectives. On the one hand, maximum, relative, and minimum adjectives all have an *optimal-threshold* reading, in which the threshold is determined in context by considering what would be optimal for the comparison distribution. For relative and minimum adjectives, this reading derives their relative interpretations, and for maximum adjectives, this reading derives their maximum interpretations. On the other hand, minimum adjectives can also have a > 0 reading, which is derived by existentially quantifying over positive degrees. This reading derives their minimum interpretations (or more accurately, > 0 interpretations, as I will argue in Chapter 4).

1.5 Outline

In sum, this dissertation makes two general contributions.

First, using case studies in the domain of degrees, it argues for the generality and importance of a distinction regarding answers to the metasemantic question of how semantic underspecification is resolved in context. Specifically, there are two types of contextual resolution mechanisms: a pragmatic one, which is a formalization of King's (2013, 2014a, 2014b) coordination account, and a semantic one. The two mechanisms are unified by functional considerations, i.e., how interlocutors should use language to best achieve their communicative goal(s). This allows us to specify models that make more concrete predictions and capture more empirical properties of context-sensitive expressions than previous accounts.

Second, it provides a new taxonomy of gradable adjectives. Instead of the traditional relative-maximum-minimum trichotomy, readings of (positive forms of) gradable adjectives are categorized into two classes based on the two possible semantic derivations: one based on considerations of an optimal threshold and the other by existentially quantifying over positive degrees. This new categorization allows us to better account for the similarities and differences between and within the classes of gradable adjectives.

The rest of the dissertation is organized as follows. In Chapter 2, I present the first case study, which concerns the directional modified numeral *up to n*. It has opposite inference patterns in two different contexts, which can seem quite puzzling at first sight. I argue that *up to n* has an underspecified lower bound in its semantics, and show that when we supplement King's (2013, 2014a, 2014b) coordination account with an explicit theory of what an idealized listener would do, we can account for such opposite inference patterns. Such a theory can be implemented qualitatively in terms of the interaction between Grice's Maxims of Quantity and Quality, or quantitatively in terms of an interaction between informativity and applicability. In Chapter 3, I present the first half of the second case study, which concerns maximum and relative gradable adjectives. I discuss the empirical properties of these gradable adjectives and review previous approaches and their limitations. In particular, I argue that the threshold of a gradable adjective is not resolved in context based on the type of the pragmatic mechanism proposed in Chapter 2. Instead, I propose it is resolved by a semantic contextual resolution mechanism, Drawing insights from previous approaches, I provide a concrete implementation of such a mechanism, and show how it can account for the different properties of relative and maximum adjectives in a unified way with the least stipulations. In Chapter 4, I present the second half of the case study of gradable adjectives, which examines the class of minimum adjectives. The main theoretical issue is whether we should analyze them in terms of ambiguity or semantic underspecification. I argue that we need both, and that minimum adjectives in fact have two readings, neither of which is characterized by the minimum degree on the scale. One reading is based on the 0 degree on the scale rather than the minimum, and the other is essentially a relative reading as analyzed in Chapter 3. For the first reading, I provide a semantic analysis which derives it by existentially quantifying over the positive degrees on the scale, and show that this provides a unified treatment of this reading of minimum adjectives and comparative constructions.

Chapter 2

Pragmatic contextual resolution: *up to n*

2.1 Introduction

In this chapter, I use English *directional modified numerals* (e.g., *up to 100*, written more generally as *up to n*) as a case study to illustrate how semantic underspecification can be introduced to analyze their context-sensitive interpretations, and how the pragmatic contextual resolution mechanism addresses the metasemantic question that arises. The main empirical data to be accounted for is the contrast in the following minimal pair (1).

(1) a. You are about to meet up to 100 people.

b. You are allowed to meet up to 100 people.

On the one hand, (1a) is mostly used in SPEAKER-UNCERTAINTY contexts, where the speaker does not know the exact number of people that the listener is about to meet.[1] A line of previous work has studied a variety of modified numerals (e.g., *superlative* such as *at most 100* and comparatives such as *fewer than 100*), focusing on how strongly each type of modified numerals is associated with such contexts and why (e.g., Geurts & Nouwen, 2007; Büring, 2008; Nouwen, 2010; Coppock & Brochhagen, 2013; Kennedy, 2015). I will not have anything new to say on this issue. Rather, in this chapter I will focus on an inference pattern that Blok (2015) observes but does not analyze: In SPEAKER-UNCERTAINTY contexts, *up to n* seems to trigger a *proximity inference*. For instance, the

[0] This chapter is based on my first qualifying paper and a corresponding conference paper (Qing, 2016).

[1] It can also be used when the speaker knows the exact number, which is close to 100, but deems it irrelevant/unimportant to report the exact number for the purpose of the conversation. The analysis proposed in this chapter can be extended to account for such cases by taking into account Grice's (1975) Maxim of Relevance.

listener of (1a) can reasonably infer that the speaker believes that the number of people is somewhere close to 100.

On the other hand, the most salient interpretation of (1b) is a permission to meet a number of people within the full range from 0 to 100, granted by a speaker who has the authority.[2] I will refer to this as the *full-range inference* in AUTHORITATIVE-PERMISSION contexts.

We can see that in SPEAKER-UNCERTAINTY and AUTHORITATIVE-PERMISSION contexts, *up to n* triggers opposite inference patterns. In SPEAKER-UNCERTAINTY contexts, the proximity inference results in a narrow range of epistemic possibilities, whereas in AUTHORITATIVE-PERMISSION contexts the full-range inference contributes to a wide range of deontic possibilities. A natural question arises: why and how does *up to n* trigger opposite inference patterns in these two contexts? As discussed in the previous chapter, an analysis of such context-sensitive interpretations of *up to n* needs to address both the semantic and the metasemantic questions (2).

(2) a. **The semantic question**: What is the context-invariant, conventional meaning of *up to n*, and what part of it allows for the contextual variability in interpretation?

b. **The metasemantic question**: What contextual features are relevant for the interpretation of *up to n*, and how exactly is the interpretation determined?

In Section 2.2, I address the semantic question and propose that *up to n* introduces an underspecified semantic lower bound. In Section 2.3, I address the metasemantic question and propose that King's (2014a, 2014b) coordination account, when supplemented with the assumption that the idealized listener assumes that the speaker follows Gricean Maxims, can account for the opposite inference patterns of *up to n* in SPEAKER-UNCERTAINTY and AUTHORITATIVE-PERMISSION contexts. In Section 2.4, I extend this qualitative analysis to a quantitative model, and show that the interaction between the Quantity and Quality maxims can be generalized to a quantitative notion of informativity-applicability interaction/tradeoff. I discuss further issues and compare the current account with Blok's (2015) previous proposal in Section 2.5.

2.2 An underspecified semantic lower bound

2.2.1 Empirical evidence for an underspecified semantic lower bound

In this section, I argue that *up to n* has an underspecified lower bound θ in its semantics that is determined in context. This allows for its variable interpretations in different contexts.

To illustrate that *up to n* has an underspecified semantic lower bound, it is helpful to consider it along with other modified numerals *at most n*, *no more than n*, and *fewer than n* (3).

[2]When (1b) is used in a context in which the speaker is known to be uncertain about the exact number of people the listener is allowed to meet, it triggers a proximity inference similar to (1a).

(3) a. You will meet up to 100 people.

b. You will meet at most 100 people.

c. You will meet no more than 100 people.

d. You will meet fewer than 100 people.

Intuitively, all the sentences in (3) introduce an upper bound on the number of people the addressee will meet, i.e., 100. If it turns out that the addressee will meet more than 100 people, all these sentences will be considered false.[3] Of course, these sentences are different in various respects, but such differences will not matter for the current discussion.[4] What is relevant for us is that these sentences can all to some extent imply that the actual number is close to 100. Given that there is no good reason to assume that there is a semantic lower bound in (3b–3d), we need to find examples where there is a contrast between *up to n* and the other modified numerals as evidence that the proximity inference in (3a) is really due to a semantic lower bound.

One such contrast is first discussed by Blok (2015), which she attributes to a conference reviewer (4).

(4) a. ? I expect to see at most ten people, but maybe no-one will show up.

b. I expect to see up to ten people, but maybe no-one will show up.

On the one hand, (4a) seems infelicitous, which is expected.[5] Given our assumption that *at most ten* does not introduce a semantic lower bound, the possibility that no-one will show up is part of the speaker's expectation, and therefore it is incoherent to use *but*, as there is no contrast between the two clauses. On the other hand, (4b) seems fine, which would be hard to explain if we assume that *up to ten* patterns with *at most ten* and similarly lacks a semantic lower bound. Blok (2015) therefore concludes that 0 should be excluded from the semantics of *up to n*. We can also check that the modified numerals in (5) pattern with *at most n*, which lends further support to this conclusion.

(5) a. ? I expect to see no more than ten people, but maybe no-one will show up.

b. ? I expect to see fewer than ten people, but maybe no-one will show up.

[3] There is a slight complication. If it turns out that the addressee will meet 101 people, while (3b–3d) are clearly false, some people may find (3a) acceptable. Blok (2015) discusses additional data and argues that the upper bound of *up to n* is pragmatically implicated as opposed to semantically imposed (which is the case for the other three modified numerals). To highlight the central claim in my proposal and simplify the exposition, for now I will ignore this complication and assume that the upper bound of *up to n* is semantically imposed, but I will return to this issue in the general discussion.

[4] For instance, the upper bound is inclusive for the first three examples, but is exclusive for the last one. Similarly, there is a general consensus in the literature (see the references in the previous section) that while the first two sentences are strongly (perhaps even obligatorily) associated with Speaker-Uncertainty contexts, the last two sentences are perfectly felicitous even if the speaker knows the exact number.

[5] The ? annotation for infelicity is Blok's, which I follow in the examples below for consistency.

There is a complication to this diagnostic: there may be other ways to construe a contrast so that (4a) and (5) would not be infelicitous. For instance, perhaps there is a contrast between the two clauses by virtue of the fact that they highlight different possibilities: *at most ten* highlights 10 and *maybe no-one* highlights 0. However, we can change the clauses to past tense and make it clearer that the intended contrast is between an expectation and an outcome. For instance, (6) sounds highly contradictory.

(6) # I expected to see at most ten people, but contrary to this expectation, it turned out that no-one showed up.

In fact, the pattern is more general than what Blok discusses. Such examples, which I will call *contrary-to-expectation* sentences, are attested even when the *but*-clause involves a non-zero number (7).

(7) a. Vernell expected up to 10 vendors but only six materialized.[6]

b. Although more funding for police training was supposed to come from a $2 fee on rental cars that went into effect this year, Bump said that fee has not generated as much as expected. State officials expected up to $10 million from the fee, but it appears that's closer to $6 million.[7]

c. Allison had expected up to 1,000 extras, but only about 60 people were in costume on the set. Later, the director would pump our small party into a mob by filming us from various camera angles as we ran in different directions. Look closely in the finished film and you may find each extra in three or four different sectors of the battle.[8]

d. Maroney expected to take up to 50 hours to make this swim, but completed the 110 mile (176 kilometer) crossing in 24 hours and 31 minutes.[9]

e. It anticipated up to 40 cases would be mediated, but realised only 12 cases.[10]

For instance, (7a) suggests that Vernell's expectation does not include the possibility that only 6 vendors materialized, which means that 6 is not part of the semantic content of *up to ten* in (7a).

These examples suggest that the range specified by *up to n* can be quite flexible and context-sensitive, and need not start from 0. Crucially, since such non-0-starting ranges can be embedded under *expected* or *anticipated*, they must be part of the compositional semantics.

Once we assume that *up to n* has an underspecified semantic lower bound, we can account for another contrast that Blok (2015) discusses (8).

[6] http://www.uphamscornernews.com/up-market-june-2014-vendors-managers-products-a-few-customers.html

[7] https://www.wbur.org/news/2019/11/18/massachusetts-report-police-training

[8] http://producingaletheia.blogspot.com/

[9] https://www.youtube.com/watch?v=KUfXM6DH-R0

[10] https://www.royalsoced.org.uk/cms/files/advice-papers/inquiry/negligence/report.pdf. Paragraph 5.17.

(8) a. Fortunately, up to 100 people will attend my wedding.

b. Fortunately, at most 100 people will attend my wedding.

Blok points out that (8a) implies that the speaker is happy that a high number of guests will attend the wedding, whereas (8b) implies that the speaker is happy that not many people (no more than 100) will be at the wedding. Similarly, we can check that the modified numerals in (9) pattern with *at most n*.

(9) a. Fortunately, no more than 100 people will attend my wedding.

b. Fortunately, fewer than 100 people will attend my wedding.

This contrast can be accounted for if we assume that in (8a) *fortunately* is targeting the semantic lower bound introduced by *up to 100*, which is close to 100, while in (8b) *fortunately* is targeting the semantic upper bound introduced by *at most 100*.[11]

In sum, we have seen evidence that unlike *at most/no more than/fewer than n*, *up to n* has an underspecified semantic lower bound, which accounts for their different interpretations when they are embedded under *expect* or *fortunately*.

2.2.2 Inquisitive semantics implementation

Below I provide a formal implementation of the proposal that *up to n* has an underspecified semantic lower bound. The analysis is implemented in the inquisitive semantics framework (e.g., Ciardelli, Groenendijk, & Roelofsen, 2009, 2012). There are two main reasons for choosing this framework. First, it simplifies the discussion of the pragmatic mechanism for the metasemantic question that I will propose in the next section, because the auxiliary assumptions that are orthogonal to my proposal are encoded in the semantics. Second, it also allows for an easy comparison with Blok's (2015) analysis (Section 2.5.2). However, the core analysis of the opposite inference patterns of *up to n* does not hinge on this choice. Also, since these motivations for using inquisitive semantics do not apply in later chapters, I will not use this framework there.

In classical semantic theories, a declarative sentence denotes a proposition, which can be represented as a set of possible worlds. In inquisitive semantics, however, a declarative sentence denotes a set of propositions, i.e., a set of sets of possible worlds. A nice feature of inquisitive semantics is that the classical denotation (truth conditions) of a sentence can be retrieved by applying set union to its denotation in inquisitive semantics. This makes inquisitive semantics compatible with classical theories in terms of truth conditions, while having more fine-grained representations to capture the different inference patterns between expressions that have the same classical truth conditions.

[11] Blok only takes this contrast to be evidence that the upper bound of *up to 100* is not semantically imposed (or in her words, not part of the asserted content). Even if she is right about the nature of the upper bound of *up to 100*, it is still crucial to assume that in this case *up to 100* has a semantic lower bound that is close to 100. For instance, *fortunately, 10 to 100 people will attend my wedding* would not have the same implication as (8a).

I propose that *up to n* has the following semantics (10).

(10) $[\![\text{up to n}]\!] = \{\lambda M.\, \max(M) = k \mid k \in [\theta, n]\}$,
where θ is a contextual lower bound ($0 \leq \theta < n$).

Let us use *up to 100* as a concrete example to unpack this definition. Its denotation is in (11), which is a set of functions.

(11) $[\![\text{up to 100}]\!] = \{\lambda M.\, \max(M) = k \mid k \in [\theta, 100]\}$
$= \{\lambda M.\, \max(M) = \theta, \lambda M.\, \max(M) = \theta + 1, \ldots, \lambda M.\, \max(M) = 100\}$
θ is a contextual lower bound ($0 \leq \theta < 100$)

In *you are about to meet up to 100 people* (1a), *up to 100* takes scope over the rest of the sentence, which is a degree property, i.e., a function M that takes a degree d (in this case, a number) and returns the proposition that the addressee is about to meet at least d people.[12] In this example, when M takes the number 3 as input, it returns the proposition that the addressee is about to meet at least 3 people. Applying the max operator transforms the function M so that the output proposition concerns the exact number the addressee is about to meet. For instance, $\max(M) = 80$ is the proposition that the addressee is about to meet exactly 80 people. The full derivation of (1a) is shown in (12).

(12) $[\![\text{you are about to meet up to 100 people}]\!]$
$= [\![\text{up to 100}]\!]\ (\lambda d.[\![\text{you are about to meet } d\text{-many people}]\!])$
$= \{p_\theta, p_{\theta+1}, \ldots, p_{100}\}$
where θ is a contextual lower bound ($0 \leq \theta < 100$) and p_i is the proposition that the addressee is about to meet exactly i people.

The classical truth conditional content of (12) is the union of all the propositions, which is the proposition that the exact number of people the addressee is about to meet, n_0, is within the range $[\theta, 100]$. In other words, the informative content conveyed by (12) is that $n_0 \in [\theta, 100]$. The speaker asserts that the number is within this range. Meanwhile, since the definition requires that $\theta < 100$, the denotation of (12) always has at least two alternatives. According to Coppock and Brochhagen's (2013) Maxim of Interactive Sincerity, a cooperative speaker should not raise multiple alternatives if she already knows which one is true. Therefore, the speaker of (12) will violate this maxim if she already knows the exact number of people. This accounts for why (12) generally requires Speaker-Uncertainty contexts.

[12] I propose such a degree-based semantics because this is the dominant approach in the more recent literature on modified numerals since Hackl (2000). However, the pragmatic mechanism that I will propose in the next section does not hinge on this choice. If desired, one can also implement the semantics of *up to 100* in the style of the Generalized Quantifier Theory (Barwise & Cooper, 1981): $[\![\text{up to 100}]\!] = \{\lambda P \lambda Q.|P \cap Q| = k \mid k \in [\theta, 100]\}$ ($0 \leq \theta < 100$). It can be easily verified that this produces the same denotations for (12) and (13) as the degree-based semantics.

To analyze *you are allowed to meet up to 100 people* (1b) in AUTHORITATIVE-PERMISSION contexts, I adopt the common assumption in the literature that the permission modal $\Diamond$ scopes above *up to n* (Büring, 2008; Coppock & Brochhagen, 2013; Kennedy, 2015). The derivation is shown in (13).

(13) $[\![\text{You are allowed to meet up to 100 people}]\!] = \{\Diamond[\![\text{you meet up to 100 people}]\!]\}$
$= \{\Diamond\{p_\theta, p_{\theta+1}, \ldots, p_{100}\}\}$
where $0 \leq \theta < 100$ and p_i is the proposition that you meet exactly i people.

Here, we have a singleton set containing the proposition $\Diamond\{p_\theta, p_{\theta+1}, \ldots, p_{100}\}$, which represents the effect of the permission modal scoping above a set of possibilities. It is well known that permission modals scoping above a set of possibilities can trigger a *free-choice* inference (14), i.e., each possibility in the set is allowed. This is independently observed in studies of the interaction between permission modals and disjunctions (e.g., Kamp, 1973, 1978; Zimmermann, 2000; Kratzer & Shimoyama, 2002).

(14) $\Diamond\{p_\theta, p_{\theta+1}, \ldots, p_{100}\} \rightsquigarrow \Diamond p_\theta \wedge \Diamond p_{\theta+1} \wedge \ldots \wedge \Diamond p_{100}$

There is no consensus in the literature on whether the nature of the free-choice inference (14) is semantic or pragmatic. My analysis does not crucially hinge on this issue, and therefore I will not take a stance here and will abstract away from the specific implementation that captures the free-choice inference. The only assumption I will make is that the free-choice inference takes place before and feeds into the pragmatic mechanism that I am going to propose in the next section, which resolves the lower bound θ. If the free-choice inference is a semantic entailment, this is just a standard assumption about the semantics/pragmatics interface. If the free-choice inference is instead taken to be pragmatic, even though it might seem unusual to assume that a pragmatic inference can feed into another pragmatic mechanism from a traditional point of view, I note that this is in fact compatible with major pragmatic accounts of the free-choice inference. Below I discuss a few representative examples. To derive the free-choice inference, Kratzer and Shimoyama (2002) assume that an expression can be used to defeat the implicature of an alternative expression. This allows for a pragmatic inference (an implicature) to feed into another layer of pragmatic reasoning to derive the free-choice inference, and the additional assumption in my analysis is perfectly compatible with it: all I need to assume is that the free-choice inference then feeds into yet another layer of pragmatic reasoning (to be spelled out in the next section) to resolve the lower bound θ. There are various ways to formally implement Kratzer and Shimoyama's account, some of which involve very different conceptions of the division of labour between semantics and pragmatics, but they all effectively allow for the result of a pragmatic inference to feed into another pragmatic process. For instance, Fox (2007) assumes that the relevant implicatures are derived in the grammar by a covert exhausitivity operator *Exh*, and the free-choice inference is the result of recursive exhaustification. Crucially, even though this account is pragmatic in that the application of *Exh* is in principle optional but subject to certain economy conditions, the free-choice inference is in fact entailed by the full semantic representation (i.e., with the recursive applications of *Exh*). All I need to assume, then, is that it is this full semantic representation that feeds into the pragmatic mechanism in the next section, which once again is perfectly in line with standard assumptions about the semantics/pragmatics interface. Alternatively, one could adopt Franke's (2011) account that is more "pragmatic" (in that it does not involve additional semantic representations), based on game-theoretic pragmatics, where speakers and listeners iteratively reason about each other. For instance, a level-1 speaker reasons about a listener that interprets expressions literally, and a level-2 listener reasons about a level-1 speaker, i.e., one who reasons about a listener that interprets expressions literally. Crucially, this chain of reasoning can go up to any level. This provides a natural way to formally capture Kratzer and Shimoyama's original formulation. For example, if a level-2 listener would derive a certain implicature, a level-3 speaker, who reasons about the level-2 listener, may use expressions in a way that avoids this implicature, which will then be reasoned about by a level-4 listener and leads to additional implicatures. If one adopts such an account of the free-choice inference, then my proposal can be seen as adding further levels of reasoning to this account. Concretely, if the free-choice inference is derived by a level-k listener, my proposal can then be seen as describing the reasoning of a level-$(k+1)$ speaker and a level-$(k+2)$ listener.[13] Therefore, the assumption that the free-choice inference can feed into another pragmatic mechanism is also compatible with major pragmatic analyses of free-choice inferences.

The proposed semantics of *up to n* allows us to represent its different interpretations in different contexts, by setting the underspecified lower bound to different values. To account for proximity inference in Speaker-Uncertainty contexts, we can say that the lower bound is resolved to a number that is close to the upper bound in these contexts. Similarly, to account for the full-range inference in Authoritative-Permission contexts, we can say that the lower bound is resolved to 0 in these contexts. This addresses the semantic question, i.e., the conventional meaning of *up to n* introduces an underspecified lower bound θ that is determined in context. However, we still need to address the metasemantic question, i.e., exactly how the underspecified semantic lower bound θ is resolved context. In particular, the answer should explain the opposite inference patterns in Speaker-Uncertainty and Authoritative-Permission contexts. In the next section, I propose a pragmatic contextual resolution mechanism to address this metasemantic question.

[13]In game-theoretic pragmatics (and its closely related approaches), there are in principle many more possible model configurations that take into account both the free-choice inference and the uncertainty about the lower bound θ. My proposal here may not be the simplest or the most natural way to integrate the two phenomena within the framework, and I suspect that not all possible configurations will lead to the same predictions, which means that ideally there should be independent motivations to justify the particular configuration in my proposal. However, note that these issues are internal to game-theoretic pragmatics. Given that the purpose of the discussion in this paragraph is to show the general compatibility between various analyses of the free-choice inference and my analysis of the contextual resolution of the lower bound θ proposed in the next sections, here I am merely showing that there is a possible model configuration within game-theoretic pragmatics that can account for both phenomena, leaving the exploration of other possibilities and comparisons between them for future research.

2.3 Coordination of the lower bound

The pragmatic contextual resolution mechanism I propose can be understood in terms of King's (2013, 2014a, 2014b) coordination account discussed in Chapter 1. Essentially, I suggest that the speaker and listener are coordinating on the underspecified semantic lower bound of *up to n*. Concretely, the coordination account for *up to n* is shown below (15).

(15) **The coordination account for *up to n***: the underspecified semantic lower bound of *up to n* is θ in context c iff the following two conditions are met

a. The speaker intends θ to be the lower bound of *up to n* in c.

b. A competent, attentive, reasonable hearer who knows the common ground of the conversation at the time of utterance would know that the speaker intends θ to be the lower bound of *up to n* in c.

As discussed in Chapter 1, a major limitation of King's coordination account is that it does not specify what the idealized listener can do, which is left entirely to our intuition. In the case of *up to n*, we do not seem to have clear intuitions about whether an idealized listener can recognize the speaker's intention to use a certain value as the lower bound. For example, consider *you are about to meet up to 100 people* in a SPEAKER-UNCERTAINTY context. If the speaker intends to use, e.g., 10 as the lower bound, would an idealized listener recognize such an intention? Or if the speaker intends to use 80 as the lower bound, would an idealized listener recognize such an intention? Or would an idealized listener be able to recognize the intention to use 80 as the lower bound more so than the intention to use 10 as the lower bound? These questions are hard to answer without knowing what an idealized listener is supposed to be capable of. Of course, to the extent that we make proximity inferences in SPEAKER-UNCERTAINTY contexts, one can *stipulate* that an idealized listener can better recognize the intention to use 80 as the lower bound than the intention to use 10 as the lower bound, but this would not be explanatory.

I suggest that this problem can be avoided if we assume that the idealized listener assumes that the speaker follows Grice's (1975) Maxims, in particular Quantity and Quality. Below I introduce a qualitative analysis based on this idea, and in the next section I show how this qualitative analysis can be generalized to a quantitative model to make more fine-grained predictions.

According to the Quantity Maxim, an utterance should be as informative as required. According to the Quality Maxim, an utterance should not be made if it is untrue or if the speaker does not have enough evidence for it. For brevity, if an utterance adheres to the Quality Maxim, we say that it is *applicable*. These maxims are what a cooperative speaker should try to adhere to, but it is possible that they cannot be satisfied at the same time, and in such cases certain inferences can be made a result. I propose that we can analyze the contextual lower bound of *up to n* by considering the interaction/trade-off between informativity (Quantity) and applicability (Quality).

Let us first consider Speaker-Uncertainty contexts, and use *you are about to meet up to 100 people* (12) as a working example. Let n_0 be the exact number of people that the addressee will in fact meet. The informative content of the utterance is $n_0 \in [\theta, 100]$. A θ that is close to 100 will result in a narrower range, and therefore will make the utterance more informative. Since in such contexts the speaker is known to be uncertain about the exact number of people, applicability concerns whether the speaker has enough evidence to support the utterance. If the lower bound θ is, e.g., 99, then the informative content of the utterance is that the actual number n_0 is either 99 or 100, and a speaker known to be uncertain about the exact number is unlikely to have enough evidence to support such a strong claim. For instance, suppose the speaker made an estimate of the number of people by taking a quick look at the room before the addressee arrives. Given our general world knowledge about perception, we know that it is highly unlikely that the speaker would have enough evidence to conclude that the exact number is either 99 or 100, or within a range that is unreasonably narrow. This means that the lower bound θ cannot be too high, due to considerations of applicability. On the other hand, if the lower bound θ is, e.g., $\theta = 10$, then the informative content of the utterance is that the actual number of people is within the range $[10, 100]$. While the speaker is quite likely to have enough evidence to support this claim, i.e., it is very likely to be applicable, it is not very informative. This means that the lower bound θ cannot be too low, due to considerations of informativity. When we take considerations of both informativity and applicability into account, we can see that the lower bound θ should be close to the upper bound n to be informative, and yet not too close so that the utterance is still applicable. As a result, if the speaker intends, e.g., 70, to be the value of the lower bound θ, an idealized listener, who assumes that the speaker follows Gricean Maxims, would be able to recognize such an intention. This explains the proximity inference of *up to n* in Speaker-Uncertainty contexts.

Now we turn to uses of *up to n* in Authoritative-Permission contexts, and take *you are allowed to meet up to 100 people* (13) as an example. Recall that I assume that the free-choice inference takes place before the current pragmatic considerations. Under this assumption, the informative content of (13) is the conjunction $\Diamond p_\theta \wedge \Diamond p_{\theta+1} \wedge \ldots \wedge \Diamond p_{100}$ (where p_i is the proposition that the addressee meets exactly i people). Consequently, a smaller θ, which corresponds to more conjuncts, will make the sentence more informative. For example, when $\theta = 10$, the addressee will learn that meeting $10, 11, \ldots, 100$ people is allowed, but when $\theta = 50$, the listener will only learn that meeting $50, 51, \ldots, 100$ people is allowed, and remains uncertain about whether meeting $10, 11, \ldots, 49$ people is allowed. We can see that the most informative θ would be 0, because it corresponds to the most conjuncts. Meanwhile, as long as the speaker has the authority to grant permission, (13) will be applicable. Moreover, in many contexts, there are no a priori reasons to think that the speaker intends to grant permission only to a partial range $[\theta, n]$ (where $\theta \neq 0$).[14] Therefore, if

[14]There are contexts where the listener can have a prior expectation that small numbers are unlikely to be allowed. For example, given our world knowledge that there tends to be a minimum-length requirement for passwords, *you are allowed to use up to 20 characters for your password* will probably not be interpreted as permitting the use of, e.g.,

the speaker intends 0 to be the lower bound, an idealized listener, who assumes that the speaker follows Gricean Maxims, would recognize such an intention. This explains the full-range inference in Authoritative-Permission contexts.

To sum up, we have seen that the opposite inference patterns of *up to n* in Speaker-Uncertainty and Authoritative-Permission contexts can be explained, if we supplement King's (2013, 2014a, 2014b) coordination account with the assumption that the idealized listener assumes that the speaker follows Gricean Maxims. Crucially, the opposite inference patterns are explained by the interaction/trade-off between informativity (Quantity) and applicability (Quality). In Speaker-Uncertainty contexts, the two factors are in tension with each other and are "pushing" the lower bound θ to opposite directions. As a result, the most salient interpretation has a lower bound that is close to the upper bound but not too much so. In Authoritative-Permission contexts, informativity prefers 0 (or a small number) to be the lower bound, and applicability is generally not against it. As a result, the default interpretation has 0 as the lower bound.

The above analysis is qualitative. For example, for Speaker-Uncertainty contexts it only predicts that the interlocutors would coordinate on a lower bound that is close (but not too close) to the upper bound, but does not provide more specific quantitative predictions about how our world knowledge and contextual information about the speaker's level of uncertainty would affect the lower bound. To address this, in the next section I introduce a probabilistic model that generalizes the notions of informativity and applicability.

2.4 The informativity-applicability tradeoff: A probabilistic model

In this section I introduce a probabilistic model and show that the previous discussion can be formalized to make quantitative predictions. This primitive model has a lot of simplifying assumptions, but it suffices to illustrate the main concept.

Consider *you are about to meet up to 100 people* in Speaker-Uncertainty contexts. Let m be the actual number of people that the addressee will meet, and G be the speaker's belief state, i.e., the set of all the numbers that the speaker considers possible. In order to measure applicability quantitatively, we need to formalize the contextual information about the speaker's level of uncertainty about the actual number m. Ideally we would like to specify a distribution over the speaker's belief state G, but in practice it can be hard to specify such a big distribution, so I will make certain simplifications.

First, I assume that the speaker's belief state G is a range $[a, b]$. The speaker chooses the upper bound n in *up to n*, and similarly intends an implicit lower bound θ, based on a and b.

---

only 2 characters.

(16) a. $p(m, a, b, \theta, n) = p(m, a, b) \cdot p(n \mid m, a, b) \cdot p(\theta \mid m, a, b, n)$

b. $p(m, a, b, \theta, n) = p(m, a, b) \cdot p(n \mid a, b) \cdot p(\theta \mid a, b, n)$

According to the chain rule, we have (16a). Since the speaker does not know m and chooses the bounds based on her own belief a, b, the bounds θ, n are conditionally independent of m given a, b, and therefore (16a) can be reduced to (16b).

The model (16b) involves the choice of the upper bound n. Since we are more interested in the conditional probabilities $p(\theta \mid n)$ and $p(m \mid n)$, where n is already given, the choice of the upper bound n is not particularly relevant. Therefore, to simplify the model and highlight the informativity-applicability trade-off in determining θ, which is our main interest, I assume that the speaker always chooses $n = b$, i.e., n in *up to n* is the maximal number that the speaker considers possible. Intuitively this is very plausible: if 100 is maximal number of people that the speaker considers possible, then uttering *you are about to meet up to 120 people* is less informative than *you are about to meet up to 100 people* and there is no reason for a cooperative speaker to do that.[15]

This simplification helps us eliminate the variable b, and according to the chain rule, we obtain (17a).

(17) a. $p(m, a, \theta \mid n) = p(m, a \mid n) \cdot p(\theta \mid m, a, n) = p(m, a \mid n) \cdot p(\theta \mid a, n)$

b. $p(a, \theta \mid n) = p(a \mid n) \cdot p(\theta \mid a, n)$

We are interested in the conditional distribution $p(\theta \mid n)$, i.e., how likely the speaker intends to use θ as the lower bound when uttering *up to n*, therefore we should marginalize over m and a. Marginalizing over m yields (17b). It has two parts: $p(a \mid n)$ encodes the contextual information about the speaker's level of uncertainty, and the second part reflects the informativity-applicability tradeoff.

Note that the Maxim of Quality requires that $[a, b] \subseteq [\theta, n]$, i.e., the semantic content of the utterance needs to be entailed by the speaker's belief state. This requires that $\theta \leq a$. Therefore, assuming that the speaker follows the Maxim of Quality, $p(\theta \mid a, n) = 0$ when $a < \theta$. When $a \geq \theta$, $p(\theta \mid a, n)$ depends on the informativity of θ, which is measured as the reduction in uncertainty (entropy) the lower bound θ contributes. Without the lower bound, the listener only knows that the number of people is between 0 and n, and therefore the entropy is $\log(n + 1)$, assuming a uniform prior. The lower bound θ narrows down the range to $[\theta, n]$, whose entropy is $\log(n + 1 - \theta)$. Therefore the reduction of entropy is $\log(n + 1) - \log(n + 1 - \theta)$, as shown in (18a). Note that the larger the lower bound θ is, the more informative it is.

(18) a. $\text{Informativity}(\theta) = \log(n + 1) - \log(n + 1 - \theta)$

[15] Unless, of course, Quantity interacts with Manner. For instance, if 98 is the maximal number that the speaker considers possible, she might sacrifice a little bit of information and use the simpler form *up to 100*. This can be modeled by assigning higher costs to expressions with non-round numbers (Kao, Wu, Bergen, & Goodman, 2014), which will not be considered here just for simplicity.

b. $p(\theta \mid a, n) \propto \delta_{a \geq \theta} \cdot \text{Informativity}(\theta)^{\lambda}$

c. $p(\theta \mid n) = \sum_a p(a \mid n) \cdot p(\theta \mid a, n)$

$\propto \sum_a p(a \mid n) \cdot \delta_{a \geq \theta} \cdot \text{Informativity}(\theta)^{\lambda}$

$= \sum_{a \geq \theta} p(a \mid n) \cdot \text{Informativity}(\theta)^{\lambda}$

$= \text{Informativity}(\theta)^{\lambda} \cdot \sum_{a \geq \theta} p(a \mid n)$

d. $\text{Applicability}(\theta) = \sum_{a \geq \theta} p(a \mid n)$

e. $p(\theta \mid n) \propto \text{Informativity}(\theta)^{\lambda} \cdot \text{Applicability}(\theta)$

Now we can define $p(\theta \mid a, n)$ as in (18b), where $\delta_{a \geq \theta}$ is a delta function, which returns 1 if $a \geq \theta$ and 0 otherwise, and λ is a parameter that captures the importance of informativity in the choice of θ. If $\lambda = 0$, it means that informativity is not considered, and when $\lambda \to +\infty$, it means that informativity is the only consideration. Basically, (18b) says that the more informative a lower bound θ is, the more likely that it will be intended by the speaker, as long as the Maxim of Quality is satisfied, i.e., $\theta \leq a$.

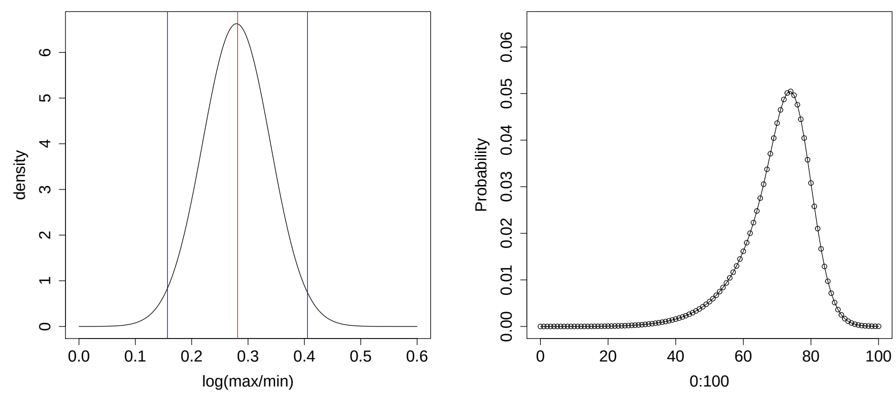
Now we can marginalize over a, and plug in the definition in (18b) to simplify $p(\theta \mid n)$ in (18c). In the end we can see that $p(\theta \mid n)$ is depends on the product of two terms. The first term is $\text{Informativity}(\theta)^{\lambda}$, which increases as θ increases. The second term is $\sum_{a \geq \theta} p(a \mid n)$, i.e., the probability of $a \geq \theta$, which is the probability that the Maxim of Quality is satisfied. We can use it to define our generalized, quantitative notion of applicability as in (18d). Crucially, note that applicability decreases as θ increases, just as discussed in the previous section. With applicability defined in this way, (18c) can be seen as the informativity-applicability tradeoff, as shown in (18e).

In order for the above model to produce actual quantitative predictions, we need to specify $p(a \mid n)$, i.e., the conditional probability of the minimal possible number considered by the speaker, given the maximal possible number n. This distribution is contextually determined, and depending on the context various assumptions can be made. Here, I will assume that this probability depends on n/a, i.e., the ratio between the maximal possible number, and that the log of this ratio is normally distributed. This assumption seems plausible in many contexts, especially those that involve number perception, because previous work in psychophysics has shown that our perception is generally sensitive to ratios (see, e.g., Dehaene, 2003 for more introduction and discussion).

As a concrete example, I assume that $p(a \mid n) = \phi(\log(n/a))$, where $\phi(x)$ is a normal distribution shown in Figure 2.1(a), which corresponds to a contextual assumption that the maximal possible number is typically 20%–50% larger than the minimal possible number and most likely around 35% larger.[16]

When $\lambda = 4$, the predicted distribution of θ for *you are about to meet up to 100 people* is shown in Figure 2.1(b), but for a reasonable range of λ, the shape of the curve is qualitatively the same: a

[16] When $n/a = (1 + p)$, $\log(n/a) = \log(1 + p)$. When $p = .2$, $\log(1 + p) = .18$ (the blue line on the left). When $p = .5$, $\log(1 + p) = .41$ (the blue line on the right). When $p = .34$, $\log(1 + p) = .29$ (the red line in the middle).



(a) Contextual prior of the log(max/min) ratio (b) Distribution of θ, $\lambda = 4$

Figure 2.1: Model for *you are about to meet up to 100 people*

larger λ will shift the curve slightly to the right and make it more concentrated, and a smaller λ will shift the curve slightly to the left and make it more flat. For instance, Figure 2.2(a) shows the prediction when $\lambda = 6$. In either case, we can see that the model predicts that it is highly unlikely that the lower bound θ is above 80 or below 50. In other words, after hearing *you are about meet up to 100 people*, the listener would generally infer that the speaker intends a lower bound between 50 and 80, which is an intuitively plausible result that accounts for the proximity inference.[17]

Note that the prediction is based on the ratio between the upper bound n and the minimal possible number a rather than the difference. For example, we can see from Figure 2.2(b) that the shape of the curve is the same for *up to 10*. This result is also intuitively plausible.

We have seen how the probability of the lower bound θ can be inferred from the upper bound, i.e., $p(\theta \mid n)$. Quite often, we would also like to reason about the actual number m. However, for the joint probability in (17a), repeated below in (19a), when we first marginalize over θ (19b) and then over a (19c), we can see that these steps do not provide us with any real information about the actual number m.

(19) a. $p(m, a, \theta \mid n) = p(m, a \mid n) \cdot p(\theta \mid m, a, n) = p(m, a \mid n) \cdot p(\theta \mid a, n)$

b. $p(m, a \mid n) = p(m, a \mid n)$ Marginalize over θ

c. $p(m \mid n) = p(m \mid n)$ Marginalize over a

This might look surprising at first, but it is actually plausible. Note that the actual number does not directly determine the lower bound or the upper bound of *up to n*. Its influence on the

[17]Keep in mind that this prediction is based on the specific contextual assumption about the speaker's level of uncertainty. If the listener believes that the speaker's level of uncertainty is lower, then the lower bound would be inferred to be even closer to the upper bound 100.

bounds is through the maximal and minimal numbers in the speaker's belief state. Therefore, if no additional assumption is made to link the actual number to the speaker's belief, there is no way to infer "backwards" and calculate $p(m \mid n)$.

To see this more clearly, note that by Bayes' rule we have (20).

(20) $p(m \mid n) \propto p(m) \cdot p(n \mid m)$

Here $p(m)$ is the prior of the actual number, and $p(n \mid m)$ is the probability that the maximal number that the speaker considers possible is n when the actual number is m. This is the link from the reality to the speaker's belief that we need to specify. In the extreme case where n and m are totally independent, e.g., in a context where the speaker chooses the absolute maximum in the common ground (e.g., the full capacity of the room) as the upper bound, the listener will gain no information at all from the utterance *you are about to meet up to 100 people* and has to rely completely on the prior expectation about m. In other cases, the speaker's belief is formed by a noisy observation of the actual number m, e.g., when the speaker took a quick look at the people in the room. In such cases, the probability $p(n \mid m)$ may be reasonably assumed to depend on the ratio n/m, whose log is normally distributed, and sometimes it might be reasonable to further assume that n/m and m/a are independently distributed with identical distributions. Under this assumption, the mean and variance of n/a is twice as much as those of n/m. In yet some other cases, it might actually be most natural to directly estimate $p(m \mid n)$. For example, if the maximal number that speaker considers possible is based on the number of people who have responded "yes" or "maybe" to the invitation, then the listener can use his general knowledge about the typical attendance rate to directly estimate $p(m \mid n)$. In any case, depending on the additional information context provides about $p(m \mid n)$ or $p(n \mid m)$, the listener's inferred distribution of the actual number m after hearing *up to n* can vary, and it is possible to infer, e.g., that 45 is most likely after hearing *up to 50*.

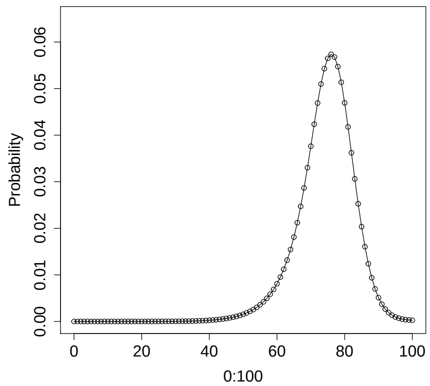
Now let us consider *you are allowed to meet up to n people* in AUTHORITATIVE-PERMISSION contexts. As discussed in the previous section, the Maxim of Quality is satisfied as long as the speaker has the authority to grant permission. Therefore, the sentence is always be applicable and without loss of generality, I assume that the applicability of θ is a constant, e.g., 1 (21a).

(21) a. $\text{Applicability}(\theta) = 1$

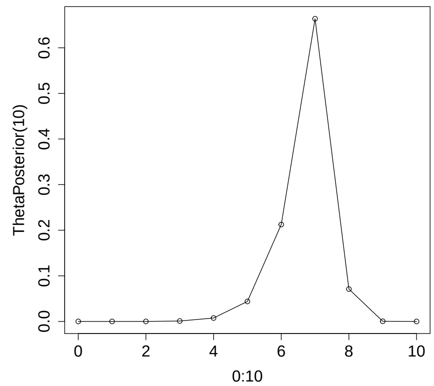
b. $\text{Informativity}(\theta) = (n + 1 - \theta) * \log_2 2 = n + 1 - \theta$

c. $\Pr(\theta) \propto \text{Informativity}(\theta)^{\lambda} \cdot \text{Applicability}(\theta)$

The informative content is the big conjunction that meeting exactly $\theta, \theta + 1, \ldots, n$ people are all allowed. I make the simplifying assumptions that the listener initially is totally ignorant about whether meeting i people is allowed for any number of i, and that they are all independent of each other. This means that for each i the entropy of the listener's belief is $\log_2 2 = 1$. After hearing the sentence, the listener has no uncertainty about the deontic status of $\theta, \theta + 1, \ldots, n$ (since the



(a) $n = 100$



(b) $n = 10$

Figure 2.2: Distribution of θ for *you are about to meet up to n people*, with $\lambda = 6$

sentence says that they are all allowed), which means the entropy of the listener's belief about each of these numbers is now 0. Therefore the total reduction of entropy is $n + 1 - \theta$, as shown in (21b). Finally, the interaction between applicability and informativity is the same as before (21c), except that now it does not involve a tradeoff because applicability is a constant.

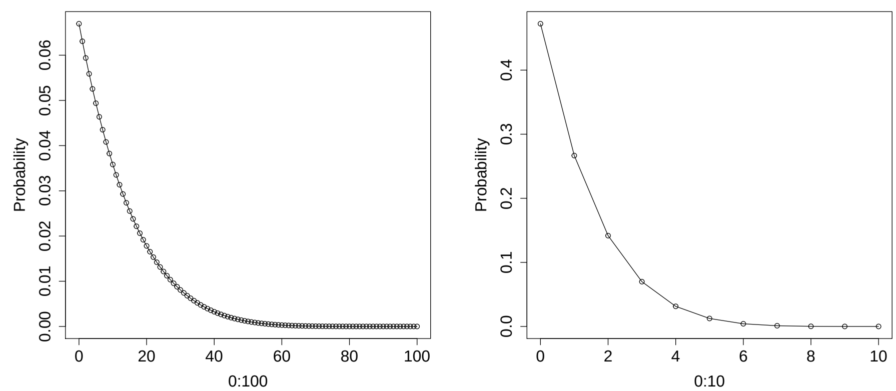
Under the above assumptions, with $\lambda = 6$, the predictions of the distribution of θ for *up to 100* and *up to 10* are shown in Figure 2.3. In both cases, we can see that the most likely θ is 0 and the distributions are monotonically decreasing. This corresponds to the intuition that the most likely interpretation is that the full range is allowed.

I should emphasize that the main purpose of introducing the probabilistic model is to illustrate that in principle we can make concrete, quantitative predictions about the contextual distribution of the lower bound θ of *up to n*. This will enable us to evaluate the extend to which the proposal that θ is determined by an interaction/trade-off between applicability and informativity captures the inference patterns of *up to n* in different contexts. Of course, as we have seen, it is not a trivial task to formalize a probabilistic model. In building the model, I need to make a lot of simplifying assumptions that may well be empirically inaccurate. Future work is needed to improve the model and test its quantitative predictions.

2.5 General discussion and comparison with previous work

2.5.1 Generalization to universal deontic modals

The proposed analysis of *up to n* can be generalized to explain its inference patterns under universal deontic modals. For instance, Nouwen (2008) observes that *up to n* is "not so happy with strong



(a) $n = 100$ (b) $n = 10$

Figure 2.3: Distribution of θ for *you are allowed to meet up to n people*, with $\lambda = 6$

modals" (22).

(22) ?? Jasper is required to invite up to 10 children to his party.

Note that Nouwen intends *up to 10* to take scope below the universal modal in (22), which corresponds to an authoritative reading (the wide-scope reading in SPEAKER-UNCERTAINTY contexts is perfectly fine here). While it is true that many native speakers tend to find (22) odd, many similar naturally-occurring examples can be found (23).

(23) a. The squad must contain up to 25 players and have no more than 17 players who do not fulfil the Home Grown Player criteria.

b. These essays should present the authors' arguments in a clear and structured manner and should include up to 5 references. "In My View" contributions are limited to 1200 words.

c. It was required that the answers link to at least one and up to 3 references from the reference module.

d. [Q & A] What do I do if I receive a "warning" within the EMPLOYEE STATE/PROVINCE field? This field may not be left blank and must contain up to 2 characters of text.

e. So Mr Speaker, we are proposing that Britain should resettle up to 20,000 Syrian refugees over the rest of this Parliament.

These examples do not seem as weird. This suggests that the oddity of (22) cannot be explained simply as semantic incompatibility between *up to n* and necessity modals. Instead, I suggest that the weirdness of (22) can be explained by considering informativity and applicability.

(24) a. $[\![(22)]\!] = \{\Box\{p_\theta, p_{\theta+1}, \ldots, p_{10}\}\}$
where p_i is the proposition that Jasper invites exactly i children to his party.

b. $\Box\{p_\theta, p_{\theta+1}, \ldots, p_{10}\} \vDash \bigwedge_{i \notin [\theta,10]} \neg\Diamond p_i$

c. $\Box\{p_\theta, p_{\theta+1}, \ldots, p_{10}\} \rightsquigarrow \Diamond p_\theta \wedge \Diamond p_{\theta+1} \wedge \Diamond p_{10}$

As with permission modals, I assume that in the authoritative reading the necessity modal scopes above *up to n*. This results in the necessity modal taking a set of alternative propositions as its argument (24a). The classical logical property of deontic necessity $\Box$ dictates that anything outside of the informative content of its complement is not allowed. Therefore, we know that (24a) entails that any number outside the range $[\theta, 10]$ is not allowed (24b). Deontic necessity modals scoping above a set of alternatives can also give rise to the free-choice inference, e.g., *you are required to eat an apple or an orange* implies that eating an apple is allowed and eating an orange is allowed. The result of the free-choice inference of (24a) is that inviting any number within the range $[\theta, 10]$ is allowed (24c). Recall that I assume the result of the free-choice inference feeds into the pragmatic mechanism that determines θ, therefore (24b) and (24c) together completely settle whether inviting exactly i children is allowed, for every number i. Therefore informativity will be the same for any θ and the choice of θ totally relies on applicability. Similar to permission modals, the Maxim of Quality is satisfied as long as the speaker has authority, which is assumed to be the case. Therefore, the sentence will always be applicable. However, this means that the interaction between informativity and applicability tradeoff does not prefer any θ: they are all equally good. As a result, (22) would be very ambiguous: if all the lower bounds are equally good, the idealized listener would not be able to recognize the speaker's intention to use a particular lower bound. For naturally-occurring examples, however, the idealized listener can resort to background world knowledge to reasonably infer the intention of the speaker. This is much harder for a decontextualized sentence such as (22): without enough background knowledge about the kind of party Jasper has, we could not tell whether inviting only a few children is allowed. I suggest that it is this great ambiguity that renders (22) odd.

2.5.2 Comparison with Blok's (2015) account

In this section I compare the current analysis of *up to n* with Blok's (2015) account.

The denotation of *up to n* in Blok's account is shown in (25).

(25) $[\![\text{up to n}]\!] = \{\lambda M.M(k) \mid k \in [s, n]\}$
s is the contextually determined starting point of the scale $(0 < s < n)$

As a concrete example, the denotation of *you are about to meet up to 100 people* is in (26a) according to Blok's account and in (26b) according to the current analysis.

(26) $[\![\text{you are about to meet up to 100 people}]\!]$
$= [\![\text{up to 100}]\!]\ (\lambda d.[\![\text{you are about to meet } d\text{-many people}]\!])$

a. $= \{q_s, q_{s+1}, \ldots, q_{100}\}$
where s is a contextual starting point of the scale ($0 < s < n$) and q_i is the proposition that the addressee is about to meet at least i people.

b. $= \{p_\theta, p_{\theta+1}, \ldots, p_{100}\}$
where θ is a contextual lower bound ($0 \leq \theta < n$) and p_i is the proposition that the addressee is about to meet exactly i people.

Even though both analyses assume a semantic lower bound, there are two crucial differences.

First, in Blok's semantics the lower bound s of *up to n* can never be 0. Since Blok also requires that $s < n$, her semantics predicts that *up to one* is *never* felicitous when 1 is the smallest non-zero number in the underlying scale. This prediction agrees with Schwarz, Buccola, and Hamilton's (2012) same descriptive generalization, which they call the *bottom-of-the-scale effect* (BotS). However, there are naturally-occurring examples of *up to one* in various linguistic environments (27), where *up to one* arguably means 0 or 1, contra Schwarz et al. (2012) and Blok (2015).

(27) a. You are allowed to bring up to one guest.

b. The committee will submit up to one application.

c. Each panel should consist of a convener, up to four presenters, and up to one respondent.

These examples suggest that 0 can be part of the semantic content of *up to n*, which means that the lower bound can be 0. BotS, which seems to hold mostly for simple episodic sentences, requires a pragmatic explanation.

Second, the contextually determined lower bound s in Blok's semantics is only intended to capture the contextual *granularity* of the scale. For instance, if we assume that eggs are minimally sold in cartons of six then the starting point of the scale s would be 6 in this context. Essentially, s is by definition always the lowest non-zero number in the contextual scale, i.e., the bottom of the scale. This means that Blok's semantics does not allow for the lower bound to be greater than the bottom of the scale and therefore provides no explanation for the proximity inference (which she leaves for future research). In contrast, in the current analysis, the contextual lower bound is allowed to range within $[0, n)$. This, together with a concrete pragmatic analysis of the contextual resolution of θ, accounts for the proximity inference.

There is another difference between Blok's proposal and the current analysis. Blok argues that, unlike *at most n*, *up to n* in fact does not impose a semantic upper bound. Rather, its upper bound is pragmatically implicated, and therefore can be, e.g., suspended (28a).

(28) a. Up to ten people died in the crash, perhaps even more.

b. # At most ten people died in the crash, perhaps even more.

Therefore, according to Blok, the denotation in (26a) contains only one-sided propositions q_i, i.e., the addressee is about to meet at least i people. This means that the informative content does not impose an upper bound. In contrast, the current analysis assumes that the denotation in (26b) contains two-sided propositions p_i, i.e., the addressee is about to meet exactly i people, which imposes an upper bound in the informative content.

However, I note that this is not a crucial difference. The assumption that the upper bound of *up to n* is semantic in this chapter is for simplicity, and can be adapted if we want to follow Blok's analysis that the upper bound of *up to n* is pragmatically implicated rather than semantically imposed. Similar to the previous discussion on free-choice inferences, here we will need to assume that the result of this pragmatic mechanism feeds into the pragmatic mechanism proposed in this chapter. This assumption is in fact very natural given Blok's analysis, because she derives the pragmatically implicated upper bound by applying an exhaustivity operator EXH defined by Coppock and Brochhagen (2013). In this respect the "pragmatic" upper bound in fact still has a semantic representation, and therefore we can reasonably assume that this semantic representation can feed into the pragmatic mechanism proposed in this chapter.

2.6 Conclusion

In this chapter, I examined the opposite inference patterns of *up to n* in SPEAKER-UNCERTAINTY and AUTHORITATIVE-PERMISSION contexts. I argued that such patterns can be accounted for by postulating an underspecified, contextually determined lower bound θ in the semantic content of *up to n*, together with a pragmatic mechanism to determine θ by considering the interaction/trade-off between applicability and informativity. I showed that this interaction can be understood qualitatively as the classic interaction between Grice's Maxims of Quantity and Quality, and that it can also be generalized to a probabilistic model to make more quantitative predictions.

This case study of *up to n* shows how King's (2013, 2014a, 2014b) coordination account can be supplemented with a concrete (qualitative or quantitative) theory of how the idealized listener infers the speaker's intention to provide a concrete pragmatic contextual resolution mechanism to answer the metasemantic question. In the next chapter, however, I argue that this is not the correct contextual resolution mechanism for positive forms of gradable adjectives, which similarly have underspecified semantic lower bounds. What we need instead is a semantic contextual resolution mechanism.

Chapter 3

Semantic contextual resolution: Gradable adjectives

3.1 Introduction

This chapter concerns gradable adjectives such as *tall* and *full*, which can participate in a variety of degree constructions (1).

(1) a. This glass is tall/full. (positive: predicative)
b. Please give me the tall/full glass. (positive: attributive)
c. This glass is taller/fuller than that one. (comparative)
d. This glass is as tall/full as that one. (equative)
e. This glass is the tallest/fullest. (superlative)
f. That glass is tall/full enough. This glass is too tall/full. (enough/too constructions)

I will focus on the positive forms (i.e., morphologically unmarked forms such as (1a) and (1b)), and for brevity I will henceforth often refer to positive forms of gradable adjectives simply as gradable adjectives. As discussed in Chapter 1, the gradable adjective *tall* has context-sensitive interpretations, and it is common to assume that its semantic representation involves an underspecified standard/threshold θ that is determined or resolved in context (2a). Moreover, given that *tall* and *full* can be analyzed uniformly in most degree constructions in (1), ideally we would like to maintain the parallel between the two adjectives in positive forms, i.e., we may expect that *full* has a semantic representation (2b) that is in parallel with (2a).

(2) a. $[\![\text{This glass is tall}]\!] = \mathbf{height}(\mathbf{g}) \geq \theta$, where θ is a contextually determined standard/threshold

b. $[\![$This glass is full$]\!] = \mathbf{fullness(g)} \geq \theta$, where θ is a contextually determined standard/threshold

Note that the threshold θ is essentially a lower bound for the height/fullness of the glass. In this respect, the underspecified semantic representations for gradable adjectives seem highly similar to the one for *up to n* as discussed in the previous chapter. A natural question is whether the same pragmatic contextual-resolution mechanism developed there can be used to answer the metasemantic question for gradable adjectives.

Despite this apparent similarity, in this chapter I argue that the underspecified thresholds in (2) are resolved semantically rather than pragmatically. The critical test cases involve gradable adjectives in definite referring expressions in referential contexts (1b), where the pragmatic contextual-resolution mechanism proposed in the previous chapter makes incorrect predictions. In contrast, I show that once we identify the relevant contextual parameter that determines the threshold, together with a theory of how this relevant contextual parameter is determined when the gradable adjective is embedded under a definite description in a referential context, we can see that the dependency between the threshold and the contextual parameter is the same whether or not the gradable adjective is embedded under a definite description. This is exactly what is expected if the underspecified threshold is determined by a semantic contextual-resolution mechanism.

On the other hand, despite the differences between the semantic and pragmatic contextual-resolution mechanisms, I will also show that they are also closely related in that they are motivated by functional considerations about how to use language in a way that best serves the interlocutors' communicative goal. In particular, the semantic contextual-resolution mechanism similarly involves an interaction/tradeoff between informativity and applicability as discussed in the previous chapter.

The rest of the chapter is organized as follows. In Section 3.2, I discuss empirical properties of gradable adjectives and the related theoretical issues they raise. Section 3.3 provides a critical review of previous approaches, in which I discuss the answers to the semantic and metasemantic questions these approaches each provide, their limitations, and the parts that motivate my own analysis. Specifically, I discuss three types of approaches: (i) a now fairly standard degree-based semantics by Kennedy (2007), (ii) analyses that assume semantics very similar to Kennedy's but provide a different answer to the metasemantic question (King, 2014a; Lassiter & Goodman, 2013, 2015), and (iii) recent delineation-based approaches (van Rooij, 2011b, 2011b). In Section 3.4, I argue that the threshold of a gradable adjective is resolved by a semantic mechanism, based on a contextual comparison distribution, which is a generalization of comparison classes. Crucially, the mechanism is sensitive to both the central tendency and dispersion of the comparison distribution. I show how this mechanism can account for the properties of gradable adjectives and address the related theoretical issues. In Section 3.5, I provide further discussion and comparisons with the previous approaches.

3.2 Empirical properties and theoretical issues

3.2.1 Context sensitivity

Many gradable adjectives, such as *tall* and *big*, have context-sensitive interpretations. For instance, suppose John is $6'2''$ and is a professional basketball player. Then (3) would (generally) be considered true if we are comparing him with the general US male adult population, but false if we are comparing him with professional basketball players.

(3) John is tall.

This naturally raises a theoretical question: Exactly how does context influence the interpretation of a gradable adjective? There is a general consensus that a contextually determined *comparison class* (CC), e.g., the general US male adult population and professional basketball players, has a major influence on the interpretation of a gradable adjective, and comparison classes can be made explicit by, e.g., *for*-PPs (4).[1]

(4) a. John is tall for a professional (adult) basketball player.

b. John is tall for a male adult.

Moreover, many approaches assume that the interpretation of a gradable adjective involves a *standard of comparison* (also called a *threshold*). For instance, (3) means that John's height reaches a certain standard/threshold. The comparison class influences the standard/threshold in a systematic way. For instance, we can plausibly infer (4b) from (4a).

Taking into account the discussion above, we can now make the first theoretical issue more specific (5).

(5) **Theoretical issue 1 (CC-sensitivity):** How does the contextual comparison class influence the interpretation of a gradable adjective? More specifically, for theoretical approaches that assume standards/thresholds, how is the standard/threshold determined? In particular, how do we account for plausible inference patterns such as the one from (4a) to (4b)?

3.2.2 Vagueness

Many gradable adjectives, including *tall* and *big*, are also vague, even after we control for the contextual comparison class. Consider again *John is tall* (3). Suppose the contextual comparison class is the general US male adult population. The sentence still does not have a clear-cut interpretation and

[1] Throughout this dissertation, unless otherwise specified, comparison classes should be understood as a descriptive notion that corresponds to our intuitive understanding of what determines the interpretation of a gradable adjective. Similarly, when I say a comparison class is made explicit, specified, or introduced by a *for*-PP, it should not be taken as a theoretical commitment that *for*-PPs denote comparison classes. It just means that the *for*-PP helps identify the comparison class appropriate for the intended interpretation.

there are borderline cases. For instance, if John is $5'11''$ or $6'$, then it is less clear whether *John is tall* is true. Any adequate theory of gradable adjectives should account for such (aspect of) vagueness.[2]

In the literature, gradable adjectives such as *tall* and *big* are called *relative* adjectives (e.g., Kennedy & McNally, 2005; Kennedy, 2007) are called *relative* adjectives and are contrasted with *absolute* adjectives that I will discuss in the next section. A gradable adjective is called a relative adjective, or is said to have a relative standard, iff its interpretation intuitively requires a contextual comparison class.[3] Note that this definition based on interpretations of gradable adjectives, so it is often more appropriate to say that a gradable adjective has a relative interpretation. However, following the literature, I will continue to call gradable adjectives such as *tall* and *big* relative (gradable) adjectives because they do not have the absolute interpretations to be introduced in the next section.

In addition, if the interpretation of a gradable adjective is not clear-cut and has borderline cases, then the gradable adjective (or more precisely, this interpretation of the gradable adjective) is said to be vague or have a vague standard. The discussion above about *tall* suggests that relative adjectives are vague. In fact, this seems to be a fairly robust empirical generalization and is often taken as a characteristic property of relative adjectives (e.g., Kennedy, 2007). Therefore, a natural theoretical challenge is to explain why this is the case.

Following Graff (2000), I assume that vagueness should be analyzed in terms of the *boundarylessness* of the interpretation, i.e., the inability for us to identify the exact set of borderline cases. For analyses that assume standards, this means that a relative adjective lacks not only an exact standard, but also an exact range of standards. The second theoretical issue is to explain why this is the case (6).

(6) **Theoretical issue 2 (rel-adj-vague)**: Why are relative adjectives vague? More specifically, for analyses that assume standards, why do they lack exact standards or even exact ranges of standards?

A satisfactory analysis of the vagueness of relative adjectives should also address both theoretical issues 1 and 2. Note that listeners can reliably infer (4b) from (4a) despite the vagueness of *tall* in each sentence, i.e., it is unclear exactly how tall would count as tall for a professional basketball player in (4a) or for a male adult (4b). Therefore, it is not enough to just provide a way to represent vagueness. We need to also specify how such representations are influenced by contextual comparison classes in a way that allows us to account for the relevant inference patterns.

---

[2]There are other diagnostics for/aspects of vagueness, most notably susceptibility to sorites paradoxes, that I will not discuss in this dissertation. As far as I can tell, my proposed analysis of gradable adjectives is compatible with major approaches to the sorites paradox.

[3]I use *intuitively* to highlight the fact that I take the classification of gradable adjectives into relative and absolute ones to be descriptive, and leave open the theoretical possibility that all gradable adjectives require a comparison class.

3.2.3 Absolute gradable adjectives

There are other classes of gradable adjectives whose interpretations are not context-sensitive and vague (in the sense that will be specified below). Such gradable adjectives are called absolute adjectives, or said to have absolute standards. Absolute adjectives are further categorized into two subclasses: *maximum* and *minimum* (gradable) adjectives. Below I use maximum adjectives as examples to discuss the relevant theoretical issues.

Adjectives such as *straight*, *full/empty*, and *flat* are called maximum adjectives (or said to have maximum standards). They participate in various degree constructions just like relative gradable adjectives, e.g., *full* in (1) and *straight* in (7). However, their positive forms receive a maximum interpretation. For instance, (7a) is true iff the stick is maximally straight. This interpretation is quite clear-cut. Also, it does not seem to require a contextual comparison class, i.e., intuitively the truth of (7a) does not depend on what we are comparing the stick with: the stick is straight, simpliciter.

(7) a. This stick is straight. (positive: predicative)

b. Please give me the straight stick (positive: attributive)

c. This stick is straighter than that stick. (comparative)

d. This stick is as straight as that stick. (equative)

e. This stick is the straightest. (superlative)

f. This stick is straight enough. That stick is too straight. (enough/too constructions)

However, maximum adjectives do not always have clear-cut, context-insensitive interpretations. First, Kennedy (2002), attributing the examples in (8) to Jeff King, observes that the *for*-PPs are felicitous and therefore concludes that it is in fact possible for a comparison class to shift the standard of a maximum adjective. Also, note that the gradable adjectives in (8) do not have exact standards or exact ranges of standards, i.e., they are vague.

(8) a. That cue is straight for a pool cue in a dive like this.

b. This theater is empty for a theater showing a popular movie.

These examples suggest that maximum adjectives can in fact have vague relative interpretations. I will set this complication aside for now and confine the discussion to sentences without overt *for*-PPs, but I will return to this issue when I discuss Kennedy's (2007) analysis of maximum adjectives.

Another source of context-sensitivity and vagueness when we interpret maximum adjectives is *imprecision* or *granularity*. In general, speakers do not always use language in a perfectly strict or precise manner. For instance, it is generally acceptable to assert (9a) in ordinary conversation if John in fact left at 3:02.[4] Similarly, (9b) is generally acceptable if the glass is in fact 95% full.

[4]Whether or not (9a) is considered true in this context will depend on exactly how one sets up the semantics and the semantics/pragmatics interface.

(9) a. John left at 3.

b. This glass is full.

The acceptable level of imprecision is often context sensitive and is often not clear-cut, which resembles the interpretation of a relative adjective. However, one crucial difference is that once we explicitly control the level of imprecision using expressions such as *strictly speaking*, maximum adjectives receive a clear-cut, maximum interpretation (10b), whereas it is infelicitous to do so for relative adjectives (10c).[5]

(10) a. Strictly speaking, John did not leave at 3. (He left at 3:02.)

b. Strictly speaking, this glass is not full. (It is 95% full.)

c. ? Strictly speaking, John is not tall. (He is 5′11″.)

The examples above might leave the impression that imprecise uses are only about the speaker being sloppy and saying things that they know to be untrue. But this need not be the case. Consider (11), which is modified from D. Lewis's (1979) example.

(11) The road is flat.

If you are driving on the road and experiences no bumps, then it seems that you can felicitously assert (11), even if the road in fact has small bumps that are too small to be felt when people are driving in a car. The idea is that our perception or measure of flatness depends on the context and has various levels of granularity. In the above example, the road is maximally flat as far as the driving experience is concerned, and the small bumps on the road do not make a difference in our perception or measure of flatness. This may change in a different context. For instance, if you are pushing a small cart on the same road, then you may notice a lot of bumps and rightfully complain that the road is not flat.

The notion of granularity can help us understand the contrast between (12a) and (12b).

(12) a. The mirror is flatter than the road, ??but both are flat.

b. The road is flat. The mirror is flatter.

On the one hand, the weirdness of the *but*-clause in (12a) suggests that *flat* has a maximum standard (cf. *John is taller than Bob, but both are tall.*). On the other hand, (12b) sounds fine. Crucially, one

[5]I should acknowledge that this is just one interpretation of the data in (10), and a crucial assumption is that *strictly speaking* is used to raise the level of precision. While this view is fairly standard in the literature, in reality people may interpret *strictly speaking* itself loosely as an expression to raise the standard and therefore find (10c) acceptable. Also, they may be imagining a context where there is some technical, precise definition of *tall*, and interpreting *strictly speaking* as referring to that definition. However, note that even if one is willing to accept (10c), one would still be unable to identify an exact standard for *tall*, whereas one can easily identify the exact standard for *full* in (10b). Therefore this contrast still needs to be explained.

way to interpret (12b) coherently is to assume a shift in the way we measure flatness. For the first sentence, since the subject is the road, we may naturally use a measure of flatness in terms of our driving experience. For the second sentence, since the subject is the mirror, we may naturally use a measure of flatness in terms of closer visual inspection or touching, which is more fine-grained than the previous measure. Under such a measure the mirror is indeed flatter than the road. To the extent that we can conceive such a shift in measure (and consequently in granularity), and to the extent that we are willing to accept it as relevant, (12b) can be coherently uttered. In contrast, when we interpret the first clause in (12a), similar to the discussion above, since the subject is the mirror, we may naturally use a measure of flatness in terms of closer visual inspection or touching. When we interpret *flat* in the *but*-clause, due to coherence we would prefer to use the same measure as in the first clause, wrt which the road is not flat. This explains the weirdness of (12a). Note that (12a) can be improved if the first clause is followed by *but both are flat in some sense.* This can explain why some people may find (12a) acceptable.

There is no consensus in the literature about how exactly imprecision and granularity are related or how to formally analyze them (see, e.g., Rotstein & Winter, 2004; Kennedy & McNally, 2005; van Rooij, 2011b; Burnett, 2014; Égré, 2017, and the references therein). I will not provide an answer to these questions, and in particular I will not model them explicitly in the formal semantics.[6] Due to imprecision and/or granularity, the interpretation of a maximum adjective is also context-sensitive and vague. However, once we factor out imprecision and granularity, we can see that a maximum adjective has an exact standard that is the maximum value (or degree) of the relevant measure. In contrast, even if we control for the level of imprecision and granularity, a relative adjective still does not have an exact standard. Therefore, I will henceforth say that maximum adjectives, or more precisely, their maximum interpretations, are not vague, in the sense that they have an exact standard. Our third theoretical challenge is to explain why this is the case (13).

(13) **Theoretical issue 3 (max-adj-nonvague)**: Why do maximum adjectives have maximum interpretations that are not context sensitive and have an exact standard (and therefore are not vague)? Moreover, why is this exact standard the maximum degree?

Let us briefly turn to the last class of gradable adjectives. Adjectives such as *bent* and *dirty* are called minimum adjectives. They are also gradable (e.g., 14a), and their positive forms receive a minimum interpretation. For instance, (14b) is true iff the stick is at least minimally bent. Again, this interpretation is generally also quite clear-cut and does not seem to require a contextual comparison class.

[6] However, I do think that there is evidence that the two notions should be separated. For instance, I can say *the road is completely flat* to emphasize that I do not feel any bumps at all while driving, i.e., I am strictly adhering to the maximum standard, but in this case I am only using a coarse-grained measure. Similarly, I can say *these gloves are totally clean* to emphasize that there is no visible stain at all, but it need not mean that the gloves are sterilized. Previous analyses tend to focus on one notion or even lump the two notions, which I think is a main reason for the existing disagreements.

(14) a. This stick is more bent than that stick.

b. This stick is bent.

Minimum adjectives share a lot of properties with maximum adjectives and they are often treated in parallel. However, there are further complications to the class of minimum adjectives, which will be discussed in detail in the next chapter. Therefore, I will focus on maximum and relative adjectives in this chapter.

3.2.4 Gradable adjectives in definite referring expressions

Many existing theories of gradable adjectives focus on the interpretation of their positive forms used descriptively in predicative positions, such as *this glass is tall/full* (1a). However, as Kennedy (2007) points out, it is important to consider uses of gradable adjectives in attributive positions in referring definite descriptions (1b) to fully evaluate different theories of gradable adjectives.

As a working example, consider a context where two glasses A and B are present. Glass A is a fairly short glass and is almost empty, and glass B is a glass with a height that does not seem tall (but is considerably taller than Glass A) and is half full. Henceforth I will call this context Two-Glasses. Intuitively (15a) is a felicitous request for the taller glass B, whereas (15b) is an infelicitous request for the fuller glass B. These patterns have been tested experimentally and are robust for adults (Syrett et al., 2010).

(15) a. Give me the tall glass.

b. # Give me the full glass.

Syrett et al. (2010) characterize such patterns in terms of the ability to shift the standard for different classes of gradable adjectives. In the working example above, the fact that neither glass A nor glass B is considered tall when judged independently means that neither meets the standard of *tall* when *tall* is interpreted in a general way. However, the fact that (15a) is a felicitous request for the taller glass B means that glass B is "the tall glass" in this referential context, whose existence presupposition requires that glass B meet the standard of *tall* in the Two-Glasses context. Therefore, the standard of *tall* in the Two-Glasses context must have been shifted somehow so that glass B can meet it. In contrast, the fact that (15b) is an infelicitous request for the fuller glass B suggests that the standard of *full* cannot be shifted in the Two-Glassess context. Using this characterization, Syrett et al.'s findings can be summarized as follows. It is generally easy to shift standards for relative adjectives in referential contexts, but generally difficult to do so for maximum adjectives. I will explain an exception for relative adjectives later, and for now let us focus on the part about maximum adjectives. The infelicity of (15b) is very robust in our working example above. In general, however, the empirical facts are more subtle. Foppolo and Panzeri (2011) found that the felicity of shifting the standard for *straight* depended on whether the objects in the referential context can evoke a clear comparison class. For instance, participants were more likely to find *choose the straight banana* felicitous and choose the straighter (but not completely straight) banana, whereas they were less likely to find *choose the straight object* felicitous when facing abstract objects, in which case they were more likely to reject the request. Even though such findings can in principle be explained in terms of imprecision, assuming that different types of objects allow for different levels of imprecision, it also seems plausible to assume that the standard can be shifted from the maximum degree to a non-maximum one when the objects evoke a clear comparison class. Therefore here I say it is "generally difficult" to shift standards for maximum adjectives, to state the empirical generalization as theory-neutral as possible. Our fourth theoretical issue is to explain the difference between relative and maximum adjectives in terms of shifting standards (16).

(16) **Theoretical issue 4 (shifting-standards)**: Why is it generally easy to shift standards for relative adjectives in referential contexts, but generally difficult to do so for maximum adjectives?

There is an exception to the above generalization for relative adjectives. As Kennedy (2007) observes, (15a) becomes infelicitous if the height difference between the two glasses are minor but still noticeable. He suggests that this is closely related to why we find the inductive premise of the sorites paradox, e.g., *a glass that is half an inch shorter than a tall glass is still tall*, plausible and hard to deny. Indeed, for a context where there are two glasses whose height difference is half an inch, the reason why *the tall glass* is infelicitous is presumably that we are unable to say both that the taller glass is tall and that the half-an-inch-shorter glass is not tall. I will call this context CRISP-JUDGMENT, borrowing Kennedy's terminology that describes the phenomenon. Exactly what counts as a minor but noticeable difference is itself vague and imprecise, but we can recast Kennedy's observation in terms of a more gradient generalization: as the degree difference between the two objects approaches 0, it becomes more and more difficult to shift the standard, and consequently the definite description with the gradable adjective becomes less and less felicitous. I will call this the *crisp-judgment* effect, and the last theoretical issue in this section is to account for it (17).

(17) **Theoretical issue 5 (crisp-judgment)**: How do we account for the crisp-judgment effect, i.e., shifting standards becomes more and more difficult as the degree difference approaches 0?

The theoretical issues discussed so far are summarized in (18).

(18) Summary of the theoretical issues so far

1. **CC-sensitivity**: How does the contextual comparison class influence the standard? (5)
2. **Rel-adj-vague**: Why do relative adjectives have vague standards? (6)
3. **Max-adj-nonvague** Why do maximum adjectives have context-invariant, clear-cut interpretations that use maximum degrees as standards? (13)

4. **Shifting-standards**: Why is it generally easy to shift standards for relative adjectives in referential contexts but impossible for maximum adjectives? (16)

5. **Crisp-judgment** Why does shifting standards become more and more difficult as the degree difference approaches 0? (17)

3.3 Previous approaches

In this section, I review and compare representative theories of gradable adjectives. For each theory, I introduce the compositional semantics, and discuss whether and how the semantics-pragmatics interface posited (explicitly or implicitly) by the theory addresses the theoretical issues discussed in the previous section, and the extent to which the explanation is satisfying. Along the way, additional theoretical issues will arise. Ultimately, I conclude that while each of the approaches cannot satisfactorily address all the theoretical issues on its own, they all provide important insights that will inform my proposal in the next section.

I will use the following type system for the theories to be discussed.

(19) a. Basic types in the system: e (individuals), t (truth values) and d (degrees)

b. For any type τ in the system, there is an intensional type $s \to \tau$, which is a function that takes an *index* i and returns a value of type τ. An index i is a pair $\langle w, v \rangle$, where w is a possible world and v is a contextual variable/parameter whose specification will differ depending on the theory.[7]. For an index i, I will use w_i and v_i to refer to its first and second element.

c. For any type τ in the system there is a *maybe* type $\tau^?$, which is just the original type τ plus a special value `undefined`. This allows us to directly represent undefinedness in the compositional semantics, which will be useful for representing borderline cases (for some theories) and presupposition failure.

d. For any two types σ, τ in the system, there is a functional type $\sigma \to \tau$

e. All the types are generated via (19a–19d).

3.3.1 Kennedy's (2007) analysis

Kennedy (2007), building on Kennedy and McNally (2005), proposes a what is now fairly standard compositional analysis of gradable adjectives. Here I present an intensionalized version, in which an index i is a world-context pair $\langle w, c \rangle$, where c is simply the context of utterance without further specification. A gradable adjective A denotes a (intensionalized) measure function, which takes an

[7]Note that I am using $\langle w, v \rangle$ exclusively to mean a pair. A functional type from σ to τ is only written as $\sigma \to \tau$ or simply $\sigma\tau$.

individual x and an index i and return x's degree of the relevant measure at the world of evaluation w_i (20). The subscript of $[\![\cdot]\!]$ indicates its type. For simplicity, I assume that the measure function is only relativized to w_i, but one could have the measure function relativized to both w_i and c_i, to encode, e.g., the contextual granularity of the measure.

(20) a. $[\![A]\!]_{esd} = \lambda x \lambda i.\mathbf{A\text{-}measure}_{w_i}(x)$

b. $[\![\text{tall}]\!]_{esd} = \lambda x \lambda i.\mathbf{height}_{w_i}(x)$; $[\![\text{full}]\!]_{esd} = \lambda x \lambda i.\mathbf{fullness}_{w_i}(x)$

The positive form of a gradable adjective A denotes an individual property, which takes an individual x and an index i and returns true iff x's degree of the relevant measure in w_i reaches the standard of comparison/threshold θ, which is determined by the measure function g denoted by the adjective A and the context c_i (21).[8]

(21) a. $[\![pos]\!]_{esd \to est} = \lambda g \lambda x \lambda i.\, g(x)(i) \geq \mathbf{s}(g)(c_i)$

b. $[\![pos\ \text{tall}]\!]_{est} = \lambda x \lambda i.\mathbf{height}_{w_i}(x) \geq \mathbf{s}([\![\text{tall}]\!])(c_i)$
$[\![pos\ \text{full}]\!]_{est} = \lambda x \lambda i.\mathbf{fullness}_{w_i}(x) \geq \mathbf{s}([\![\text{full}]\!])(c_i)$

When used in a predicative position, the positive form is composed with the subject, yielding a contextual proposition, i.e., the exact propositional content depends on the context (22a, 22c).

(22) a. $[\![\text{John is } pos \text{ tall}]\!]_{st} = \lambda i.\mathbf{height}_{w_i}(\mathbf{j}) \geq \mathbf{s}([\![\text{tall}]\!])(c_i)$

b. $[\![\text{John is } pos \text{ tall}]\!]^c = \lambda w.\mathbf{height}_{w}(\mathbf{j}) \geq \mathbf{s}([\![\text{tall}]\!])(c)$

c. $[\![\text{Glass B is } pos \text{ full}]\!]_{st} = \lambda i.\mathbf{fullness}_{w_i}(\mathbf{b}) \geq \mathbf{s}([\![\text{full}]\!])(c_i)$

d. $[\![\text{Glass B is } pos \text{ full}]\!]^c = \lambda w.\mathbf{fullness}_{w}(\mathbf{b}) \geq \mathbf{s}([\![\text{full}]\!])(c)$

For example, for *John is pos tall*, if we fix the context to c, then it denotes the proposition that John's height exceeds the threshold $\mathbf{s}([\![\text{tall}]\!])(c)$, which is the correct denotation (22b), similarly for *Glass B is pos full* in (22d).

The remaining problem is how to specify the function $\mathbf{s}$. In particular, it should account for the difference between relative and maximum adjectives. In effect, Kennedy (2007) specifies $\mathbf{s}$ in the following way (23).

(23) a. For a measure function g (type esd), think of it as a bivariate function taking type e and type s inputs, and consider its range, i.e., the set $\{d \mid \exists x \exists i.g(x)(i) = d\}$. If this set has a maximum (or minimum) value, then we say that the measure function has a maximum (or minimum) degree and call this degree $g_{\max}$ (or $g_{\min}$).

[8] Kennedy (2007) formulates $\mathbf{s}$ as "a context-sensitive function from measure functions to degrees that returns a standard of comparison based both on properties of the adjective g (such as its domain) and on features of the context of utterance" (p. 16). Here I am simply using c_i to explicitly represent such contextual features.

b. $\mathbf{s}$ is such that, if $g_{\max}$ (or $g_{\min}$) exists, then $\mathbf{s}(g)(c) \equiv g_{\max}$ (or $g_{\min}$) regardless of c, and otherwise $\mathbf{s}(g)(c)$ returns a standard/threshold θ that "stands out" in the context c in some way.[9]

Given that ⟦tall⟧ does not have a maximum degree whereas ⟦full⟧ does (i.e., 100%), (23) predicts a context-dependent denotation for *John is pos tall* and a context-independent denotation for *Glass B is pos full* (24).

(24) a. $[\![\text{John is } pos \text{ tall}]\!]^c = \lambda w.\mathbf{height}_w(\mathbf{j}) \geq \theta_c$, where θ_c is the minimum degree that "stands out" in c in some way

b. $[\![\text{Glass B is } pos \text{ full}]\!]^c = \lambda w.\mathbf{fullness}_w(\mathbf{b}) \geq 100\%$

This captures the contrast between relative and maximum adjectives in terms of context sensitivity and vagueness when they are used descriptively in predicative positions. It addresses theoretical issues 1–3. The influence of the contextual comparison class on the interpretation of a relative adjective is encoded in the notion of "standing out" in context (**CC-sensitivity**). Relative adjectives are vague because we do not know exactly what is the minimum degree that "stands out" in context (**rel-adj-vague**). The definition of $\mathbf{s}$ ensures that a maximum adjective has a context-invariant, clear-cut interpretation that uses the maximum degree as the standard (**max-adj-nonvague**).

Now we turn to uses of gradable adjectives in a definite description *the pos A N* (e.g., *the pos tall/full glass*) in referential contexts. Assuming intersective modification, after composing the positive form with the noun, we have (25).

(25) a. $[\![pos \text{ tall glass}]\!]_{est} = \lambda x \lambda i.\mathbf{glass}_{w_i}(x) \wedge \mathbf{height}_{w_i}(x) \geq \mathbf{s}([\![\text{tall}]\!])(c_i)$

b. $[\![pos \text{ full glass}]\!]_{est} = \lambda x \lambda i.\mathbf{glass}_{w_i}(x) \wedge \mathbf{fullness}_{w_i}(x) \geq \mathbf{s}([\![\text{full}]\!])(c_i)$

The denotation of the definite article *the* in referential definite descriptions is defined in (26).[10] The definition captures the existence and uniqueness presuppositions of a definite description.[11]

(26) a. $\imath$ is a function of type $et \to e^?$ defined as follows: for any input Q of type et, if there is a unique individual x s.t. $Q(x) = 1$, then $\imath(Q) = x$, otherwise $\imath(Q) = \texttt{undefined}$

b. $[\![\text{the}]\!]_{est \to se^?} = \lambda P_{est} \lambda i.\imath(\lambda x.P(x)(i))$

[9] If both $g_{\max}$ and $g_{\min}$ exist, then the threshold can potentially be either one, but not anything else.

[10] I treat $\imath$ as a function instead of using the more popular notations such as $\imath x.P(x)$ and $\iota x.P(x)$, so that we do not need to add extra syntax to the formal language, and also to avoid potential confusion caused by people's previous experience of how those terms are interpreted.

[11] I follow Syrett et al. (2010) in assuming that definite descriptions have a uniqueness presupposition according to which *the P*, where P is a singular noun phrase, requires that there be at most one P. However, the arguments and analyses in this chapter do not actually hinge on this assumption. Rather, they rely on the existence presupposition. Therefore, if preferred, one could follow, e.g., D. Lewis (1979), and assume a weaker version of the uniqueness presupposition, which only requires that there be at most one P that is the most salient.

This allows us to derive the denotation of *the pos tall glass* in (27a), whose type is $e^?$ when relativized to the world of evaluation w (omitted for brevity) and the context c (27b). Since *tall* is a relative adjective, a context-dependent threshold θ_c is used (27c). In the TWO-GLASSES context, where $\mathbf{height}(\mathbf{a}) < \mathbf{height}(\mathbf{b})$, (27c) is defined iff $\mathbf{height}(\mathbf{a}) < \theta_c \leq \mathbf{height}(\mathbf{a})$, and its value is **b** when it is defined.

(27) a. $[\![\text{the } pos \text{ tall glass}]\!]_{se^?} = \lambda i.\iota(\lambda x.\mathbf{glass}_{w_i}(x) \wedge \mathbf{height}_{w_i}(x) \geq \mathbf{s}([\![\text{tall}]\!])(c_i))$

b. $[\![\text{the } pos \text{ tall glass}]\!]^c = \iota(\lambda x.\mathbf{glass}(x) \wedge \mathbf{height}(x) \geq \mathbf{s}([\![\text{tall}]\!])(c))$

c. $[\![\text{the } pos \text{ tall glass}]\!]^c = \iota(\lambda x.\mathbf{glass}(x) \wedge \mathbf{height}(x) \geq \theta_c)$, where θ_c is the minimum degree that "stands out" in c

Since Kennedy does not provide a formal account of what "standing out" means, we can only resort to our intuitions. Presumably, if there are only two glasses in the context and glass B is taller than A (and the height difference is big enough), then glass B would "stand out" in such a context and glass A would not, i.e., $\mathbf{height}(a) < \theta_c \leq \mathbf{height}(b)$. Therefore, *the pos tall glass* felicitously refers to the taller glass B in such a context. This accounts for the ability to shift standards for relative adjectives. Similarly, when the heights of the two glasses are too close, presumably neither glass would stand out in such a context, and therefore *the pos tall glass* is undefined (and thus infelicitous). This addresses theoretical issue 5 (**crisp-judgment**).

As for maximum adjectives, the denotation of *the pos full glass* relativized to the actual world (omitted) and a context c is in (28a). Given that $[\![\text{full}]\!]$ has 100% as the maximum degree, (28a) amounts to (28b).

(28) a. $[\![\text{the } pos \text{ full glass}]\!]^c = \iota(\lambda x.\mathbf{glass}(x) \wedge \mathbf{fullness}(x) \geq \mathbf{s}([\![\text{full}]\!])(c))$

b. $[\![\text{the } pos \text{ full glass}]\!]^c = \iota(\lambda x.\mathbf{glass}(x) \wedge \mathbf{fullness}(x) \geq 100\%)$

This predicts that *the pos full glass* is defined iff there is a unique glass that is maximally full and if this is the case the definite description denotes that glass. Therefore shifting standards is predicted to be impossible (or difficult when we also take into account imprecision and granularity) for maximum adjectives, but generally easy for relative adjectives (from the discussion before). This addresses theoretical issue 4 (**shifting-standards**).

Summing up, Kennedy's (2007) analysis addresses all the theoretical issues discussed so far. However, despite the impressive coverage, Kennedy's account has several limitations, both conceptual and empirical.

Conceptually, the obligatory use of the maximum (minimum) degree as the threshold can seem stipulative without independent motivation. Kennedy (2007) is aware of this concern and proposes as motivation a principle of *Interpretive Economy* (IE), which requires that interlocutors "maximize the contribution of the conventional meanings of the elements of a sentence to the computation of its truth conditions" (p. 36). This principle, while intuitively plausible, is not without problems.

First, it requires independent evidence. Kennedy (2007) assumes that this is a constraint on semantic processing. The idea is presumably that conventional meanings are easier to process whereas it takes more effort to compute meanings that are specific to the context, and therefore it is economical for interlocutors to use conventional meanings as much as possible. While this idea intuitively sounds plausible, Kennedy does not provide independent evidence. Indeed, he acknowledges that "this is clearly a hypothesis that should be further tested and developed in an experimental context; here I will focus on providing empirical support for it by showing the role it plays in explaining the facts discussed in this paper" (fn. 31, p. 36). Second, this principle is not formalized, and therefore it is not entirely clear how it should be understood and what auxiliary assumptions are needed to derive its intended effect. For instance, as Potts (2008) points out, in order to apply IE to maximum/minimum adjectives, we need to assume that "endpoints are conventionalized meanings in some sense." Note that technically, the conventional meaning of a gradable adjective is a measure function. Endpoints as well as interior points are just parts of the range of this measure function, and therefore it is prima facie unclear in what sense endpoints are part of the conventional meaning of a gradable adjective but interior points are not.

To facilitate further discussion below, let us introduce the terminology Kennedy uses. Depending on whether the measure function g associated with a gradable adjective has maximum and minimum endpoints in its range, there are four logical possibilities for the *scale structure* of the gradable adjective (29), as schematically illustrated in (30).

(29) a. If the measure function g has neither a maximum nor a minimum in its range, the adjective has a (totally) open scale.

b. If the measure function g has a minimum but not a maximum in its range, the adjective has an lower-closed scale.

c. If the measure function g has a maximum but not a minimum in its range, the adjective has an upper-closed scale.

d. If the measure function g has both a maximum and a minimum in its range, the adjective has a (totally) closed scale.

(30) *A typology of scale strucures*

a. (TOTALLY) OPEN: ○————————○

b. LOWER CLOSED: ●————————○

c. UPPER CLOSED: ○————————●

d. (TOTALLY) CLOSED: ●————————●

According to Kennedy, endpoints are special because they mark "natural transitions" on the scale. He further links natural transitions to his notion of "standing out" by assuming that "what it means to stand out relative to the measure expressed by a closed scale adjective is to be on the upper end of a natural transition based on the scale" (p. 35). With these assumptions, IE will ensure that absolute adjectives always have endpoints as standards in the following way. The semantics of *pos* always requires that the threshold "stand out." In principle, the threshold can stand out relative to some contextually determined set of degrees, or relative to the entire scale of the adjective. The latter is part of the conventional meaning of the gradable adjective, and therefore IE requires that we choose this option whenever possible. Given the assumptions about natural transitions above, only endpoints stand out relative to the entire scale. Therefore, absolute adjectives must use endpoints as standards. In contrast, relative adjectives do not have endpoints on their scales, and therefore nothing stands out relative to their scales and they have to resort to contextually determined standards.

Note that the crucial assumption is that only endpoints mark natural transitions, or equivalently, only endpoints stand out relative to the entire scale. But why would this be the case? Why would an interior point (e.g., 50% full) not mark a natural transition? Note that this is not the same question as why an interior point cannot be the standard for *full*. The latter question is what Kennedy's IE is designed to answer, which crucially relies on the assumption that interior points are not natural transitions. It is important that this assumption about natural transitions not be circular. For instance, if we say that the maximum endpoint 100% full is a natural transition because it marks the transition from not full to full, then we seem to be begging the question. Since Kennedy does not provide an independent definition of natural transitions, it is hard to know whether IE provides a non-circular explanation.

Such an independent definition is possible. Potts (2008) proposes that natural transitions are to be understood as *Schelling points*, i.e., points that are independently cognitively prominent to us and able to facilitate coordination. The idea is basically the following. Imagine that two people are playing a game. In front of them is an array of glasses ranging from containing no water at all to completely full. The rule is that each of them independently choose a glass, without the other person knowing what they choose, and they win iff their choices are the same. Intuitively, it seems plausible that people would show a preference (which need not be large) for the endpoints, i.e., the completely empty or full glasses. Assuming this is indeed the case, Potts (2008) shows that using endpoints is the only evolutionarily stable strategy.

Potts's proposal has some nice features. First, it is in principle empirically testable (though I am not aware of any experimental work on this, and I should also acknowledge that Potts does not explicitly say that the setup of the game that I described above is the right way to test his theory). Second, it is not circular, because the game can in principle be played by people who have never heard *full/empty* before. Finally, there is no need to explicitly postulate Interpretive Economy to achieve its desired effects, because strategies that always choose Schelling points will also outperform strategies that make context-dependent choices.

However, note that Potts's proposal in fact does not rule out the possibility that an interior point marks a natural transition. According to Potts, if an interior point is salient enough to be a Schelling point, then it can also serve as the standard. In fact, Potts argues that this is supported by the following example. Suppose we are at a bar where the glasses have a marked line indicating how much the bartender are supposed to fill them. The marked line makes the corresponding non-maximal degree salient enough to be a Schelling point, and therefore we can use it as a standard and truthfully say *the glass is full* as long as the glass is filled to that line. McNally (2011) similarly points out that a glass of wine is considered full if it is filled with half of its capacity with wine, and argues that in this case it is best to analyze *full* as having an exact, non-maximal standard at 50% full.[12]

I do not know whether Kennedy would endorse this definition of natural transitions. The reason I discuss Potts's proposal is to illustrate below that even though it addresses the worry about circularity, it does not address the other limitations.

For example, another conceptual concern is about open-scale adjectives. Kennedy's analysis only predicts *that* their standards are context sensitive and vague, and says little about *how* they are determined beyond the requirement that they "stand out" in the context. Since he does not formally spell out what "standing out" means, the empirical properties of relative adjectives (e.g., **CC-sensitivity**) are explained by appealing to our intuitions. McNally (2011) expresses essentially the same concern as follows.

> "Nonetheless, at a deeper level, it is difficult to see how the value returned by **s** could be characterized in any truly unified terms across absolute and relative adjectives other than as 'the degree that makes the adjectival predicate truthfully hold.' Moreover, in the case of relative adjectives, I do not see any way to derive this standard degree except in an *a posteriori* fashion on the basis of the way the members of the comparison class are sorted in any given context.[13] But if this is the case, it would seem that, as mentioned above, the identification of the standard presupposes that we are able to successfully use the adjective. This is of course not a problem for a strictly formal account of adjective semantics, but it lends support to the criticisms of such a semantics as a useful model of our semantic competence." (p. 166)

Essentially, the worry is that Kennedy's degree-based analysis of positive forms of gradable adjectives might be circular, which would be the case if the notion of "standing out" in a context, which allows us to identify the standard, is actually entirely based on our knowledge about the positive forms. Ideally, it should be the other way around, i.e., an explanatory theory should be able to predict what the standard of a positive form is, based on independent information that does not require knowledge about the meaning of the positive form. Note that since Potts (2008) does not provide an analysis for relative adjectives, his proposal does not address this concern.

Having discussed these conceptual concerns about Kennedy's analysis and a possible way to address

[12] Further contextual variability is possible. Perhaps 50% full is the standard for white wine only, and for red wine glasses the standard is 1/3 full. And different places can have different conventions.

[13] The verb *sort* is ambiguous. It can mean to classify, but it can also mean to order or rank (especially in computer science). Here McNally (2011) is using it in the first sense.

some but not all of them, let us turn to some more serious concerns about its empirical predictions. By design, Kennedy's analysis predicts Kennedy and McNally's (2005) empirical generalizations (31).

(31) a. Open-scale adjectives are relative adjectives, i.e., they always have relative standards.

b. Lower-closed-scale adjectives are minimum adjectives, i.e., they always have minimum standards.

c. Upper-closed-scale adjectives are maximum adjectives, i.e., they always have maximum standards.

d. Closed-scale adjectives can be minimum or maximum adjectives, but not relative adjectives, i.e., they can have maximum or minimum standards but never relative standards.

The above generalizations capture the strong correlation between scale structure and the standard or a gradable adjective. However, these generalizations are too strong because the correlation is not perfect.

To begin with, at least some closed-scale adjectives can have relative standards when there is an overt *for*-PP (8). Below is another example (32).

(32) The theater is full for a Thursday afternoon.

Crucially, (32) is different from the examples we discussed about beer or wine glasses, in that the standard in (32) is sensitive to the comparison class and vague (i.e., we cannot identify an exact standard).

According to Kennedy's (2007) treatment of *for*-PPs, the effect of the *for*-PP in (32) is to restrict the domain of the measure function of *full* so that it is defined only when we are measuring the fullness of a theater on a Thursday afternoon. However, as far as the range is concerned, this restricted measure function still ranges from 0 to 100% full, because it is possible for a theater to be completely empty or full on a Thursday afternoon (even though the latter might be unlikely). In other words, the restricted measure function still corresponds to a closed scale, and therefore IE predicts that it is impossible for the adjective to have a relative standard. Note that Potts's (2008) proposal also makes the same prediction.

One potential way for Kennedy to avoid this problem is to instead treat the *for*-PP as a semantic argument of *pos* that denotes the comparison class (although he explicitly argues against this move). In this case, since the comparison class is supplied by the conventional meaning of the *for*-PP, there is a construal of IE that prefers the relative standard based on this comparison class. The relative standard is determined based on the conventional meanings of all the parts in the sentence, whereas if we choose a standard that stands out wrt the scale, we would be discarding the contribution of the conventional meaning of the *for*-PP. If we take IE to require that the conventional meanings of all the parts in the sentence be used, then the relative standard is preferred when there is an overt *for*-PP. Note, however, that this move is not available for Potts's analysis, since it does not assume IE.

Alternatively, Kennedy can maintain his analysis of *for*-PPs, and assume that IE can interact with other pragmatic principles such as Grice's (1975) Maxim of Manner. Under this assumption, IE is not an inviolable constraint, but rather a default preference that can be overridden when it is in tension with the Maxim of Manner because of the overt *for*-PP. In (32), using the maximum degree as the standard would violate "be brief" because the sentence would mean the same thing without the *for*-PP, and therefore the requirement of IE is lifted in this case and (32) has a relative standard.

In general, defenders of Kennedy's account (or some variant thereof) can postulate that overt *for*-PPs have additional semantic/pragmatic effects that make relative standards possible. However, this strategy would not work in general because relative standards are possible even without overt *for*-PPs.

Kennedy himself acknowledges that *bald* is an exception to his generalization, but argues that "this is the exception that proves the rule" (p. 35). The implication is that if we cannot identify systematic patterns in the exceptions, then the generalization can be maintained. Below I will discuss *bald* in detail. The goal is not to argue that it is particularly problematic for Kennedy's generalization. On the contrary, I think that defenders of IE can in fact make a plausible case that *bald* is not really an exception to Kennedy's generalization. However, discussing *bald* in detail will help illustrate its contrast with *dark*, which I will argue presents a much more serious challenge to IE.

Intuitively, the scale of *bald* has a maximum degree, i.e., having no hair at all. This maximum is also linguistically accessible. For instance, we can say *completely/perfectly bald*. The reason why such adverbial modifications are relevant is that defenders of Kennedy's account often assume that the linguistically relevant scale structure that determines the effect of IE may differ from the intuitive one and the former should be determined strictly based on linguistic diagnostics. For instance, in light of Lassiter's (2010, 2011) observation that the relative adjective *likely* (and similarly *probable*) presents a counterexample to IE because intuitively its scale has a maximum degree (i.e., 100%), Klecha (2012, 2014) argues that despite this intuition, the linguistically relevant scale structure does not have a maximum degree, because it is infelicitous to say, e.g., *completely/perfectly likely* (intending them to mean *100% likely* or *certainly*). An immediate problem with Klecha's analysis is that it is in conflict with standard assumptions about scalar implicatures. Standard tests of implicatures suggest that *often*, *common*, *rare*, *few* are semantically compatible with *always*, *ubiquitous*, *extinct*, and *no*, respectively. For instance, the tests for *often* are shown in (33).

(33) a. John often comes to this place for lunch. In fact, he always does. (cancellation)

b. John often comes to this place for lunch, and perhaps he always does. (suspension)

c. John often, but not always, comes to this place for lunch. (reinforcement)

d. Everybody who often comes to this place for lunch knows how consistently good the food is. $\models$ Everybody who always comes to this place for lunch knows how consistently good the food is.

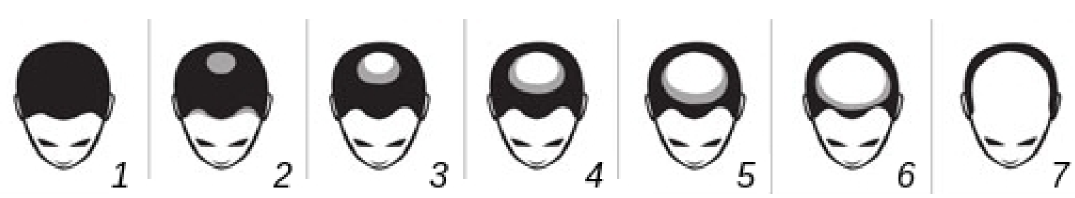
(implicature disappears under downward-entailing environments)

The adjectives/adverbs in the first group are all gradable and intuitively the adjectives/adverbs in the second group correspond to the maximum degrees of their respective scales. However, under Klecha's analysis, since the adjectives/adverbs in the first group cannot be modified by *completely* or *perfectly*, we are forced to conclude that their scales do not have maximum degrees and therefore they are semantically incompatible with the adjectives/adverbs in the second group (e.g., *John often comes to this place for lunch* is undefined if he always does). In order to account for the sentences in (33), Klecha (2012) has to further stipulate that "scales with no inherent maxima or minima can be coerced to include such endpoints in the ordering to the extent that it is intuitive to do so" (p. 371). As a result, we have two types of explanations for what seem to be run-of-the-mill cases of scalar implicatures. On the one hand, if we replace *often* with *sometimes* in (33), the felicity of the sentences has a standard explanation, i.e., *sometimes* is semantically compatible with *always*. On the other, the felicity of the sentences in (33) is explained by an optional coercion operation.

For the sake of the argument, for now I will set aside the issue of whether it is really worth abandoning standard assumptions about scalar implicatures just to save IE and the additional issues discussed by Lassiter (2017), and strictly follow Klecha's linguistic diagnostics, which are largely based on Kennedy and McNally (2005) and Kennedy (2007). Specifically, maximum degrees are targeted by *completely/perfectly/totally/absolutely*, minimum degrees are targeted by *slightly/a little*, and if the scale has both maximum and minimum degrees, the adjective can be modified by proportional modifiers such as *half*, *three quarters*, and *10%*.

The felicity of *completely/perfectly/totally/absolutely* bald suggests that *bald* has a maximum degree on its scale and therefore should not have a vague standard. However, *bald* is famously a vague adjective. This is why Kennedy (2007) considers it an exception.

Defenders of IE may argue that even though *bald* is vague in the general sense that its interpretation is not clear-cut, has borderline cases, and is susceptible to the sorites paradox, it does not mean that it necessarily has a vague standard in the sense that I use in this chapter. It is possible that its vagueness is due to imprecision and/or granularity, as well as other factors. For concreteness, consider the people with various amounts of hair in (34).

(34) 

Intuitively, the scale of *bald* also has a minimum endpoint (person 1 with a full head of hair), which is also confirmed by the felicity of *slightly/a little bald*.[14] This means that IE in fact predicts

[14]Crucially, due to granularity, this endpoint need not correspond to an exact amount of hair. Somebody with 1 hair less than person 1 is not *visibly* balder than person 1, and therefore is not balder as far as the natural level of granularity for baldness in everyday conversation is concerned.

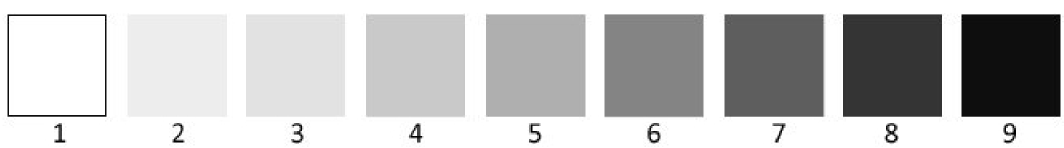
that *bald* can have a minimum standard, according to which persons 2 to 7 are all bald, as well as a maximum standard, according to which nobody in (34) is *bald*. However, perhaps persons 5 to 7 are close enough and therefore at a certain level of imprecision we might loosely say that they are bald. This level of imprecision is often inexact in the context. If the level is a little lower, then perhaps persons 4 to 7 loosely count as bald. If it is a little higher (but still not maximal), maybe only person 7 loosely counts as bald. These factors can easily muddy our interpretation of *bald*, and therefore it might look like we can still maintain Kennedy's account.

However, *dark* (when used to describe objects or materials) presents a more serious challenge. Similar to *bald*, intuitively, the scale of *dark* has a maximum degree (i.e., total absorption of light). Moreover, this is confirmed by linguistic diagnostics, as shown in the following naturally-occurring examples (35).[15]

(35) a. Hair is naturally reflective, so black hair is not completely dark in bright light.[16]

b. Researchers have recently discovered that a microscopic "forest" of vertically aligned single-wall carbon nanotubes of varying heights applied to a surface has extremely low reflectance across a wide range of wavelengths of visible light, the closest scientists have come thus far to creating a perfectly dark material.[17]

Therefore, Kennedy's theory predicts that *dark* should have a maximum standard (unless its scale also has a minimum degree). However, this is arguably not the case. Consider the squares in (36).

(36)


Even though square 2 is darker than square 1, it is clearly not dark (when all the squares are taken into account). This is very different when compared with person 2 in (34). Person 2 is balder than person 1, and in some sense he is bald, or at the very least, it is not the case that he is clearly not bald. This is further confirmed by the linguistic diagnostics. Whereas *person 2 is slightly/a little bald* is true, it is infelicitous to say *square 2 is slightly/a little dark*. Also, it is infelicitous to use *half/three quarters/10% dark* to talk about the darkness of these squares (note that they can in principle be used to talk about what proportion of a square is dark, but this interpretation is obviously ruled out in this case). Therefore, we can plausibly conclude that *dark* does not have a minimum degree on

[15] Note that (35a) once again highlights the need to take into account granularity. What we call black hair looks black under normal lighting conditions, because the amount of light it reflects is not enough for our eyes to detect. This is no longer the case in bright light, which is exactly what (35a) points out.

[16] https://en.wikipedia.org/wiki/Black_hair

[17] *5 lb. Book of GRE Practice Problems*, by Manhattan Prep, retrieved from Google Books.

the scale, or in any event, it clearly does not have a minimum interpretation. As a result, IE indeed predicts that *dark* should have a maximum interpretation.

However, square 6 is visibly less dark than 7 to 9, so clearly it does not have a maximum degree of darkness, but it is still quite reasonable to say that square 6 is dark (when all the squares are taken into account). Can it be that *dark* has a maximum standard and square 6 only loosely counts as dark? I find this analysis implausible, because it does not seem possible to use *strictly speaking* to raise the level of precision (37, cf. 10).[18]

(37) a. ?Strictly speaking, square 6 is not dark.

b. A: John is wearing a dark T-shirt today.
B: ?His T-shirt is charcoal grey, and therefore strictly speaking, it is not dark.

Perhaps a charitable listener would find the examples in (37) acceptable and infer that the speakers have higher standards, but it is still not clear that the speakers are using maximum standards. In this respect, (37) patterns with (10c) more than (10a, 10b). Therefore, it seems plausible to assume *dark* has a relative standard. Moreover, it is quite easy to shift standards for *dark*. For instance, even though square 4 is (presumably) not dark, if we are only looking at squares 2 and 4, we can felicitously use *the dark one* to refer to square 4. Relatedly, we can say *square 4 is dark compared to square 2*. This provides further evidence that *dark* has a relative interpretation.

This is problematic for IE, and its defenders would need to make further stipulations to rule out the possibility of *dark* having a maximum interpretation. One obvious option is that perhaps the maximum interpretation of *dark* is blocked because it is already lexically realized by *black*. While this sounds plausible, it immediately raises a whole series of questions about the scope of this blocking mechanism, i.e., which expression can potentially block the maximum or minimum interpretations of which expression. For instance, why are the maximum interpretations of *full/safe/clean* not blocked by *filled up/risk-free/dirt-free*? Simply appealing to a notion of complexity is unlikely to provide the full answer. For instance, why is the maximum interpretation of *impartial* (e.g., *the judge is impartial*) not blocked by *fair*? Similarly, why are the minimum interpretations of *unsafe/unclean* not blocked by *risky/dirty*? Given all these questions, it seems that introducing a blocking mechanism to account for the relative interpretation of *dark* creates more problems than it fixes.

Summing up the discussion so far about the relation between the scale structure and standard of a gradable adjective, I conclude that (38) provides more accurate generalizations (focusing on relative and maximum adjectives).

(38) a. Open-scale adjectives have relative standards

[18] Note that *bald* is not a proper comparison in this case due to the interference from its minimum standard interpretation.

b. Closed-scale adjectives can have maximum standards, but they can also have relative standards, which may involve additional linguistic material, e.g., *full for a Thursday afternoon* but does not need to, e.g., *dark*.

Similar observations are made by McNally (2011) and Lassiter and Goodman (2013). In fact, Kennedy (2007, p. 35) himself notes that Kennedy and McNally's (2005) generalization that "gradable adjectives that use totally or partially closed scales have absolute interpretations" is "not quite exceptionless," which means that he essentially agrees with the generalizations in (38). However, since his IE predicts a perfect correlation between the scale structure and the interpretation of the positive form of a gradable adjective, the exceptions are left unexplained. Therefore, if a theory can explain such a correlation as well as why the exceptions are exceptions, it should be preferred to Kennedy's analysis.

Summing up the discussion so far, Kennedy's analysis has several limitations: (i) the principle of Interpretive Economy and the notion of a natural transition are not fully spelled out and independently motivated (ii) it says relatively little about how context affects the interpretation of relative adjectives, and (iii) the correlation between the scale structure and the standard of a gradable adjective that it predicts is too strong.

Despite these limitations, Kennedy's analysis provides many important insights. In particular, even though the correlation between the scale structure and the standard of a gradable adjective is imperfect, it is quite strong and therefore still worth explaining (39).

(39) **Theoretical issue 6 (correlation-with-SS)**: How do we account for the imperfect correlation between the scale structure (SS) and the standard of a gradable adjective summarized in (38)? More specifically, what factors account for the differences between closed-scale adjectives that have maximum standards and those that have relative standards?

Before I introduce the next two closely-related approaches, I should highlight that Kennedy's analysis provides a semantic contextual resolution mechanism for the metasemantic question. Specifically, the threshold of a gradable adjective is resolved by **s**, which is part of the conventional meaning of *pos*. This function **s** corresponds to a notion of "standing out," and it derives the threshold based on two factors, i.e., the conventional scale structure of the adjective and the context. In Section 3.4, I will propose an analysis that also assumes a semantic contextual resolution mechanism for the threshold. However, the new analysis will spell out the notion of "standing out" in a different way and identify a single contextual parameter that encompasses the two factors above.

In contrast, the two approaches I will discuss in the next sections provide a different answer to the metasemantic question, according to which the threshold is resolved in context pragmatically. I will first review King's (2014a) coordination account and discuss the basic assumptions and problems in a qualitative way, before discussing Lassiter and Goodman's (2013) independently developed quantitative model that shares the same basic assumptions and problems. However, despite its problems, we will see that Lassiter & Goodman's model provides an important ingredient for the analysis I will propose.

3.3.2 King's (2014) coordination account

King (2014a) proposes a slight modification of Kennedy's (2007) semantics. According to King, a relative gradable adjective still denotes a measure function as in Kennedy's analysis (repeated in 40a), but the contextual variable v is taken to be the threshold θ (40b).

(40) a. $[\![A]\!]_{esd} = \lambda x \lambda i.\textbf{A-ness}_{w_i}(x)$
$[\![\text{tall}]\!]_{esd} = \lambda x \lambda i.\textbf{height}_{w_i}(x)$;

b. $[\![pos]\!]_{esd \to est} = \lambda g_{esd} \lambda x \lambda i.\, g(x)(i) \geq \theta$
$[\![pos \text{ tall}]\!]_{est} = \lambda x \lambda i.\textbf{height}_{w_i}(x) \geq \theta$

As a result, utterances that use gradable adjectives in a predicative position have parallel denotations (41a), or (41b) when it is relativized to a threshold θ.

(41) a. $[\![\text{John is } pos \text{ tall}]\!]_{st} = \lambda i.\textbf{height}_{w_i}(\textbf{j}) \geq \theta$

b. $[\![\text{John is } pos \text{ tall}]\!]^{\theta} = \lambda w.\textbf{height}_{w}(\textbf{j}) \geq \theta$

King then applies his coordination account to address the metasemantic question of how the threshold is determined in context (42).

(42) The threshold of a gradable adjective A is θ in context c iff the following two conditions are met

a. The speaker intends θ to be the threshold of A in c

b. A competent, attentive, reasonable hearer who knows the common ground of the conversation at the time of utterance would know that the speaker intends θ to be the value of A in c

A feature of this proposal is that it is not specific to gradable adjectives as it provides a unified metasemantics for all context-sensitive expressions. However, the account has several limitations.

On the one hand, this account is attractive because of its uniformity, on the other hand, it does not directly apply to absolute adjectives. Glanzberg (2020) uses Kennedy's (2007) analysis of absolute adjectives to argue that speaker's intention is not always relevant to determining the threshold of a gradable adjective. King's (2020) response is that the coordination account does not apply to absolute adjectives because they are not context sensitive as they do not involve a contextual threshold. Note that this essentially gives up a unified semantics for different classes of gradable adjectives. In order to derive the context-invariant interpretation for a maximum adjective such as *full*, King would need to postulate a compositional semantics different from (40b), such as (43).[19]

[19] King can also postulate a semantics like (22c) instead of (43b), but this would still not be a unified semantics for gradable adjectives because King assumes that relative adjectives do not have **s** as part of the semantics.

(43) a. $[\![pos_{\text{max}}]\!]_{esd \to est} = \lambda g_{esd} \lambda x \lambda i.\, g(x)(i) = g_{\text{max}}$

b. $[\![\text{Glass B is } pos_{\text{max}} \text{ full}]\!]_{st} = \lambda i.\mathbf{fullness}_{w_i}(\mathbf{b}) = 100\%$

Moreover, he would need to block the combination between *pos* as defined in (40b) (call it pos_{rel}) and a maximum adjective in some cases, because maximum adjectives do not always have relative interpretations. This presumably can only be done in the syntax, because semantically pos_{rel} is perfectly compatible with the measure function denoted by, e.g., *full*, and we will see below that a pragmatic account will fail to account for the difficulty of shifting standards. The problem, however, is that there seems to be little (if any) independent evidence that relative and maximum adjectives are different syntactically.

Furthermore, as discussed in previous chapters, since this account is not formal, it is not always clear what the idealized listener in (42b) would know about the speaker's intention in a given context. As a result, it can be hard to evaluate its predictions in certain cases.

This is particularly problematic when we consider uses of gradable adjectives in referential contexts. Consider the Two-Glasses context. According to the semantics in (40b), we can derive the semantics of the definite description *the pos tall glass* as in (44).

(44) $[\![\text{the } pos_{\text{rel}} \text{ tall glass}]\!]^{\theta} = \iota(\lambda x.\mathbf{glass}(x) \wedge \mathbf{height}(x) \geq \theta)$

In this context, the definite description is felicitous iff $\mathbf{height}(\mathbf{a}) < \theta \leq \mathbf{height}(\mathbf{b})$ (assuming B is taller than A). Suppose that the speaker intends a threshold θ that satisfies this condition. Would an idealized listener (i.e., competent, attentive, etc.) know this intention? The answer to this question crucially depends on how an idealized listener is to be understood. In the previous chapter, I suggest that we think of the idealized listener as making inference about the speaker's intention by assuming that the speaker follows Grice's (1975) Maxims, or generalizations thereof. We can similarly develop an analysis by making a minor addition to Grice's Maxim of Quality, i.e., do not use a definite description if it is undefined.[20]

This requirement can be motivated independently. Consider a scenario where we are at a party and you tell me that you are interested in gradable adjectives. After hearing this, I point in a certain direction and utters (45) to you.

(45) His wife is a leading expert on gradable adjectives.

Suppose that there are only two men in that direction, John and Bob. John is a famous bachelor, and Bob is wearing a wedding ring. Then you can infer that the possessor of the possessive pronoun *his* should be resolved to Bob, because even though the conventional meaning of the possessive pronoun *his* alone only requires that the possessor be male, the conventional meaning of the referring

[20] Alternatively, we can think of this requirement as included in the Maxim of Manner, i.e., avoid obscurity. Presumably, if a definite description is undefined, it would be obscure.

expression *his wife* requires that the possessor have a wife in order for it to be defined and therefore Bob is the only possibility.

Let us now consider what the coordination account with this assumption about the idealized listener predicts about the interpretation of the definite description (44) in the TWO-GLASSES context. Given that the definite description is defined iff $\mathbf{height}(\mathbf{a}) < \theta \leq \mathbf{height}(\mathbf{b})$, and that the idealized listener assumes that the speaker would not use a definite description if it is undefined, it follows that the idealized listener would indeed know that the speaker intends to use a threshold that satisfies this condition, regardless of the height difference between A and B (as long as it is still noticeable).[21]

While this accounts for the fact that shifting standards is generally easy for relative adjectives, it makes problematic predictions about shfiting standards for maximum adjectives and the crisp-judgment effect.

If pos_{rel} can be freely combined with maximum adjectives, we will similarly derive the following semantics of the definite description *the* pos_{rel} *full glass* (46).

(46) $[\![\text{the } pos_{\text{rel}} \text{ full glass}]\!]^{\theta} = \iota(\lambda x.\mathbf{glass}(x) \wedge \mathbf{fullness}(x) \geq \theta)$

Applying the same reasoning as in the case of *tall*, the coordination account predicts that it should be just as easy to shift standards for maximum adjectives, contrary to fact. Therefore the coordination account will have to block the combination between pos_{rel} and *full* when the gradable adjective is in a definite description. However, as discussed before, there is little independent evidence for why this should be the case.

Moreover, in a CRISP-JUDGMENT context, i.e., where the height difference between the two glasses is very small but still noticeable, the coordination account wrongly predicts that the definite description is felicitous.[22] This is because the idealized hearer can apply the same reasoning to infer the speaker's intention, and if anything, the speaker's intention can be better narrowed down to a smaller range of thresholds.

We have seen that if we assume that the idealized listener infers the speaker's intention by assuming that the speaker follows Grice's Maxims, the coordination account makes wrong predictions about uses of gradable adjectives in referential contexts. Below I consider two strategies defenders of the coordination account may adopt to avoid the problems above. Both strategies are used by King (2020) to (convincingly, in my opinion) counter most of Glanzberg's (2020) criticisms of the coordination account. However, I suggest that neither seems applicable here. I will focus on the crisp-judgment effect below, but the same argument can be made about the inability to shift standards for maximum adjectives.

[21] King (2014a) allows for the speaker to intend a range of thresholds, and therefore it is not a problem that the idealized listener can only know the intention to use a range of thresholds.

[22] In fact, Kennedy (2007) discusses crisp judgments to provide evidence for the view that "standing out" is a matter of semantics, although he does not explicitly use it to argue against the coordination account.

The first strategy is to admit that an idealized listener can infer the speaker's intention in a Crisp-Judgment context, and explain the infelicity of the utterance *give me the tall glass* not in terms of the undefinedness of the definite description but rather by appealing to pragmatic irrelevance. It is hard to see how this strategy could work in such a context. The conversational goal is presumably for the speaker to request the glass they want. If the idealized listener can infer the speaker's intended threshold and hence identify the referent of the definite description, the utterance *give me the tall glass* would exactly communicate the request for the taller glass that the speaker wants and is therefore perfectly relevant.

The second strategy is to deny that an idealized listener would know the speaker's intention a Crisp-Judgment context. How plausible this strategy is would depend on the extent to which one can plausibly deny that an idealized listener would know the speaker's intention. This is hard to evaluate without an explicit theory about what an idealized listener would know. Below I just provide a case that I think can be quite challenging for this strategy.

Imagine the following scenario. I am a foreigner who does not know a lot about English, and you are teaching me "the true semantics" of the positive form of *tall* in (40b). Despite my limited knowledge of English, I do know the meanings of *the* and *glass* and therefore I know the meaning of *the pos tall glass* (44). We are in a Crisp-Judgment context. I want the taller glass, and based on the semantics in (44), I say to you *please give me the tall glass*, intending the height of the taller glass to be the threshold. It seems that it is perfectly felicitous for you to say (47) to me in order to teach me how to correctly use gradable adjectives in definite descriptions in referential contexts.

(47) I know that you intended the height of the taller glass to be the threshold for *tall*. However, what you just said is not felicitous in this context. You really have to say "please give me the taller glass."

For defenders of the coordination account who deny that the idealized listener would know the speaker's intention in a Crisp-Judgment context, the challenge is to explain why (47) nevertheless seems felicitous. Of course, the critical issue is whether the knowledge claim should be taken at face value. However, the burden of proof is on defenders of the coordination account if they want to claim that (47) should not be taken at face value. They need to spell out the relevant notion of knowledge in the definition of the coordination account and explain why it is different from the knowledge claim in (47).

3.3.3 Lassiter & Goodman's (2013, 2015) probabilistic model

As mentioned in the previous section, Lassiter and Goodman (2013, 2015) independently developed a probabilistic model of the descriptive uses of gradable adjectives in predicative positions, which can be seen as a formalization of King's (2014a) coordination account and has some further good properties. Below I discuss their proposal in detail. I will show that while the functional considerations in their proposal, i.e., how interlocutors use language to best serve their communicative goals can provide insights into the properties of gradable adjectives, their analysis cannot be straightforwardly extended to account for uses of gradable adjectives in referential contexts.

Like King (2014a), Lassiter and Goodman (2013, 2015) use a threshold-as-free-variable semantics for relative adjectives (48a). Furthermore, they treat maximum adjectives in a parallel fashion (48b).

(48) a. $[\![\text{John is } \textit{pos} \text{ tall}]\!]_{st} = \lambda i.\textbf{height}_{w_i}(\textbf{j}) \geq \theta$
$[\![\text{John is } \textit{pos} \text{ tall}]\!]^{\theta} = \lambda w.\textbf{height}_{w}(\textbf{j}) \geq \theta$

b. $[\![\text{Glass B is } \textit{pos} \text{ full}]\!]_{st} = \lambda i.\textbf{fullness}_{w_i}(\textbf{b}) \geq \theta$
$[\![\text{Glass B is } \textit{pos} \text{ full}]\!]^{\theta} = \lambda w.\textbf{fullness}_{w}(\textbf{b}) \geq \theta$

This provides the basis of defining a *literal listener* L_0, who hears an utterance u and updates his belief about the world w by conditioning on the truth of u, under the assumption that θ is the threshold (49).

(49) $L_0(w \mid u, \theta) \propto \Pr(w) \cdot \delta_{w \in [\![u]\!]^{\theta}}$

a. $\Pr(w)$ is the prior probability of w (listener's prior belief)

b. $\delta_{\phi} = 1$ iff ϕ is true, 0 otherwise

We can then define a speaker S_1, who knows the state of the world w (e.g., John's height) and tries to convey it through an utterance u, which we assume can be either the positive form of the adjective (u_A) or just silence (u_N).

The speaker S_1, under the assumption that θ is the threshold, makes her choice by maximizing the utility, which consists of the informativity of u (measured by $\log(L_0(w \mid u, \theta))$) and its cost (50).

(50) $S_1(u \mid w, \theta) \propto \exp(\alpha \cdot (\log(L_0(w \mid u, \theta) - \text{Cost}(u)))$

a. α is a parameter that controls the level of optimization. If $\alpha = 0$ the choice of utterance is totally random, and when $\alpha \to +\infty$ the speaker always chooses the utterance that strictly maximizes the utility $U(u, w, \theta) = \log(L_0(w \mid u, \theta) - \text{Cost}(u)$

b. Assume a small positive $\text{Cost}(u_A)$, and assume $\text{Cost}(u_N) = 0$

Now we can define a pragmatic listener $L_1(w, \theta \mid u)$, who hears an utterance u and jointly infers the world w the speaker is trying to convey and the threshold θ she is using.

(51) $L_1(w, \theta \mid u) \propto \Pr(w) \cdot \Pr(\theta) \cdot S_1(u \mid w, \theta)$

a. $\Pr(w)$ is the prior probability of w (listener's prior belief)

b. $\Pr(\theta)$ is a uniform prior over the range of degrees (listener does not have any prior preference for a degree to be the threshold)

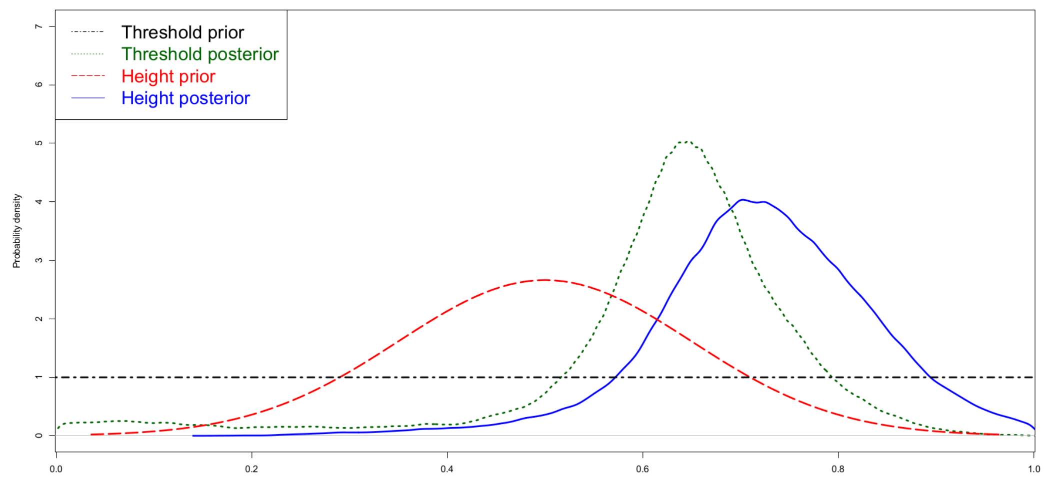


Figure 3.1: Lassiter & Goodman's (2013) pragmatic listener; *John is pos tall*

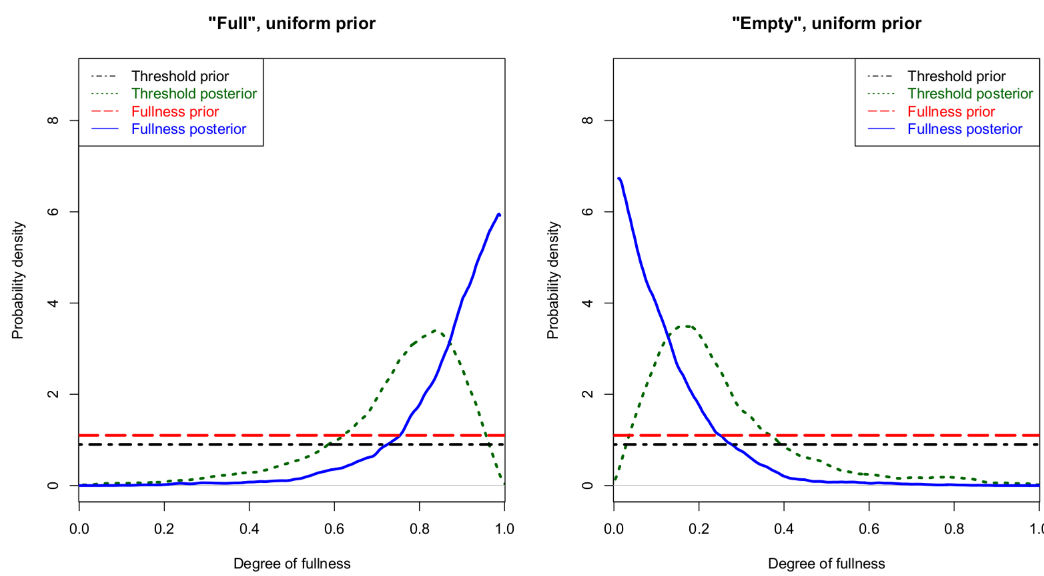


Figure 3.2: Lassiter & Goodman's (2013) pragmatic listener; *Glass B is pos full/empty*

Let us first consider the descriptive use of a gradable adjective A in the predicative position (e.g., *John is tall*) and see a couple of examples of the predictions made by this model. In this case, the relevant aspect encoded in the possible world is the degree the individual possesses. Assuming a Gaussian prior of height and a uniform prior over degrees of fullness, with reasonable choice of the α and cost parameters, Lassiter and Goodman (2013) predict the interpretations of *John is pos tall* and *Glass B is pos full/empty* in Fig. 3.1 and Fig. 3.2. For relative adjectives such as *tall*, the threshold posterior (green) is shifted from the degree prior (red) to the higher end of the scale, but not too much, i.e., the pragmatic listener believes the threshold that the speaker was using is most likely a little above the mean of the prior degree distribution, which is intuitively correct (e.g., the threshold for *tall* is a little above the average height). In contrast, for maximum adjectives such as *full*, the threshold posterior (green) is more concentrated around the higher end of the scale (around 85%). Lassiter and Goodman (2013) argue that this explains why maximum adjectives tend to have maximum standards (threshold posterior is closer to the upper endpoint) and why they are less vague than relative adjectives (threshold posterior has lower variance).

Lassiter & Goodman's free-variable model overcomes some of the limitations of Kennedy's and King's analyses discussed earlier. First, it specifies how contextual degree distribution affects the interpretation of relative adjectives. As a result, the inference from *John is tall for a professional basketball player* to *John is tall for a male adult* is derivable from the model. Once we input our world knowledge about the height distributions for professional basketball players and male adults, the model automatically predicts that the standard of tall for professional basketball players is generally greater than the standard of tall for male adults. This better addresses the first theoretical issue (**CC-sensitivity**) than Kennedy's and King's analyses.

Moreover, the model is conceptually appealing in that it unifies the interpretation of both relative and maximum adjectives based on general principles of communication and reasoning. There is no need to stipulate a context-independent interpretation for maximum adjectives. The difference between relative and maximum adjectives follow from their different degree priors.[23]

Since Lassiter & Goodman's model can be seen as a formalization of King's coordination account, it is worth spelling out the core assumptions about the idealized pragmatic listener. Essentially, the model makes two assumptions about the pragmatic listener. First, the pragmatic listener assumes that the speaker's intended threshold is consistent with the constraints imposed by the conventional meaning of the sentence. For instance, if the speaker utters *John is tall*, the intended threshold must not exceed John's height, for otherwise the utterance would be false. Second, the pragmatic listener

[23] However, I should note that Lassiter and Goodman (2013) do not end up deriving a qualitative difference between relative and maximum adjectives. For instance, for the adjective *full*, the standard that the pragmatic listener infers is most likely close to the maximum rather than at the maximum, and this result can be quite sensitive to the values of the rationality and cost parameters. In light of the difficulties of extending their model to referential cases to be discussed below, I will set aside the issue of whether this really captures the differences between relative and maximum adjectives in a satisfactory way, and refer interested readers to some further discussions of Lassiter & Goodman's model in Qing & Franke, 2014.

assumes that the speaker intends a threshold that is good for the communicative goal. In descriptive uses of gradable adjectives in predicative positions, this means that the speaker would prefer a high threshold (as long as it still makes the utterance true) because it would make the utterance more informative.

However, as I will show below, under such assumptions, this free-variable model cannot be straightforwardly extended to account for referential uses of gradable adjectives in definite descriptions (which Lassiter and Goodman did not consider in their papers).

We will once again consider the TWO-GLASSES context. Under the free-variable analysis, the denotations of *the pos tall/full glass* are in (52a) and can be relativized to the commonly known actual world (omitted) and a free threshold variable θ as in (52b).

(52) a. $[\![\text{the } pos \text{ tall glass}]\!]_{se^?} = \lambda i.\iota(\lambda x.\textbf{glass}_{w_i}(x) \wedge \textbf{height}_{w_i}(x) \geq \theta)$
$[\![\text{the } pos \text{ full glass}]\!]_{se^?} = \lambda i.\iota(\lambda x.\textbf{glass}_{w_i}(x) \wedge \textbf{fullness}_{w_i}(x) \geq \theta)$

b. $[\![\text{the } pos \text{ tall glass}]\!]^{\theta} = \iota(\lambda x.\textbf{glass}(x) \wedge \textbf{height}(x) \geq \theta)$
$[\![\text{the } pos \text{ full glass}]\!]^{\theta} = \iota(\lambda x.\textbf{glass}(x) \wedge \textbf{fullness}(x) \geq \theta)$

In this referential context, the communicative goal is to identify the intended referent r of a definite description u_A (of the form *the pos A N*).

Parallel to (51), we can define a pragmatic listener $L_1(r, \theta \mid u)$ in (53).

(53) $L_1(r, \theta \mid u_A) \propto \Pr(r) \cdot \Pr(\theta) \cdot S_1(u_A \mid r, \theta)$

a. $\Pr(r)$ is the prior probability of r (listener's prior belief about the referent), which is assumed to be uniform

b. $\Pr(\theta)$ is a uniform prior over the range of degrees (listener does not have a prior preference for a degree to be the threshold)

We then need to define a speaker $S_1(u \mid r, \theta)$, who has an intended referent r and tries to convey it by using an utterance u_A that contains a definite description with the positive form of the adjective (e.g., *give me the tall/full glass*) to pick out the intended referent, under the assumption that θ is the threshold.

Given the semantics in (52b), we know that *the tall/full glass* denotes glass B iff $d_A < \theta \leq d_B$, and undefined otherwise. Therefore the utterance u_A *give me the tall/full glass* picks out glass B as the referent iff $d_A < \theta \leq d_B$ and is infelicitous otherwise. As a result, we can conclude the following for a rational, competent speaker S_1 in this scenario (54).

(54) a. For any θ, $S_1(u_A \mid g_A, \theta) = 0$, because the utterance u_A never picks out glass A g_A as the referent

b. $S_1(u_A \mid g_B, \theta) = 0$ when $\theta \notin (d_A, d_B]$, because the utterance u_A is infelicitous when $\theta \notin (d_A, d_B]$

Combining (53) and (54), the model makes the following predictions (55).

(55) a. For any θ, $L_1(g_A, \theta \mid u_A) = 0$, and therefore the marginal probability $L_1(g_A \mid u_A) = 0$

b. For any $\theta \notin (d_A, d_B]$, $L_1(g_B, \theta \mid u_A) = 0$

c. The marginal probability $L_1(g_B \mid u_A) = 1 - L_1(g_A \mid u_A) = 1$

In sum, after hearing *give me the tall/full glass*, the pragmatic listener in this free-variable model concludes that (i) the intended referent is the taller/fuller glass (glass B), and that (ii) the threshold θ used by the speaker falls within the interval $(d_A, d_B]$. In particular, these conclusions hold regardless of whether the glasses are judged to be tall/full in isolation or how different their heights/degrees of fullness are.

For relative adjectives and scenarios without crisp-judgment effects, the model correctly predicts that the pragmatic listener would accommodate the presupposition of the definite description by shifting the standard. However, the problem is that the model predicts presupposition accommodation *across the board*, regardless of the adjective class and the degree difference between the two objects. Therefore it fails to account for maximum adjectives or the crisp-judgment effect.

The reason for the failure, I suggest, is the assumption that the threshold for the adjective is a free variable, i.e., the speaker is totally free to choose any threshold to best serve the communicative goal at hand, and that the listener's task is to infer such a threshold together with the speaker's communicative intent, assuming that the speaker is rational in the choice of the threshold. As a result, the inference about the free variable θ in referential uses has nothing to do with the descriptive uses, which should be a fundamental prediction made by the free-variable approach. Indeed, if θ is truly a free variable that can be arbitrarily set by speakers to serve whatever communicative goal at hand, then the dissociation between descriptive and referential uses would be expected.

However, this does not seem to be the case empirically. In order to use *the tall/full glass* felicitously to refer to the taller/fuller glass, the descriptive counterpart *glass B is tall/full* and *glass A is not tall/full* need to be true in some sense. The challenge is to spell out what this sense is and capture the relation between descriptive and referential uses. In the next section, I review delineation-based approaches, which will motivate the assumptions in my proposal that address this challenge.

3.3.4 Delineation-based approaches

Another highly influential framework for analyzing gradable adjectives is Delineation Semantics (DelS) (e.g., Klein, 1980; van Benthem, 1982; van Rooij, 2011a, 2011b; Burnett, 2014), according to which gradable adjectives denote individual properties that are relativized to contextual comparison classes.

Formally, the contextual variable is a comparison class K, which is a set of individuals. Given a comparison class K, a gradable adjective such as *tall* divides it into up to three subsets: the set of individuals in the positive extension (i.e., those who are definitely tall), the set of individuals in the negative extension (i.e., those who are definitely not tall), and the set of individuals in neither (i.e., borderline cases).

(56) a. $[\![\text{tall}]\!]_{est^?} = \lambda x \lambda i.\mathbf{tall}_{w_i}(K_i)(x)$

b. $[\![\text{John is tall}]\!]^{w,K} = \mathbf{tall}_w(K)(x)$

One nice feature of this approach is that positive forms can be straightforwardly composed with the subject to yield the proper truth conditions, without the need to postulate any additional silent material such as *pos* in degree-based approaches. The meaning of comparatives can also be derived compositionally. The basic idea is that *John is taller than Mary* is true iff there exists a comparison class in which John is tall but Mary is not (57).

(57) a. $[\![\text{-er}]\!]_{est^? \to eest} = \lambda F \lambda y \lambda x \lambda i.\exists K(F(x)(\langle w_i, K\rangle) \wedge \neg F(y)(\langle w_i, K\rangle))$

b. $[\![\text{taller}]\!]_{eest} = \lambda y \lambda x \lambda i.\exists K(\mathbf{tall}_{w_i}(K)(x) \wedge \neg \mathbf{tall}_{w_i}(K)(y))$

c. $[\![\text{John is taller than Mary}]\!] = \lambda i.\exists K(\mathbf{tall}_{w_i}(K)(\mathbf{j}) \wedge \neg \mathbf{tall}_{w_i}(K)(\mathbf{m}))$

In order to ensure that the truth conditions for comparatives accord with our intuitions about ordering and comparisons, additional constraints need to be imposed on the interpretations of positive forms. The most commonly used ones are due to van Benthem (1982) (58).

(58) a. No Reversal: If x is tall in comparison class K and y is not, then it is impossible to find another comparison class K' in which y is tall and x is not.

b. Upward Difference: If x is tall in comparison class K and y is not, then for any larger comparison class $K' \supseteq K$, there exist x' and y' such that x' is tall in comparison class K' and y' is not.

c. Downward Difference: If x is tall in comparison class K and y is not, then for any smaller comparison class $K' \subseteq K$, there exist x' and y' such that x' is tall in comparison class K' and y' is not.

When the three constraints in (58) are met, van Benthem (1982) proves that the ordering relation *is taller than* derived from the compositional semantics of comparatives in (57) is a strict weak order.[24]

Early delineation-based approaches only consider relative gradable adjectives. Given that maximum adjectives are also gradable, it is natural to expect a parallel treatment of, e.g., *full*, in (59).

(59) a. $[\![\text{full}]\!]_{est^?} = \lambda x \lambda i.\mathbf{full}_{w_i}(K_i)(x)$

b. $[\![\text{Glass B is full}]\!]^{w,K} = \mathbf{full}_w(K)(\mathbf{b})$

[24] A strict weak order R is a binary relation that is (i) irreflexive, i.e., $\forall x.\neg xRx$, (ii) transitive, i.e., $\forall xyz$ if xRy and yRz then xRz, and (iii) almost-connected, i.e., $\forall xyz$, if xRy then xRz or zRy.

However, this quickly runs into trouble when we consider the comparative construction, whose denotation is predicted to be (60).

(60) $[\![\text{Glass B is fuller than glass A}]\!] = \lambda i.\exists K(\mathbf{full}_{w_i}(K)(\mathbf{b}) \wedge \neg\mathbf{full}_{w_i}(K)(\mathbf{a}))$

Here is the problem. If glass A is almost empty, and glass B is half full, then the comparative construction (60) is true, which means that there is a comparison class K in which glass B is full (and glass A is not). However, this seems like a problematic prediction. Given that glass B is only half full, intuitively it would never count as full in any comparison class!

In order to fix this problem, van Rooij (2011b) proposes that the positive form of an absolute adjective must always be interpreted wrt the whole domain. However, as Burnett (2014) points out, the problem is fixed "by essentially severing the link between the actual use of the positive form (which is comparison-class independent) and the construction of the comparative relation (which proceeds through comparison-class variation)," and therefore this aspect of his proposal "undercuts one of the more interesting and distinctive hypotheses of DelS, namely that the use of the comparative form should be a function of the use of the positive form."[25]

Another issue arises when we consider the crisp-judgment effect, i.e., if there are two glasses A and B whose heights are extremely close (and B slightly taller), then it is felicitous to say (61a) but not so for (61b).

(61) a. Glass B is taller than glass A.

b. ? Compared to glass A, glass B is tall.

However, given that the comparative sentence (61a) is true, the delineation semantics predicts that there exists a comparison class K such that glass B is tall and glass A is not. By Downward Difference, in the 2-element comparison class K_{AB} that contains just A and B, one glass is tall and the other is not. By No Reversal, it has to be that glass B is tall K_{AB} and glass A is not. But then what is wrong with (61b)?

To address this issue, van Rooij (2011a) proposes that we abandon the assumption that any subset of the domain individuals can serve as a comparison class. In particular, he suggests that if the 2-element comparison class K_{AB} is not an admissible comparison class if the degree difference

[25] Burnett (2014) proposes her own delineation-based analysis, which crucially builds on imprecise uses of absolute gradable adjectives. However, it seems to me that her proposal is not completely immune to her own criticism. In her analysis, in order for *glass B is fuller than glass A* to be true, there must be a context in which glass B is tolerantly full (and glass A is not). However, suppose that glass B is 10% full and glass A is completely empty. It seems highly implausible to assume that a 10% full glass will be loosely considered full in some context. Of course, one could respond that such a context can exist in principle no matter how implausible or unlikely. But then this still seems to sever the link between the actual use of the positive form (which hardly ever allows for this much imprecision) and the comparative construction.

between A and B are too small, and this is what accounts for the infelicity of (61b).[26]

What about the similar contrast between (62a) and (62b) in referring definite descriptions, which van Rooij (2011a) does not discuss?

(62) a. Give me the taller glass.

b. ? Give me the tall glass.

The main difference is that unlike (61b), the comparison class is not explicit in (62b). Therefore, we need to first make an assumption about what comparison classes are admissible in this case (63).

(63) In evaluating the positive form in (62b) in the context with two glasses A and B, a comparison class is admissible only if (i) it is a comparison class for general descriptive uses (e.g., the domain of glasses/artifacts) or (ii) it is the 2-element comparison class K_{AB} containing only the objects in the immediate referential context (i.e., glasses A and B) and its two elements are different enough.

Here is why we cannot allow any arbitrary set to be the comparison class. Given that glass B is taller than glass A, the delineation semantics predicts that there exists a comparison class K with respect to which glass B is tall and glass A is not. However, this means that glass B is the only glass among the two glasses that is tall wrt this comparison class K and as a result (62b) would be wrongly predicted to be felicitous if K were an admissible comparison class. Therefore it is not enough to just make comparison class K_{AB} inadmissible when the degree difference between the two objects are too small. Given that we have little control over what K might turn out to be, it seems that the more principled approach would be to specify what comparison classes are admissible, and hence the assumption in (63).[27]

Given this assumption about admissible comparison classes, in crisp-judgment contexts, the comparison class K_{AB} is not admissible because the two glasses are not different enough in height, and therefore the gradable adjective can only be interpreted wrt a comparison class for general descriptive uses (e.g., the domain of glasses/artifacts). When there is not a unique glass that is tall under such general interpretations (which is highly likely given that the heights of the two glasses are very close), the definite description is infelicitous. In this way, the crisp-judgment effect is accounted for.

[26] Strictly speaking, this is one of the analyses (the pragmatic solution) he proposes. He also proposes a semantic solution, according to which neither A nor B is tall in the 2-element comparison class K_{AB}, and shows how van Benthem's constraints can be modified to achieve comparable results. My criticism to this semantic solution is essentially the same as the one to the pragmatic solution that I will discuss below.

[27] Note that (63) really is just a minor modification of what already is generally assumed (explicitly or implicitly) by delineation approaches. If (62b) is uttered in a context with two glasses A and B, delineation approaches would generally simply take the relevant comparison class to be K_{AB}, and clause (ii) in (63) just adds the requirement that the two elements are different enough to capture the crisp-judgment effect.

To sum up the discussion so far, classic delineation semantics of gradable adjectives faces problems with absolute adjectives and the crisp-judgment effect, and one way to address such problems is by imposing constraints on admissible comparison classes. However, such constraints are directly stipulated and lack independent motivations. For maximum adjectives, the requirement that they always be interpreted relative to the whole domain undercuts the delineation approaches' thesis that comparatives are derivable from the use of the positive forms. Similarly, the crisp-judgment effect is directly stipulated rather than derived. Indeed, van Rooij's focus is on how the rest of a delineation analysis needs to be modified once we incorporate the crisp-judgment effect. The resulting system does not explain why there should be a crisp-judgment effect in the first place: it just assumes that there is and treats it as a primitive. These limitations suggest that the constraints on admissible comparison classes should ideally be derived by independently motivated assumptions rather than stipulated. In the next section, I suggest that this can be done within a degree-based framework, with a contextual parameter that corresponds to a generalized notion of comparison classes.

Besides the limitations above, existing delineation approaches also say little about exactly how contextual comparison classes affect the interpretation of relative adjectives. Indeed, the constraints on the interpretations of positive forms across different comparison classes, such as the standard ones (58), are fairly weak. This is taken to be a feature of such accounts because it allows for a wide range of possible interpretations of a gradable adjective. However, the problem is that such constraints are allowing too much variability. As a result, it is very difficult for such approaches to derive or explain the plausible inference from *John is tall for a professional basketball player* to *John is tall for a male adult*.

To better illustrate the problem, let us consider a related but simpler example. Suppose there are 4 sticks whose lengths satisfy $l_1 < l_2 < l_3 < l_4$. Then *long* can be in principle interpreted in the following way. With respect to the comparison class $\{s_2, s_3, s_4\}$, both s_3 and s_4 are long, but with respect to the comparison class $\{s_1, s_3, s_4\}$, only s_4 is long.[28] However, this pattern seems highly implausible: given that s_3 is considered long in $\{s_2, s_3, s_4\}$, why would it not be long when the shortest member in the comparison class is replaced by something even shorter? This suggests that the constraints in (58) are not enough and the current system is over-generating interpretation patterns. While it is in principle possible to add more constraints to rule out such implausible patterns, it is not obvious how to do so in a general way so that we have a set of constraints that fully characterize possible contextual variability. In contrast, such patterns can be easily ruled out in degree-based approaches, if we assume that the standard of comparison is influenced by the average degree.

To be clear, I am not claiming that this problem cannot be solved within the delineation semantics framework. After all, degree-based and delineation-based approaches can by and large be translated

[28] The full meaning specification is as follows: $\mathbf{long}(\{s_1, s_2, s_3, s_4\}) = \{s_3, s_4\}$, $\mathbf{long}(\{s_2, s_3, s_4\}) = \{s_3, s_4\}$, $\mathbf{long}(\{s_1, s_3, s_4\}) = \{s_4\}$, $\mathbf{long}(\{s_1, s_2, s_4\}) = \{s_4\}$, $\mathbf{long}(\{s_1, s_2, s_3\}) = \{s_3\}$, and for any $x < y$, $\mathbf{long}(\{s_x, s_y\}) = \{s_y\}$. It is easy to verify that all of the constraints in (58) are satisfied and the derived ordering is indeed $l_1 < l_2 < l_3 < l_4$.

from one to the other. However, from the discussion above, it seems that degree-based approaches directly provide insights into the interpretations of positive forms of gradable adjectives, which are not obvious from the perspectives of delineation-based approaches alone. For this reason, my proposal in the next section is stated within a degree-based framework.

3.4 Semantic contextual resolution: Optimal threshold model

In this section I propose my analysis of the interpretation of gradable adjectives, covering both their descriptive and referential uses.

Here is a quick overview of my analysis, which is related to the approaches discussed above in significant ways. It can be seen as a further specification of the contextual parameter in Kennedy's (2007) analysis and a formalization of the notion of "standing out" in context. Similar to Lassiter and Goodman's (2013) analysis, I propose a unified semantics for both relative and maximum adjectives, and take into account the communicative goal of using gradable adjectives descriptively (although the details differ). Finally, I impose constraints on the contextual parameter similar to those in the delineation approach, and show that they can be better motivated and therefore are more explanatory.

My proposal has three major components. First, I generalize the traditional comparison classes, which are sets of individuals, to *comparison distributions*, which are probability distributions over degrees, and use them as the contextual parameter in the semantics of gradable adjectives. Second, I propose constraints on admissible comparison distributions that allow us to account for the differences between maximum and relative adjectives, as well as their properties in definite descriptions in referential contexts. Third, the precise formal semantics that outputs the threshold given a contextual comparison distribution is motivated by Lassiter and Goodman's (2013) analysis and similarly takes into account the interlocutors' communicative goal when they use gradable adjectives.

The second and the third components of the proposal are to some extent independent. Therefore, below I will first introduce a simplified semantics in Section 3.4.1. This simplified semantics highlights the main intuition behind the analysis while abstracting away from the technical details of the third component. I discuss how this analysis accounts for the basic difference between maximum and relative adjectives Section 3.4.2. Then I discuss the role of the second component in analyzing gradable adjectives in definite descriptions in Section 3.4.3. The full version of the semantics is introduced in Section 3.4.4.

3.4.1 Comparison distribution as the contextual parameter

The proposed semantics is closely related to Bartsch and Vennemann's (1972) proposal, according to which the threshold is computed relative to a contextual comparison class K using some **norm** function (often assumed to be the average) (64).

(64) a. $[\![pos]\!]_{esd \to est} = \lambda g_{esd} \lambda x \lambda i.\, g(x)(i) \geq \mathbf{norm}_{w_i}(g)(K_i)$

b. $[\![pos \text{ tall}]\!]_{est} = \lambda x \lambda i.\mathbf{height}_{w_i}(x) \geq \mathbf{norm}_{w_i}([\![\text{tall}]\!])(K_i)$

c. $\mathbf{norm}_{w_i}([\![\text{tall}]\!])(K_i)$ returns the average height of the individuals in K_i in world w_i

This account, while intuitively plausible, is rejected by Kennedy (2007). One main objection, which he attributes to Bogusławski (1975), is that (64) does not derive the correct truth conditions. For instance, this account predicts that (65) should be a contradiction, contrary to fact.

(65) Nadia's height is greater than the average height of a gymnast, but she is still not tall for a gymnast.

This means that standards are not averages. What are they, then?

My proposal is very similar to (64). However, one crucial difference is that, instead of using a *comparison class* K, which is a set of individuals, I propose that the threshold is computed relative to a *comparison distribution* κ, which is a probability distribution over degrees (I will motivate this move later). The threshold θ is determined by applying an operator **Opt** to this contextual parameter κ (66).

(66) a. $[\![pos]\!]_{esd \to est} = \lambda g_{esd} \lambda x \lambda i.\, g(x)(i) \geq \mathbf{Opt}(\kappa_i)$

b. $[\![pos \text{ tall}]\!]_{est} = \lambda x \lambda i.\mathbf{height}_{w_i}(x) \geq \mathbf{Opt}(\kappa_i)$

c. **Opt** takes a probability distribution κ_i over degrees and returns the sum of its *mean* (i.e., average) μ and *standard deviation* σ.[29]

d. $[\![pos \text{ tall}]\!]^{\kappa} = \lambda x \lambda w.\mathbf{height}_w(x) \geq \mu_\kappa + \sigma_\kappa$

The second crucial difference is that the threshold depends not only on the mean/average of the comparison distribution κ, but also on its standard deviation (in the simplified version above, the threshold is simply the sum of the two). To help illustrate the intuition behind this second move, let us consider two scenarios. In the first, we live in a society where the average height of adult males is $5'9''$ and the heights are concentrated between $5'6''$ and $6'$, and in the second, the average height is also $5'9''$ but the heights are concentrated between $5'2''$ and $6'4''$. Intuitively, it seems that a man who is $5'11''$ could be plausibly considered tall in the first scenario but less so in the second, even though the average height is the same in both scenarios. This suggests that the standard of a gradable adjective should be sensitive to not only the average but also how dispersed or spread out the degrees are from the average. Note that dispersion should not be seen as measuring the height difference between the tallest and shortest people, i.e., the *range* of heights. Rather, it measures the level of concentration of heights around the average/mean. This motivates the use of the standard deviation σ, which is a common measure of dispersion of a distribution.

This observation is not new. Solt (2011) essentially makes the same point. The simplified semantics proposed here is very similar to her proposal, with two differences. The first difference

[29] For now we only consider adjectives such as *tall*, whose scales do not have an upper bound.

is that I use mean and standard deviation instead of the *median* and *median absolute deviation* in Solt's proposal. This is just a simplified version of the semantics that I will propose in Section 3.4.4. As will become clear there, this difference is not crucial. The second difference is more important. I assume that the contextual parameter is a probability distribution, whereas Solt assumes that *pos* takes a comparison class as a semantic argument, which can be introduced by a *for*-PP. Crucially, later in this section I will argue that even if Solt is correct in that a *for*-PP is a syntactic argument of *pos* and provides a set as a semantic argument of *pos* to derive a presupposition, we should not conclude that this set is what is directly responsible for the calculation of the threshold. I will provide evidence for why the semantics proposed above should be preferred as far as the calculation of the threshold is concerned.

But before I dive into that, let us first see how Kennedy's objection is avoided. Given the new semantics, (65) is no longer a contradiction. Moreover, the new analysis automatically accounts for the inconsistency of (67), the flip side of (65).

(67) # Nadia is tall for a gymnast, but her height is no greater than the average height of a gymnast.

Together, (65) and (67) suggest that while being tall does not mean the same as taller than average, the former entails the latter. This asymmetric entailment relation is a theorem according to our formal semantics, since the standard deviation is by definition greater than zero.

The proposed semantics can be seen as a way to formalize Kennedy's (2007) notion of "standing out in a context," or similar ones in the literature such as Graff's (2000) notion of "significantly greater than the norm," as "at least one standard deviation above the mean of the contextual comparison distribution." Below I further motivate such a formalization and address some related issues.

First, one might wonder whether we really gain anything by formalizing Kennedy's notion of "standing out." Sure, we have seen that the proposed semantics can account for (67), but so can Kennedy's account, according to which *Nadia is tall for a gymnast* means her height "stands out" among the heights of gymnasts. Intuitively, if a height "stands out" among some heights, it should at least exceed the average of those heights, and therefore (67) is a contradiction. Note that this explanation crucially relies on our intuitive understanding of the notion "standing out," which is not part of the formal model (or trivially part of it in the form of a meaning postulate). In contrast, the explanation provided by my proposal is a formally derivable theorem. Therefore, the question is essentially whether or why a formally derivable explanation should be preferred to one that simply appeals to our intuitive understanding of the relevant concept.

But this is a very general issue, and in fact one major goal of the enterprise of formal semantics is to provide formally derivable explanations. For example, consider the study of modals, and more specifically the fact that (68a) entails (68b).

(68) a. John must be in the meeting.

b. John might be in the meeting.

How do we explain such an entailment pattern? In principle, one could provide the following semantics for *must* and *might* (69).

(69) a. $[\![\text{must}]\!] = \lambda p \lambda i.\mathbf{Necessity}(c_i)(w_i)(p)$

b. $[\![\text{might}]\!] = \lambda p \lambda i.\mathbf{Possibility}(c_i)(w_i)(p)$

According to (69), *must p* is true iff p is a necessity (the type of which is contextually determined and can be epistemic, deontic, etc.) at the world of evaluation. Similarly, *might p* is true iff p is a possibility at the world of evaluation. Given our intuitive understanding of the concepts of necessity and possibility, we know that if p is a necessity of some type then it is also a possibility of the same type. This explains why (68a) entails (68b).

Now consider a Kripean semantics (70).

(70) a. $[\![\text{must}]\!] = \lambda p \lambda i.\forall w \in f_i(w_i).p(w) = 1$

b. $[\![\text{might}]\!] = \lambda p \lambda i.\exists w \in f_i(w_i).p(w) = 1$

According to (70), *must p* is true iff p is true in all of the worlds that are contextually accessible from the world of evaluation (as given by $f_i(w_i)$). Similarly, *might p* is true iff p is true in some of the worlds that are contextually accessible from the world of evaluation. Assuming seriality, i.e., $f_i(w_i)$ is never empty, we can explain why (68a) entails (68b) in terms of the logical properties of the universal and existential quantifiers.

Presumably, most of us would not find the semantics in (69) and the corresponding explanation of the entailment pattern satisfactory, and would prefer the Kripean semantics (70). Why?

One reason is that the Kripean semantics distills the crucial logical properties shared by various types of necessity/possibility in terms of quantificational force, and therefore provides us a deeper understanding of the core meaning of the modal expressions and a deeper explanation of why (68a) entails (68b) (Kratzer, 1977). In general, our semantic theorizing should aim to provide such deeper explanations (Napoletano, 2019).

Another reason is that the Kripean semantics is precise enough for us to explore its potential limitations. As is well known, the Kripean semantics is in fact not the correct semantics for *must* (Kratzer, 1977), and can be improved by introducing *ordering sources* (Kratzer, 1981). For modals in general, one may further refine the Kripean semantics, e.g., by allowing for multiple ordering sources (von Fintel & Iatridou, 2008), or propose a more radically different framework (Lassiter, 2017). Crucially, we would not be able to make any such progress if we were satisfied with the semantics in (69), whose correctness is extremely difficult (if possible at all) to dispute. The moral, then, is that a precise theory, even if it turns out to be incorrect, may nevertheless be useful in that it can serve as the basis for further refinements or help motivate the need for a more radically different approach.

I take the above reasons to be methodological motivations for the proposed semantics of gradable adjectives (66). To further illustrate that the proposed semantics indeed captures the crucial logical properties shared by various senses of "standing out" or "significantly greater than the norm," consider the various kinds of norms discussed by Graff (2000). According to Graff, a norm could be about (i) what is typical, e.g., (71a) can be true in the sense that there is significantly more bear than one typically finds in the fridge, even when the speaker knows that the beer is still not enough for the large party they are planning tonight, (ii) what is wanted or needed, e.g., (71b) can be true in the sense that the stock is significantly more than what is wanted or needed by shoppers on any given day, even though the stock is actually typical and expected for a supermarket, or (iii) what is expected, e.g., (71c) can be true in the sense that there is significantly more corn than the speaker expected to come upon at the moment, even when the speaker is driving in Iowa, where corn fields are typical. (And this list is not exhaustive.)

(71) a. There is a lot of beer in the fridge!

b. Our local supermarket has a lot of milk in stock.

c. That's a lot of corn!

In each case, we can find a plausible comparison distribution. For (71a) it is the probability distribution over the amount of beer that one actually finds in the fridge (over a period of time). For (71b) it is the distribution over the amount of milk that shoppers want or need on any given day. For (71c) it is the distribution over the amount of corn the speaker expects to come upon at the moment. What these examples have in common is that the shape of the comparison distribution (more specifically, the mean and standard deviation) determines their truth conditions, and their differences are due to the different ways the comparison distributions are constructed (just like modals can have various flavors such as *metaphysical*, *bouletic*, and *doxastic* depending on the way the modal base and ordering source are chosen).

Moreover, the proposed semantics can account for the following example discussed by Stanley (2003).

(72) Mount Everest is tall for a mountain.

As Stanley (2003) points out, (72) has an interpretation whose truth or falsity does not depend on whether anybody finds Mount Everest's height significantly greater than some norm. Under perhaps its most salient interpretation, (72) is true simply by virtue of Mount Everest being the tallest mountain in the world. In other words, its truth condition can be satisfied based on purely distributional criteria (Kennedy, 2007). Stanley (2003) argues that this is a problem for Graff's (2000) account, according to which the interpretation of a gradable adjective is always relativized to some agent. Regardless of whether this is indeed a problem for Graff's account, (72) presents no problem for my proposal: (72) is true when the comparison distribution is the distribution of heights of actual mountains, which makes no reference to any agent.

Now let us return to an earlier issue: what motivates the move from comparison classes to comparison distributions? Traditionally, comparison classes are taken to be sets of individuals in the world of evaluation. This traditional notion can be generalized so that comparison classes can contain counterparts of individuals (Toledo & Sassoon, 2011), times and situations (Solt, 2011), but crucially such generalized comparison classes are still sets. For brevity, henceforth generalized comparison classes are simply referred to as comparison classes. Note that the notions of mean and standard deviation can be applied to a set of degrees as well as a probability distribution, which means that the **Opt** function in our proposed semantics (66) can be also applied to the set of degrees determined by a comparison class K and a measure function g, i.e., the set of the degrees that the members of K possess $\{g(y) \mid y \in K\}$. Given this, one might naturally wonder whether there is really a need to use comparison distributions instead of comparison classes as the contextual parameter. Moreover, the presuppositionality of some *for*-PPs might seem to be evidence against the use of comparison distributions. Consider the examples below (Solt, 2011).

(73) a. Fred is tall for an 8-year-old — *Presupposition:* Fred is an 8-year-old

b. Sara reads difficult books for an 8-year-old — *Presupposition:* Sara is an 8-year-old

c. The store is crowded for a Tuesday — *Presupposition:* The utterance time is Tuesday

Intuitively, the *for*-PPs in (73) indicate that their denotations are the relevant comparison classes. In addition, they introduce a presupposition that the subject is a member of the set denoted by the *for*-PP. For instance, (73a) introduces a presupposition that Fred is an 8-year-old, as can be shown by Gennaro and McConnell-Ginet's (1990) *family-of-sentences* test in (74).

(74) a. Fred is not tall for an 8-year-old. $\rightsquigarrow$ Fred is an 8-year-old

b. Is Fred tall for an 8-year-old? $\rightsquigarrow$ Fred is an 8-year-old

c. Fred might be tall for an 8-year-old. $\rightsquigarrow$ Fred is an 8-year-old

d. If Fred is tall for an 8-year-old, so is George. $\rightsquigarrow$ Fred is an 8-year-old

Henceforth I will call such presuppositions *membership presuppositions*. Note that the membership presupposition is of the form $x \in P$, where the set P is the denotation of the *for*-PP (henceforth called the *membership class*) and x can be an individual or a time (or perhaps even a situation), which is often, but not always (cf. 73c), the denotation of the subject.

Solt (2011) proposes that we can capture this dual role of the *for*-PP as the standard-setter and the trigger of the membership presupposition by treating it as an argument of *pos*. Abstracting away from differences not crucial to the current discussion, a concrete implementation of her analysis is shown in (75a).[30]

(75) a. $[\![pos]\!]_{\tau t \to \tau d \to \tau st} = \lambda K \lambda g \lambda x : \underline{x \in K}.\ \lambda i.g(x) \geq \textbf{Opt}(\{g(y) \mid y \in K\})$

b. $[\![pos]\!]_{\tau t \to \tau d \to \tau st} = \lambda K \lambda g \lambda x : \underline{x \in K}.\ \lambda i.g(x) \geq \textbf{Opt}(\kappa_i)$

κ_i a contextually determined comparison distribution (which should be appropriate in the context, taking into account the presupposition that $x \in K$)

If (75a) is indeed correct, i.e., the denotation of the *for*-PP is all we need to determine the threshold, then my proposal of using comparison distributions (75b) would be dispreferred due to parsimony, as there is simply no need to introduce a probability distribution as an additional contextual parameter. However, note that this appeal to parsimony crucially relies on the assumption that the membership class, i.e., the denotation of the *for*-PP, is always identical to the comparison class, i.e., the set being used to help determine the threshold.

But this is not the case. For example, (73a) can be interpreted as *Fred is tall for an 8-year-old boy*. That is, the comparison class is the set of 8-year-old boys, which is not the same as the membership class/the denotation of the *for*-PP.[31] One might try to maintain the identity between the membership class and the comparison class by allowing for the denotation of the *for*-PP to be contextually restricted, and point out that in this case the membership presupposition may well be that Fred is an 8-year-old boy. However, this strategy would not work in general. For instance, suppose that the speaker of (73a) is from the US and is meeting his Dutch relatives. It is possible (in fact, quite plausible) that the speaker is using the set of 8-year-old American boys as the comparison class.[32] For example, Fred's parents can reply: "Yes, if you are comparing him with 8-year-old American boys, but actually Fred's height is quite normal for Dutch boys of his age." In this case, the comparison class cannot possibly be identical to the membership class, since Fred is not an American boy.[33] This example shows that, even though the membership class constrains and perhaps even tends to be the comparison class, they are not always identical. Therefore, no matter how we derive the membership presupposition introduced by the *for*-PP, we need an additional contextual parameter for the determination of the threshold. Crucially, the fact that the denotation of the *for*-PP is a set does not provide direct evidence about whether this additional contextual parameter should be

[30] The underlined part highlights the presupposition that $x \in K$.

[31] The height difference between average 8-year-old boys and girls is relatively small, so whether the comparison class is the set of 8-year-old children or the set of 8-year-old boys does not make a big difference. However, for *John is tall for a professional basketball player*, whether the comparison class is the set of all professional basketball players or the set of all male professional basketball players will make a bigger difference.

[32] A related example that illustrates the possibility of using 8-year-old American boys as the comparison class is for the speaker to say *I find Fred (quite) tall for an 8-year-old.*

[33] Note that it would not work to assume that the comparison class is instead the set of 8-year-old boys in the US, because Fred does not need to be in the US (e.g., the conversation can well take place in Australia, where the speaker and Fred's family are attending a wedding), and therefore the comparison class still cannot be identical to the membership class.

a set or a probability distribution. All we know is that the set denoted by the *for*-PP constrains this parameter, but such a constraint can be stated whether this parameter is a set or a probability distribution. If the parameter is a set, then the constraint imposed by *for an 8-year-old* can be that the parameter must be a subset of the set of 8-year-olds.[34] If the parameter is a probability distribution, as in (75b), then the constraint can be that the probability distribution must be such that the height of an 8-year-old can plausibly be seen as a sample from this distribution. Note that this constraint is derived from a general constraint on which comparison distribution(s) would be appropriate in context. In general, a comparison distribution is appropriate in context if the degrees in the intuitive comparison class, i.e., relevant entities under discussion, can be plausibly seen as samples from the comparison distribution. For instance, suppose the contextually relevant entities under discussion are basketball players, then it would not be appropriate to use the height distribution of jockeys as the comparison distribution. In the special case where an overt *for*-PP, e.g., *for an 8-year-old* is used, the membership presupposition it introduces is part of the context, and therefore will similarly constrain the comparison distribution. As a result, only height distributions of 8-year-olds or contextually salient subsets of them are appropriate comparison distributions.

The above discussion sheds new light on the discussion in the literature about whether the *for*-PP should be treated as an argument of *pos*. On the one hand, Kennedy (2007) argues against this treatment, based on the observation that the comparison classes introduced by *for*-PPs and modified nominals (e.g., *elephant* in *a big elephant*) have different presupposition statuses (76). In (76b), the comparison class for *big* can be the set of elephants, i.e., the denotation of the modified nominal. However, as shown by its felicity, (76b) does not presuppose that the subject is an elephant, unlike (76a). If the comparison class is a semantic argument of *pos* that can be provided by any expression of the appropriate type, then this difference between (76a) and (76b) cannot be accounted for.

(76) a. ?? That mouse is (obviously) not big for an elephant.

b. That mouse is (obviously) not a big elephant.

Based on this, Kennedy (2007) argues for a domain-restriction analysis of for-PPs, according to which the role of a for-PP is to restrict the domain of the measure function denoted by the gradable adjective to the denotation of the for-PP (77).

(77) $[\![\text{big for an elephant}]\!] = \lambda x : \underline{\mathbf{elephant}(x)}.\, \mathbf{size}(x)$

On the other hand, Solt (2011) provides semantic considerations against Kennedy's analysis. For example, (73b), repeated below in (78), presupposes that Sara is an 8-year-old. However, this presupposition cannot be captured by restricting the domain of the measure function denoted by the gradable adjective *difficult*.

[34] Note that here we are simply stating the constraint imposed by the *for*-PP per se. There can be general pragmatic constraints at play so that an admissible parameter probably needs to be a salient subset of 8-year-olds such as 8-year-old boys and girls.

(78) Sara reads difficult books for an 8-year-old *Presupposition:* Sara is an 8-year-old

First, there is an obvious obstacle, i.e., the denotation of the for-PP is the set of 8-year-olds, but intuitively the domain of the measure function denoted by *difficult* should be restricted to the set of books read by 8-year-olds. As Solt (2011) points out, even though one can bypass this obstacle by assuming that the domain of the measure function may be restricted to a function of the denotation of the for-PP (79), this is still not enough to capture the presupposition of (78).

(79) $[\![\text{difficult for an 8-year-old}]\!] = \lambda x : \underline{x \in \textbf{read-by}(\{y \mid \textbf{8-year-old}(y)\})}.\textbf{difficulty}(x)$

The domain restriction in (79) only ensures that Sara reads books read by 8-year-olds, but this is not enough to guarantee that Sara herself is an 8-year-old. Essentially, the problem with the domain-restriction analysis is that it analyzes (80a) the same way as (80b), and therefore cannot explain why only (80a) presupposes that Sara is an 8-year-old.[35]

(80) a. Sara reads difficult books for an 8-year-old *Presupposition*: Sara is an 8-year-old

b. Sara reads books (that are) difficult for an 8-year-old *No presupposition*

Solt's (2011) analysis does not suffer from this problem the domain-restriction analysis faces.[36] However, it does not provide a satisfying account of the difference between *for*-PPs and modified nominals and thus does not address Kennedy's (2007) original objection, as Solt (2011) herself acknowledges. Therefore, a tension remains about whether for-PPs and/or comparison classes should be treated as an argument of *pos*.

This tension can be resolved when we separate the dual roles of a *for*-PP. We can assume that *pos* takes a *for*-PP as an optional syntactic argument, and semantically only uses it to introduce the membership presupposition and impose constraints on the contextual parameter that helps determine

[35] Of course, (80a) often implies that Sara is an 8-year-old, but this is arguably not semantically encoded, as can be shown by examples such as *Sara reads papers (that are) difficult for an undergraduate, now that she is in graduate school*.

[36] Note that the above domain-restriction analysis is just Solt's (2011) extrapolation of Kennedy's (2007) analysis. While Kennedy (2007) does take *for*-PPs to be restricting the domain of gradable adjectives, since he does not consider structurally more complex examples like (78), it is possible that his actual position is different from this rather literal construal. For example, I think one reasonable construal is that *for*-PPs are modifiers of measure functions and introduce domain restrictions, but the measure function being modified by a *for*-PP is whatever is in its scope, which need not be the denotation of a gradable adjective. Under this construal, in (78), the *for*-PP modifies the measure function denoted by the VP *reads difficult books*, i.e., $\lambda x.\textbf{difficulty}(\textbf{books-read-by}(x))$, and restricts the domain to the set of 8-year-olds. This also correctly captures the presupposition of (78), i.e., Sara is an 8-year-old. The main difference between this construal and my analysis is that under this construal, the *for*-PP itself modifies the measure function in its scope and the resulting measure function becomes the argument of *pos*, whereas in my analysis the *for*-PP is an argument of *pos*. The only reason I do not adopt this construal is that there is syntactic evidence that *for*-PPs behave like arguments of *pos* rather than adjuncts (Fults, 2006; see also Bylinina, 2014). However, if there is further evidence suggesting otherwise, we can adopt this construal of the domain-restriction analysis, without affecting my proposal about how the standard is calculated based on comparison distributions.

the threshold. A modified nominal is not a syntactic argument of *pos* (since it must be realized as a *for*-PP), and therefore does not introduce the membership presupposition. Crucially, since in my analysis what determines the standard, i.e., the comparison distribution, is a contextual parameter, it is perfectly possible that the denotation of a modified nominal constitutes the intuitive comparison class, even though the modified nominal itself is not a syntactic (or semantic) argument of *pos*. In this respect, my analysis agrees with Kennedy's (2007) in that the comparison class (or whatever plays the standard-setting role) is not an argument of *pos* that can be overtly realized as a *for*-PP or a modified nominal, but still maintain that *for*-PPs are syntactic and semantic arguments of *pos*.

The discussion so far motivates the need for a contextual parameter separated from the membership class to serve the standard-setting role (whether it is a comparison class or a comparison distribution). The main motivation for using a comparison distribution instead of a comparison class as the contextual parameter is that it allows us to account for empirical properties of gradable adjectives in a more natural way. For instance, attributing the observation to Graff (2000), Kennedy (2007) notes that (81) can still be vague even in a context where we know the rent of every single apartment on the street.

(81) A rent of $725 is expensive for an apartment on this street.

This is expected when the contextual parameter is a comparison distribution. Even if we know the rent of every single apartment on this street, there are still multiple probability distributions that are appropriate for generating these actual rents, i.e., the actual rents can be plausibly taken to be samples from such probability distributions. As a result, we will always have uncertainty about the comparison distribution, and this uncertainty can lead to uncertainty about the threshold, which explains the vagueness of (81). This explanation is closest to Graff's (2000) notion of "ignorance of the context" discussed at the end of her paper. However, I think it is in principle also compatible with other approaches to vagueness. For epistemicism (e.g., Williamson, 1994), we can similarly use the uncertainty about the comparison distribution to explain why we do not know the exact threshold. For supervaluationism (e.g., Fine, 1975), we can think of comparison distributions as a way to provide precisifications of the threshold, i.e., instead of taking the set of admissible precisifications as a primitive, we derive it from admissible comparison distributions in context.

In contrast, it is less straightforward to provide a similar explanation if the contextual parameter is a comparison class. First, if we assume that the comparison class is simply the set of actual apartments on this street, our semantics would produce a precise threshold, or a precise range of thresholds if we adopt Solt's (2011) version. Either way, given the assumption that we know the rent of every single apartment on the street, we would have a difficult time accounting for the vagueness of (81) in this context. Second, we might instead assume that the comparison class is always an intensionalized one and therefore we will always have uncertainty about it even in this context where we know every actual apartment. But then we need to decide which possible apartments (in addition to the actual ones) can be included in the comparison class. It seems to me that any plausible criterion would effectively make reference to a rent distribution in some way. Therefore, it would be simpler if we just use the rent distribution as the contextual parameter.

3.4.2 Threshold stability under uncertainty about the comparison distribution

In this section, I generalize the proposed semantics in (66) to gradable adjectives with different scale structures and discuss how it can account for the imperfect correlation between the scale structure and the standard of a gradable adjective and the difference between maximum and relative adjectives in terms of vagueness.

For ease of comparison, I follow Kennedy and McNally (2005) in assuming that degrees are values that are isomorphic to the real numbers between 0 and 1. For simplicity, I assume that the comparison distribution κ is a truncated Gaussian distribution $\mathcal{N}_{[0,1]}(\mu_\kappa, \sigma_\kappa^2)$.[37] Consequently, the definition of **Opt** in (66) is updated to (82) to reflect the fact that the threshold cannot possibly be greater than 1.

(82) $\mathbf{Opt}(\kappa) = \min\{\mu_\kappa + \sigma_\kappa, 1\}$

I focus on (truncated) Gaussian distributions for two main reasons. First, intuitively they can represent (at least approximately) our knowledge about the distribution of degrees for many gradable adjectives. Second, their shapes are completely determined by the mean μ_κ and standard deviation σ_κ parameters, and these parameters are easy to estimate from a set of samples.[38] (The relevance of this will become clear later when I discuss referential uses of gradable adjectives).

The scale structure of a gradable adjective constrains admissible comparison distributions. Here I focus on the maximum endpoint since this is the relevant factor to account for the differences between relative and maximum adjectives.

If the scale does not have a maximum endpoint (i.e., 1 is not part of the scale), then the comparison distribution should approach 0 as the degree approaches 1 (since it is impossible for the degree to be 1). This requirement can be approximated by assuming, e.g., $\mu_\kappa + 3\sigma_\kappa < 1$. Crucially, this means that σ_κ is relatively low for an open-scale adjective. As a result, we know from (82) that for an open-scale adjective, the threshold is $\mu_\kappa + \sigma_\kappa$. Note that this threshold is sensitive to μ_κ and σ_κ. This means that when we have uncertainty about the exact comparison distribution κ, we will also

[37] A truncated Gaussian distribution $\mathcal{N}_{[0,1]}(\mu_\kappa, \sigma_\kappa^2)$ is the result of restricting a Gaussian distribution $\mathcal{N}(\mu_\kappa, \sigma_\kappa^2)$ so that the value of the random variable can only range over $[0, 1]$ instead of the set of all real numbers. In other words, if a random variable X has a Gaussian distribution with mean μ and variance σ^2, then X conditional on $X \in [0, 1]$ has a truncated Gaussian distribution $\mathcal{N}_{[0,1]}(\mu_\kappa, \sigma_\kappa^2)$.

[38] Strictly speaking, σ_κ is not the standard deviation of the truncated Gaussian distribution, but rather the standard deviation of the Gaussian distribution before truncation. Note that what we need in our semantics is some measure of dispersion, which need not be a standard deviation. For brevity and convenience I will continue to call σ_κ the standard deviation parameter.

be uncertain about the exact threshold. In light of the discussion in the previous section, we can see that this accounts for the vagueness of open-scale adjectives.

In contrast, if the scale has a maximum endpoint, there is no general constraint on σ_κ. On the one hand, the comparison distribution can be quite dispersed, i.e., σ_κ is relatively high, such that $\mu_\kappa + \sigma_\kappa > 1$. In this case, we know from (82) that the threshold is the maximum degree 1. In addition, even if we have slight uncertainty about the exact comparison distribution, we can still be sure that the threshold is the maximum degree. This explains why upper-closed-scale adjectives can have maximum standards and such standards are not/less vague, e.g., *this glass is full.* In this case, since we can easily make a glass full to any degree, the comparison distribution is not concentrated in any part of the scale. In other words, the comparison distribution is quite dispersed, i.e., with a relatively high standard deviation. On the other hand, σ_κ can still be relatively low, just like the case of open-scale adjectives discussed above. Therefore upper-closed-scale adjectives can also have vague standards, e.g., *this feather is dark* and *this theater is full for a Thursday afternoon.* In the former case, the comparison distribution is such that pure black or pure white feathers are much rarer than feathers that have medium levels of darkness (brown, grey, etc.) and therefore the distribution is relatively concentrated in the middle of the scale, which means that σ_κ is relatively low. Similarly, in the latter case, the *for*-PP indicates that the comparison distribution is over degrees of fullness of the theater on Thursday afternoons. Given our world knowledge that it is quite rare for theaters to be completely full on Thursday afternoons, the comparison distribution is relatively concentrated in the middle (or maybe even the lower part) of the scale, which again means that σ_κ is relatively low.

In sum, the proposed analysis provides a unified semantics for maximum and relative adjectives. The crucial parameter that accounts for the differences between the two classes in terms of vagueness is the standard deviation σ_κ. When σ_κ is relatively low (so that $\mu_\kappa + \sigma_\kappa < 1$), the adjective has a vague, relative standard. When σ_κ is relatively high (so that $\mu_\kappa + \sigma_\kappa \geq 1$), the adjective has a maximum standard and is not (or less) vague. In this way the imperfect correlation between the scale structure and the standard is captured. Only closed-scale adjectives can have a σ_κ high enough for them to have an absolute maximum standard. Meanwhile, some closed-scale adjectives have a low σ_κ and therefore have a vague relative standard just like open-scale adjectives. So far, the analysis has addressed theoretical issues 1–3 and 6. Comparison classes correspond to comparison distributions, which determine the threshold (**CC-sensitivity**). The level of dispersion of the comparison distribution, formally implemented as the standard deviation, influences the threshold stability under uncertainty about the comparison distribution, which accounts for the differences between maximum and relative adjectives in terms of vagueness (**rel-adj-vague** and **max-adj-nonvague**). The imperfect correlation between the scale structure and the standard of a gradable adjective is also accounted for. An open-scale adjective necessarily has a vague relative standard because the standard deviation of its comparison distribution is relatively low given that the distribution approaches 0 towards the bounds the scale. In contrast, a (upper-)closed-scale adjective may or may not have a comparison distribution with a relatively high standard deviation. If it does, it has a maximum standard, and if it does not, it has a relative standard (**correlation-with-SS**). In Section 3.5.1, I will further discuss factors that influence whether the comparison distribution has a high or low standard deviation.

3.4.3 Referential uses in definite descriptions

Now we turn to referential uses of gradable adjectives in definite descriptions (83).

(83) a. $[\![\text{the } pos \text{ tall glass}]\!]^{\kappa} = \iota(\lambda x.\mathbf{glass}(x) \wedge \mathbf{height}(x) \geq \mathbf{Opt}(\kappa))$

b. $[\![\text{the } pos \text{ full glass}]\!]^{\kappa} = \iota(\lambda x.\mathbf{glass}(x) \wedge \mathbf{fullness}(x) \geq \mathbf{Opt}(\kappa))$

The main task is to specify which comparison distributions can be used in such referential contexts when evaluating the positive form.

I propose the following constraints on admissible comparison distributions (84), drawing insights from van Rooij's (2011a, 2011b) delineation approach (cf. 63). I will first explain their effects and then discuss their motivations. Also note that these are only necessary conditions and there can be additional constraints.

(84) In evaluating a positive form in a definite description in a referential context, a comparison distribution is admissible only if (i) it is a comparison distribution for general descriptive uses, e.g., the height distribution for glasses/artifacts (henceforth called a *generally admissible comparison distribution*), or (ii) it is a comparison distribution κ suitable for generating the set of objects in the immediate referential context and not too different from the generally admissible comparison distributions in (i).

For concreteness, consider the Two-Glasses context. Part (i) in (84) allows for the positive form to be interpreted in general, i.e., *give me the tall/full glass* can be interpreted as *give me the glass that is tall/full in some general sense* (e.g., tall for a glass/an artifact). Part (ii) allows for the positive form to be interpreted in a way that is specific to the current referential context. Let us go through its details.

It helps to start by considering the counterpart of part (ii) in the delineation approach (63), where the relevant contextual parameter is the comparison class K_{AB} (i.e., the set $\{\mathbf{a}, \mathbf{b}\}$). What we want here is a comparison distribution counterpart of K_{AB}. Conceptually, the general relationship between comparison classes and comparison distributions is that a comparison class is generated by sampling from a comparison distribution.[39] Therefore it is natural to assume that the comparison distribution counterpart of K_{AB} is one that is suitable for generating this set. A natural candidate is

[39] Strictly speaking, sampling from the comparison distribution generates the relevant degrees possessed by the members of the comparison class. Just for brevity, here and henceforth I simply say that it generates the comparison class.

κ_{MLE}, the distribution inferred from K_{AB} using Maximum Likelihood Estimation (MLE) (85), since this is the distribution that maximizes the likelihood of generating the set K_{AB}.

(85) Maximum Likelihood Estimations for μ and σ[40]
$\mu_{\text{MLE}} = (d_A + d_B)/2$, $\sigma_{\text{MLE}} = (d_B - d_A)/2$

Note that the optimal threshold for κ_{MLE} is $\mu_{\text{MLE}} + \sigma_{\text{MLE}}$, which is d_B, the higher degree of the two. Given (83), we know that wrt κ_{MLE}, *the tall/full glass* denotes the taller/fuller glass B, regardless of whether it is considered tall/full in general. Therefore, if κ_{MLE} is an admissible comparison distribution then the definite descriptions in (83) are felicitous. This means that shifting standard is predicted to be possible in general, barring further constraints.

The second half of (ii) in (84) is designed to limit the availability of κ_{MLE} as an admissible comparison distribution, to account for the fact that shifting standard is not always possible. The intuition behind this further constraint can be better illustrated by treating the process of determining an admissible comparison distribution as a form of Bayesian inference (86).

(86) $p(\kappa \mid K_{AB}) \propto \Pr(\kappa) \cdot p(K_{AB} \mid \kappa)$

We want to define the admissibility of a comparison distribution κ in the context with two glasses A and B, i.e., how likely the listener would consider κ as the comparison distribution in the context. According to the Bayes' rule, this is influenced by two factors (86). The first factor is the prior $\Pr(\kappa)$, i.e., the probability that κ is considered the comparison distribution a priori. This prior comes from language users' general world knowledge and past linguistic experiences. The distributions that have high prior probabilities are those that are likely to be used in general descriptive contexts, i.e., the generally admissible distributions. The more different κ is from such distributions, the lower the prior $\Pr(\kappa)$. The second factor is the likelihood $p(K_{AB} \mid \kappa)$, i.e., how likely the comparison class K_{AB} is generated by sampling from the distribution κ. As discussed before, the distribution that maximizes this term is by definition κ_{MLE}. Combining these two factors, we can see that since κ_{MLE} already maximizes the likelihood term, it is admissible only if its prior is high enough, i.e., it does not deviate too much from the generally admissible comparison distributions.

The discussion above shows that the constraints in (84) can be motivated from the intuitive connection between comparison classes and comparison distributions (i.e., the former are generated by sampling from the latter) and the Bayes' rule, a general principle of probabilistic reasoning, and therefore are not just stipulations. Once we have these constraints, we can provide a unified explanation for the difficulty/impossibility to shift standards for maximum adjectives and the crisp-judgment effect, in terms of the difference between κ_{MLE} and the generally admissible distributions.

[40] Strictly speaking, these are the Maximum Likelihood Estimation for the Gaussian distribution before truncation. However, σ_{MLE} is relatively low in cases that involve standard shifting and therefore the Gaussian distribution before and after truncation are relatively close to each other.

From the discussion in the previous section, we know that for maximum adjectives, the standard deviations for the generally admissible distributions are relatively high (when $\mu = 1/2$, σ is at least $1/2$ and in general can be much higher). Note that $\sigma_{\text{MLE}} = (d_B - d_A)/2$ is relatively low (it is by definition at most $1/2$, and generally much lower). Therefore κ_{MLE} is quite different from the generally admissible distributions, and hence it is generally difficult to shift the standard of a maximum adjective. However, if there is a generally admissible distribution with a lower standard deviation (e.g., the distribution of straightness of bananas), or if σ_{MLE} is close to the upper bound $1/2$ (e.g., for a context where there is an 80% full glass and a completely empty glass), then κ_{MLE} is closer to the generally admissible distributions and shifting the standard would be easier in these cases.

In contrast, for relative adjectives, the standard deviations for the generally admissible distributions are relatively low (when $\mu = 1/2$, σ is less than $1/6$), which means that κ_{MLE} is often similar to generally admissible distributions. Therefore it is generally easy to shift the standard for a relative adjective. This, together with the discussion above for maximum adjectives, accounts for theoretical issue 4 (**shifting-standards**). That said, in a Crisp-Judgment context, where the degree difference between the two objects is very small, $\sigma_{\text{MLE}} = (d_B - d_A)/2$ will also be very small, which makes κ_{MLE} too different from the generally admissible distributions to be admissible. Therefore the definite description is infelicitous in this case. In general, as the degree difference approaches 0, σ_{MLE} would be further and further away from the standard deviations of the generally admissible distributions, and therefore shifting the standard is more and more difficult. This accounts for the crisp-judgment effect for relative adjectives, which addresses theoretical issue 5 (**crisp-judgment**).

Below I discuss how the above analysis improves on van Rooij's delineation approach. Let us start by critically examining the crucial ingredients of van Rooij's analysis.

First, van Rooij (2011b) assumes that maximum adjectives are interpreted wrt the maximal comparison class, i.e., the whole domain. However, being interpreted wrt the whole domain does not guarantee that they would receive a maximum interpretation (cf. *dark*, which has a non-maximum interpretation wrt the whole domain). Therefore, the interpretation of a maximum adjective wrt the whole domain is purely stipulated and there is no deeper explanation of why it has a maximum interpretation.

Second, in order to account for the impossibility to shift standards for maximum adjectives, van Rooij needs to further stipulate that the overt use of a maximum adjective (as opposed to when it is embedded in a comparative) cannot be interpreted wrt a smaller comparison class. Again, there is no explanation of why this should be the case.

Last, in order to account for the crisp-judgment effect, van Rooij needs to stipulate that a two-object comparison class is inadmissible if the degree difference is too small. Once again, this is a stipulation and no deeper explanation is provided.

In contrast, under my analysis, the crucial difference between maximum and relative adjectives is whether the comparison distribution κ has a relatively low or high standard deviation σ_κ. In many cases the choice of κ can be plausibly independently motivated and therefore the current analysis is less stipulative. However, I acknowledge that there is not yet a fully general theory to predict which comparison distributions can be used for a gradable adjective, and therefore the choice of κ and particularly its standard deviation parameter is not completely stipulation-free. I will return to this issue in Section 3.5.1. For now, let us see how the current analysis avoids the stipulations discussed above.

First, once we assume that maximum adjectives are interpreted wrt a comparison distribution κ that has a high standard deviation σ_κ (which is high enough so that $\mu_\kappa + \sigma_\kappa > 1$), our semantics guarantees a maximum standard. In other words, no further stipulation is needed to derive the maximum interpretation.

Second, the difficulty of shifting standards for maximum adjectives is due to the fact that κ_{MLE} is less admissible for maximum adjectives, which in turn is due to the large deviation between κ_{MLE} and generally admissible comparison distributions in terms of the standard deviation. The reason that such a large deviation would result in lower admissibility is derived from the conceptual connection between comparison classes and comparison distributions and the general principle of Bayesian inference. Therefore, no further stipulation is needed.

Last, the crisp-judgment effect is explained in a parallel way. In a Crisp-Judgment context, i.e., when the degree difference between the two objects is very small, κ_{MLE} is less admissible, due to the large deviation between κ_{MLE} and generally admissible comparison distributions in terms of the standard deviation. Once again, no further stipulation is needed.

In sum, once we make the standard deviation of a comparison distribution the key parameter that distinguishes between relative and maximum adjectives, we can further derive the properties of their referential uses in definite descriptions in a unified way without further stipulations.

3.4.4 Optimal threshold for a comparison distribution

We have seen that the proposed semantics can account for the various properties of relative and maximum adjectives in both descriptive and referential uses and improve on previous approaches. However, as Solt (2011) points out, one aspect of this proposal that is not entirely satisfying is that the semantics seems to be rather complicated, in that it makes reference to two statistical measures, i.e., the mean and the standard deviation. In fact, similar to Solt's implementation, a more empirically accurate semantics would have an even more complex form $\textbf{Opt}(\kappa) = \mu_\kappa + \beta \cdot \sigma_\kappa$, where β is a constant and $0 < \beta \leq 1$.[41] While this complexity is not in itself a problem (as long as it is empirically accurate), it does suggest that it would be nice if we can find a deeper explanation of

[41] Note that the constant β does not affect the qualitative properties discussed so far, and the crucial parameter that distinguishes between relative and maximum adjectives is still the standard deviation. The constant β only changes the quantitative details of the analysis.

the quantitative relation between the threshold and comparison distribution, and in particular why the semantics of a positive form should be sensitive to both the mean and the standard deviation of the comparison distribution.

To address this issue, in this section we will take a step back and reconsider our definition of the operator **Opt**. The strategy is that instead of directly specifying what **Opt** is, we do it indirectly by specifying what it should be, based on considerations of communicative efficiency (Qing & Franke, 2014).

For concreteness, suppose that there is a speech community that often need to communicate people's heights to one another by using the adjective *tall*.[42] Given the comparison distribution κ, we consider the following question. What would be the best θ to use as the conventional threshold for *tall*, such that speakers and listeners who produce and interpret gradable adjectives following this convention would be in the long run most successful in communicating people's heights drawn from this comparison distribution κ?

To formalize this question, we define the utility of a threshold θ (for a comparison distribution κ) in (87).

(87) $$U_\kappa(\theta) = \sum_{d<\theta} \kappa(d) \cdot p(\text{L correctly guesses } d) + \sum_{d\geq\theta} \kappa(d) \cdot p(\text{L correctly guesses } d)$$
$$= \sum_{d<\theta} \kappa(d) \cdot \kappa(d) + \sum_{d\geq\theta} \kappa(d) \cdot \kappa(d \mid d \geq \theta)$$

Here is intuitively what (87) says. Suppose that a speech community has decided to use a particular threshold θ for a gradable adjective A. Then for a degree d randomly drawn (with probability $\kappa(d)$) from the comparison distribution κ, there are two types of scenarios. (i) If $d < \theta$, then given the semantics, the gradable adjective cannot be truthfully applied. Therefore the speaker will stay silent and the listener can only guess the degree based on the prior information, i.e., the comparison distribution κ. The probability that the listener will correctly guess the intended degree d is thus $\kappa(d)$. Such scenarios correspond to the first summand $\sum_{d<\theta} \kappa(d) \cdot \kappa(d)$ in (87). (ii) If $d \geq \theta$, then given the semantics, the gradable adjective can be truthfully applied. Therefore the speaker will use the gradable adjective to convey to the listener the additional information that the intended degree d reaches the threshold θ. Hence the probability that the listener will correctly guess the intended degree d in this case is the conditional probability $\kappa(d \mid d \geq \theta)$. Such scenarios correspond to the second summand $\sum_{d\geq\theta} \kappa(d) \cdot \kappa(d \mid d \geq \theta)$ in (87). The utility of a particular threshold θ, with respect to a comparison distribution κ, is the sum of these two terms, which measures the probability of the listener correctly guessing the intended degree in the long run, i.e., the *expected* probability of communicative success.

Note the definition (87) involves an informativity-applicability tradeoff similar to the one discussed in the previous chapter. To see this more clearly, we can subtract the constant $\sum_d \kappa(d) \cdot \kappa(d)$ from (87), which yields (88).

[42] For simplicity, we assume that this is the only expression they can use.

(88) $U_\kappa(\theta) - \sum_d \kappa(d) \cdot \kappa(d) = \sum_{d \geq \theta} \kappa(d) \cdot (\kappa(d \mid d \geq \theta) - \kappa(d))$

On the one hand, the higher θ is, the higher the conditional probability $\kappa(d \mid d \geq \theta)$ is, which means that each term that is being summed up is higher. Intuitively, if the threshold θ is high, the positive form will be informative. However, at the same time, the higher θ is, the fewer terms are actually being added, because there are fewer degrees that reach the threshold. Intuitively, if the threshold θ is high, the positive form will not be very applicable, because few individuals will actually have a degree that reaches this threshold. Therefore, (88) can be understood as a tradeoff between informativity and applicability.

Now that the utility of each threshold θ is defined, we define the function **Opt** to pick out the threshold that maximizes the utility (89). Henceforth I will call this model the *Optimal Threshold Model* (OTM).[43]

(89) $\mathbf{Opt}(\kappa) = \arg\max_\theta U_\kappa(\theta)$

Let us consider some concrete examples.

(90) a. John is tall for a male adult.

b. John is tall for a male professional basketball player.

First, consider (90a). Our prior world knowledge about how male adults' heights are distributed corresponds to a Gaussian distribution, and let us assume that it is represented as the orange solid curve in Fig. 3.3. According to the definition in (87), the predicted utility of using each height as the threshold for *tall* is plotted as the orange dashed curve in Fig. 3.3. Note that the maximum utility is achieved when the threshold θ is above the average height (but not too much so). This correctly predicts that *John is tall for a male adult* should be semantically stronger than *John is taller than the average height of male adults* (but not too much so).

We observe similar results for (90b). The prior distribution of male professional basketball players' heights is plotted as the blue solid curve in Fig. 3.3, and the predicted utility of using each height as the threshold for *tall* is plotted as the blue dashed curve on the right. Again, the maximum utility is achieved when the threshold θ is above the average height of male professional basketball players

[43]The main difference between the OTM and Qing and Franke's (2014) Speaker-Oriented Model (SOM) is that the OTM uses strict maximization to pick out the threshold with the highest utility, whereas the SOM uses a softmax function that returns a probability distribution over thresholds, where thresholds with higher utilities are preferred. However, I want to emphasize that even though Qing and Franke (2014) used a softmax function to derive a probability distribution over thresholds, based on which they defined a probabilistic speaker rule, they only used it as a way to model loose talk, and they also "interpreted vagueness as the stability of the optimal threshold under uncertainty about the exact prior distribution of degrees in the comparison class" (p. 33; see also their discussion on p. 38 on the difference between vagueness and loose talk). Therefore, the OTM presented here, together with the discussion in Section 3.4.2, is really just a more explicit formulation of Qing and Franke's analysis of vagueness, with imprecision factored out to avoid confusion.

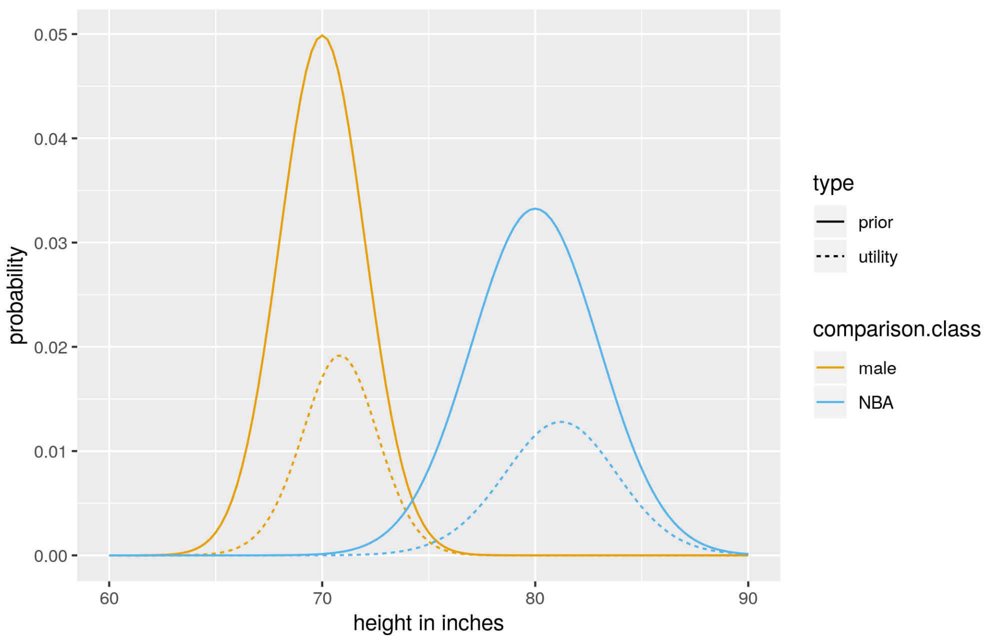


Figure 3.3: Optimal Threshold Model for *John is pos tall for a male adult (orange) / for a male professional basketball player (blue)*

(but not too much so). Moreover, the optimal threshold for male professional basketball players is predicted to be greater than the one for male adults. Therefore, similar to Lassiter and Goodman's (2013, 2015) model, the inference from (90b) to (90a) is derivable in the formal model.

Now let us consider some examples of closed-scale adjectives (91).

(91) a. The glass is full.

b. The feather is dark.

Following Lassiter and Goodman (2013), I assume that the prior/comparison distribution for *full* is a uniform distribution (the orange solid line in Fig. 3.4). Under this assumption, the predicted utility of using each degree of fullness as the threshold for *full* is plotted as the orange dashed line. We can see that the optimal threshold that maximizes the utility is the maximum degree, i.e., 100%. This correctly predicts that *full* has a maximum standard in its interpretation.

However, a closed-scale adjective does not necessarily have a maximum (or minimum) standard in this model. Take (91b) for example. The darkness of a feather presumably corresponds to a closed scale ranging from no darkness (i.e., white) to maximum darkness (i.e., black). However, given that there are many shades in between, it is plausible to use a truncated Gaussian distribution (the blue solid curve in Fig. 3.4)) as the comparison distribution. Such a distribution takes into account the existence of black and white feathers, by assigning non-negligible probabilities to the endpoints

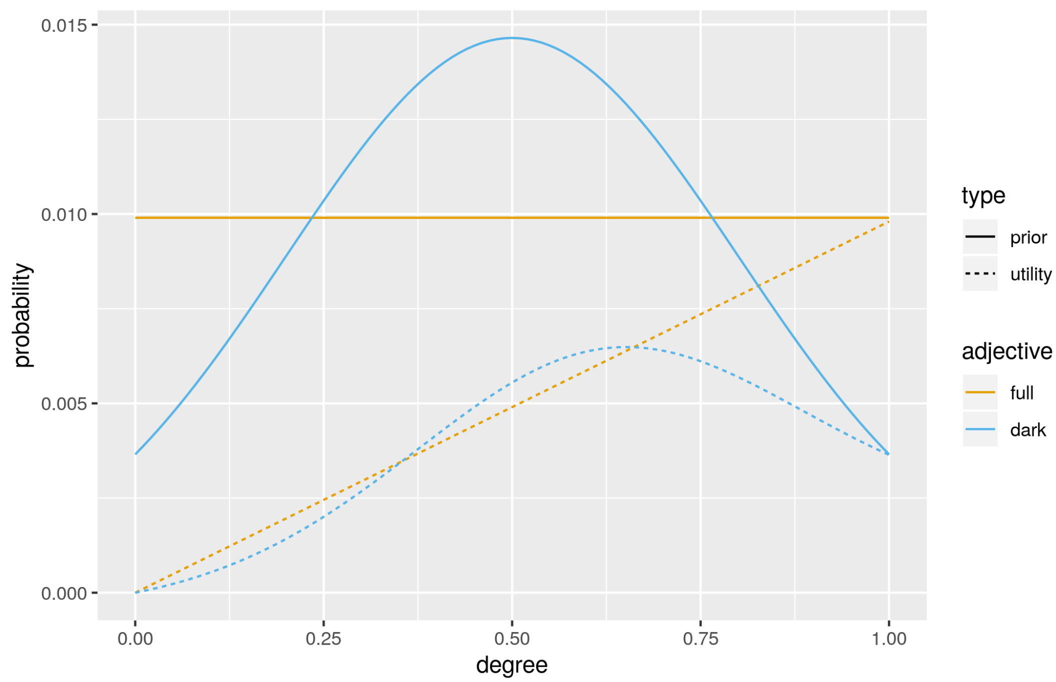


Figure 3.4: Optimal Threshold Model for *the glass is full* (orange) and *the feather is dark* (blue)

of the scale, and is also in accord with our knowledge that most feathers have some intermediate darkness. With respect to this comparison distribution, the predicted utility of using each degree of darkness as the threshold for *dark* is plotted as the blue dashed curve. We can see that the optimal threshold is predicted to be above the average darkness but not too much so.[44] This patterns with open-scale adjectives and crucially the optimal threshold is not a maximum (or minimum) standard. This correctly predicts that *dark* has a relative standard despite having a closed scale.

In sum, we have seen that the model accounts for the standards for open-scale as well as closed-scale adjectives. Just as the previous version of our semantics, the model makes precise predictions about how the contextual comparison distribution affects the interpretation of a gradable adjective, and can formally derive certain inference patterns across comparison classes. In particular, the central feature of the previous version of our analysis is preserved, i.e., a closed-scale adjectives have a maximum standard if the comparison distribution is flat or dispersed enough (as is the case for *full*) but can also have a relative standard if the comparison distribution is concentrated enough in the interior of the scale (as is the case for *dark*). This allows us to capture the (imperfect) correlation between scale structure and interpretation of the positive form and preserve our earlier analysis of the various properties of relative and maximum adjectives in descriptive and referential uses.

Crucially, the definition of this updated version of the semantics does not explicitly make reference

[44] Here the average darkness is assumed to be the midpoint between black and white just for concreteness.

to the mean or standard deviation of the comparison distribution. Rather, the definition is based on considerations of communicative efficiency. The fact that the threshold is influenced by the mean and the standard deviation of the comparison distribution is a consequence of such considerations of optimal language use. This provides a deeper explanation of why the mean and the standard deviation are relevant for determining the threshold, and addresses Solt's (2011) worry about the complexity of the earlier version of our semantics. Also, this model allows us to state the central thesis of our analysis in more general terms, which is that the threshold of a gradable adjective is determined by both the central tendency and the dispersion of the comparison distribution, abstracting away from the specific statistical measures (e.g., whether to use mean or median to measure central tendency).

3.5 General discussion

3.5.1 Comparison distributions for maximum adjectives

We have seen that relative and maximum adjectives can have a unified semantics, according to which the threshold is determined by both the central tendency and the dispersion of the comparison distribution. The crucial parameter that accounts for the differences between these two classes of adjectives is the dispersion. Maximum adjectives have comparison distributions that are flat or dispersed enough, while relative adjectives have ones that are relatively concentrated. Naturally, a remaining question is whether this assumption can be independently motivated or justified.

Given that most relative adjectives have an open scale, which means that their comparison distributions by definition will be relatively concentrated (because they will have to approach 0 as the degree increases), the question is mainly about whether maximum adjectives can be reasonably assumed to have comparison distributions that are flat or dispersed enough.

Fully addressing this question requires a complete theory of how comparison distributions are determined. As mentioned before, unfortunately I cannot provide such a theory and have to leave it for future research. However, below I will provide some considerations that motivate this assumption about maximum adjectives.

One intuitive way to estimate the dispersion of the comparison distribution of a gradable adjective is to consider the amount of decline in likelihood when the degree moves towards the upper end of the scale (or approaches infinity) from a central, non-extreme degree. To better illustrate this, consider *tall* (for a US adult male) and *full*. Given our world knowledge, we know that as the height increases from a central non-extreme height (e.g., 5′9″), the likelihood of someone having that height will decline quite a lot, e.g., it is much rarer for someone to have an extreme height (e.g., 7′) than 5′9″. In contrast, as the degree of fullness increases from half full, the likelihood of a glass having that amount of fullness does not seem to drop all that much. For instance, a glass that is 95% full is in a way not that much rarer than one that is half full, considering the fact that we can easily make a half-full glass to 95% full and vice versa. (If you do not already share this intuition, there is some further discussion below.) This provides some initial evidence that *full* has a comparison distribution that is overall quite flat or dispersed and therefore can be reasonably approximated by a uniform distribution.

Note that in the discussion above, we are appealing to two types of intuitions when we compare the relative likelihood of two degrees. In the case of *tall*, we are comparing different individuals, e.g., how likely one would encounter someone who is $7'$ as opposed to $5'9''$. In the case of *full*, we are essentially comparing different stages of the same individual, e.g., how likely a glass would be in the state of 95% full as opposed to 50% full.

This suggests that there is a correlation between individual-/stage-level predicates and the way their comparison classes/distributions are constructed (Toledo & Sassoon, 2011). On the one hand, since individual-level predicates denote stable properties, it only makes sense to construct their comparison classes/distributions by comparing different individuals. On the other hand, for a stage-level predicate, in principle it makes sense to construct the comparison class/distribution from between-individual or within-individual comparisons. For instance, we can say *this theater is full for a theater in this area* (between-individual comparison) and *this theater is full for a Thursday afternoon* (within-individual comparison).

However, it is often unclear which probability distributions should be used as the comparison distribution, even if there is a clear intention about the comparison class. This is why it is generally strange to say *this circle is big for a circle* out of the blue, because interlocutors generally do not have a shared probability distribution of sizes of circles in the common ground. Or as Bierwisch (1989) puts it, circles do not have an *intrinsic norm* wrt the size dimension. He defines an intrinsic norm as a norm that can be established based on features of the comparison class, and provides *bed*, *house*, and *city* as examples that have intrinsic norms for size. Although he does not specify what counts as a feature or how the norm is supposed to be established, it seems plausible to assume that the reason such examples have intrinsic norms is that interlocutors generally have some shared world knowledge about the distribution of sizes for beds, houses and cities.

Similarly, when I point to a glass of water and say to you *this glass is full*, there is probably little shared world knowledge about the distribution of degrees of fullness for various glasses (between-individual comparison) or various stages of this particular glass (within-individual comparison) to allow us to coordinate on a comparison distribution (based on world knowledge). Fortunately, however, for closed-scale adjectives, there is another way for us to coordinate on a comparison distribution, i.e., by appealing to the *principle of maximum entropy*. According to this principle, the probability distribution that best represents our knowledge (or the lack thereof) is the one with largest entropy. In the case of closed-scale adjectives, if there is no prior knowledge, then the maximum entropy distribution is the uniform distribution, and the optimal threshold for this distribution is the maximum degree.[45]

[45] If the scale is only upper-closed, e.g., $(a, b]$, we can truncate it to $[a', b]$, where a' is close to a. This would be a

To sum up, there are two main motivations for the assumption that maximum adjectives have comparison distributions that are rather flat or dispersed. First, they tend to be stage-level predicates, and the transient nature means that it is more likely that the various degrees are more or less equally likely to be manifested, which corresponds to a rather flat or dispersed comparison distribution. Second, for (upper-)closed-scale adjectives, if there is no prior world knowledge shared in the common ground, the principle of maximum entropy allows the interlocutors to coordinate on the uniform distribution as the comparison distribution, from which our semantics derives a maximum interpretation.

3.5.2 Scale structure, comparison distributions, and degree modification

Under the current analysis, the scale structure of a gradable adjective indirectly influences its threshold by constraining the comparison distributions (in particular the dispersion). A major advantage is that we can account for the imperfect correlation between the scale structure and the threshold of a gradable adjective, without having to stipulate counter-intuitive scale structures. However, this also complicates the analysis of degree modification.

Traditionally, degree modifiers such as *completely* and *slightly* are analyzed as directly targeting the maximum or minimum endpoint of the scale structure (Kennedy & McNally, 2005; Kennedy, 2007). In the case of *slightly*, Solt (2012) provides convincing arguments that it does not directly target the minimum endpoint of the scale. Rather, it only requires that the gradable adjective being modified has a non-vague standard in its positive form, which need not be the minimum endpoint. In the next chapter, we will see examples of such gradable adjectives. Since this chapter concerns relative and maximum adjectives, I will focus on maximizers such as *completely* and *perfectly*. Lassiter (2017) suggests that we may similarly analyze them as imposing constraints on the interpretations of the positive forms of the gradable adjectives being modified. For instance, if we assume that *completely* and *perfectly* require that the gradable adjective being modified have a maximum standard in the positive form, then we can account for why they can modify *certain* but not *likely/probable* despite them intuitively having closed scales: only the former has a maximum standard in the positive form.

However, what remains to be explained is why adjectives such as *likely/probable* do not have a maximum standard in the positive form. According to the current analysis, this amounts to explaining why such adjectives cannot have comparison distributions that are dispersed enough, while adjectives such as *certain* and *dark* can. Once again, since I do not have a complete theory of how comparison distributions are determined, I am not able to provide such an explanation and have to stipulate the incompatibility between adjectives such as *likely/probable* and comparison

good approximation of the original scale, because the probability mass between a and a' would be close to 0 as long as a' is sufficiently close to a. We can then appeal to the principle of maximum entropy and use the uniform distribution over $[a', b]$ as the comparison distribution. Note that the optimal threshold would always be the maximum degree b regardless of the exact value of a'. Therefore we also get a stable, maximum standard for upper-closed-scale adjectives.

distributions that are dispersed enough to prevent them from having a maximum standard. Note that it is in principle possible that comparison distributions are not completely determined based on world knowledge, but rather have lexically specific constraints, which can be the result of lexical competition and/or blocking. If we adopt this assumption, then the competition between *certain* and *likely/probable* may result in them having different constraints on comparison distributions, whereas *dark* is more flexible wrt comparison distributions due to lack of competitors. Admittedly, this is still not entirely satisfactory, but note that the traditional analysis essentially makes the same type of assumptions, by stipulating that *likely/probable* has a scale structure without a maximum degree and *certainly* has one with a maximal degree. The only difference here is that we assume a range of admissible comparison distributions instead of scale structure as part of the lexical meaning of a gradable adjective. However, an advantage of this analysis is that we no longer need to further stipulate that the scale structure of *likely/probable* can be coerced into one with a maximal degree in order to account for the data involving comparatives.

3.5.3 Comparison with the free-variable model

Finally, let us compare the Optimal Threshold Model with Lassiter & Goodman's free-variable model.

Both models assume that the threshold of a gradable adjective is underspecified and needs to be resolved in context in some way. For descriptive uses, the two models make very similar predictions, and they are similarly motivated by functional considerations about how to use language in a way that best serves the interlocutors' communicative goal. However, there is a crucial conceptual difference between the two models regarding the nature of the contextual-resolution mechanism.

In the free-variable model, the threshold of a gradable adjective is treated as a variable that is semantically unconstrained, i.e., the speaker can in principle choose whatever degree they want as the threshold, and gets resolved in context pragmatically. The main idea is that not every degree will be equally good to serve the communicative goal, and therefore a rational speaker would only choose the better ones. Such choices can be reasoned about by a pragmatic listener so that interlocutors can coordinate on the threshold reasonably well (despite the fact that this is not the primary communicative goal, which is to convey a degree). As a consequence of the assumption that the threshold is semantically unconstrained, one would expect that when the gradable adjective is embedded in a definite description and the communicative goal is different (i.e., to convey a referent), a rational speaker would choose those thresholds that better serve this new referential goal, which might be different from the general descriptive uses. However, empirically this is not the case, as illustrated by the difficulty of shifting standard for maximum adjectives and the crisp-judgment effect.

In contrast, in the Optimal Threshold Model, the threshold is not semantically unconstrained. Rather, it is the output of a semantic function (i.e., **Opt**) that takes as input a contextual parameter, i.e., the comparison distribution. In this sense, the threshold is resolved semantically in context.

This treatment has different consequences. On the one hand, we can still use considerations of optimal language use to inform the specification of such a semantic function, i.e., for a contextual comparison distribution, what threshold should **Opt** output such that the speech community can be most successful in communicating degrees drawn from this distribution? On the other, since such a function is part of the conventional meaning of the positive form, we would expect that when the positive form is embedded in a definite description, the very same function will be used in the interpretation and the threshold will be determined in the same way, and therefore we would not expect the function from comparison distributions to thresholds to change, even if that might turn out to be better for the communicative goal in that particular referential context. That said, there is a complication when we analyze referential uses of gradable adjectives. Since a referential context is different from a descriptive context, it is possible that a different comparison distribution is used in the referential context, and therefore we need to specify which comparison distributions are admissible in referential contexts (in particular those in addition to the admissible ones in descriptive contexts). Under reasonable assumptions about admissible comparison distributions, the Optimal Threshold Model correctly predicts the possibility of shifting the threshold, limited by the admissibility of the additional comparison distribution κ_{MLE}.

I should emphasize that I am taking Lassiter & Goodman's free-variable model as a formal implementation of King's coordination account, which is not necessarily a commitment they want to make when it comes to modeling referential uses of gradable adjectives in definite descriptions. As a result, I acknowledge that there could be ways to extend Lassiter & Goodman's free-variable model to account for such referential uses. However, it seems that in order to capture the link between descriptive and referential uses, any feasible extension would need to first specify which degree distribution should be used as the prior distribution over degrees, and presumably it would need to integrate our general world knowledge about the degree distribution and the specific information about the degrees of the relevant objects in the immediate visual context. I suspect that the end result would be similar to my definition of admissible comparison distributions, and thus such models will not end up being substantially different from my proposal.

Finally, I note that my proposal is in a way compatible with King's coordination account. The only adjustment that needs to be made is that the relevant conceptual parameter that the interlocutors are coordinating on is the comparison distribution, as opposed to the threshold in King's original proposal. The speaker who uses a gradable adjective in a context intends a certain degree distribution as the comparison distribution, and if the communication is to be successful, an idealized listener should be able to recognize such an intention. Whether or not such an intention is recognizable is characterized by the notion of admissible comparison distributions. When the interlocutors successfully coordinate on the comparison distribution, they can use **Opt**, the conventional mapping from comparison distributions to thresholds, to obtain the corresponding threshold. As far as I can tell, nothing about this adjustment goes against the general spirit of King's coordination account.

3.6 Conclusion

In this chapter, I propose a unified semantics for (positive forms of) relative and maximum gradable adjectives, according to which they have an underspecified threshold that is resolved in context. Crucially, I argue that this contextual-resolution mechanism is semantic, in that there is a conventional mapping from the contextual parameter, i.e., the comparison distribution, to the threshold. This conventional mapping is defined by considerations of communicative efficiency, from which it follows that the threshold of a gradable adjective is sensitive to both the central tendency and the dispersion of the comparison distribution. Supplementing this conventional mapping with a notion of admissible comparison distributions, I further illustrate that the dispersion of the comparison distribution is the key parameter that accounts for the differences between relative and maximum adjectives in descriptive as well as referential uses.

In the next chapter, I will argue that even though the current analysis can be applied to minimum adjectives to derive their relative interpretations, minimum adjectives are in fact systematically ambiguous and their minimum interpretations need to be derived from a different mechanism.

Chapter 4

Underspecification vs. ambiguity: Minimum adjectives

4.1 Introduction

This chapter examines the class of minimum(-standard) gradable adjectives (e.g., *wet* and *bent*), whose positive forms have an interpretation that is traditionally linked to the minimum degree on the adjective's scale (e.g., Kennedy & McNally, 2005; Kennedy, 2007). For instance, the scale of *wet* has a minimum degree, which corresponds to zero amount of wetness (i.e., completely dry). Correspondingly, *wet* has a minimum standard: something is wet iff it has non-zero amount of wetness.[1] Within a degree-semantic framework, the interpretation of the positive form of a minimum adjective in context can be formally represented in (1). Note that I intentionally represent only the surface form, so that I am not presupposing any specific analysis that assumes silent materials.

(1) $[\![x \text{ is } \mathrm{A}_{\mathrm{min}}]\!]^c = \textbf{A-measure}(x) > d_{\mathrm{min}}$

Comparing (1) with the interpretations of positive forms of relative (2a) and maximum (2b) adjectives, we can see that there is a clear parallel between the three classes of gradable adjectives. Moreover, given my proposal in the last chapter that the maximum-standard interpretation (2b) results from an underspecified threshold being resolved to the maximum degree in context, which unifies (2a) and (2b), a natural question to ask is whether the same can be said about the minimum-standard interpretation (1) and thus we can have a unified semantics for all three classes of gradable adjectives.[2]

[1] Recall from Section 3.2.3 that we need to factor out imprecision/granularity.

[2] There is an obvious technical problem, i.e., the inequality in (1) but the ones in (2a) and (2b) are not. For the sake of the argument, I will set this issue aside and assume that it can be addressed by, e.g., incorporating granularity so that

(2) a. $[\![x \text{ is } \mathrm{A_{rel}}]\!]^c = \textbf{A-measure}(x) \geq \theta_c$

b. $[\![x \text{ is } \mathrm{A_{max}}]\!]^c = \textbf{A-measure}(x) \geq d_{\max}$

This question is related to a broader theoretical issue, i.e., how to distinguish between semantic underspecification and ambiguity. In general, if we observe variability in the interpretations of an expression, or in this case a class of expressions, a theory that posits semantic underspecification needs to be supplemented with a contextual-resolution mechanism that accounts for the observed variability, and a theory that posits ambiguity should ideally be supported by independent evidence. In this chapter, I propose an ambiguity analysis of minimum adjectives, according to which the minimum interpretations of their positive forms are derived from a mechanism different from the one used for relative and maximum adjectives proposed in the previous chapter. I argue for this ambiguity analysis by both illustrating that the semantic-underspecification analysis in the previous chapter is unable to account for the minimum interpretation of the positive form of a minimum adjective, and providing independent evidence for the ambiguity by considering the parallel between positive forms of minimum adjectives and comparative constructions.

Before getting into the details of my proposal and arguments, I want to emphasize a seemingly minor issue that will turn out to be very important for the later discussions. Note that the formal representation of the interpretation of the positive form of a minimum adjective makes reference to the minimum degree (1). Meanwhile, recall that we also say that something is wet iff it has non-zero amount of wetness. In the case of *wet*, these two descriptions are equivalent because the minimum degree is also a meaningful, non-arbitrary zero point on the scale, i.e., it indicates the absence of the relevant property. In fact, this is the case for so many of the minimum adjectives, that zero and minimum degrees are often used interchangeably in the literature. However, in this chapter, I argue that zero and minimum degrees are not always identical, and that the standard of the so-called "minimum interpretation" is in fact based on the zero degree rather than the minimum. As a result, even though I will maintain the class label *minimum gradable adjectives*, I will replace *minimum interpretations* with > 0 *interpretations* (read: greater-than-zero interpretations) to highlight the fact that the standard is the zero degree rather than the minimum.

In this next section, I will use the gradable adjective *profitable* as a case study, and argue that while it does have a meaningful zero degree as its standard, the zero degree is arguably not a minimum degree on the scale. This suggests that it is descriptively more accurate to call such an interpretation a > 0 interpretation. Moreover, I argue that this interpretation cannot be derived from the contextual-resolution mechanism proposed in the previous chapter, which instead derives a relative interpretation also attested for *profitable*. This provides some initial evidence in favor of an ambiguity analysis. In Section 4.3, I discuss more cases where a minimum adjective can have both relative and > 0 interpretations, and provide independent evidence for an ambiguity analysis by

the non-zero degrees on the scale of a minimum adjective also have a minimum degree (written as $d^+_{\min}$ to distinguish it from the absolute minimum/zero degree $d_{\min}$) to serve as the standard, i.e., $[\![x \text{ is } \mathrm{A_{min}}]\!]^c = \textbf{A-measure}(x) \geq d^+_{\min}$.

noting the parallel between positive forms of minimum adjectives and comparative constructions as observed by Sawada and Grano (2011). Given this parallel, in Section 4.4, I provide a compositional semantic analysis of the > 0 interpretation, based on Schwarzschild and Wilkinson's (2002) and Schwarzschild's (2005) analyses of comparative constructions. I discuss the implications of this analysis and remaining issues in Section 4.5.

4.2 A case study: *profitable*

As a case study, let us consider the adjective *profitable*. This adjective is clearly gradable, as can be seen from the examples in (3).

(3) a. This company is more profitable than that company.

b. How profitable was this company last year?

c. This industry is so profitable that everybody wants to get a share.

As for the positive form, *the company was profitable* has an interpretation which is true iff the company at least made a minimum/non-zero amount of profit. For example, consider the following naturally occurring example (4).

(4) [Headline] Spotify, the leading music streaming app, is finally profitable.

[Main text] [...] Today, for the very first time, the company is reporting that it's turned a profit.

That's right: some 13 years and 96 million paid subscribers later, Spotify is finally making money.[3]

In the headline, the author asserted that Spofify is profitable, and from the main text it is clear that the intended interpretation is that the company has turned a profit or in other words is making money.

In this respect, this interpretation of *profitable* patterns with run-of-the-mill minimum adjectives such as *wet* and *bent*: something is wet/bent iff it has at least a minimum amount of wetness/bend. Moreover, just as we can use *slightly* to modify *wet* or *bent*, we can do so to modify *profitable*, e.g., in the naturally occurring example below (5).

(5) "If they [Tesla] hit that number, its going to equate to 48,000 model 3s produced in the September quarter. That should get them to profitability, slightly profitable," Munster said.[4]

[3] https://www.theverge.com/2019/2/6/18214331/spotify-earnings-financial-announcement-profits-music-streaming-podcast

[4] https://www.cnbc.com/2018/07/02/tesla-will-be-profitable-by-september-says-gene-munster.html

Therefore, as far as the interpretation of the positive form is concerned, there is good evidence that *profitable* belongs to the class of minimum adjectives. The remaining question is whether such an interpretation should be characterized in terms of the zero degree on its scale or in terms of the minimum degree, and this question requires us to answer whether there is a minimum degree on the scale in the first place. Note that unlike Kennedy (2007), who takes *slightly*-modification to be a diagnostic of whether a scale has a minimum degree, I follow Solt (2012) and assume that the felicity of *slightly Adj* only requires that the adjective have a non-vague standard, which can but need not be the minimum degree of the scale. As a result, the possibility of *slightly*-modification (5) does not entail that the scale has a minimum degree. Therefore we need to independently determine the scale structure for *profitable*, by considering how the profitability of a company should be measured. Intuitively, the profitability of a company can be measured by the amount of the profit it makes.[5] However, there is a complication (besides the complicated actual accounting process to determine profit). It is possible for companies to lose money. For concreteness, suppose that the performances of six companies in 2019 are listed as follows (6), and assume that these are the relevant facts for evaluating the examples used throughout this section.

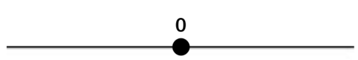
(6)

Company	A	B	C	D	E	F
Profit/Loss	-\$100M	-\$10M	0	\$5K	\$1M	\$1B

In order to compositionally derive interpretations of various degree constructions that contain the gradable adjective *profitable*, we need to first specify the measure function it denotes (and in doing so, we would also have specified its scale structure). The critical question is what should be the output of this measure function for companies that lost money (i.e., A and B in this example).

I propose that we take (6) at face value, i.e., we simply treat losses as negative degrees of profitability. The resulting scale structure is shown schematically in (7). I will call this the full-range analysis.

(7)



This full-range analysis is highly intuitive. Later I will compare it with two alternatives proposed in the literature (not specifically for *profitable*, but for similar adjectives) and argue against those alternatives. For now let us assume that (7) is indeed the scale structure of *profitable* and discuss its consequences.

Note that (7) is an open scale and does not have a minimum degree. This means that the > 0 interpretation of *profitable*, e.g., in (4), cannot be based on a minimum degree. Can this

[5]Of course, this is not the only way to measure profitability. For example, a commonly used measure is *profit margin*, which is profit divided by revenue. If we use this measure of profitability, then it is possible for a company to be more profitable than another even if the two companies made the same amount of profit. I choose profit as the measure for *profitable* just for simplicity. Nothing I say in this section crucially hinges on this choice.

> 0 interpretation be derived as a relative interpretation, whose standard happens to be resolved to 0 in context? Arguably not, for two reasons.

First, deriving the > 0 interpretation as a relative interpretation requires a contextual comparison distribution that is generally implausible. Recall that according to the contextual-resolution mechanism proposed in the previous chapter for relative interpretations, the standard is always greater than the mean of the contextual comparison distribution. As a result, in order for the standard of *profitable* to be 0, the mean of the comparison distribution must be less than 0. In other words, in order to derive the > 0 interpretation as a relative interpretation, we would have to assume that companies on average lost money. This seems implausible. Intuitively, a speaker can truthfully say *this company was profitable*$_{>0}$ *last year*, without there being a belief in the common ground that companies on average lost money. Now, one might think that in (4) perhaps we do have a general expectation that music streaming services lose money. But this is still not enough to derive the > 0 interpretation as a relative interpretation. Note that the standard of the relative interpretation is not too much greater than the mean. Therefore, we also need to assume that music streaming services on average do not lose too much money. This once again is not very plausible. If it has been challenging for the leading app to make money, there are good reasons to believe that other music streaming services are losing a lot of money, and therefore 0 would probably be too much above the mean to be a relative standard.

Second, and more importantly, the standard of an open-scale adjective is predicted to be vague according to the mechanism proposed in the previous chapter (as well as other major approaches discussed there). This means that even if we stipulate the contextual comparison distribution in such a way that its relative standard happens to be equal to 0, we would still predict that this interpretation is vague. In contrast, the > 0 interpretation of *profitable* is clear-cut, and therefore it cannot be a relative interpretation.

The above discussion only shows that the > 0 interpretation arguably cannot be derived as a relative interpretation, which does not rule out the possibility that *profitable* can have a relative interpretation. Intuitively, *profitable* does have a relative interpretation, which is stronger than the > 0 interpretation. Indeed, we can observe the two interpretations in one sentence. For example, suppose that Company D is an oil company. Given the common conception that oil companies are very profitable (which need not be in fact true), it is coherent to say (8).

(8) Company D was (technically) profitable$_{>0}$, but not profitable$_{\geq\theta}$ for an oil company.

Here, the second instance of *profitable* is interpreted wrt the (perceived) profit distribution of oil companies and has a vague, relative standard θ that is presumably much higher than the mere \$5K Company D made. In contrast, the first instance of *profitable* has a clear-cut interpretation that uses 0 as the standard.

The ambiguity between > 0 and relative interpretations can also be shown with polar questions. For example, the polar question in (9) can be interpreted as whether the company made any profit at all, or whether the company's profit was high compared to other companies in the same industry, companies in general, etc. The answer addresses both interpretations. The first part confirms the > 0 interpretation, and the *but*-clause implies that the relative interpretation is not true.

(9) Q: How was your company doing last year? Was it profitable$_{>0/\geq\theta}$?

A: Well, technically it was, but it only made $5K in profit.

Summing up the discussion so far, if the full-range analysis of the scale structure of *profitable* (7) is correct, i.e., *profitable* has an open scale, then we would have evidence that the > 0 interpretation is not derived from a minimum degree (since there is no such a degree) or as a special case of the relative interpretation and therefore needs to be independently derived. To further support the full-range analysis, below I consider two alternative scale structures that are lower-closed. They are proposed in the literature not specifically for *profitable* but for similar adjectives. I will show the major problem each of these two analyses has and conclude that the full-range analysis indeed provides the best analysis of the scale structure of *profitable*.

The examples I will use involve various comparative constructions, including bare comparatives (10a) or modified ones (10b). The modifier can be a measure phrase such as $5K, or an adverb such as *slightly* or *much*. It is represented as DIFF because intuitively its meaning concerns the difference between the profits of x and y. Note that different analyses can compositionally derive different semantic representations that are truth-conditionally equivalent to (10a) and (10b), and for the first analysis discussed below, (10a) and (10b) will be undefined if $\mathbf{profit}(x)$ or $\mathbf{profit}(y)$ is undefined.

(10) a. $[\![x \text{ is more profitable than } y]\!] = \mathbf{profit}(x) > \mathbf{profit}(y)$

b. $[\![x \text{ is DIFF more profitable than } y]\!] = [\![\text{DIFF}]\!](\mathbf{profit}(x) - \mathbf{profit}(y))$

I will call the first alternative, proposed by Sawada and Grano (2011), the *truncation* analysis. When applied to *profitable*, it stipulates that the measure function **profit** is undefined for companies that lost money. The resulting scale structure is shown schematically in (11).

(11) 

A major motivation for the truncation analysis is the infelicity of (12). According to the truncation analysis, *John was later than Bill* is defined only when neither of them was early. The *but*-clause contradicts this definedness condition and therefore is infelicitous.

(12) John was later than Bill, #but they were both early.

It is not entirely clear whether the truncation analysis provides a satisfying account even for *early* and *late*. As Sawada and Grano (2011) acknowledge, the infelicity of (12) is weak and some speakers may think it is fine. Therefore, even though I use their # annotation for this example, I will use ? to better indicate weak infelicity for future examples. In addition, I note that the infelicity seems even weaker when we use the adverb *late* instead. Multiple native speakers I have consulted reported that (13) sounded better than (12), and some found (13) perfectly fine.

(13) John arrived later than Bill, ?but they were both early.

Moreover, there are naturally occurring examples similar to (13).

(14) a. There was traffic around the hotel and an accident on the freeway. I got there a little later than normal, not late. My workday starts at 4:30 p.m., in my opinion. I got to the stadium at 4:04 p.m.[6]

b. Finance Manager Marlene Kelleher said that due to the implementation of the District's new finance software, the audit was being presented slightly later than usual but still on time.[7]

For example, from (14a) we know that the speaker got to the stadium early, and he normally arrived even earlier. Similarly, from (14b) we can infer that the audit was usually presented early. Given that the measure function of *late* is undefined for any time point before "on time" according to the truncation analysis, it will have trouble accounting for (14). Sawada and Grano (2011, fn. 4) tentatively suggest that speakers who accept (12) are resetting the minimum value of the scale so that the scale includes enough time points before "on time," and presumably they would say the same for (14).

Setting aside whether this additional stipulation really provides a satisfying explanation, let us focus on consequences of applying the truncation analysis to *profitable*. Since Company A lost money, the measure function of *profitable* is undefined for it, and therefore all sentences in (15) are predicted to be infelicitous due to undefinedness (their truth conditions are shown on the right for easy reference).

(15) a. ? Company B was more profitable than A. $\mathbf{profit}(\mathbf{b}) > \mathbf{profit}(\mathbf{a})$

b. Company E was more profitable than A. $\mathbf{profit}(\mathbf{e}) > \mathbf{profit}(\mathbf{a})$

c. Company E was much more profitable than A. $\mathbf{profit}(\mathbf{e}) - \mathbf{profit}(\mathbf{a}) \geq \theta_{\text{much}}$

d. Company E was \$101M more profitable than A. $\mathbf{profit}(\mathbf{e}) - \mathbf{profit}(\mathbf{a}) \geq \$101\text{M}$

However, while there might be something slightly weird about (15a), (15b–15d) are clearly true.[8] Therefore, the truncation analysis wrongly predicts (15b–15d) to be infelicitous. Note that the option

[6] http://www.startribune.com/valentine-threat-to-punch-radio-host-was-a-joke/168706046/

[7] https://www.vidwater.org/files/8d80b12d0/2015_01_07_minutes.pdf

[8] For those who need more convincing that (15b) is true and felicitous. Below is a naturally occurring example.

(1) Did you know that your business is more profitable than Amazon.com? Literally.

Sure, Amazon is one of the most well-known companies in the world. And there's no denying that their current of resetting the minimum value of the scale will not explain the contrast between (15a) and (15b). In order for (15b–15d) to be felicitous, we need to be able to reset the minimum value of the scale low enough so that the measure function is defined for Company A. However, if we do that, then the measure function will also be defined for Company B, which means that (15a) is now defined and we lose the initial explanation of why it is infelicitous.

In comparison, the full-range analysis correctly predicts that (15b–15d) are true. It also predicts that (15a) is true. While this leaves the infelicity or weirdness of (15a) unexplained, we can reasonably assume that this is in fact the correct prediction as far as the semantics is concerned. Similar to the discussion above with *early/late* and as indicated by the ? annotation, the infelicity of (15a) is weak, and the dominant intuition reported by the native speakers I have consulted is that (15a) is technically true but very misleading. This strongly suggests that the infelicity of (15a) should be explained in the pragmatics rather than in the semantics. Therefore, I conclude that the full-range analysis, when supplemented with a proper pragmatic theory, should be preferred to the truncation analysis.

Now I turn to the second alternative, proposed by Bierwisch (1989), and call it the *compression* analysis. When applied to *profitable*, it stipulates that the measure function **profit** maps all the companies that lost money (as well as those that broken even) to the zero degree. The resulting scale structure is shown schematically in (16).

(16) 

Bierwisch (1989) proposes the scale structure in (16) to analyze what he calls *evaluative* adjectives such as *pretty/ugly* and *industrious/lazy*. According to him, an evaluative adjective is not directly associated with a measure function. For instance, *pretty* simply denotes an individual property P (which may be context-sensitive), and it becomes gradable only relative to a comparison class. The individuals in the comparison class are ordered wrt the degree or extent to which they satisfy P. In particular, all the individuals that do not satisfy P are mapped (or compressed) to the same zero degree.

The compression analysis correctly predicts that (15b) is true. However, it predicts that (15a) is false. Given the discussion above, this does not seem to be the right way to capture its (weak)

$148 billion market capitalization is likely just a wee bit higher than yours. Or that they employ a few more people (about 110,000) than you or have higher revenues ($74 billion in their last fiscal year). But really and truly, your business is making more money than Amazon.com. In fact, just last quarter, the company lost $126 million and expects to have an operating loss of $810 million this year. Why? According to the company's chief financial officer, a price war over cloud services, the opening of additional warehouses and significant spending on new content were just a few of the reasons.

So congratulations: your business is profitable and has been for years. Amazon's is not.

https://www.inc.com/gene-marks/4-reasons-why-your-business-is-more-profitable-than-amazon.html infelicity, but it is not obviously wrong, either. Its predictions for (15c), however, is much worse. To highlight the problem, note that it predicts that the sentence pairs in (17) and (18) have the same truth values, because the measure function maps A and C to the same value 0. However, these predictions are clearly wrong. We can reasonably say that (17a) is true and (17b) is false, and similarly that (18a) is false and (18b) is true. Finally, (15d) is true but the compression analysis wrongly predicts that it is false.

(17) a. Company D was (only) slightly more profitable than C. $0 < \mathbf{profit}(\mathbf{d}) - \mathbf{profit}(\mathbf{c}) \leq \theta_{\text{slightly}}$

b. Company D was (only) slightly more profitable than A. $0 < \mathbf{profit}(\mathbf{d}) - \mathbf{profit}(\mathbf{a}) \leq \theta_{\text{slightly}}$

(18) a. Company D was much more profitable than C. $\mathbf{profit}(\mathbf{d}) - \mathbf{profit}(\mathbf{c}) \geq \theta_{\text{much}}$

b. Company D was much more profitable than A. $\mathbf{profit}(\mathbf{d}) - \mathbf{profit}(\mathbf{a}) \geq \theta_{\text{much}}$

We have seen that the compression analysis as stated above makes incorrect predictions. However, the compression analysis is actually not the full theory proposed by Bierwisch (1989). According to him, while an evaluative adjective primarily has the compressed scale structure in (16), some evaluative adjectives can also have the full-range scale structure in (7) by conjoining the two compressed scales of the evaluative adjective and its antonym at the zero point. In effect, Bierwisch is proposing a hybrid account that assumes both the full-range scale and the compressed scale.

Nevertheless, in the case of *profitable*, I argue that the full-range analysis (supplemented with a proper pragmatic analysis) should still be preferred to such a hybrid account. First, since the hybrid account also needs to be supplemented with a pragmatic analysis to spell out which scale structure to use in any given context, all else being equal, the full-range analysis should be preferred because it is semantically more parsimonious. Second, note that the full-range scale structure in (7) is needed in order to derive the correct truth values for (15d), (17a), and (18b), and that the compressed scale structure generates wrong truth values for such cases. It would be very difficult to explain why the full-range scale structure is obligatorily used for such cases, in particular if the compressed scale structure is assumed to be the primitive.[9]

Therefore, I conclude that the scale structure of *profitable* is best analyzed as the full-range scale in (7).

To sum up, we have seen that *profitable* arguably has a full-range scale structure that does not have a minimum degree. Nevertheless, it has a clear-cut > 0 interpretation, which uses the zero degree as the standard. This > 0 interpretation is arguably not derived from the mechanism proposed in the previous chapter, which is instead responsible for a relative interpretation of *profitable*. Therefore, we have some initial evidence that minimum adjectives in fact have two types of interpretations, a

[9] This is particularly so for (17b), which is predicted to be true based on the compressed scale structure, but is nevertheless false. This means that we cannot say that the full-range scale is used to salvage an otherwise false/infelicitous sentence.

clear-cut > 0 interpretation and a vague, relative interpretation, and the two interpretations are derived from different mechanisms, i.e., they are due to ambiguity rather than underspecification.

4.3 Further evidence for ambiguity

In the previous section, we have seen that *profitable* provides some initial evidence that minimum adjectives in fact have two types of interpretations and they are derived from different mechanisms. In this section, I provide further evidence for this ambiguity. I will discuss two types of evidence. In Section 4.3.1, I discuss more examples of minimum adjectives and argue that they have different interpretations that can be understood in terms of an ambiguity between relative and > 0 interpretations. In Section 4.3.2, I provide independent evidence for deriving the > 0 interpretation using a different mechanism, by noting the similarities between minimum adjectives and comparative constructions.

4.3.1 Relative and > 0 interpretations

In this section, I discuss more examples of minimum adjectives having different interpretations that can be understood in terms of an ambiguity between relative and > 0 interpretations. However, I note that not all examples provide evidence as decisive as *profitable*. Depending on one's initial theoretical preference, one might be tempted to derive one interpretation from the other with some auxiliary assumptions, and this can be done for most of the examples. However, what I hope to illustrate is that assuming an ambiguity between relative and > 0 interpretations provides a simple explanation that has the best coverage of the empirical data, without the need for further stipulations (even if such stipulations are plausible in many cases).

The first class of examples includes *early/late*, *fast/slow* (for clocks and watches), and *sharp/flat* (for music instruments). We have already seen in the previous section that there are reasons to think that *early* and *late* have a full-range scale. Similarly, the examples below provide evidence to think that *fast/slow* and *sharp/flat* have full-range scales (19).

(19) a. My watch is 2 minutes fast. This clock is 1 minute slow. Therefore, my watch is 3 minutes faster than this clock.

b. The A string on this guitar is two semitones sharp. The one on that guitar is one semitone flat. Therefore, The A string on this guitar is three semitones sharper than the one on that guitar.

However, these examples are different from *profitable* in that the zero standards in their interpretations are more contextual. For example, the interpretations of *fast* and *slow* make reference to a zero point only when we are talking about instruments that measure time, where there is a standard clock whose speed serves as the zero point. Therefore, one might try to derive the > 0 interpretations of these cases in a special way, while assuming a single scale structure. For example, according to

Kennedy (2001), the scale structure of *fast/slow* is always a full range from the absolute zero (i.e., motionless) to infinity, regardless of whether we are talking about clocks or cars. In general, *fast* maps an individual to the interval on the scale from 0 to the individual's speed, and *slow* maps an individual to the interval from the individual's speed to infinity. The interpretations of *fast* and *slow* in (19a) are special in that they map their argument to intervals that extend from the "on time" point to the actual speed, and that they include an additional ZERO function as part of their meanings to shift such intervals to intervals that start from the absolute zero. This is so that they are comparable with measure phrases, which Kennedy (2001) assume denote an interval starting from the absolute zero. Formally, we have semantic representations such as in (20).

(20) $[\![\text{My watch is 2 minutes fast}]\!] = \text{ZERO}(\mathbf{fast}_\delta(\mathbf{w})) \succeq [\![\text{2 minutes}]\!]$,
which is equivalent to $\text{ZERO}(\mathbf{fast}_\delta(\mathbf{w})) \supseteq [0, 2\text{min}]$

Even though Kennedy (2001) uses the name ZERO and talks about zero points of the scale in various places in the prose, he actually defines the function ZERO in terms of the minimal element of the scale, presumably under the assumption that the intuitive notion of a zero point can be formally characterized as the minimal element of the scale.[10] As a result, in order for the ZERO function to be defined, the scale structure needs to have a minimal element. While this is the case for *fast/slow* and *sharp/flat*, it is not obviously true for *early/late*. Sure, it is not uncommon for people to hold mythological, religious, or even scientific beliefs that there is a starting point of time, but it seems a little strange to conclude from this that the scale of *early/late* has a minimal element. I acknowledge that the response to this concern can simply be "well, people just use *early/late* in a way as if the scale had a minimal element, regardless of whether or not they actually believe there is a starting point of time." However, if there is an analysis that has the same or even better empirical coverage and does not have to rely on this specific stipulation about the scale structure of *early/late*, then such an analysis should be preferred.

I suggest that such an analysis does exist. Instead of characterizing the zero point in terms of the minimal element of a scale, we can simply treat it as a primitive. That is, when we specify a scale, we can optionally specify a degree that is considered to be the zero point of that scale. In many cases, such a zero point is lexical and based purely on the property being measured, e.g., the zero point for *tall* is the degree that corresponds to zero amount in height, and the zero point for *profitable* is the degree that corresponds to zero amount of profit. In other cases, the zero point can be more contextual and based on our knowledge about the relevant conventional standard. Here the convention can be widely shared and stable over time, e.g., how fast an accurate clock is supposed to be, but it

[10] There is a further complication. Note that except when we are talking about instruments that measure time, *fast* is a relative adjective. Given the strict correlation between scale structure and interpretation of the positive form assumed by Kennedy and McNally (2005) and Kennedy (2007), they would probably assume that the scale of *fast* does not have a minimal element, but rather a greatest lower bound. However, whether or not one tries to characterize the zero point in terms of a minimal element or a greatest lower bound would not affect the discussion below.

can also be extremely local and highly variable, e.g., the time you and your friend are supposed to meet on various occasions.[11] Different adjectives (or different interpretations of an adjective) differ in whether they allow for a scale with a contextual zero point, which can be implemented in terms of semantic selection rules. Under this analysis, the clear-cut interpretations of *profitable*, *early/late*, *fast/slow*, and *sharp/flat* are all instances of the > 0 interpretation, with the only difference being whether the zero point is lexically determined (*profitable*) or contextually determined (the other ones). This analysis does not make more stipulations than Kennedy's. Whenever the interpretation of an adjective is assumed to include ZERO as part of the meaning in Kennedy's analysis, this analysis assumes that it can select for a scale that has a corresponding contextual zero point. Crucially, this analysis does not need to stipulate that the scale of *early/late* has a minimal element, since any degree can in principle be the contextual zero point. In addition, this analysis has better empirical coverage than Kennedy's (2001), which does not and in fact cannot account for *profitable* (since its zero point cannot be characterized as a minimal element). Therefore I conclude that this analysis should be preferred to Kennedy's (2001).[12]

Note that under this analysis, the zero points of *early/late*, *fast/slow*, and *sharp/flat* are interior points of the scales. Therefore, similar to *profitable*, the > 0 interpretations of these adjectives, which are clear-cut, cannot be derived as relative interpretations. We can also observe the two interpretations in one sentence (21). Here I use 0_c to emphasize that the zero point is contextually determined.

(21) a. John was late$_{>0_c}$, but it was not late$_{\geq\theta}$ for John. (He would normally be at least 20 minutes late, but this time he was only 5 minutes late.)

b. This clock is slow$_{>0_c}$, but not slow$_{\geq\theta}$ for a clock that has not been wound up for a week. (It is only half a minute slow, and typically a clock would be one or two minutes slow after a week.)

Therefore, this first class of adjectives provides further evidence for the ambiguity between relative and > 0 interpretations.

The remaining examples to be discussed below all have a minimal element on the scale that can be seen as a zero point. Given the discussion so far, I will continue to use the term > 0 interpretation instead of minimum interpretation, even though technically the two cannot be teased apart for these examples.

So far we have been using *for*-PPs to introduce relative interpretations that differ from > 0 interpretations. However, it is not always natural to do so with *for*-PPs, and another way to show

[11] Note that a zero point is contextual iff it cannot be determined based on the property being measured (i.e., it is not lexical) and therefore needs to be supplied by context. This definition does not entail that the zero point will vary in different contexts. When the relevant convention is widely shared and stable over time, the zero point will be highly stable across contexts.

[12] Despite this difference, as we will see in Section 4.4, the insights of Kennedy's (2001) analysis are still preserved.

the two different interpretations is by considering polar questions (22).

(22) a. Is John still active$_{>0/\geq\theta}$ (as an actor)?

b. Is this treatment effective$_{>0/\geq\theta}$?

c. Is this disease curable$_{>0/\geq\theta}$?

For instance, (22a) has two interpretations. It can be asking whether John is retired, which is the > 0 interpretation, i.e., whether John is engaged in acting activities at all.[13] It can also be asking whether John has a lot of acting activities compared to other actors, which is the relative interpretation. Similarly, (22b) also has two interpretations. Under the > 0 interpretation, it is asking whether the treatment has any real positive effect at all, while under the relative interpretation, it is asking whether such an effect is good enough (measured in terms of, e.g., cure rate) relative to a comparison class. Note that the comparison class need not be the set of all treatments for the disease under discussion, but can be, e.g., the set of the best available treatment for each disease. Therefore, even if a treatment is currently the best available treatment for a particular disease, it might still fail to count as an effective treatment. This can account for the naturally occurring example below (23a).

(23) a. With a cure rate of only 25.9%, current treatments for eumycetoma, the fungal form of mycetoma, are ineffective and have many side effects![14]

b. Current treatments for eumycetoma are not effective$_{\geq\theta}$

Intuitively, (23a) entails (23b). Since the treatments have a positive (though small) cure rate, the interpretation of *effective* in (23b) must be relative. This example can also help illustrate the ambiguity in (22c). Given the information above, is eumycetoma curable? It seems that technically it is. After all, roughly one in four patients can be cured. However, intuitively it also seems that one could feel reluctant to simply say that eumycetoma is curable.

Depending on one's initial theoretical preference, one might be tempted to deny that there is an ambiguity and derive the different interpretations based on a single primitive reading. On the one hand, if one wants to maintain the tight correspondence between the scale structure and the interpretation of a positive form assumed by Kennedy and McNally (2005) and Kennedy (2007), one can say that the > 0 interpretation is primitive and the relative interpretation is due to imprecision/granularity. According to such an analysis, *eumycetoma is curable* and *its treatments are effective* are both true semantically, and the reason why we may feel reluctant to say them is because the contextual level of imprecision/granularity is high. On the other, recent probabilistic models that attempt to provide a unified semantics for different classes of gradable adjectives assume that

[13] In fact, we can make this > 0 interpretation more salient by explicitly using *at all* in the polar question. I will return to this relation between the > 0 interpretation and the NPI *at all* later when I discuss the examples in (28).

[14] https://www.facebook.com/dndi.org/photos/with-a-cure-rate-of-only-259-current-treatments-for-eumycetoma-the-fungal-form-o/10155829476961455/

the relative interpretation is primitive and the > 0 interpretation is derived when the comparison distribution is highly concentrated on the minimum degree (Lassiter & Goodman, 2013; Qing & Franke, 2014). According to such an analysis, the > 0 interpretations are derived wrt a prior belief that the vast majorities of diseases are not curable at all and that the vast majorities of treatments are not effective at all. While it is difficult to provide knockdown arguments against such approaches, it is also difficult to provide principled, independently motivated justifications for the stipulations such approaches need to make. For approaches that try to derive the relative interpretation in terms of imprecision/granularity, the difficulty is to justify the relatively high level of imprecision/granularity it sometimes needs to assume. Note that under the relative interpretation, a treatment with a cure rate of 70% or 80% can still be considered not effective if there are alternatives with much higher cure rates,e.g., 95%. The level of imprecision/granularity would not to be unreasonably high in order to account for such interpretations.[15] Approaches that try to derive the > 0 interpretation as a relative interpretation also seem to make assumptions that can be difficult to justify. When we say a disease such as Alzheimer's is not curable, intuitively what we mean is that there is no known cure for it, and it seems irrelevant whether we believe other diseases have a cure or how good the treatments for other diseases are. Therefore, it seems a little too strong to require that we believe the vast majority of diseases do not have a cure. In general, it is not clear whether similar requirements would be plausible for other minimum adjectives.

For instance, consider adjectives that describe tastes such as *sweet*, *salty*, and *spicy*. I will call such adjectives *taste adjectives*. This unfortunately can be a little confusing because such adjectives also belong to the class of *predicates of personal tastes* (e.g., Lasersohn, 2005), which are often called *taste predicates* for short. It helps to keep in mind that here I am using *taste adjectives* in a very literal and narrow sense, and if needed, I will always use the full term *predicates of personal tastes*. The literature on predicates of personal tastes has mostly focused on the subjective nature of such predicates. Taste adjectives are indeed subjective. Our taste buds can be different and so we may have different levels of sensitivity. As a result, we can have different experiences when we are tasting the same food, which means that we can disagree on how something tastes without one of us being wrong. However, this subjectivity is not the focus of the discussion below, and to abstract away from it as much as possible, I will assume that the interlocutors have identical tasting experiences and fully agree on how something tastes.

Intuitively, taste adjectives have a minimum degree on the scale, which corresponds to a complete lack of the relevant taste. However, they readily have relative interpretations. For instance, *the maple*

[15] In this case, there is a complication due to the fact that cure rates have a maximum degree, i.e., 100%, and therefore one might think that this relative interpretation of *effective* is actually a maximum interpretation. However, this strategy would not work in general as the scale of *effective* need not have a maximum degree. For instance, suppose we are talking about various steroids and the effectiveness is measured in terms of the amount of muscle gain after a period of time with no training. In this case, since the amount of muscle gain has no maximum, *effective* does not have a maximum interpretation. However, a steroid that allows you to grow, e.g., 6 pounds of muscles can still be considered not effective if the steroids in the comparison class generally allow you to grow even more.

syrup is sweet presumably involves a higher standard than *the wine is sweet*. We can also make the relative interpretation more explicit by using a *for*-PP, as shown in the naturally occurring examples in (24).

(24) a. Solid radler, sweet$_{\geq\theta}$ for a [S]tone beer [Stone is the brewery's name] but not sweet$_{\geq\theta'}$ for a radler, very drinkable in summer and lemony, nice color[16]

b. The chicken is really good, and not salty$_{\geq\theta}$ for a Mexican restaurant [-] a problem I often face.[17]

c. The food ordered was a little tasteless and not spicy$_{\geq\theta}$ for a South Indian cuisine. We had requested the food to be medium hot — but all dishes were served mild.[18]

Such relative interpretations are problematic for approaches that assume a perfect correspondence between the scale structure and the interpretation of the positive form.

Given that the relative interpretation entails the > 0 interpretation when 0 is the minimum degree of the scale, the > 0 interpretation can be hard to detect, particularly when the adjective is in a predicative position, e.g., *the dish is salty*. Moreover, such uses often have an excessive interpretation, i.e., *the dish is salty* can be interpreted as *the dish is too salty*.

However, I think there are still good reasons to think that the > 0 interpretation exists for taste adjectives. Note that for now I focus on showing that the standard of the positive form can be very low so that anything with a non-zero degree will be in its extension, and remain neutral about whether this > 0 interpretation can be derived as a relative interpretation. Also, I assume that the relevant measures in the examples below are about how the food actually tastes, which encode the appropriate level of granularity. For instance, if you add a tiny pinch of salt to a large bowl of soup, the soup is saltier than before in the sense that it contains a larger amount of salt. However, as far as the actual taste is concerned, the soup is presumably not saltier because it is unlikely that one can taste the difference. In other words, under such a measure, something must actually taste saltier than something that is completely tasteless in order to have a degree of saltiness greater than 0.

With this assumption in mind, we can see that the attributive use of *salty* in the headline in (25) arguably has a > 0 interpretation.

(25) [Headline] 6 Ways to Add Salty$_{>0}$ Flavor Without Salt

[Main text] The level of salt in olives runs a wide gamut, from mildly salty fresh green [C]astelvetrano olives to deeply salty oil-cured black olives, which gives you lots of options.[19]

---

[16] https://untappd.com/user/SophieSteiner/checkin/756698546

[17] https://www.yelp.com/biz/chipotle-mexican-grill-santa-cruz?hrid=eNXek1LdReEqDrunMNGsCw&utm_campaign=embed_body&utm_medium=embedded_review

[18] https://www.tripadvisor.ca/LocationPhotoDirectLink-g186338-d2699789-i289812476-Anjanaas-London_England.html

[19] https://www.epicurious.com/expert-advice/how-to-add-salty-flavor-to-food-without-using-salt-article

Intuitively, to add *salty flavor* only requires that the flavor has a non-zero degree of saltiness. This is indeed supported by the main text, where the author suggests that the mildly salty fresh green Castelvetrano olives provide an option to add salty flavor. Just how salty the mildly salty Castelvetrano olives are? The description in (26) suggests that their saltiness can be very low. Therefore, we have reasons to believe that the author of (25) intends the > 0 interpretation of *salty* in the headline.

(26) These Castelvetrano Olives offer a mild and buttery flavor. Whole and un-pitted, their mild flavor comes across as slightly sweet with a hint of salt.[20]

Essentially, the above discussion shows that *mildly salty* entails *salty* in some sense, i.e., when *salty* has the > 0 interpretation. However, if we try to evaluate the entailment pattern directly by judging whether something is mildly salty is salty, our judgments can be easily interfered by the fact that the more salient interpretation of *salty* is the relative one that has a much higher standard.

Similarly, the following Reddit thread (27) suggests that *spicy* also has a > 0 interpretation.[21]

(27) Does mild mean not spicy or less spicy[?][22]

a. Less spicy, very minor amount of heat.

b. Nowadays mild is practically no spice, or too many people complain. Used to be slightly spicy.

c. Depends on what you're buying and where.

Curry in a restaurant that's in a Little India district where Indian people eat? Mild won't blow your skull off but the spice will be detectable. I've been to places that advertise hot, medium, mild, and *no hot pepper* [emphasis the author's].

Salsa in a jar in a grocery store in a really white little American town? Tastes like watery tomato soup and I don't believe them when the list of ingredients says "jalapeno".

Given that the thread title is an alternative question, the second disjunct is presumably interpreted as *less spicy but still spicy*, where the positive form (the second occurrence of *spicy*) has the same standard as the *spicy* in the first disjunct. What should this standard be? The responses all suggest that the standard is the zero degree. The author of (27a) considers "very minor amount of heat" good enough for the second alternative. The author of (27b) thinks *mild* means "practically no spice" nowadays and used to mean "slightly spicy." It is natural to understand these as the contrast between the two alternatives in the question, and therefore the relevant standard is again 0. This is made more explicit by the author of (27c). The last paragraph corresponds to the first alternative, according to which the Salsa "tastes like watery tomato soup" and therefore has zero spiciness. The

[20] https://www.pasolivo.com/castelvetrano-olives.html

[21] However, given the nature of Reddit threads, it is hard to know whether the authors are native speakers of English.

[22] https://www.reddit.com/r/NoStupidQuestions/comments/973a5j/does_mild_mean_not_spicy_or_less_spicy/

second paragraph corresponds to the second alternative, according to which *mild* means that "the spice will be detectable." Once again, the contrast between the two alternative answers suggests that the standard being used for *spicy* is the zero degree.

The discussion above is based on direct intuitions. Below I will provide some further evidence based on theoretical considerations. The natural occurring examples in (28) use *almost not Adj* and the latter two also have an NPI (*whatsoever*, *at all*).

(28) a. So I searched online and found a perfect recipe for that — coconut-based strawberry mousse with nut crumble. The mousse itself turned out to be very light, fluffy and almost not $\text{sweet}_{>0}$, but the sweetness and crunchiness of the crumble complemented it just perfectly.[23]

b. As usual, the Kikkoman provided the intense bomb of saltiness that I was accustomed to. Little Soya, on the other hand, is almost not $\text{salty}_{>0}$ whatsoever.[24]

c. I know I can have rice, and plain Naan bread. I just don't really handle spicy foods very well and tend to avoid them as much as possible. My choice would have to be almost not $\text{spicy}_{>0}$ at all, or extremely mild.[25]

In general, *almost* is not very compatible with vague predicates, whether or not they are negated (29).

(29) a. ? This is almost tall/fast/expensive. . .

b. ? This is almost not tall/fast/expensive. . .

The sentences in (29) sound a little odd, but not completely uninterpretable. This can be explained basically in the same way as Solt's (2012) analysis of *slightly*. The meaning of *almost* requires that the degree be close to the standard of the adjective. However, if the adjective has a vague standard, then we may not have a consistent way to tell whether a degree is close to the standard. For instance, if we consider a difference of $1''$ to be close but the standard of *tall* varies from $5'11''$ to $6'1''$, then we cannot consistently tell whether someone who is $5'10''$ is close to the standard. If this analysis is correct, it means that the taste adjectives in (28) are not vague. Given that their interpretations are clearly not excessive or maximum interpretations (which are also not vague), they must be > 0 interpretations.

The examples with NPIs provide further evidence for the > 0 interpretation. Take the NPI *at all* for instance. At first sight, it seems compatible with all three traditional classes of gradable adjectives (30).

(30) a. This is not full/straight/dry at all. (maximum)

[23] https://www.thrivingbeet.com/blog/strawberry-coconut-mousse-with-nut-crumble

[24] https://marymakesdinner.typepad.com/marymakesdinner/2012/10/little-soya-soy-sauce.html

[25] https://www.quora.com/What-are-some-safe-non-spicy-Indian-food-to-order/answer/Lisa-Marie-Tew?ch=10&share=da0f2c30&srid=kUUC

b. This is not tall/expensive/fast at all. (relative)

c. This is not wet/bent/dirty at all. (minimum)

However, note that if an adjective does not have a > 0 interpretation, *not Adj at all* does not require that the relevant degree be at most 0 (if 0 is the minimum degree then the requirement is simply that the degree be 0). For instance, *not expensive at all* does not mean *free*, as shown by the felicity of (31a). Similarly, *not full at all* does not mean *(completely) empty* (31b). Also, note that *not Adj at all* is vague in this case.

(31) a. This brand new laptop is not expensive at all. It only costs $200.

b. The flight was not full at all. Only one third of the seats were taken.

In contrast, for adjectives that have a > 0 interpretation, *not Adj at all* has an interpretation which requires the relevant degree be at most 0, and this interpretation is not vague. This is the case for (30c), as well as for open-scale adjectives that have a > 0 interpretation (32). Finally, note that when such an adjective also has an easily accessible relative interpretation (which may even be the more salient one), *not Adj at all* can be used to make clear that the > 0 interpretation of Adj is the intended one (32a), otherwise *not Adj at all* has the effect of reducing the permissible level of imprecision (30c, 32b).

(32) a. This company was not profitable at all.

b. John was not late at all.

We can draw two conclusions from the above observations. First, *at all* seems to be targeting the interpretation of the positive form, rather than the scale structure of the gradable adjective. This is because even though *expensive* and *full* do have a zero degree on the scale, the corresponding *not Adj at all* construction does not require the relevant degree be (at most) 0. Second, the relative and the > 0 interpretations are two distinct readings. This is because the effect of *not Adj at all* is sensitive to whether *Adj* has a > 0 interpretation and how salient its relative interpretation is. On the one hand, if the adjective has a relative interpretation (recall from the previous chapter that I treat a maximum interpretation as a special case of a relative interpretation), *at all* effectively lowers its relative standard to a vague standard, which produces a strengthened meaning after composing with negation.[26] On the other hand, if the adjective has a > 0 interpretation, then *at all* does not change this zero standard, and the strengthening under negation is due to either choosing the > 0 interpretation over the relative one, or lowering the permissible level of imprecision. Note that it might be tempting to treat the second case as a special case of the first one, i.e., *at all* simply lowers the relative standard of the adjective and in the case of minimum adjectives the standards

[26] To be clear, I am not suggesting that *at all* literally lowers the standard of the adjective. Rather, I am only describing its effect in an intuitive way. Ultimately, the analysis should be integrated in a general theory of NPI-licensing.

are "squished" to the minimum endpoint, creating the appearance of a non-vague > 0 interpretation and/or a lowered permissible level of imprecision. However, this strategy would fail to derive a non-vague interpretation when the zero degree is on an open scale: (32) would be wrongly predicted to only have a vague interpretation just as (31).[27]

The interpretation of *Adj at all* in a polar question similarly exhibits a contrast between the relative and > 0 interpretations. On the one hand, when the adjective has a relative interpretation, it is generally a bit odd to use *Adj at all* in a polar question out of the blue (33). Without additional contextual support, it is unclear what exactly is being asked.

(33) a. ? Is John tall at all?
  b. ? Is this laptop expensive at all?
  c. ? Was the flight full at all?

On the other hand, when the adjective has a > 0 interpretation, it is clear what *Adj at all* in a polar question means: once again, the effect of *at all* is either to reduce the level of permissible imprecision or to make clear that the > 0 interpretation of the adjective is the intended one (34).

(34) a. Is this wet/bent/dirty at all?
  b. Was John late at all?
  c. Was this company profitable at all?
  d. Is this vaccine/treatment effective at all?
  e. Is this disease curable at all?

Now let us return to (28c). Given that *not spicy at all* there does mean "no taste of hotness," which is then modified by *almost* (another indicator that its meaning is not vague), we can conclude that *spicy* has a > 0 interpretation. This is further supported by the fact that *is this dish spicy at all?* has a clear interpretation asking whether the dish has any taste of hotness.

We have seen reasons to think that taste adjectives do have > 0 interpretations. In order to derive them as relative interpretations, we need the comparison distribution to highly concentrate on the zero degree. However, since multiple taste adjectives can have > 0 interpretations in an alternative question (35), we would need to assume that the vast majority of foods and condiments are completely tasteless, which seems implausible. This provides further evidence that the > 0 interpretation is not derived from the relative one.

(35) Put small portions of five to 10 foods and condiments into bowls. Then, blindfold your child. Ask him to identify the food first by smell and then by taste. Sometimes, the smell and taste

[27] Note that (32a) does have a vague interpretation, which can be seen more clearly in the following example. *Company D only made $5K last year, which was not profitable at all (for an oil company).* The issue is that (32a) also has a non-vague, > 0 interpretation, which cannot be derived by simply lowering the (vague) relative standard.

answers may differ. Then, ask him to describe the food. Is it sweet$_{>0}$ or salty$_{>0}$, sour$_{>0}$ or spicy$_{>0}$?[28]

To conclude, we have seen more examples of minimum adjectives having both relative and > 0 interpretations, which are arguably different readings. Below, I discuss relevant data concerning comparative and excessive *too* constructions, and argue that existing analyses of such data can be easily extended to account for the > 0 interpretation of a minimum adjective.

4.3.2 Comparative and excessive *too* constructions

In English, measure phrases such as *3 inches* can appear with a plain gradable adjective, as well as in comparative and excessive *too* constructions (36).[29]

(36) a. This glass is 3 inches tall. (plain)

b. This glass is 3 inches taller than that glass. (comparative)

c. This glass is 3 inches too tall. (excessive *too*)

Whether a measure phrase can appear with a plain gradable adjective is generally idiosyncratic both within a language and crosslinguistically. For instance, Schwarzschild (2005) observes that in English we cannot say *5 pounds heavy* (cf. 36a), but we can say so in other languages such as Italian and German. In contrast, measure phrases are generally more compatible with comparative and excessive *too* constructions. Crucially, as Schwarzschild (2005) observes, there seems to be a universal that if a language allows measure phrases to appear with a plain adjective (he calls such measure phrases *direct* measure phrases), then it also allows them in the corresponding comparative and excessive *too* constructions (he calls such measure phrases *indirect* measure phrases).

Note that the reverse of this universal generally does not hold, For instance, in English we can say *this bag is 5 pounds more/too heavy* but not *this bag is 5 pounds heavy*. However, Sawada and Grano (2011) observes that this reversal does seem to hold for the class of minimum adjectives. For instance, even though direct measure phrases are uniformly banned for plain lower-open-scale adjectives in Japanese, Spanish, Korean and Russian, they are compatible with comparatives, as well as plain minimum adjectives such as *bent*.[30]

[28] https://www.baltimoresun.com/features/bal-the-dreaded-snow-day-20130115-story.html

[29] Here I use the term "plain" instead of "positive form of" to avoid the implication that a silent morpheme *pos* is present.

[30] There is a further point of variation that Sawada and Grano (2011) discuss. Measure phrases can in fact appear with plain adjectives in Japanese, but in such cases they must be interpreted as comparatives with implicit standards (unlike the other three languages, where even such interpretations are impossible). In fact, similar phenomena can be observed in English. For example, while it is generally weird to say *John is* 5′5″ *short*, we can say *the sleeves are 1 inch short*. However, it does not mean that the length of the sleeves is 1 inch, but rather that the sleeves are 1 inch too short. I will follow Sawada and Grano (2011) in assuming that in such cases the plain adjectives are coerced to comparative interpretations.

These two universals suggest that there is something about minimum adjectives that is shared by comparative and excessive *too* constructions, which lower-open-scale adjectives do not share. This provides some initial motivation to look into comparative and excessive *too* constructions more closely to see whether there is something that can inspire our analysis of minimum adjectives.

Intuitively, the measure phrase *3 inches* in (36b) is measuring the *difference* between the degree of the subject and the degree of the *than*-phrase. For this reason, comparatives that are modified by measure phrases are also called *differential comparatives*. Similarly, the measure phrase in (36c) is measuring the degree difference between the subject and some implicit standard that is considered appropriate in the context. Given that excessive *too* constructions can be seen as a type of comparative constructions that have implicit standards, from now on I will only discuss comparative constructions.

This provides us with a way to think of a plain (or bare) comparative such as (37) in terms of differential comparatives. Intuitively, (37) is true iff John's height is greater than Mary's, and this is equivalent to saying that the height difference between John and Mary is greater than 0. In this respect, we can say that plain comparatives and plain minimum adjectives are similar in that they both have > 0 interpretations.

(37) John is taller than Mary

Meanwhile, plain comparatives and plain minimum adjectives have a crucial difference. In the previous sections, we saw a lot of evidence that minimum adjectives have both relative and > 0 interpretations. In contrast, plain comparatives, whether or not they are formed by relative or minimum adjectives, never have relative interpretations. For instance, (37) does not have an interpretation that requires the height difference between John and Mary be large enough wrt a comparison class, even if we try to induce such an interpretation by making the comparison class explicit (38).

(38) a. # John is taller than Mary, but not taller than Mary for a professional basketball player.
(cf. John is tall, but not tall for a professional basketball player.)

b. # Compared to other professional basketball players, John is not taller than Mary.
(cf. Compared to other professional basketball players, John is not tall.)

Similarly, plain comparatives formed by minimum adjectives do not have relative interpretations, either (39).

(39) a. # Company D was more profitable than C, but not more profitable than C for an oil company.
(cf. Company D was profitable, but not profitable for an oil company.)

b. # Compared to other oil companies, Company D was not more profitable than C.
(cf. Compared to other oil companies, Company D was not profitable.)

Note that here I am merely extrapolating the relative interpretations of plain gradable adjectives to plain comparatives to spell out what relative interpretations of plain comparatives would be like if they did exist, and asserting that they in fact do not exist. At this point it might be unclear why one would expect such relative interpretations of plain comparatives to exist in the first place. This issue depends on how one analyzes plain gradable adjectives and comparatives, and will become clearer when I discuss Sawada and Grano's (2011) analysis later. For now, all that matters is the fact that plain comparatives do not have relative interpretations whereas plain minimum adjectives do.

In the previous sections, I provided arguments that the relative and > 0 interpretations of plain minimum adjectives are derived from different mechanisms. Given that plain comparatives pattern with plain minimum adjectives in terms of measure phrase modification but only have > 0 interpretations, it is natural to suspect that the mechanism for the > 0 interpretations of plain comparatives is also responsible for deriving the > 0 interpretations of plain minimum adjectives. In the next section, I argue that this is indeed the case, and show that the mechanism(s) independently developed for comparatives in the literature can be straightforwardly extended to account for the > 0 interpretations of plain minimum adjectives.

4.4 A compositional analysis of the > 0 interpretation

In this section, I introduce a compositional analysis of the > 0 interpretation. The analysis is largely based on Schwarzschild and Wilkinson's (2002), Schwarzschild's (2005), and Sawada and Grano's (2011) analyses, with slight adaptations.

The central assumption in Schwarzschild's (2005) analysis is that a comparative construction such as *John is taller than Mary* represents a *gap* between John and Mary's heights, which he writes as $[h_m \rightarrow h_j]$. Consequently, a measure phrase such as *2 inches* is a predicate over gaps. There are many ways to formally represent such a gap based on more primitive mathematical constructions. I will follow Schwarzschild and Wilkinson (2002) in assuming that the gap is represented as a degree interval. Concretely, the compositional derivation of *John is 2 inches taller than Mary* is shown as follows (40).

(40) John is 2 inches taller than Mary.

a. $[\![\text{-er}]\!] = \lambda g \lambda y \lambda x.[g(y), g(x)]$, where $[g(y), g(x)]$ is defined as the set $\{d \mid g(y) \leq d \leq g(x)\}$

b. $[\![\text{John is \_\_ taller than Mary}]\!] = [\textbf{height}(\textbf{m}), \textbf{height}(\textbf{j})]$

c. $[\![\text{2 inches}]\!] = \lambda I.\, \mu(I) \geq 2''$, where μ is the appropriate measure of the size of the interval

d. $[\![\text{John is 2 inches taller than Mary}]\!] = \mu([\textbf{height}(\textbf{m}), \textbf{height}(\textbf{j})]) \geq 2''$

For simplicity, I only consider phrasal comparatives, for which the construction of the relevant interval is quite straightforward, as shown in (40a) and (40b). In (40b) I use the underline as a placeholder of the measure phrase, which will take scope over the rest of the sentence. This is to abstract away from the specific details about scope taking to highlight the core idea of this analysis. As an analogy, even though in *John did not see Mary*, the word *not* syntactically composes with the VP rather than the full sentence, there is still a sense in which the propositional negation $\neg$ is the core meaning of *not*. The rest is a matter of scope taking, and can be implemented in a number of different ways.

The measure phrase *2 inches* takes scope over the rest of the sentence, which denotes the degree interval from Mary's height to John's, and returns true iff the size of this interval argument at least $2''$ (40c).[31] When the measure phrase is composed with the rest of the sentence, we obtain the truth condition that the size of the interval is at least $2''$ (40d). Note that $\mu([\mathbf{height}(\mathbf{m}), \mathbf{height}(\mathbf{j})]) = \mathbf{height}(\mathbf{j}) - \mathbf{height}(\mathbf{m})$ when $\mathbf{height}(\mathbf{j}) \geq \mathbf{height}(\mathbf{m})$. This equality holds because according to measurement theory, **height** corresponds to an interval scale (which in turn is because it in fact corresponds to a ratio scale) (e.g., Krantz, Luce, Suppes, & Tversky, 1971). When $\mathbf{height}(\mathbf{j}) < \mathbf{height}(\mathbf{m})$, according to our definition, the interval $[\mathbf{height}(\mathbf{m}), \mathbf{height}(\mathbf{j})]$ is the empty set, which means that its measure would be 0. Therefore, (40d) is equivalent to $\mathbf{height}(\mathbf{j}) - \mathbf{height}(\mathbf{m}) \geq 2''$, which is intuitively the correct truth condition.

When there is no overt measure phrase in the comparative construction, Schwarzschild and Wilkinson (2002) assume that a silent SOME, as defined in (41), is applied to the degree interval.

(41) SOME(J) is true iff $\mu(J) \geq \delta$, where δ is determined by context

According to Schwarzschild and Wilkinson (2002), this definition is motivated by the mass quantifier *some*, and they use the contrast in (42) to motivate the introduction of the contextual parameter δ, i.e., whether a very tiny piece of wood is significant or relevant enough to be taken into account depends on the context.

(42) a. There is some wood in my eye.

b. There is some wood in my truck.

Essentially, the δ parameter is intended to capture the contextual level of imprecision/granularity. Since granularity is encoded in the measure function in my system and imprecision is treated as a pragmatic phenomenon, I will remove the δ parameter from the definition of SOME and use (43) instead.

(43) SOME(J) is true iff $\mu(J) > 0$

Also, note that even though Schwarzschild and Wilkinson (2002) use *some* to motivate SOME, bare mass nouns perhaps provide an even better motivation for this silent operator (44). There is no

[31] I will set aside the issue of the well-known ambiguity between the one-sided (i.e., at least) and two-sided (i.e., exactly) interpretations. Here I choose the one-sided interpretation in the semantics just for illustration.

overt mass quantifier in (44), but the sentence is interpreted in parallel with (42). Therefore, it is reasonable to assume that there is a silent counterpart of *some*, written as $\emptyset_{\text{SOME}}$, that has the same semantics as defined in (43).

(44) There is water in the glass.

With these assumptions in place, the derivation of *John is taller than Mary* is shown in (45).

(45) John is taller than Mary.

a. $[\![$John is __ taller than Mary$]\!] = [\mathbf{height}(\mathbf{m}), \mathbf{height}(\mathbf{j})]$

b. $[\![$John is $\emptyset_{\text{SOME}}$ taller than Mary$]\!] = \mu([\mathbf{height}(\mathbf{m}), \mathbf{height}(\mathbf{j})]) > 0$,
which is equivalent to $\mathbf{height}(\mathbf{j}) > \mathbf{height}(\mathbf{m})$

This silent operator can similarly help us derive the > 0 interpretation of the positive form of a minimum adjective. I will use *profitable* as a concrete example. Note that it denotes a measure function **profit**, whose type is $e \to d$. After composing it with a type e subject, we will end up having a degree, rather than a degree interval expected by $\emptyset_{\text{SOME}}$.

To avoid this type mismatch, we need some way for *profitable* to return an appropriate degree interval. Following Svenonius and Kennedy (2006) and Sawada and Grano (2011), I assume that *profitable* in this case is first composed with a silent head *Meas*, and then composed with the measure phrase (which is the specifier of *Meas*). My definition of *Meas* in (46a) is different from theirs, but as will be discussed later, the definition can easily be modified to capture the core insights in their analyses.

(46) Company A was profitable$_{>0}$.

a. $[\![Meas]\!] = \lambda g : \underline{g^{-1}(0_g) \neq \emptyset}.\ \lambda x.[0_g, g(x)]$

b. $[\![Meas$ profitable$]\!] = \lambda x.[0_{\mathbf{profit}}, \mathbf{profit}(x)]$

c. $[\![$Company A was __ $Meas$ profitable$]\!] = [0_{\mathbf{profit}}, \mathbf{profit}(\mathbf{a})]$

d. $[\![$Company A was $\emptyset_{\text{SOME}}$ $Meas$ profitable$]\!] = \mu([0_{\mathbf{profit}}, \mathbf{profit}(\mathbf{a})]) > 0$,
which is equivalent to $\mathbf{profit}(\mathbf{a}) > 0_{\mathbf{profit}}$

The basic role of *Meas* is to turn a measure function g into a function that maps an individual x to a degree interval starting from the zero point of the scale 0_g to the individual's measure $g(x)$ (46a). It also has a general selectional restriction on the measure function, underlined in (46a), according to which the inverse image of the zero point must be non-empty.[32] This requirement is adapted from Sawada and Grano (2011). Intuitively, $g^{-1}(0_g) \neq \emptyset$ means that the zero point of the scale must be

[32] $g^{-1}(0_g)$ is defined to be the set $\{x \mid g(x) = 0_g\}$. Strictly speaking, this is the inverse image of the singleton set $\{0_g\}$, or the *fiber* of the element 0_g. However, I will continue to call $g^{-1}(0_g)$ the inverse image of 0_g because the term is more familiar.

realizable. This will rule out adjectives such as *tall* and *heavy*, because nothing can have zero height or weight.[33] However, following Sawada and Grano (2011), I assume that this selectional requirement can be idiosyncratically overridden. As a result, in English *tall* is compatible with measure phrases but *heavy* is not, whereas in German and Italian *heavy* is also compatible with measure phrases. In contrast, this requirement is always satisfied by a minimum adjective, and therefore it can always be composed with *Meas* (46b) and then the subject (46c) to obtain a degree interval, which is the same type of meaning representations as comparatives (cf. 40b). This accounts for why minimum adjectives pattern with comparatives in terms of measure phrase modification. Once we have the degree interval, we can apply the silent $\emptyset_{\text{SOME}}$ operator to derive the > 0 interpretation (46d). We can also compose it with an overt measure phrase and derive the correct truth condition (47). This accounts for the naturally occurring examples in (48).[34]

(47) $[\![$Company A was 3 million dollars *Meas* profitable$]\!] = \mu([0_{\mathbf{profit}}, \mathbf{profit}(\mathbf{a})]) \geq \$3\text{M}$, which is equivalent to $\mathbf{profit}(\mathbf{a}) \geq \$3\text{M}$

(48) a. The arts help to make Indianapolis profitable — about $468 million dollars profitable as the Arts in American Profitability study has shown.[35]

b. "I, uh. . . " Williams cleared his throat. "Did you get my telegram, Sheri–uh, Mr. Staley?" Staley turned toward him. "Yeah, I got your telegram," he said.

"Then you know that I have a proposition for you."

"I believe what you said was that you had a *profitable* proposition for me," Staley said, emphasizing the word "profitable."

"Yes. Indeed, it could be very profitable," Williams replied.

"How profitable?"

"Five thousand dollars profitable," Williams said.[36]

In sum, we have seen that the proposed analysis provides a unified compositional account for comparative constructions and the > 0 interpretation of minimum adjectives. It also accounts for why minimum adjectives pattern with comparative constructions in terms of measure phrase modification. Comparative constructions by definition denote degree intervals and therefore they are automatically

[33] While this is presumably true for *tall*, it is more complicated for *heavy*. Given the way gravity and buoyancy work, it is possible for, e.g., a balloon to be completely weightless if the relevant measure is how much effort we need to lift it from the ground. Therefore it is not totally obvious that nothing can have zero weight. In general, I think the compatibility between measure phrases and plain gradable adjectives is a complex issue that has not been solved completely. The selectional requirement proposed here is to maintain maximal compatibility with Sawada and Grano's (2011) analysis.

[34] Note that these examples may be special in that the positive forms have already been separately used before the measure phrases are introduced.

[35] https://saveindyarts.wordpress.com/2008/08/04/vote-here/

[36] *Rampage of the Mountain Man*, by William W. Johnstone, retrieved from Google Books.

compatible with measure phrases (to the extent that there is a conventional measure, e.g., there is no conventional measure of how much dirtier something is than something else). Minimum adjectives have a realizable zero degree on the scale, which allows them to be composed with *Meas* so that we can obtain the degree interval to be predicated on by the measure phrase. In contrast, adjectives that do not have a realizable zero degree (either because there is simply no zero degree on the scale or because the zero degree is not realizable by any individual) are generally incompatible with measure phrases, but we can assume that some languages may allow for an idiosyncratic set of such adjectives to override this restriction.

The current analysis draws insights from two lines of approaches in the literature. The treatment of comparatives is adapted from Schwarzschild and Wilkinson (2002) and Schwarzschild (2005), and the treatment of plain gradable adjectives is adapted from Sawada and Grano (2011), whose analysis is in turn adapted from Svenonius and Kennedy (2006). The two lines of approaches are similar in many respects but also different in some. In the remaining part of this section, I will discuss how the current analysis can be modified so that it is more in line with Sawada and Grano's and Svenonius and Kennedy's analyses. The goal is to illustrate that some of the apparent differences between the current analysis and their analyses are not important for my purposes. In the next section, however, I will discuss how my analysis is crucially different from theirs in terms of how the > 0 interpretation is derived, and show the problem with the unified analysis of plain gradable adjectives and comparatives that Sawada and Grano (2011) and Svenonius and Kennedy (2006) propose.

Under the current analysis, measure phrases are introduced in two different ways. On the one hand, measure phrases are directly introduced in comparative constructions, since they are semantically compatible with each other. On the other, when modifying a plain gradable adjective, the measure phrase is introduced as the specifier of *Meas*. However, Sawada and Grano (2011) and Svenonius and Kennedy (2006) provide arguments for a more unified syntax, according to which a measure phrase is always introduced as the specifier of *Meas*. This can be achieved by a slight modification of the current analysis. Instead of treating the comparative morpheme as introducing intervals, we follow Sawada and Grano (2011) and Svenonius and Kennedy (2006) in assuming that it simply "resets" the zero point of scale to the degree of the *than*-phrase. However, there is a crucial difference. Since Sawada and Grano (2011) and Svenonius and Kennedy (2006) assume that the zero point is characterized by the minimum degree of the scale, they will have to truncate or compress the original scale so that the degree of the *than*-phrase is the new minimum degree. In contrast, since I take the zero point as the primitive, all I need to do is change the specification of the zero point. Formally, the scale that corresponds to a measure function g is specified as a tuple $\langle D_g, \leq_g, 0_g\rangle$, where D_g is a set of degrees and $\leq_g$ is an ordering on D_g, just as in classic approaches.[37] The third component 0_g

[37]Note that a degree in D_g need not to be realizable, and therefore in principle we can always assume that D_g is the set of extended real numbers (i.e., real numbers and $\pm\infty$). The set of realizable degrees can be retrieved as Range(g). Since the unrealizable degrees are generally useless, it is often enough to think of D_g as Range(g) together with its

is an element in D that is considered to be the zero point of the scale. For a measure function g and a degree $d \in D_g$, we define $g_{d\to 0}$ as the measure function that corresponds to the scale $\langle D_g, \leq_g, d\rangle$ and satisfies $g_{d\to 0}(x) = g(x)$ for any x in the domain. Intuitively, $g_{d\to 0}$ is the result of resetting the scale so that d is considered to be the zero point of the new scale. This new scale is identical to the original one except that a new degree is considered to be the zero point. Crucially, as far as the output values are concerned, $g_{d\to 0}$ is the same as g. Nothing gets truncated or compressed.

With these modifications in place, the derivation of a differential comparative is shown in (49).

(49) John is 2 inches taller than Mary.

a. $[\![\text{-er}]\!] = \lambda g \lambda y \lambda x. g_{g(y)\to 0}(x)$

b. $[\![\text{taller than Mary}]\!] = \lambda x.\mathbf{height}_{\mathbf{height(m)}\to 0}(\mathbf{x})$

c. $[\![\textit{Meas}\text{ taller than Mary}]\!] = \lambda x.[\mathbf{height}(\mathbf{m}), \mathbf{height}(\mathbf{x})]$

d. $[\![\text{John is }\_\_\ \textit{Meas}\text{ taller than Mary}]\!] = [\mathbf{height}(\mathbf{m}), \mathbf{height}(\mathbf{j})]$

e. $[\![\text{John is 2 inches }\textit{Meas}\text{ taller than Mary}]\!] = \mu([\mathbf{height}(\mathbf{m}), \mathbf{height}(\mathbf{j})]) \geq 2''$,
which is equivalent to $\mathbf{height}(\mathbf{j}) - \mathbf{height}(\mathbf{m}) \geq 2''$

In this new system, the comparative morpheme resets the zero of the scale to the degree of the *than*-phrase (49a). As a result, *taller than Mary* denotes a measure function that is basically the same as **height** except that the height of Mary is considered to be zero point of the scale (49b). We need to compose it with *Meas* to obtain the relevant degree interval (49c). This composition is always allowed because the measure function denoted by *taller than Mary* has a realizable zero point on the scale, i.e., Mary's height, and the result is a function that maps an individual x to a degree interval starting from Mary's height (since it is the zero point of this new scale) to the height of x (since the measure function is otherwise identical to **height**). Now that we obtain the relevant degree interval, the rest of the composition (49d, 49e) can proceed just as in the original analysis. Plain comparatives can be derived similarly (50b).

(50) John is taller than Mary.

a. $[\![\text{John is }\_\_\ \textit{Meas}\text{ taller than Mary}]\!] = [\mathbf{height}(\mathbf{m}), \mathbf{height}(\mathbf{j})]$

b. $[\![\text{John is }\emptyset_{\text{SOME}}\ \textit{Meas}\text{ taller than Mary}]\!] = \mu([\mathbf{height}(\mathbf{m}), \mathbf{height}(\mathbf{j})]) > 0$,
which is equivalent to $\mathbf{height}(\mathbf{j}) - \mathbf{height}(\mathbf{m}) > 0$

In the current version, the measure phrase takes scope over the rest of the sentence, and *Meas* only helps construct the relevant degree interval. This again is not a crucial assumption. The analysis can be modified so that a measure phrase just denotes a degree, and *Meas* not only helps construct the degree interval but also introduces the degree argument, which will be closer to what Svenonius

greatest lower bound and least upper bound.

and Kennedy (2006) and Sawada and Grano (2011) propose. Concretely, the new denotations of *Meas* and *2 inches* are shown in (51c) and (51e), and the derivation (51) is otherwise very similar to the derivation before.

(51) John is 2 inches taller than Mary.

a. $[\![\text{-er}]\!] = \lambda g \lambda y \lambda x . g_{g(y) \to 0}(x)$

b. $[\![\text{taller than Mary}]\!] = \lambda x . \mathbf{height}_{\mathbf{height(m)} \to 0}(x)$

c. $[\![\textit{Meas}]\!] = \lambda g : \underline{g^{-1}(0_g) \neq \emptyset}. \lambda d \lambda x . \mu([0_g, g(x)]) \geq d$

d. $[\![\textit{Meas} \text{ taller than Mary}]\!] = \lambda d \lambda x . \mu([\mathbf{height}(\mathbf{m}), \mathbf{height}(x)]) \geq d$

e. $[\![\text{2 inches}]\!] = 2''$,

f. $[\![\text{2 inches } \textit{Meas} \text{ taller than Mary}]\!] = \lambda x . \mu([\mathbf{height}(\mathbf{m}), \mathbf{height}(x)]) \geq 2''$,

g. $[\![\text{John is 2 inches } \textit{Meas} \text{ taller than Mary}]\!] = \mu([\mathbf{height}(\mathbf{m}), \mathbf{height}(\mathbf{j})]) \geq 2''$,
which is equivalent to $\mathbf{height}(\mathbf{j}) - \mathbf{height}(\mathbf{m}) \geq 2''$

In the case of plain adjectives, the denotation of the silent operator $\emptyset_{\text{SOME}}$ will change accordingly. Again, abstracting away from the details about scope taking, I assume that the rest of the sentence denotes a degree property that $\emptyset_{\text{SOME}}$ takes scope over (52a, cf. 51g) . This degree property is about the length of the relevant degree interval and what $\emptyset_{\text{SOME}}$ does is state that this interval has a positive length (52b). When the two parts are composed, we obtain the correct truth condition in (52c).

(52) John is taller than Mary.

a. $[\![\text{John is \_\_ } \textit{Meas} \text{ taller than Mary}]\!] = \lambda d . \, \mu([\mathbf{height}(\mathbf{m}), \mathbf{height}(\mathbf{j})]) \geq d$

b. $[\![\emptyset_{\text{SOME}}]\!] = \lambda f_{dt} . \exists d > 0 \, (f(d))$

c. $[\![\text{John is } \emptyset_{\text{SOME}} \textit{ Meas} \text{ taller than Mary}]\!] = \exists d > 0 \, (\mu([\mathbf{height}(\mathbf{m}), \mathbf{height}(\mathbf{j})]) \geq d)$,
which is equivalent to $\mathbf{height}(\mathbf{j}) - \mathbf{height}(\mathbf{m}) > 0$

Finally, we need not assume that the silent $\emptyset_{\text{SOME}}$ is a counterpart of measure phrases. Instead, we can analyze it as a counterpart of *Meas*. Both *Meas* and $\emptyset_{\text{SOME}}$ help construct the degree interval. However, whereas *Meas* introduces a degree argument to measure the length of the interval (51c), $\emptyset_{\text{SOME}}$ simply states that the interval has a positive length (53b).

(53) John is taller than Mary.

a. $[\![\text{taller than Mary}]\!] = \lambda x . \mathbf{height}_{\mathbf{height(m)} \to 0}(x)$

b. $[\![\emptyset_{\text{SOME}}]\!] = \lambda g : \underline{g^{-1}(0_g) \neq \emptyset}. \lambda x . \, \mu([0_g, g(x)]) > 0$

c. $[\![\text{John is } \emptyset_{\text{SOME}} \text{ taller than Mary}]\!] = \mu([\mathbf{height}(\mathbf{m}), \mathbf{height}(\mathbf{j})]) > 0$,
which is equivalent to $\mathbf{height}(\mathbf{j}) - \mathbf{height}(\mathbf{m}) > 0$

The >0 interpretation of a minimum adjective can be similarly derived (54).

(54) Company A was profitable.

a. $[\![\emptyset_{\text{SOME}}]\!] = \lambda g : \underline{g^{-1}(0_g) \neq \emptyset}.\lambda x.\, \mu([0_g, g(x)]) > 0$

b. $[\![\emptyset_{\text{SOME}}\ \text{profitable}]\!] = \lambda x.\, \mu([0_{\mathbf{profit}}, \mathbf{profit}(x)]) > 0$

c. $[\![\text{Company A was } \emptyset_{\text{SOME}}\ \text{profitable}]\!] = \mu([0_{\mathbf{profit}}, \mathbf{profit}(\mathbf{a})]) > 0$,
which is equivalent to $\mathbf{profit}(\mathbf{a}) > 0_{\mathbf{profit}}$

4.5 General discussion

4.5.1 Comparison with Sawada and Grano's (2011) analysis

The analysis proposed in the previous section has a lot in common with the analyses proposed by Svenonius and Kennedy (2006) and Sawada and Grano (2011). However, they are crucially different in terms of how the >0 interpretations of comparatives and minimum adjectives are derived. In my analysis, such interpretations are derived by applying the silent $\emptyset_{\text{SOME}}$. This is different from the silent *pos* discussed in the previous chapter, which is responsible for the relative (and maximum) interpretations. In contrast, Svenonius and Kennedy (2006) and Sawada and Grano (2011) assume that *pos* (implemented differently) is responsible for all the interpretations of plain gradable adjectives and comparatives. In this section I will focus on Sawada and Grano's version, because its details are slightly more explicit and elaborated, and show that such a unified analysis is problematic.

Sawada and Grano (2011) propose that minimum adjectives and comparatives both have scales with a minimum degree. Specifically, they propose that comparative constructions have a scale structure with a minimum degree because the *than*-phrase transforms the original scale of the gradable adjective to one that starts with the degree introduced by the *than*-phrase (55).

(55) $[\![\text{taller than Mary}]\!] = \lambda x.\mathbf{height_{height(m)}}^{\uparrow}(x)$, where $\mathbf{height_{height(m)}}^{\uparrow}$ is a transformed scale of height whose starting point is Mary's height.

The semantic type of (55) is $e \rightarrow d$. Since Sawada and Grano (2011) assume that the type of a measure phrase such as *3 inches* is d, (55) cannot be directly composed with the measure phrase. To resolve this type mismatch, Sawada and Grano (2011) assume that a null degree morpheme *Meas*, defined in (56a), first transforms (55) to type $d \rightarrow et$ (56b). This is the correct type to allow for the composition with the measure phrase, which yields the individual property in (56c).

(56) a. $[\![\mathit{Meas}]\!] = \lambda g_{ed} : \underline{g \text{ has a minimum degree on the scale}}.\, \lambda d \lambda x. g(x) \geq d$

b. $[\![\mathit{Meas}\ \text{taller than Mary}]\!] = \lambda d \lambda x.\mathbf{height_{height(m)}}^{\uparrow}(x) \geq d$

c. $[\![\text{3 inches } \mathit{Meas}\ \text{taller than Mary}]\!] = \lambda x.\mathbf{height_{height(m)}}^{\uparrow}(x) \geq 3'$

Sawada and Grano (2011) further assume that *Meas* has a selectional restriction so that it is only compatible with measure functions that have a minimum degree on the scale. Comparatives and minimum adjectives satisfy this requirement, and therefore they are compatible with measure phrases. Lower-open-scale adjectives do not satisfy this requirement, and therefore they are generally not compatible with measure phrases. This accounts for the fact that in some languages measure phrases are totally incompatible with open-scale adjectives. However, other languages, such as English, may allow for some idiosyncratic set of open-scale adjectives to override this restriction. In this way, Sawada and Grano (2011) can account for the crosslinguistic generalizations regarding the compatibility between measure phrases and different types of degree constructions. My analysis is directly adapted from theirs and therefore is very similar in this respect. The only difference is that since I take zero points to be the primitive and assume that *Meas* makes reference to the zero point of a scale rather than the minimum degree (which may not even exist), my analysis can straightforwardly account cases such as *profitable* and *early/late* without stipulating that their scales have minimum degrees.

A minor problem with Sawada and Grano's analysis of differential comparatives is that, since they intend $\mathbf{height_{height(m)}}^{\uparrow}$ to be a truncation of the height scale, this function is undefined for people shorter than Mary (recall the earlier discussion about their analysis of *early/late*). However, if John is shorter than Mary, then *John is 3 inches taller than Mary* should be false rather than a presupposition failure. Therefore, (56c) is in fact incorrect given how Sawada and Grano define $\mathbf{height_{height(m)}}^{\uparrow}$.

For the sake of the argument, let us suppose that they switch to a compression analysis, according to which $\mathbf{height_{height(m)}}^{\uparrow}(x)$ returns the height difference $\mathbf{height(x)} - \mathbf{height(m)}$ if x is taller than Mary, and returns 0 otherwise. With this new definition, if John is shorter than Mary, then *John is 3 inches taller than Mary* is correctly predicted to be false, since $\mathbf{height_{height(m)}}^{\uparrow}(\mathbf{j})$ returns 0, which is less than $3'$. Note that this new definition still results in a scale with a minimum degree, and therefore does not affect the remaining part of Sawada and Grano's analysis.

The remaining issue Sawada and Grano need to address is how to derive the truth conditions for plain comparatives, e.g., *John is taller than Mary.* They propose that the derivation is totally in parallel with the positive form construction *John is tall*, where a silent *pos*, as defined by Kennedy (2007), is used to introduce an appropriate threshold (57a).

(57) a. $[\![pos]\!]^C = \lambda g.\lambda x.g(x) \geq \mathbf{s}(g)(C)$, where $\mathbf{s}(g)(C) = g_{\min}$ ($g_{\max}$) if $g_{\min}$ ($g_{\max}$) exists, otherwise it is a contextually determined threshold θ_C, which "stands out" in C

b. $[\![pos \text{ tall}]\!]^C = \lambda x.\mathbf{height}(x) \geq \theta_C$

c. $[\![pos \text{ taller than Mary}]\!]^C = \lambda x.\mathbf{height_{height(m)}}^{\uparrow}(x) > 0$,
which is equivalent to $\lambda x.\mathbf{height}(x) > \mathbf{height}(\mathbf{m})$

Since *tall* does not have a minimum (or maximum) degree on its scale, it has a relative standard that "stands out" in the context (57b). In contrast, since *taller than Mary* has a minimum degree in its scale, 0 is used as the standard (57c), and we derive the correct truth condition for the comparative construction.[38]

This analysis would have been exactly correct if the scale structure and the standard of a positive form were perfectly correlated. However, we saw in Section 4.3 that minimum adjectives can in fact have relative interpretations, e.g., *salty*$_{\geq\theta}$ *for a Mexican restaurant* (58).

(58) The chicken is not salty$_{\geq\theta}$ for a Mexican restaurant.

Clearly, the semantics stated in (57a) is too inflexible to allow for relative interpretations of minimum adjectives, and needs to be revised. In fact, in the previous chapter (Section 3.3.1) we have already encountered similar cases for maximum adjectives, e.g., *full for a Thursday afternoon* (59), which has a relative interpretation.

(59) The theater is full$_{\geq\theta}$ for a Thursday afternoon.

As discussed there, a potential way to account for cases like (59) is to relax the definition of *pos* to (60), which allows for both maximum/minimum and relative interpretations in principle.

(60) $[\![pos]\!]^C = \lambda g.\lambda x.g(x) \geq \mathbf{s}(g)(C)$, where $\mathbf{s}(g)(C)$ either "stands out" wrt g (which effectively means it is either $g_{\min}$ or $g_{\max}$) or "stands out" in C

Moreover, we can interpret Kennedy's (2007) Interpretive Economy in such a way that the relative interpretation is in fact the one that maximizes the conventional meanings of the sentence (59). The idea is that since the *for*-PP is explicitly mentioned and has a conventional meaning that can intuitively correspond to a contextual comparison class (regardless of how exactly this effect is implemented formally), the relative interpretation, which uses this information as well as the measure function denoted by *full*, in a sense utilizes more conventional meanings than a maximum/minimum interpretation, which only uses the measure function and totally discard the contribution of the *for*-PP. Therefore IE would prefer the relative interpretation in this case.

However, even under this construal of IE, this analysis is still inadequate, because there are counterexamples that do not involve *for*-PPs. For instance, *dark* arguably only has a relative interpretation, despite having a totally-closed scale. Furthermore, we have seen that in polar questions, minimum adjectives such as *active*, *effective*, *salty* have both relative and > 0 interpretations, even without explicit *for*-PPs.

[38] Note that the inequality sign (57c) is a strict one, which is different from the non-strict one in (57a). This is a minor technical problem with Kennedy's (2007) formulation to unify all three classes of gradable adjectives: for maximum adjectives, the degree is required to be at least the maximum degree (it is impossible to be strictly greater than that), whereas for minimum adjectives, the degree is required to be strictly greater than the minimum degree. Just to be clear, I do not take this to be a serious criticism of Kennedy's analysis, but I do note that this problem does not arise when we assume that minimum interpretations are in fact derived from a different mechanism than applying *pos*.

Moreover, this relaxation of IE creates a new problem of over-generation for plain comparatives. Note that according to the revised definition of *pos* (60), a plain comparative in principle can have a minimum as well as a relative interpretation (61).

(61) $[\![\text{John is taller than Mary}]\!]^C = \mathbf{height_{height(m)}}^{\uparrow}(\mathbf{j}) \geq \mathbf{s}(\mathbf{height_{height(m)}}^{\uparrow})(C)$,
which can be $\mathbf{height_{height(m)}}^{\uparrow}(\mathbf{j}) > 0$ or $\mathbf{height_{height(m)}}^{\uparrow}(\mathbf{j}) \geq \theta_C$

Given the above construal of IE, the relative interpretation should be available and preferred when there is an overt *for*-PP or a *compared to* phrase. The relative interpretation $\mathbf{height_{height(m)}}^{\uparrow}(\mathbf{j}) \geq \theta_C$ says that the height difference between John and Mary "stands out" in context. This predicts that the sentences in (38), repeated below in (62), have an interpretation which says that the height difference between John and Mary does not "stand out" among the set of height differences between basketball players and Mary.

(62) a. # John is not taller than Mary for a basketball player.
(cf. John is not tall for a basketball player.)

b. # Compared to other basketball players, John is not taller than Mary.
(cf. Compared to other basketball players, John is not tall.)

However, such an interpretation is never attested. For instance, suppose that Mary is $5'10''$ and John is a $6'2''$ basketball player. Given that basketball players are generally quite tall, the 4-inch height difference between John and Mary probably will not stand out among the set of height differences between basketball players and Mary. Therefore, the relative interpretation $\mathbf{height_{height(m)}}^{\uparrow}(\mathbf{j}) \not\geq \theta_C$, if exists, would be true in this case. However, the sentences in (38) are not true in this case. In fact, they are infelicitous regardless of what John and Mary's heights are, because it simply does not make sense to try to introduce a comparison class for a plain comparative. We may call such infelicity ungrammaticality, but it is hard to appeal to syntax to explain the infelicity/ungrammaticality. Even though there may be reasons to believe that *for*-PPs are syntactic arguments, it seems hard to find a way to block them from appearing in plain comparatives (62a), because intuitively a *for*-PP is the syntactic argument of *pos* but *pos* is assumed to appear in plain comparatives as well. It seems even more unlikely that the *compared to* phrase in (62b) is an argument rather than an adjunct, so it is not clear why syntactically it cannot appear with plain comparatives. Furthermore, we can specify a comparison class by using a conditional, which is clearly an adjunct (63).

(63) # If we are only talking about basketball players, John is not taller than Mary.
(cf. If we are only talking about basketball players, John is not tall)

This means that we probably cannot syntactically rule out the ungrammatical sentences. We probably cannot rule them out pragmatically, either, e.g., by appealing to the unnaturalness of the intended meanings, because the intended meanings are presumably the same as the comparatives modified by *much* (64). Even though the intended meanings are still somewhat unnatural, the sentences (64) sound much better.

(64) a. John is not much taller than Mary for a basketball player.

b. Compared to other basketball players, John is not much taller than Mary.

c. If we are only talking about basketball players, John is not much taller than Mary.

Therefore, I conclude that plain comparatives are semantically incompatible with comparison classes and do not have relative interpretations. In contrast, the sentences in (64) are fine because the positive form *much* allows for the specification of a comparison class for its relative interpretation.

To sum up, unlike plain minimum adjectives, plain comparatives *never* have a relative interpretation and it is impossible to force such an interpretation by introducing a comparison class. This difference makes it difficult to use a single mechanism of *pos* plus IE to account for *all* of their possible interpretations: either we would under-generate relative interpretations for minimum adjectives, or we would over-generate relative interpretations for comparatives. In comparison, the similarity and difference between plain minimum adjectives and comparatives are totally expected under my analysis, because the relative and > 0 interpretations are derived by two different mechanisms. The relative interpretation is derived by *pos*, which only combines with positive forms. As a result, only plain minimum adjectives can have relative interpretations. Meanwhile, the > 0 interpretation is derived by $\emptyset_{\text{SOME}}$ (perhaps also *Meas*), which is available when the scale has a realizable zero point. Therefore, both plain minimum adjectives and comparatives have > 0 interpretations.

4.5.2 Silent morphemes and a potential worry of over-generation

According to my analysis, there are three silent morphemes responsible for generating the interpretation of a positive form. The first is *pos*, which is assumed to be a silent counterpart of *very* and is responsible for the relative interpretation. The remaining two are *Meas* and $\emptyset_{\text{SOME}}$, which together generate the > 0 interpretation.

Given that measure phrases and $\emptyset_{\text{SOME}}$ have the same semantic type, i.e., they are both predicates of intervals, one consequence of this analysis is that, without further assumptions about $\emptyset_{\text{SOME}}$, positive forms that are compatible with measure phrases are predicted to have a > 0 interpretation.

In some cases, this does not cause a major problem. For instance, given that *John is* $6'$ *tall* is grammatical, the current analysis predicts that *John is tall* can have a > 0 interpretation, i.e., $\mathbf{height}(\mathbf{j}) > 0$. While this interpretation is not attested, we can presumably rule it out by noting that it is a tautology, which means that it is totally uninformative and therefore is almost never picked up by pragmatic language users.

However, this explanation is not always applicable. For instance, given that *this glass is 60% full* is grammatical, the current analysis predicts that *full* has a > 0 interpretation, which would mean *non-empty*. Since *this glass is non-empty* is not a tautology, we cannot say that the predicted

> 0 interpretation of *full* is not attested because it is totally uninformative. Similar cases exist for relative adjectives as well. For instance, *this outcome is 40% likely* is grammatical, but *likely* does not have a > 0 interpretation, which would mean *possible*.

Therefore, it seems that the current analysis faces a problem of over-generation, which cannot always be explained by pragmatics. Below I discuss another solution, but I acknowledge that it is not completely satisfying, either.

The idea is to assume that $\emptyset_{\text{SOME}}$ has further selectional requirements. Recall our earlier example *there is water in the glass*, which motivates our definition of $\emptyset_{\text{SOME}}$. Now consider the ungrammatical sentence **there is student in the classroom*. It seems reasonable to assume that the ungrammaticality is due to a syntactic restriction according to which a bare singular count noun cannot have an existential interpretation. If we adopt a similar strategy, we would say that the syntactic distribution of $\emptyset_{\text{SOME}}$ is further restricted so that it does not combine with *full* or *likely* for pure syntactic reasons.

Of course, in order for this solution to be more than a pure stipulation, we should look for morpho-syntactic evidence that helps predict whether the > 0 interpretation is available. To some extent, we can indeed find such evidence. There are morphemes that are good indicators of the > 0 interpretation (65).

(65) a. *-y*: *salty*, *spicy*, *dirty*, *windy*, ...

b. *-ive*: *effective*, *supportive*, *active*...

c. *-ful*: *helpful*, *harmful*, *colorful* ...

d. *-able*: *profitable*, *curable*, *noticeable*, *accessible*, ...

e. *-ed*: *spotted*, *striped*, *bent*, *curved*...

The list is certainly not exhaustive, but it does seem to provide some hope that maybe we can restrict the syntactic distribution of $\emptyset_{\text{SOME}}$ to rule out the over-generation cases. However, this solution is not entirely satisfying, either, because these morphemes are not completely reliable indicators of the > 0 interpretation, and there can still be an over-generation problem. For instance, *expensive/pricey/closed*, despite having the relevant morphemes, do not seem to have a > 0 interpretation, which would mean *non-free/not fully open*.

That said, I should note that previous accounts of gradable adjectives also face this problem of over-generation. An IE-based analysis says nothing about whether a gradable adjective with a fully closed scale would use the maximum or minimum degree as the standard. Probabilistic models can generate the the > 0 interpretation for some comparison distributions. Setting aside the problem that they wrongly predict the the > 0 interpretation to be vague for *profitable* (which does not arise for closed-scale adjectives), since they do not have a full theory of how the comparison distribution is determined, they do not have a principled way to avoid over-generating the > 0 interpretation, either. Therefore, determining whether an adjective has a > 0 interpretation is still an open problem, which I will leave for future research.

Chapter 5

Conclusion

5.1 Semantic vs pragmatic contextual resolution mechanisms in the domain of degrees

In this dissertation, I investigated context sensitivity in the domain of degrees. Using directional modified numerals *up to n* and gradable adjectives as two case studies, I showed how we can address the semantic and metasemantic questions in a principled way. I argued that, even though the two types of expressions have very similar underspecified semantic representations, in that both have a contextually determined semantic lower bound, their contextual resolution mechanisms are different. On the one hand, *up to n* has a pragmatic contextual resolution mechanism, i.e., its contextual lower bound is determined by the listener's pragmatic reasoning of the speaker's intention. This pragmatic mechanism is in the spirit of King's (2013, 2014a, 2014b) general account of supplementives, and has concrete qualitative as well as quantitative implementations that avoid the limitations of King's proposal. On the other hand, gradable adjectives, or more precisely, the maximum and relative readings of their positive forms, have a semantic contextual resolution mechanism, i.e., the mapping from the context to the contextual lower bound is part of the conventional meaning.

The contrast between *up to n* and gradable adjectives in terms of their contextual resolution mechanisms is in parallel with that between the pronouns *they* and *I*. This provides evidence for the generality of the distinction between the two types of contextual resolution mechanisms. I showed how the type of the contextual resolution mechanism can be determined in a principled way, by examining how the underspecified semantic representation is resolved in context under embedded environments. If the contextual dependency is the same under embedded environments, then the contextual resolution mechanism is semantic, otherwise it is pragmatic.

In addition to identifying both types of contextual resolution mechanisms in the domain of degrees, I also unified the two types of mechanisms based on functional considerations. Concretely, both mechanisms involve an interaction between informativity and applicability, which are quantitative generalizations of Grice's Maxims of Quantity and Quality. This allows us to make more precise, quantitative predictions about the inferential properties of context-sensitive expressions.

5.2 A new taxonomy of positive forms

Another novel contribution of this dissertation is a new taxonomy of positive forms of gradable adjectives.

Traditionally, gradable adjectives are classified descriptively into three major classes: maximum, relative, and minimum adjectives, based on the interpretations of their positive forms. A central theoretical question is to answer how the three classes are related in a way that accounts for their similarities and differences.

Previous approaches generally focus on the contrast between relative adjectives and absolute (i.e., maximum or minimum) adjectives in terms of vagueness, and essentially presuppose that there is a single explanation for the seemingly clear-cut interpretations of maximum and minimum adjectives.

In contrast, the analyses proposed in this dissertation suggest a new taxonomy of positive forms. According to this taxonomy, the positive form of a gradable adjective can have two kinds of interpretations, which are derived from two different compositional devices.

The first interpretation is the relative interpretation. This interpretation is derived by composing the gradable adjective with *pos*, which introduces a threshold based on a contextual comparison distribution. Crucially, the threshold is sensitive to not only the central tendency of the distribution, but also the dispersion. Consequently, a maximum interpretation is an extreme case of the relative interpretation, i.e., when the comparison distribution has a high level of dispersion.

The second interpretation is the > 0 interpretation. This interpretation is derived when the gradable adjective is first composed with *Meas*, which results in a degree interval, and then with a silent operator $\emptyset_{\text{SOME}}$, which states that the size of this interval is greater than 0. This amounts to saying that the relevant degree of the individual under discussion is greater than 0, which is exactly the clear-cut interpretation of a minimum adjective.

Compared with previous approaches, this proposal has two major advantages: (i) it accounts for the imperfect correlation between scale structure and interpretations of positive forms, and (ii) it accounts for the connections between positive forms and comparative constructions. According to this proposal, it comes as no surprise that relative interpretations are possible for positive forms of all three classes of gradable adjectives, but not for comparative constructions.

More generally, this case study of gradable adjectives highlights the importance of addressing both the semantic and metasemantic questions in order to distinguish between semantic underspecification and ambiguity in a principled way. The conclusion that the > 0 interpretation of a gradable adjective needs a different semantic derivation can be drawn only because there is a concrete and precise account of how the underspecified relative reading gets resolved in context.

5.3 Remaining challenges and future directions

There are, however, still quite a few remaining challenges and open issues for future research.

Theoretically, one important issue is how exactly the contextual parameters should be understood and determined from the intuitive notion of the context. For instance, the contextual parameter of the relative reading of a gradable adjective is a probability distribution, i.e., the comparison distribution. Such distributions are part of the interlocutors' subjective beliefs, but they are not completely arbitrary and need to be heavily constrained by factors such as interlocutors' world knowledge and prior discourse. However, it is not entirely clear which factors are relevant and exactly how they influence the comparison distribution. For instance, as discussed in Section 3.5.2, certain constraints on comparison distributions might be specific to certain lexical items and cannot be reduced to other factors such as world knowledge. Whether or not this is indeed necessary remains unknown.

Furthermore, there are computational and empirical/experimental issues when we want to test the more precise, quantitative predictions of the analysis. For instance, for many adjectives it is intuitively difficult to assign probability distributions to, e.g., *calorie-rich* and *sweet* (for its relative reading). There are many factors that can contribute to such difficulties. For example, there is no well-known conventional measure of sweetness (partly because the degree of sweetness is subjective), and most people probably do not know how much calorie each type of food has, even if they are aware that there is a conventional way to measure it. Examples like these do not pose immediate challenges to the qualitative part of the proposal. In fact, the uncertainty about the comparison distribution is a major component in the proposed account. This predicts that there should be a lot of uncertainty and variability regarding the interpretations of such gradable adjectives, which seems plausible. However, this does make it very difficult to empirically test the quantitative predictions of the model.

Addressing such challenges likely requires many small, incremental steps. We can start with gradable adjectives such as *tall*, whose comparison distributions are fairly intuitively accessible and have relatively low individual variability, to provide some initial test of the model's predictions, before moving on to adjectives whose comparison distributions are more difficult to assign.

Finally, it is also important to look for other instances of the contrast between semantic and pragmatic contextual resolution mechanisms in other domains, and investigate to what extent the present case study in the domain of degrees can shed new light there.

References

Bartsch, R., & Vennemann, T. (1972). The grammar of relative adjectives and comparison. In *Linguistische Berichte 20* (pp. 19–32).

Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy*, *4*(2), 159–219. doi: 10.1007/BF00350139

Bierwisch, M. (1989). The semantics of gradation. In M. Bierwisch & E. Lang (Eds.), *Dimensional adjectives: Grammatical structure and conceptual interpretation* (pp. 163–192). Berlin Heidelberg: Springer-Verlag.

Blok, D. (2015). The semantics and pragmatics of directional numeral modifiers. In *Semantics and linguistic theory (SALT)* (Vol. 25). LSA and CLC Publications.

Bogusławski, A. (1975). Measures are measures: In defence of the diversity of comparatives and positives. *Linguistiche Berichte*, *36*, 1–9.

Büring, D. (2008). The least *at least* can do. In C. B. Chang & H. J. Haynie (Eds.), *West Coast Confererence on Formal Linguistics (WCCFL) 26* (pp. 114–120). Somerville, MA: Cascadilla Proceedings Project.

Büring, D. (2011). Pronouns. In K. Von Heusinger, C. Maienborn, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning* (pp. 971–996). Berlin: Mouton De Gruyter.

Burnett, H. (2014). A delineation solution to the puzzles of absolute adjectives. *Linguistics and Philosophy*, *37*(1), 1–39. doi: 10.1007/s10988-014-9145-9

Bylinina, L. (2014). *The grammar of standards: Judge-dependence, purpose-relativity, and comparison classes in degree constructions* (PhD thesis). Utrecht University.

Ciardelli, I., Groenendijk, J., & Roelofsen, F. (2009). Attention! Might in inquisitive semantics. In S. I. Ed Cormany & D. Lutz (Eds.), *Semantics and linguistic theory SALT 19* (pp. 91–108).

Ciardelli, I., Groenendijk, J., & Roelofsen, F. (2012). Inquisitive semantics. *NASSLLI 2012 lecture notes*.

Coppock, E., & Brochhagen, T. (2013). Raising and resolving issues with scalar modifiers. *Semantics and Pragmatics*, *6*(3), 1–57. doi: 10.3765/sp.6.3

Dehaene, S. (2003). The neural basis of the WeberFechner law: a logarithmic mental number line.

Trends in Cognitive Sciences, *7*(4), 145–147. doi: http://dx.doi.org/10.1016/S1364-6613(03)00055-X

Égré, P. (2017). Vague judgment: a probabilistic account. *Synthese*, *194*(10), 3837–3865. doi: 10.1007/s11229-016-1092-2

Fine, K. (1975). Vagueness, truth and logic. *Synthese*, *30*(3/4), 265–300.

Foppolo, F., & Panzeri, F. (2011). *When straight means 'relatively straight' and big means 'absolutely big'.* Paper presented at the 31st Incontro di Grammatica Generativa, Rome, Italy, February 24-26.

Fox, D. (2007). Free choice disjunction and the theory of scalar implicatures. In U. Sauerland & P. Stateva (Eds.), *Presupposition and implicature in compositional semantics* (pp. 71–120). London: Palgrave Macmillan UK. doi: 10.1057/9780230210752_4

Franke, M. (2011). Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics*, *4*(1), 1–82. doi: 10.3765/sp.4.1

Fults, S. (2006). *The structure of comparison: an investigation of gradable adjectives.* (PhD thesis). University of Maryland.

Gennaro, C., & McConnell-Ginet, S. (1990). *Meaning and grammar: An introductionto semantics.* MIT Press.

Geurts, B., & Nouwen, R. (2007). 'At Least' et al.: The semantics of scalar modifiers. *Language*, *83*(3), 533–559. doi: 10.1353/lan.2007.0115

Glanzberg, M. (2007). Context, content, and relativism. *Philosophical Studies*, *136*(1), 1–29. doi: 10.1007/s11098-007-9145-5

Glanzberg, M. (2009). *Not all contextual parameters are alike.* (Ms., UC Davis)

Glanzberg, M. (2020). Indirectness and intentions in metasemantics. In T. Ciecierski & P. Grabarczyk (Eds.), *The architecture of context and context-sensitivity.* Springer.

Graff, D. (2000). Shifting sands: An interest-relative theory of vagueness. *Philosophical Topics*, *28*(1), 45–81.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3: Speech acts* (pp. 41–58). New York: Academic Press.

Hackl, M. (2000). *Comparative quantifiers* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Kamp, H. (1973). Free choice permission. *Aristotelian Society*, *74*, 57-74.

Kamp, H. (1978). Semantics versus pragmatics. In F. Guenthner & S. J. Schmidt (Eds.), *Formal semantics and pragmatics for natural languages* (pp. 255–287). Dordrecht: Reidel.

Kao, J. T., Wu, J., Bergen, L., & Goodman, N. D. (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences*, *111*(33), 12002–12007.

Kaplan, D. (1989a). Afterthoughts. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from kaplan* (pp. 565–614). Oxford University Press.

Kaplan, D. (1989b). Demonstratives: An essay on the semantics, logic, metaphysics and epistemology of demonstratives and other indexicals. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from kaplan* (pp. 481–563). Oxford University Press.

Kennedy, C. (2001). Polar opposition and the ontology of 'degrees'. *Linguistics and Philosophy*, *24*, 33–70. doi: 10.1023/A:1005668525906

Kennedy, C. (2002). *The landscape of vagueness.* (Ms., Northwestern University)

Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, *30*, 1–45.

Kennedy, C. (2015). A "de-Fregean" semantics (and neo-gricean pragmatics) for modified and unmodified numerals. *Semantics and Pragmatics*, *8*(10), 1–44. doi: 10.3765/sp.8.10

Kennedy, C., & McNally, L. (2005). Scale structure, degree modification, and the semantics of gradable predicates. *Language*, *81*(2), 345–381.

King, J. C. (2001). *Complex demonstratives: A quantificational account.* Cambridge, MA: MIT Press.

King, J. C. (2013). Supplementives, the coordination account and conflicting intentions. *Philosophical Perspectives*, *27*(1), 288-311. doi: 10.1111/phpe.12028

King, J. C. (2014a). The metasemantics of contextual sensitivity. In A. B. . B. Sherman (Ed.), *Metasemantics: New essays on the foundations of meaning* (pp. 97–118). Oxford University Press.

King, J. C. (2014b). Speaker intentions in context. *Nos*, *48*(2), 219–237. doi: 10.1111/j.1468-0068.2012.00857.x

King, J. C. (2020). Speaker intentions and objective metasemantics. In T. Ciecierski & P. Grabarczyk (Eds.), *The architecture of context and context-sensitivity.* Springer.

Klecha, P. (2012). Positive and conditional semantics for gradable modals. In A. Aguilar-Guevara, A. Chernilovskaya, & R. Nouwen (Eds.), *Sinn und bedeutung (SuB)* (Vol. 16, p. 363-376).

Klecha, P. (2014). *Bridging the divide: Scalarity & modality* (PhD thesis). University of Chicago.

Klein, E. (1980). A semantics for positive and comparative adjectives. *Linguistics and Philosophy*, *4*(1), 1–45.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement: Additive and polynomial representations* (Vol. 1). New York: Academic Press.

Kratzer, A. (1977). What 'must' and 'can' must and can mean. *Linguistics and Philosophy*, *1*(3), 337–355. doi: 10.1007/BF00353453

Kratzer, A. (1981). The notional category of modality. In H.-J. Eikmeyer & H. Rieser (Eds.), *Words, worlds, and contexts: New approaches in word semantics* (pp. 38–74). Berlin: de Gruyter.

Kratzer, A., & Shimoyama, J. (2002). Indeterminate pronouns: The view from japanese. In Y. Otsu (Ed.), *Third tokyo conference on psycholinguistics* (pp. 1–25). Tokyo, Japan: Hituzi Syobo Publishing Company.

Lasersohn, P. (2005). Context dependence, disagreement, and predicates of personal taste. *Linguistics and Philosophy*, *28*, 643–686. doi: 10.1007/s10988-005-0596-x

Lassiter, D. (2010). Gradable epistemic modals, probability, and scale structure. In N. Li & D. Lutz (Eds.), *Semantics and linguistic theory (SALT)* (Vol. 20, pp. 197–215). LSA and CLC Publications. doi: 10.3765/salt.v20i0.2557

Lassiter, D. (2011). *Measurement and modality* (PhD thesis). New York University.

Lassiter, D. (2017). *Graded modality: Qualitative and quantitative perspectives.* Oxford University Press. doi: 10.1093/oso/9780198701347.001.0001

Lassiter, D., & Goodman, N. D. (2013). Context, scale structure, and statistics in the interpretation of positive-form adjectives. In T. Snider (Ed.), *Semantics and linguistic theory (SALT) 23* (pp. 587–610). LSA and CLC Publications.

Lassiter, D., & Goodman, N. D. (2015). Adjectival vagueness in a Bayesian model of interpretation. *Synthese*, 1–36. doi: 10.1007/s11229-015-0786-1

Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, *8*(1), 339–359.

Lewis, K. S. (2019). The speaker authority problem for context-sensitivity (or: You can't always mean what you want). *Erkenntnis*. doi: 10.1007/s10670-018-0089-2

McNally, L. (2011). The relative role of property type and scale structure in explaining the behavior of gradable adjectives. In R. Nouwen, R. van Rooij, U. Sauerland, & H.-C. Schmitz (Eds.), *Vagueness in communication* (pp. 151–168). Springer.

Napoletano, T. (2019). How important are truth-conditions for truth-conditional semantics? *Linguistics and Philosophy*. doi: 10.1007/s10988-019-09261-y

Nouwen, R. (2008). Directionality in numeral quantifiers: the case of "up to". In T. Friedman & S. Ito (Eds.), *Semantics and linguistic theory (SALT)* (Vol. 18, pp. 569–582). LSA and CLC Publications.

Nouwen, R. (2010). Two kinds of modified numerals. *Semantics and Pragmatics*, *3*(3), 1–41. doi: 10.3765/sp.3.3

Potts, C. (2008). *Interpretive Economy, Schelling Points, and evolutionary stability.* (Ms., UMass Amherst)

Qing, C. (2016). Up to n: Pragmatic inference about an optimal lower bound. In M. Moroney, C.-R. Little, J. Collard, & D. Burgdorf (Eds.), *Semantics and linguistic theory (SALT)* (Vol. 26, pp. 526–545). LSA and CLC Publications. doi: 10.3765/salt.v26i0.3801

Qing, C., & Franke, M. (2014). Gradable adjectives, vagueness, and optimal language use: A speaker-oriented model. In T. Snider, S. D'Antonio, & M. Wiegand (Eds.), *Semantics and linguistic theory (SALT) 24* (pp. 23–41). Ithaca, NY: CLC.

Rotstein, C., & Winter, Y. (2004). Total adjectives vs. partial adjectives: Scale structure and higher-order modifiers. *Natural Language Semantics*, *12*(3), 259–288. doi: 10.1023/B:NALS.0000034517.56898.9a

Sawada, O., & Grano, T. (2011). Scale structure, coercion, and the interpretation of measure phrases in Japanese. *Natural Language Semantics*, *19*(2), 191–226. doi: 10.1007/s11050-011-9070-1

Schwarz, B., Buccola, B., & Hamilton, M. (2012). Two types of class B numeral modifiers: A reply to nouwen 2010. *Semantics and Pragmatics*, *5*(1), 1–25. doi: 10.3765/sp.5.1

Schwarzschild, R. (2005). Measure phrases as modifiers of adjectives. *Recherches linguistiques de Vincennes*, *30*(2), 207–228. doi: 10.4000/rlv.1401

Schwarzschild, R., & Wilkinson, K. (2002). Quantifiers in comparatives: A semantics of degree based on intervals. *Natural Language Semantics*, *10*(1), 1–41. doi: 10.1023/A:1015545424775

Solt, S. (2011). Notes on the comparison class. In R. Nouwen, R. van Rooij, U. Sauerland, & H.-C. Schmitz (Eds.), *Vagueness in communication* (pp. 127–150). Springer.

Solt, S. (2012). Comparison to arbitrary standards. In A. Aguilar-Guevara, A. Chernilovskaya, & R. Nouwen (Eds.), *Sinn und Bedeutung (SuB)* (Vol. 16, pp. 557–570).

Stanley, J. (2003). Context, interest relativity and the sorites. *Analysis*, *63*(280), 269–281. doi: 10.1111/1467-8284.00436

Svenonius, P., & Kennedy, C. (2006). Northern Norwegian degree questions and the syntax of measurement. In M. Frascarelli (Ed.), *Phases of interpretation* (pp. 129–157). The Hague: Mouton de Gruyter.

Syrett, K., Kennedy, C., & Lidz, J. (2010). Meaning and context in children's understanding of gradable adjectives. *Journal of Semantics*, *27*(1), 1–35. doi: 10.1093/jos/ffp011

Toledo, A., & Sassoon, G. (2011). Absolute vs. relative adjectives: Variance within vs. between individuals. In N. Ashton, A. Chereches, & D. Lutz (Eds.), *Semantics and linguistic theory (SALT) 21* (pp. 135–154). Ithaca, NY: CLC. doi: 10.3765/salt.v21i0.2587

van Benthem, J. (1982). Later than late: On the logical origin of the temporal order. *Pacific Philosophical Quarterly*, *63*(2), 193–203. doi: 10.1111/j.1468-0114.1982.tb00099.x

van Rooij, R. (2011a). Implicit versus explicit comparatives. In P. Égré & N. Klinedinst (Eds.), *Vagueness and language use* (pp. 51–72). Basingstoke, UK: Palgrave Macmillan.

van Rooij, R. (2011b). Vagueness and linguistics. In G. Ronzitti (Ed.), *Vagueness: A guide* (pp. 123–170). Dordrecht: Springer Netherlands. doi: 10.1007/978-94-007-0375-9_6

von Fintel, K., & Iatridou, S. (2008). How to say ought in foreign: The composition of weak necessity modals. In J. Guéron & J. Lecarme (Eds.), *Time and modality* (pp. 115–141). Dordrecht: Springer Netherlands. doi: 10.1007/978-1-4020-8354-9_6

Williamson, T. (1994). *Vagueness*. London: Routledge.

Zimmermann, T. E. (2000). Free choice disjunction and epistemic possibility. *Natural Language Semantics*, *8*, 255-290.