Very, Many, Small, Penguins

Vaguely Related Topics

Harald A. Bastiaanse

Very, Many, Small, Penguins

Harald A. Bastiaanse

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Very, Many, Small, Penguins

Vaguely Related Topics

Harald A. Bastiaanse

Very, Many, Small, Penguins

Vaguely Related Topics

ILLC Dissertation Series DS-2014-03

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

For further information about ILLC-publications, please contact

Institute for Logic, Language and Computation
Universiteit van Amsterdam
Science Park 904
1098 XH Amsterdam
phone: +31-20-525 6051
fax: +31-20-525 5206
e-mail: illc@uva.nl
homepage: http://www.illc.uva.nl/

Very, Many, Small, Penguins

Vaguely Related Topics

Academisch Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college voor
promoties ingestelde commissie, in het openbaar
te verdedigen in de Agnietenkapel
op dinsdag 25 maart 2014, te 10.00 uur

door

Harald Andreas Bastiaanse

geboren te Purmerend.

Promotiecommissie:

Promotor: Prof. dr. F.J.M.M. Veltman

Overige leden:
Prof. dr. J.F.A.K. van Benthem
Dr. J.K. Szymanik
Prof dr. R.A.M. van Rooij
Prof. dr. J.F. Horty
Dr. R.A. Muskens

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

The investigations were supported by the Netherlands Organization for Scientific Research (NWO), as part of the project *On vagueness - and how to be precise enough* (project NWO 360-20-202).

Cover design by the author.
Photo on front cover by M. Murphy [Public domain], via Wikimedia Commons
Printed and bound by PrintPartners Ipskamp

ISBN: 978–94–6259–069–4

Contents

Acknowledgments **ix**

1 Introduction **1**

2 The Rationality of Round Interpretation **5**
2.1 Introduction 5
2.2 Conditional Probability 6
2.3 Game Theory 8
2.4 Discussion 12
2.4.1 Generalization to Vagueness 12
2.4.2 Schelling Points and Evolutionary Game Theory; a problem? 15
2.5 Appendix: Continuous Scale and k on Probability 17

3 The Intensional Many - Conservativity Reclaimed **19**
3.1 Introduction 19
3.2 An Intensional Framework 21
3.2.1 Framework 21
3.2.2 Intensional Conservativity 22
3.2.3 On Scandinavians and the Reverse Reading 24
3.2.4 Other key properties 25
3.2.5 Relation with Extensional Properties 28
3.2.6 General form 28
3.3 Other readings of Many 30
3.4 Appendix A: Reductions 35
3.4.1 Lifting Theorem 35
3.4.2 Extensional Intensional Quantifiers 37
3.5 Appendix B: Characterizing the General Form 39
3.5.1 NUM and Intensional ISOM 42
3.6 Appendix C: Logical Properties of Many - A Brief Glimpse 42

3.6.1 Additivity . . . 43
3.6.2 Reflexivity and anti-Reflexivity . . . 45
3.6.3 Desirability . . . 46
3.6.4 Negative theorems . . . 46

4 A Syllogistic for Subsective Adjectives 51
4.1 Introduction . . . 51
4.2 The Syllogistics . . . 52
4.2.1 Logic . . . 53
4.2.2 An aside: Predicative use . . . 53
4.2.3 General Models . . . 54
4.2.4 Weakly Acceptable Models . . . 55
4.2.5 Acceptable Models . . . 61
4.3 Conclusion . . . 65

5 A Syllogistic Characterization of Gradable Adjectives 67
5.1 Introduction . . . 67
5.2 The Syllogistics . . . 67
5.2.1 Formalism . . . 68
5.3 Characterizing Gradable Adjectives . . . 69
5.3.1 No Reversal . . . 70
5.3.2 No Cycles . . . 72
5.3.3 The Consequences of No Cycles . . . 73
5.4 The Commensurability of Adjective Groups . . . 75
5.4.1 Polar antonyms . . . 77
5.4.2 Evaluative Adjectives and Personal Taste . . . 80
5.4.3 Strictenings and Loosenings . . . 81
5.4.4 The Logic of Commensurability . . . 82
5.5 Vagueness, Distinguishability and The Sorites . . . 83
5.5.1 The Sorites Paradox and Incomplete Judgement . . . 84
5.6 Appendix: Nouns as Contexts . . . 87

6 Making the Right Exceptions 89
6.1 Introduction . . . 89
6.2 Naive Circumscription . . . 90
6.3 Exemption and Inheritance . . . 93
6.3.1 The Exemption Principle . . . 94
6.3.2 The Inheritance Property . . . 96
6.3.3 Some more examples . . . 97
6.3.4 Coherence . . . 99
6.4 Networks . . . 100
6.4.1 Making inferences in inheritance nets . . . 101
6.4.2 Examples . . . 103

6.4.3 An algorithm 106
6.4.4 Completeness 108
6.5 Conclusion 110
6.6 Appendix: Completeness of Networks relative to the Semantics . . 110
6.6.1 New constraints, same consequences 110
6.6.2 Completeness Proof 118
6.6.3 Coherence 122

Bibliography **125**

Index **129**

Samenvatting **131**

Abstract **135**

Acknowledgments

An old law states that it always takes longer than you expect, even if you take that law into account. (Hofstadter 1979, p152) This is certainly how this project has felt for me. That I have managed to complete it successfully despite a variety of setbacks both personal and professional is due in no small part to the help I have received from many people.
First of all I would like to thank my promotor Frank Veltman: for his patience and understanding, for the numerous times he has provided valuable advice and support, but most of all for his solid confidence in my ability. At the time when I most doubted myself, his belief in me was essential in motivating myself to go on. I might not have seen this project through to the end without it.
I would also like to thank Robert van Rooij who at times was almost like a co-promotor, often offering not only useful suggestions on the research but also useful advice on some very practical matters, as well as helping to convince me of the quality and value of, in particular, my first piece.

Through our institute's Vagueness Circle and LeGO Seminar meetings as well as through private e-mail and conversations, many other colleagues in the LoLa group have provided useful comments on both the content and the presentation of the various papers that went into this thesis. For this I would like to thank Bert Baumgaertner, Johan van Benthem, Inés Crespo, Raquel Fernández, Fred Landman, Simon Pauw, Galit Sassoon, Maria Spychalska, Jakub Szymanik, Lavi Wolf, Lucian Zagan and any others I may have forgotten to list here.
I would also like to thank the anonymous peer reviewers who have given their takes on these papers. While some were more fair and insightful than others, each and every one has contributed to improving the end result.

I would like to thank Cian Chartier, Inés Crespo, Paula Henk, Johannes Marti, Simon Pauw and Ben Rodenhauser for regularly getting me to have lunch in a different building in Cafe Polder and for all the discussions we had, especially the

non-work-related ones.
Finally, I want to thank my family and friends for putting up with and supporting me during my darker hours, and for making sure I got out of the house enough.

Amsterdam Harald A. Bastiaanse
January, 2014.

Chapter 1

Introduction

The main title of this dissertation is a curious thing. Without the commas, it would at least be a noun phrase, which is a common element in such titles -though as we can see not quite a sufficient one.
However, removing the commas would actually greatly reduce the amount of sense the title makes as a title for this work. Rather than being important in combination, each of these words forms a key example of an issue we will confront: *Very* illustrates the evolution of non-vague words into vague ones, *Many* shows that an intensional approach to Generalized Quantifiers is appropriate, *Small* is a typical gradable adjective, a class of adjective we shall characterize and explore through a natural logic fragment; finally, *Penguins*, as non-flying birds, are used in the classic 'Tweety Triangle' example in the literature on default rules. These, then, are the issues dealt with in this dissertation.
Now, some elaborate mental gymnastics could be performed to come up with ways these topics are connected. But this would not be a very appropriate thing to do. As the subtitle suggests, the connections between these topics are actually not particularly strong, and those who go in expecting nice cross-references and interconnections leading to grandiose overarching insights shall be disappointed: these chapters stand alone.

Chapter 2 concerns the habit of interpreting the use of certain numbers as 'round', which is to say as an expression which encompasses not only that exact number but also other numbers which are close enough that they would be rounded to that number when rounding. Through the use of game theory and Bayesian statistics, this chapter shows that round interpretation can generally be defended as a rational decision.
The same mechanism also contributes to a loose interpretation of other words. When such a loose interpretation then becomes standard, the same loosening can then happen to this looser standard. If this happens repeatedly enough, a word which was not originally vague can end up becoming vague over time.

A key example of this is the word *Very*, which originally meant 'true, genuine, really' (cf. Ger. *wahr*, Du. *waar*), and turned into a booster in the Middle English period.[1]
Thus, this mechanism offers a (partial) explanation of the origin of (some) vagueness in natural language, and suggests that every natural language will eventually come to contain traces of vagueness.

Chapter 3 concerns the word *Many*, a vague quantifier. In the theory of Generalized Quantifiers, *Many* has long been a problematic case, since there did not appear to be an appropriate formal interpretation of it satisfying Conservativity, a property virtually all other natural language determiners do possess.
This chapter argues that there is a problem with one of the most important examples long used to conclude that *Many* is a problematic case, specifically that *Many* requires an intensional approach, which is otherwise hardly found in the literature. By using an intensional system and an intensional notion of Conservativity, *Many* is no longer problematic.
Beyond this, this chapter adresses intensional versions of several other key properties, provides a general form for intensional quantifiers which guarantees compliance with these properties, and offers a brief look at the logical properties of both *Many* specifically and intensional quantifiers in general.

Chapter 4 offers a syllogistic logic for subsective adjectives, an important category of which *Small* is a key example. Chapter 5 uses this logic to investigate the properties of gradable adjectives, a category containing many standard examples of vagueness (including *Small*). It shows that, when gradable adjectives are defined as those subsective adjectives which are based on an underlying weak order, they can be characterized based solely on their extensions, without having to know the underlying order per se.
Following up on this, it defines and characterizes the notion of a set of gradable adjectives being commensurable, which means roughly that they are based on the same underlying order. This allows a further look into how antonyms, personal taste adjectives, degree modifiers and boolean connectives fit into the framework. Finally, a means is discussed to extend the system to deal with vagueness.

While not particularly concerned with vagueness in the specific sense the other chapters touch on, chapter 6 deals with another vague issue: when we use a bare plural in a construction like "Birds fly", what do we mean? These constructions, referred to as default rules, cannot be taken to simply hide a universal quantification. Penguins (hence the last part of the title) and various other kinds of birds cannot fly, but these counterexamples are not considered to invalidate the truth, such as it is, of the general statement that birds fly.

[1]See Section 2.4.1 for further examples and citations.

Nor can they be interpreted as simply being about a majority. The sentence "It is not the case that most Dutchmen are blond" implies "Most Dutchmen are not blond", but "It is not the case that Dutchmen are blond", with the latter part being a default rule, does not in any way license a conclusion like "Dutchmen are not blond" Furthermore, having default rules of the form "A's are B" and "A's are C" allows the conclusion that "A's are B and C"[2], while "Most A are B" and "Most A are C" do not jointly imply "Most A are B and C".
A more apt interpretation of "Birds fly" would be along the lines of "All normal birds fly" or "All good examples of birds fly", statements which are rather vague indeed. The way we analyze defaults in chapter 6 is to look at what effect default rules (should) have on the reasoning of those who accept them as true. The main question there is what if anything may be concluded when multiple default rules appear to contradict each other. Based on a single underlying principle about the meaning of default rules, we provide a systematic answer to this question.
In the second half of the chapter, the same answer is given in terms of inheritance networks, which are a way of codifying and analyzing sets of default rules without using models of specific objects. The inheritance network approach is proven to give the same results as the model-theoretic approach in cases where either may be used, and furthermore gives rise to a convenient algorithm by which to determine the correct exceptions to make.

Sources of the Chapters. The material in chapter 2 previously appeared in (Bastiaanse 2011). A preliminary version of the material in chapter 3 appeared in (Bastiaanse 2013). For both of these, the final publication is available at http://link.springer.com.
The material in chapters 4 and 5 has not yet appeared elsewhere at the time of writing, but is to be published separately at a later date. Chapter 6 is based on joint work with Frank Veltman, and the material therein is also to be published separately at a later date.

[2]Or at least, the conclusion that a given A of which we know nothing else is (presumably) B and C.

Chapter 2

The Rationality of Round Interpretation

2.1 Introduction

This chapter is about why round numbers are seen as round; that is, as an approximation that can be used to refer to other numbers close to them. Much has been said about round numbers already, but other work has mostly focused on explaining the distribution of round numbers (which I will not be getting into at all) and why a speaker would want to use round numbers.

Instead, we will look at things from the perspective of someone hearing a round number being used. The point will be to show that in addition to what other good reasons there may be, round meaning can also in large part be explained just by the mathematics of the situation and people making rational decisions when interpreting things. After that, we apply the analysis to vagueness.

Despite this difference in approach, I should mention that the idea for this analysis comes from the following remark in (Krifka 2007):

```
(17) a. 0---------------------------------60----------------------...--120-...
     b. 0---------------30----------------60----------------90----...--120-...
     c. 0------15-------30-------45-------60-------75-------90----...--120-...
     d. 0-5-10-15-20-25-30-35-40-45-50-55-60-65-70-75-80-85-90-95-...--120-...
```

> Let the a-priori probability on hearing *forty-five minutes* that one of the scales (17.c) or (17.d) be used be the same, say s. Then on hearing *forty-five minutes* the probability that the more fine-grained scale (17.d) is used is 5rs, and the probability that the more coarse-grained scale (17.c) is used is double the value of that, 10rs. Hence the hearer will assume the more coarse-grained scale.

This is almost a throwaway remark in the piece in question, but it suggests an underlying principle worth far more attention.

Now the central question I will look into in the next sections is: why is it rational for a hearer to interpret a round number as a rounding? I'll investigate this by looking into several questions and the mathematics behind them. The first question is a matter of conditional probability. Some game theory will follow later.

2.2 Conditional Probability

The first question is: *Given that a round number was used, what is the chance that it was meant roundly?* In Bayesian statistics there is a straightforward answer to this question: the probability of A given B is given by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

If A means it was meant roundly and B that a round number was used, then the formula is as above, so we are looking for the chance of both happening divided by the (prior) chance a round number gets used. Keeping in mind that our A is in B and therefore $P(A \cap B) = P(A)$, we obtain

$$P(\textit{meant roundly}|\textit{round number is used}) = \frac{P(\textit{meant roundly})}{P(\textit{round number is used})}$$

Let us look into these chances using an example.

The example we're going to use is as follows: First we take a round number, say, 30. Now there are a bunch of numbers close-by enough that you might round them to 30. We will use the simplifying assumption that only integers are relevant sufficiently close ones have a chance of being rounded to 30. (See Section 2.5 for notes on how to drop both of these assumptions.) Suppose these sufficiently close ones are 25-34, or 10 numbers in total.
Now one of these numbers is randomly selected (with equally distributed probability) and the speaker wants to talk about that number. Finally, the speaker may or may not decide to round that number. Since we are interested in the hearer's side of things, we are going to just assign a value x to the chance that the speaker will choose to round to 30. For this example let us suppose $x = 50\%$. (This is perhaps on the high side, but not much depends on this; the point is to show how much larger than x the final conditional probability is. Also, Section 2.3 will show that a much smaller x can in fact suffice.) Let us see what happens given this situation.

	30	25-34 but not 30
Speaker rounds	$0,5 \cdot \frac{1}{10}$	$0,5 \cdot \frac{9}{10}$
Speaker does not round	$0,5 \cdot \frac{1}{10}$	$0,5 \cdot \frac{9}{10}$

This table outlines the probabilities of the four (a priori) possible situations. In the left column are the situations where the randomly selected number was exactly 30, in the right the ones where it was close but not 30 itself. Similarly, in the top row are the situations where the speaker chooses to round, while in the bottom are the ones where he does not.[1]

Now to get from these numbers to the conditional probability we want, the main thing to do is to apply the condition we were using. That condition was *Given that a round number was used.* Of course, if the number is not actually 30 and the speaker does not round to 30, then he will not say 30. Thus the lower-right corner is irrelevant for us. That is a lot of the total chance we're throwing out, so we can already see where this is going. But let us take a look.

	30	25-34 but not 30
Speaker rounds	$0,05$	$0,45$
Speaker does not round	$0,05$	$0,45$

$$\begin{aligned} P(\textit{Speaker rounded}) &= P(\textit{rounded}; 30) + P(\textit{rounded}; \textit{not 30}) \\ &= 0,05 + 0,45 = 0,5 \\ P(\textit{"30" is used}) &= P(\textit{rounded}) + P(\textit{didn't round}; \textit{30}) \\ &= 0,5 + 0,05 = 0,55 \end{aligned}$$

The other steps are straightforward. To get the chance the speaker rounded, take the chance he rounded and it was 30 and the chance he rounded and it was not and add them together. These are the ones in the top row, and the result is 50% again. For the chance a round number was used, we add to that the chance that the number was 30 and he did not round it, so we get 0,55.

Now we simple divide these, as per the formula. This gives

$$P(\textit{Speaker rounded} | \textit{"30" is used}) = \frac{P(\textit{both})}{P(\textit{"30" is used})} = \frac{0,5}{0,55} = \frac{10}{11} > 90\%$$

Thus, while the chance of the speaker rounding was just 50%, the chance that 30 was meant as round and should be interpreted like that is over 90%.

For the general picture, we replace our 50% chance by x, use an arbitrary round number R, and let k be the number of numbers that could be rounded to it (i.e.

[1] Keep in mind that when the actual number is exactly 30, "rounding" it still makes a difference: 30 meant sharply is not the same as 30 meant in a loose way that encompasses nearby numbers. Note also that the hearer cannot simply hear the difference between the two; indeed, figuring out how the hearer best deals with that is the point here.

10 in the above example). As mentioned before, the exact values of x and k will prove not to be too important.[2]

	Actually R	Merely close to R
Speaker rounded	$x\frac{1}{k}$	$x\frac{k-1}{k}$
Speaker didn't round	$(1-x)\frac{1}{k}$	$x\frac{k-1}{k}$

$$\begin{aligned} P(\textit{Speaker rounded}) &= P(\textit{rounded}; 30) + P(\textit{rounded}; \textit{not 30}) \\ &= x\frac{1}{k} + x\frac{k-1}{k} = x \\ P(\textit{"R" is used}) &= P(\textit{rounded}) + P(\textit{didn't round;}\mathrm{R}) \\ &= x + (1-x)\frac{1}{k} = \frac{k-1}{k}x + \frac{1}{k} \end{aligned}$$

Given these probabilities, the chance the speaker meant the number R as round is as follows:

$$P(\textit{Speaker rounded}|\textit{"R" is used}) = \frac{x}{\frac{k-1}{k}x + \frac{1}{k}} = \frac{kx}{(k-1)x+1} = \frac{k}{k-1+\frac{1}{x}}$$

With k on the large side, this is going to be close to 1. The only problem is if x is low, but for that to get problematic it has to get low enough to be inversely proportional to k.

Thus, just by the mathematics of it understanding numbers as round is the correct choice far more often than one might expect. It would seem to be the **rational** interpretation –and indeed we will be able able to say this with more confidence after section 2.3.

And it would be wrong to think that this will stay limited to hearers only. If round numbers are likely to be interpreted as such, a speaker is likely to anticipate and modify a round number if he actually means it non-roundly. But that makes round interpretation even more rational, since participants can expect this anticipation. This creates a self-reinforcing loop that makes round numbers get interpreted more and more as simply having a round meaning; in appropriate contexts, at any rate.

2.3 Game Theory

For the next part, we are going to look more closely into the rationality angle. The previous question was necessarily a bit indirect; but Game Theory is based on concepts like strategies and making the rational choice between them. Thus, it allows us to specifically ask *When is it rational to assume a round number was*

[2]See Appendix 2.5 for a treatment on how to generalize away from the discrete scale and even probability distribution.

meant roundly?, and to get an exact answer in the form of a value x has to exceed (where, as before, x is the chance of the speaker rounding). Furthermore, we will also be able to find out the exact importance of contextual factors.
To answer this question, Game Theory works by assigning so-called utility values to understanding and misunderstanding each other. Each outcome gets a value: the higher it is the better for everyone involved. These are just numbers, like the example values below. Each of the two hearer strategies then has an expected utility depending on the other player, and round interpretation simply is rational if the expected utility is higher than for non-round interpretation.

For this example, suppose the speaker has asked the hearer to show up for an appointment at 2 o'clock. This could be meant sharply, or could be meant to allow about five minutes either way. Obviously it would be preferable for the hearer to correctly understand the speaker's intent, so these outcomes get a higher value than the rest. We also assume that a greater need for precision gives rise to some inconvenience for one or both parties, so the correctly interpreted strict appointment has a slightly lower score.
Furthermore, showing up sharply on a loosely meant appointment is obviously not as bad as taking a sharply meant appointment loosely, so the values are fixed accordingly.[3,4]

	Round interpretation	*Non-round int*
Round intention	3	1
Non-round intention	0	2

Now as before we are interested in the hearer's point of view and simply let x be the chance that the speaker will round a given number. The better strategy is picked by maximizing expected utility, so round interpretation is rational if and only if

$$\begin{aligned} & P(\textit{Round intention}) \cdot 3 + P(\textit{Non-round intention}) \cdot 0 \\ > \; & P(\textit{Round intention}) \cdot 1 + P(\textit{Non-round intention}) \cdot 2 \end{aligned}$$

Filling in x, this becomes

$$3x + 0(1 - x) > 1x + 2(1 - x)$$

which simplifies to $2x > 2(1 - x)$ which is if and only if $x > \frac{1}{2}$. This result does not actually look all that good, but there is something very important being

[3]There will also be some convenience in the fact that $3 - 1 = 2 - 0$, but this is not part of the story.

[4]Note that while the choice of payoffs here is convenient, it does *not* itself offer an advantage to round interpretation, as should become clear from the calculations as well as the generalized case later one.

overlooked here.
The thing we are overlooking is not unlike the condition we posed earlier. Essentially, if the speaker uses a non-round number, there is no way it can be misinterpreted as round. So the real strategies the hearer chooses from are not round and non-round interpretation; they are to interpret roundly if a round number is used or to never interpret roundly. This changes the analysis considerably.

	Round int [if a round number]	*Non-round int*
Round intention	3	1
Non-round intention	1, 8	2

In the lower-left corner instead of 0 we get 0-if-it's-round-and-two-if-it-isn't. That comes out to $0 \cdot \frac{1}{10} + 2 \cdot \frac{9}{10} = 1,8$.[5,6] This makes round interpretation look a lot better, yielding all the advantage and only a fraction of the disadvantage. As the calculation below shows, x need only be $\frac{1}{11}$ for round interpretation to be

[5] Assuming we are being precise to the minute, resulting in what amount to a $k = 10$ as before.

[6] Readers trying to interpret in terms of signaling games should note that the type t has two independent parameters here: one is the preferred time (even distribution over ten options), the other is the importance of showing up on the minute, which also governs the payoffs. The latter has a probability of x of corresponding to the upper row and $(1 - x)$ of corresponding to the lower one.
Now formalize as follows:

$$\begin{aligned}
t_{1i} &: \textit{preferred time is 14.00} \\
t_{2i} &: \textit{preferred time not 14.00} \\
S_1 &: t_{1i}, t_{2i} \rightarrow \textit{"two o-clock"} \\
S_2 &: t_{1i} \rightarrow \textit{"two o-clock"} \\
&\quad t_{2i} \rightarrow \textit{specific other time} \\
H_1 &: \textit{"two o-clock"} \rightarrow \textit{interpret as round} \\
&\quad \textit{specific other time} \rightarrow \textit{interpret as precise} \\
H_2 &: \textit{any} \rightarrow \textit{interpret as precise}
\end{aligned}$$

That it is rational for the sender to pick S_1 iff showing up on the minute is unimportant is left to the reader. Given this relationship the second parameter and the sender's strategy are both governed by x, and the rest of the analysis follows.

rational.

$$\begin{aligned}
3x + 1,8(1-x) &> x + 2(1-x) \\
3x + 1,8 - 1,8x &> x + 2 - 2x \\
1,2x + 1,8 &> 2 - x \\
2,2x + 1,8 &> 2 \\
2,2x &> 0,2 \\
x &> \frac{0,2}{2,2} = \frac{1}{11}
\end{aligned}$$

The general picture again is similar. In the general case we use not specific numbers but the following arbitrary game:

	Round interpretation	*Non-round int*
Round intention	a	b
Non-round intention	c	d

Any good example will of course have $a > b$ and $d > c$, but the numbers are otherwise open to be chosen freely. Of course, as before the factor k marginalizes the difference between c and d, so that this arbitrary game is transformed into the following actual game:

	Round int [if a round number]	*Non-round int*
Round intention	a	b
Non-round intention	$d - \frac{d-c}{k}$	d

The condition for round interpretation to be rational thus becomes

$$\begin{aligned}
ax + \left(d - \frac{d-c}{k}\right)(1-x) &> bx + d(1-x) \\
(a-b)x &> \frac{d-c}{k}(1-x) \\
(a-b)kx &> (d-c)(1x) \\
((a-b)k + (d-c))x &> d-c \\
x &> \frac{d-c}{(a-b)k + (d-c)}
\end{aligned}$$

Thus because of the generally largish k at the bottom, x can safely be quite small. Usually the breaking point is where it gets inversely proportional to k. If $(d-c) = (a-b)$ (that is, if the cost for misunderstanding is the same either way) then x need only be as little as $\frac{1}{k+1}$ for round interpretation to be the rational choice.

Now context can matter a lot, and that will work its way into what a, b, c and d really are, but clearly the factor k strongly pushes things towards round interpretation.

2.4 Discussion

This chapter shows that even a weak inclination to round can be enough to explain why rounding is [rationally] assumed: even if the chance the speaker chooses to round is low, round interpretation is still likely to be rational, and then people adapt and it gets more and more standard until it is a standard meaning. Roundness is a rational and natural outcome.
It does not purport to −and cannot− explain why speakers should have even a small inclination to round to begin with, but in this it should be favorably combinable with existing arguments focusing on the speaker side or on inherent benefits to rounding (eg arguments from irrelevance, high cost of precision, uncertainty on the part of the speaker, manipulation or mental restrictions). Such other arguments need no longer account for a preference for rounding, just for a sufficiently significant probability.
It also does not go into why such inclinations are limited to "round" numbers. In my opinion that matter is better dealt with through other methods of investigation, eg (Dehaene and Mehler 1992, Jansen and Pollmann 2001).

2.4.1 Generalization to Vagueness

Generalizing the results about round numbers to vagueness is often surprisingly straightforward. While vagueness doesn't have much to do with numbers as such, vague terms often do have an underlying scale that's numerical –or an underlying situation that is easily numerizable, so that the same arguments apply.
This is most clearly seen with absolute adjectives (using the term absolute adjective as used in (Kennedy 2007)). Take for example the word "bald". Loose use of the strictest sense of the word could be interpreted as rounding the number of hairs to zero. But then, given the number of hairs on a normal person's head, the k −the number of hairs that can be rounded to zero− for this situation can easily be in the hundreds or even thousands. The required prior chance of rounding x is thus so low that it can be accounted for even with just the various kinds of uncertainty. In this analysis that is obviously not a stable situation, so the word will quickly get used more and more loosely.
Importantly, this process does not stop. As soon as the meaning has changed (and stabilized), it is again subject to the same analysis. There is a slight difference in that more than one case counts as strictly bald now, but this can be accommodated by replacing k with a factor dividing the number of cases of the looser meaning by that of the new 'strict' meaning. k will be smaller and x may or may not change as well, but even looser interpretation is likely to be rational several more times, and further and further loosening will occur so long as this is so.

So just how loosely will it get used and where does the repeated loosening stop? That question gets hard to answer. Even if we and the people involved are pursuing a rational answer, just how loosely people should use and interpret the word soon depends on all kinds of factors nobody really knows; matters like how loosely everyone else is, should be, has been and should have been using it. Given that people might not use words equally loosely there will be much uncertainty and legitimate disagreement about such things, and this becomes more and more relevant as the process of loosening goes on. Eventually, the word becomes vague.[7] (Some people may prefer the following line of reasoning instead: if precise loose use is rational, there is also support for vague loose use, especially if people aren't actually capable of the former but can manage the latter. In this way we get a reduction of other vagueness to the vagueness inherent in loose use. When loosening stops, then, it is not so much because the term has become vague but because it has become vague enough/too vague, with further loosening making no difference: [current] vague terms are fixpoints of the loosening operator.)

What we have here then is a possible explanation for a lot of vagueness. Loose interpretation is often rational, this makes loose use become the norm over time, and therefore things eventually get vague.
There are a number of reasons to hypothesize that this is indeed the origin of much vagueness. The context-dependence of most vague terms can be explained in terms of the context-dependence of loose use. It also correctly predicts that vagueness occurs mostly for cases where there is an associated measurable property on a continuous or extremely fine scale, as these are the cases the argument is most naturally and easily applied to.[8] A number of vague terms do indeed have an associated "literal" or "absolute" meaning, e.g. "bald", "flat", "full".[9]
Furthermore, if we think absolute adjectives like "flat" and "full" as having prototypes, then the suggestion in prototype theory that the prototypes are by and large clear and universal across while the boundaries between concepts are not is consistent with an account where modern concepts are the result of repeated loosening of concepts that originally coincided with these prototypes far more

[7]There is also another possible reason, which I will not expand on here. If the loosenings of two related words start to overlap, the extensions may stop expanding there, since it remains more rational to use the "closer" word. Still, for the reasons above one would not expect the boundaries this results in to be sharp.

[8]Loose use can involve situations where no clear measurable property is involved –e.g. "I need a Kleenex." (where in fact any tissue would suffice) (Wilson and Sperber 2002)– but in such cases it cannot easily be argued that *repeated* loose use occurs often enough to achieve vagueness.

[9]In some cases, words that don't may have such a meaning at one point only for it to be evolved away or taken over by another word. See also the section on "very". Also, some vague terms may have evolved from other vague terms with the vagueness itself still coming about in the proposed way.
I wouldn't go so far as to propose that this process underlies *all* vagueness, though.

strictly. One example of such a suggestion is made in (Wierzbicka 1990) and supported in (Tribushinina 2008, p58-78)

When we are investigating a word like "bald", one might object that even if it is commonly used to refer to more than just an endpoint, the endpoint still remains and can be referred to with modifiers like "completely" and "absolutely". There would seem to be a difference between the loose use of absolute adjectives and the vagueness of other adjectives such as "tall". However, the section below outlines a big problem with such a view, further suggesting that repeated loosening can in fact produce vagueness.

On *very*, and the futility of remaximizing

It is well-known that many kinds of expressions can be vague, including adjectives, nouns, quantifiers and modifiers. This also includes the word "very", which may in fact be an even better example of this theory than "bald". I suggested just now that modifiers like "completely" and "absolutely" can refer to the endpoint of words like "bald", but is this really the case? In modern times nobody associates the word "very" with any specific endpoint. It is simply a strengthener. But in earlier centuries, they did. There is a paragraph about this in Elena Tribushinina's work(Tribushinina 2008) which is worth quoting at length.

> It is also worth noting that *extremely* is probably undergoing a semantic change from a maximizer to a booster. A similar development has taken place for *quite* and *very*. In the times of Chaucer, *quite* was only used in the sense of 'entirely' (e.g. *quite right*). The weaker sense of 'fairly' (as in *quite tall*) is attested from mid 19^{th} century (Paradis 1997: 74). Similarly, *very* originally meant 'true, genuine, really' (cf. Ger. *wahr*, Du. *waar*), and turned into a booster in the Middle English period (Cuzzolin & Lehmann 2004; Lorenz 2002; Mendez-Naya 2003; Peters 1994; Stoffel 1901).[10]

So as we can see here "very" originally meant something along the lines of "truly" or "completely", until it succumbed to the kind of pressures we have been talking about, which are also affecting "extremely", "totally", "completely" and pretty much every maximizer you can think of.[11] The phenomenon is well documented[12],

[10] The papers she cites are (Cuzzolin and Lehmann 2004), (Lorenz 2002), (Méndez-Naya 2003), (Peters 1994) and (Stoffel 1901), respectively.

[11] Indeed, many people have been annoyed at the way even "literally" gets (ab)used these days. From a discussion on the internet:

> A: I literally ROFL'd.
> B: You literally rolled over the floor laughing? Ouch.

People who understand both "literally" and "ROFL" can be hard to come by.

[12] See also (Ito and Tagliamonte 2003).

and is entirely natural and perhaps because of these arguments also fairly predictable.
And of course, if even "very" can turn out to have come about in this way, so can any other word.

2.4.2 Schelling Points and Evolutionary Game Theory; a problem?

During the course of writing this chapter it has come to my attention that Christopher Potts has done a related game-theoretical analysis on a related phenomenon. (Potts 2008) While his subject matter is different, one of his predictions contradicts an important one of my own. Before I mention how I account for this, a brief introduction of it is in order.
In (Potts 2008), Potts seeks to derive Kennedy's Interpretive Economy principle (Kennedy 2007), or rather, a substitute with the same practical consequences (in particular, solving Kennedy's puzzle) as that principle, from basic assumptions about cognitive prominence and evolutionary stability. This of course has little to do with general vagueness, much less round numbers, but his analysis would still be problematic for my own ideas discussed above.
Potts's argument rests on the notion that amongst the possible ways to interpret an adjective related to a scalar endpoint, the most strict one stands out as a so-called Schelling point, making it initially (at least marginally) more likely to be selected than other ways. The extent to which this is so is what he refers to as the strength of the "Schelling assumption". Insofar as the Schelling assumption is fairly weak, I will not argue against it here.
He then combines the Schelling assumption with evolutionary game theory, arguing that even a slight preference will result in strict interpretation becoming standard. This is a fairly straightforward application of evolutionary game theory, and I will mostly not argue against it either.
However, it does go against my own notions: in Section 2.4.1 in particular I argued that the evolution is likely to go the other way around, with vague words possibly being a result of repeated loosening of previously much sharper words. So how do I account for this? Naturally, the answer lies in doing what I have been doing in this chapter.
Potts's most important analysis starts from the following basic game:

	$[\![\textbf{full}]\!]_.$	$[\![\textbf{full}]\!]_d$
$[\![\textbf{full}]\!]_.$	10	9.9
$[\![\textbf{full}]\!]_d$	9.9	10

In this example, $[\![\textbf{full}]\!]_.$ represents the maximum (ie sharp) interpretation of "full" while $[\![\textbf{full}]\!]_d$ represents a looser interpretation. In order to let this conform more to the examples I have been using myself I will flip the table here, as follows:

	$[\![\textbf{full}]\!]_d$	$[\![\textbf{full}]\!]_.$
$[\![\textbf{full}]\!]_d$	10	9.9
$[\![\textbf{full}]\!]_.$	9.9	10

Now using evolutionary mechanics Potts shows that when a coordination game like this is repeated, even a very weak Schelling assumption will make the population evolve towards overwhelmingly favoring the Schelling point –in this case strict interpretation.
There is nothing inherently wrong with this analysis, except that it ignores the point I have been making in this chapter. Stay with this example, a loose usage of the word "full" can be used in more situations than strict use. Following the analyses of this chapter, we should assign a discrete scale or use the continuous analysis in Appendix 2.5 to find the appropriate number k for the amount/ratio of situations sufficiently close to be loosely referred to as "full".[13]
Assuming either an even distribution or one taken included as part of k as per Appendix 2.5, we should then follow Section 2.3 and replace the 9.9 in the lower-left by $9.9 \cdot \frac{1}{k} + 10 \cdot \frac{k-1}{k} = 10 - \frac{0.1}{k}$, thus replacing the basic game above by the following:[14]

	$[\![\textbf{full}]\!]_d$	$[\![\textbf{full}]\!]_.$
$[\![\textbf{full}]\!]_d$	10	9.9
$[\![\textbf{full}]\!]_.$	$10 - \frac{0.1}{k}$	10

By the math in the earlier Section 2.3, it follows that loose interpretation is rational if $x > \frac{1}{k+1}$. In this example the population distribution provides this x, and if loose interpretation is rational at the initial time t_0 it will only get more so, so the condition for loose interpreation to be the end result of evolution becomes $P^{t_0}([\![\textbf{full}]\!]_d) > \frac{1}{k+1}$. Therefore a weak Schelling assumption (where it suffices for $P^{t_0}([\![\textbf{full}]\!]_.)$ to be just barely higher than 50%) is nowhere near enough. To win, strict use would have to start out at more than $\frac{k}{k+1}$.
Given everything I've argued here, a factor benefiting strict use needs to be strong, not merely minimal, to be of much use against the k factor.

[13] The value of k in this depends on what specific d is being used, but since the stricter reading consists of a single point it depends even more on how fine the scale is. Indeed, increasingly fine scales can render k arbitrarily high.

[14] or in Potts's notation,

	$[\![\textbf{full}]\!]_.$	$[\![\textbf{full}]\!]_d$
$[\![\textbf{full}]\!]_.$	10	$10 - \frac{0.1}{k}$
$[\![\textbf{full}]\!]_d$	9.9	10

2.5 Appendix: Continuous Scale and k on Probability

It has been convenient to use the simplifying assumption of a discrete scale, but it is straightforward enough and interesting to drop this notion, especially in light of the discussion in section 2.4.1.

Starting from the general case scenario in Section 2.2, let R be some round number and as before let x be the prior chance that a sufficiently close number will be rounded to it. Let C be a set of real numbers sufficiently close to R to be rounded to it in this fashion. In order to avoid dividing by zero later on, we also let $A \subset C$ be a set of numbers so close to R as to be considered identical, or at least indistinguishable.[15]

Now let $B = C - A$, assume that the actual number is picked randomly with probability distributed evenly over C, and assume that $|\cdot|$ is an appropriate measure on $\mathbb{R}$.[16] Then we can "divide out"/ignore the probability part to obtain the following familiar-looking table:

	Actually R	Merely close to R
Speaker rounded	$x\|A\|$	$x\|B\|$
Speaker didn't round	$(1-x)\|A\|$	$(1-x)\|B\|$

I have not yet mentioned how k should be defined here, but by looking at the table it should surprise no one that the definition is simply $k = \frac{|C|}{|A|} = \frac{|A|+|B|}{|A|}$[17]. This leads to the following:

$$\begin{aligned} P(\textit{Speaker rounded}|\text{"}R\text{" is used}) &= \frac{x|A|+x|B|}{x|A|+x|B|+(1-x)|A|} = \frac{|A|+|B|}{|B|+|A|/x} \\ &= \frac{(|A|+|B|)/|A|}{\left(\frac{|A|+|B|}{|A|}-\frac{|A|}{|A|}\right)+1/x} = \frac{k}{k-1+\frac{1}{x}} \end{aligned}$$

which is of course the same result as in the discrete case.

Taking the probability distribution out in this way may seem suspect, and in any case it is interesting to consider the impact of non-even distributions. The resulting formula threatens to get convoluted, but this is easily avoided through cheating: redefine k as

$$k = \frac{P(A \cup B)}{P(A)}$$

[15] Of course in more general situations A may also simply be whatever R refers to sharply, so long as that has non-zero measure.

[16] In the more general case, pick an appropriate measure on at least C.

[17] In the general case, the equality obtains because A and B are disjoint and we picked an appropriate measure function.

Then it is clear that we can just combine area and distribution into probability to get the following table:

	Actually R	Merely close to R
Speaker rounded	$xP(A)$	$xP(B)$
Speaker didn't round	$(1-x)P(A)$	$(1-x)P(B)$

Thus the results are exactly as before[18,19] except that now the effect of a change in probability distribution is a straightforward impact on k: for instance, the k in the above example could end up much fairly small if the distribution were a bell curve around R, with details depending on σ and the size of A.

[18]In this case the equality $P(A \cup B) = P(A) + P(B)$ follows from the laws of probability.

[19]Reobtaining the exact results from the sections involving game theory is not too difficult –and left to the reader.

Chapter 3

The Intensional Many - Conservativity Reclaimed

3.1 Introduction

In the theory of Generalized Quantifiers, much weight is given to the property of Conservativity, which for a binary quantifier Q can be paraphrased as

$$QAB \text{ if and only if } QA(A \text{ and } B)$$

Conservativity is often suggested as a linguistic universal (eg (Barwise and Cooper 1981)(Keenan and Stavi 1986)), as it seems almost trivially true for virtually every natural language determiner. For instance, all of the following seem obvious enough:

No man is perfect.	$\Leftrightarrow$	No man is a perfect man.
Seven women are running.	$\Leftrightarrow$	Seven women are women who are running.
All good philosophers are wise.	$\Leftrightarrow$	All good philosophers are good philosophers who are wise.
Many men smoke.	$\Leftrightarrow$	Many men are men who smoke.

The last one, however, is actually problematic.

The problem

Westerståhl (Westerståhl 1985) coined the following classic example to demonstrate the problem. In a certain class at a certain college 10 out of the 30 students got the highest grade on a certain exam, which is unusually many. Those same 10 students are the only ones in the class who are right-handed, which is unusually few. Let A be the set of students in the class, B_1 the set of students at the college

who got the highest grade in their class, and B_2 the set of right-handed students at the college.
Thus, the assumptions from the example are expressed roughly as follows:

$$\textbf{many}(A, B_1), \textit{not } \textbf{many}(A, B_2)$$

If Conservativity were true of many, from this we could then conclude.

$$\textbf{many}(A, A \cap B_1), \textit{not } \textbf{many}(A, A \cap B_2)$$

But of course $A \cap B_1$ and $A \cap B_2$ are in fact the same set. Hence "many" can not be Conservative, or at least not without giving it two different interpretations to arbitrarily fix the problem.

Issues

It is hard to argue with the formal part of this argument, but it does leave something to be desired. For as soon as we translate the result back into natural language, serious problems with our intuitions arise. If we give up on Conservativity for this case and reject the conclusion that $\textbf{many}(A, B_1)$ and *not* $\textbf{many}(A, A \cap B_2)$, then we have to in turn accept the opposite of at least one of these. Hence we would be forced to accept one of the following natural language sentences:

- *Not many students in the class are students in the class who got the highest grade on the exam.*
  (While we at the same time accept many students in the class did get the highest grade.),

or

- *Many students in the class are right-handed students in the class.*
  (While we at the same time accept not many students in the class are right-handed.)

Neither of these is a particularly attractive statement to endorse, and then there is the question of which of the two we should pick. Westerståhl offers no answer to this question, and it is hard to see how anyone could; they seem equally counterintuitive and "resolve" the inconsistency equally well. So how do we get out of this problem?

I would say that rather than a straightforward case against Conservativity for many, what the example really provides is a complication arising from a different problem.
We saw before that $A \cap B_1$ and $A \cap B_2$ were the same set. Let us call this set

C. Do we now have any intuitions about the sentence "Many students in the class are C"? Of course not. There is no obviously correct way to parse C as something we would have intuitions about. To have an idea about whether 10 students in a class being C is many or not, we need to know not the set itself but the *property* it is representing -and hence presenting the same set as an instantiation of different properties leads to different intuitions. This, then, is our problem. The theory of Generalized Quantifiers as formalized by Barwise & Cooper (Barwise and Cooper 1981) and van Benthem (van Benthem 1984) is inherently extensional: while it involves possible universes and how quantifiers deal with them, it does not allow properties to be identified as more than subsets of a specific universe. We can use it to talk about "right-handed students at the college, in this particular world/situation", but not of right-handedness as a property in its own right identified independent of any one universe. We are limited to identifying properties by their local extensions, whereas many requires an intensional approach.
This, of course, is not a particularly new thought. The fact that many is intensional has been generally agreed upon after being pointed out by Keenan and Stavi (Keenan and Stavi 1986). What *is* interesting here is that we shall see that when it is treated in this way, Conservativity is reclaimed.

In the next section we will construct an intensional framework for generalized quantifiers and create an intensional version of Conservativity. We will then show that this move resolves the issues created by the example, and further support this position by providing a specific reading of many which works well for it and is (Intensionally) Conservative.
The point of doing this is not to suggest that this is the single best reading of many, or even that it is the single best framework in which to consider such readings. Rather, the point is to demonstrate that when cast into a proper intensional form, Conservativity can be reclaimed as an important standard by which to judge quantifiers, even previously problematic ones like many.
In Section 3.3, we take a look at some other partly intensional readings of many that have been proposed and see to what extent they can meet this standard.

3.2 An Intensional Framework

3.2.1 Framework

3.2.1. Definition. Where L is a set of predicates closed under Boolean combination, a *structure S for L* is a triple $< W, D, [\![\cdot]\!] >$ where W is a non-empty set of *possible worlds*[1], D assigns to each world $m \in W$ a non-empty set $D(m)$

[1]The set of worlds W serves as a basis from which to derive intensional standards that are not (heavily) dependent on the interpretations in any one world. The idea here is *not* that W would include every logical possibility, but rather that it is made up of worlds which are much

referred to as the *domain* of that world, and $[\![\cdot]\!]$ assigns to each predicate $A \in L$ its *intension* $[\![A]\!]$, which in turn for each world m determines the *extension* $[\![A]\!]^m$ of A. We demand that $[\![A]\!]^m \subseteq D(m)$ and that intensions satisfy the following rules:

$$\begin{aligned} [\![A \wedge B]\!]^m &= [\![A]\!]^m \cap [\![B]\!]^m \\ [\![A \vee B]\!]^m &= [\![A]\!]^m \cup [\![B]\!]^m \\ [\![\neg A]\!]^m &= D(m) - [\![A]\!]^m \end{aligned}$$

More generally, a *property on* S is a function which assigns to each $m \in W$ a subset of $D(m)$.

Thus we may identify each $m \in W$ with the first-order model $< D(m), [\![\cdot]\!]^m >$ (where the derived interpretation function $[\![\cdot]\!]^m$ simply assigns to each predicate its extension in m, as previously defined). From now on we will refer to possible worlds as models. Also, we will use capital letter from the beginning of the alphabet (A, B, C) for predicates and boldface capitals from the end of the alphabet $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ for properties and write quantifiers in boldface.

We now get to the essential non-cosmetic change, which is that quantifiers are applied to properties rather than extensions.

3.2.2. DEFINITION. An *intensional quantifier* $\mathbf{Q}$ is a function whose input consists of two properties on the same S and a model in W and whose output is the evaluation true or false.
We will write $\mathbf{Q}_m\mathbf{XY}$ to denote that this evaluation is true -and hence $\mathbf{Q}_m[\![A]\!][\![B]\!]$ when the properties in question are the intensions of the predicates A and B.[2]

3.2.2 Intensional Conservativity

For the sake of generality, we define Intensional Conservativity in terms of arbitrary properties, rather than only those which are the intensions of predicates.
To do this, we first need to define a property-conjunction operation, which obviously is just to say that $\mathbf{X} \wedge \mathbf{Y}$ is the unique property satisfying

$$\forall m : (\mathbf{X} \wedge \mathbf{Y})^m = (\mathbf{X}^m) \cap (\mathbf{Y}^m)$$

like the actual world except (possibly) for the issues at hand, for which they will by and large correspond to our expectations and the things we consider normal and plausible.

[2] In more traditional intensional semantics, the thing we call a *structure* above is referred to as a *model*, and $\mathbf{Q}_m\mathbf{XY}$ would be expressed as $S \models QXY[m]$.

It is now a straightforward task to rephrase the definition of Conservativity into Intensional Conservativity, which we define as follows:

For all S, for all properties $\mathbf{X}, \mathbf{Y}$ on S, for all $m \in W$,

$$\mathbf{Q}_m\mathbf{XY} \Leftrightarrow \mathbf{Q}_m\mathbf{X}(\mathbf{X} \wedge \mathbf{Y})$$

To see that Conservativity is now possible, let us take another look at the earlier example. Let the predicate C stand for students at the particular college in question, and A for students at the particular class. Let R stand for right-handedness and H for getting the highest grade in class.
Now the complex predicates $B_1 = C \wedge H$, $B_2 = C \wedge R$ are appropriate to express the assumptions of the example, which amount to

$$\mathbf{many}_m[\![A]\!][\![B_1]\!],\ \textit{not}\ \mathbf{many}_m[\![A]\!][\![B_2]\!]$$

The question is: can **many** be interpreted in a way that satisfies the above while also being intensionally conservative?
It can. From these assumptions, Intensional Conservativity merely lets us conclude that

$$\mathbf{many}_m[\![A]\!][\![A \wedge B_1]\!], \neg\mathbf{many}_m[\![A]\!][\![A \wedge B_2]\!]$$

Since $[\![A \wedge B_1]\!]$ and $[\![A \wedge B_2]\!]$ are not the same properties, this does not lead to a contradiction.

A sample reading

While technically the above is enough to conclude the argument, it will carry more weight when we have an actual single interpretation $\mathbf{Q}$ that is a reasonable reading of many and satisfies these conditions.
For this we use just one further simplifying assumption, that W is finite.[3]

Given this, consider the following definition, which says roughly that many students have property $\mathbf{Y}$ iff the relative number of students who have that property is larger than the average of that same number taken over all models:[4]

$$\mathbf{Q}_m\mathbf{XY} \Leftrightarrow \left(\frac{|\mathbf{Y}^m \cap \mathbf{X}^m|}{|\mathbf{X}^m|} > \frac{1}{|W|} \sum_{n \in W} \frac{|\mathbf{Y}^n \cap \mathbf{X}^n|}{|\mathbf{X}^n|} \right)$$

[3]This assumption may sometimes be undesirable, but keep in mind that this reading is merely an illustrative example. We shall see in Section 3.2.6 that there is a broad general form such that any reading of that form will possess Conservativity and other key properties.
Thus, for certain infinite W the average could be generalized using series summation $\left(\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \frac{|\mathbf{Y}^{w_i} \cap \mathbf{X}^{w_i}|}{|\mathbf{X}^{w_i}|}\right)$ or integration $\left(\int_W h(w) \frac{|\mathbf{Y}^w \cap \mathbf{X}^w|}{|\mathbf{X}^w|} \mathrm{d}w,\ \textit{where}\ \int_W h(w)\mathrm{d}w = 1\right)$, or be replaced by an intensional standard based on a probability function on W, a subset of particularly 'normal' or normative worlds, or some other notion (see also Section 3.3). For any of these, the desirable properties remain attainable.

[4]To get around division by zero, we may harmlessly use $\frac{0}{0} = 1$.

The example establishes that $|[\![A]\!]^m| = 30$, while $|[\![B_1]\!]^m \cap [\![A]\!]^m| = |[\![B_2]\!]^m \cap [\![A]\!]^m| = 10$. Since it is rare for as many as a third of students to get a top grade, we may expect $\frac{|[\![B_1]\!]^n \cap [\![A]\!]^n|}{|[\![A]\!]^n|}$ to be lower on average, and thus we obtain $\mathbf{Q}_m[\![A]\!][\![B_1]\!]$. At the same time, since right-handedness is commonplace, we may expect $\frac{|[\![B_2]\!]^n \cap [\![A]\!]^n|}{|[\![A]\!]^n|}$ to average significantly higher than one-third, so that we do *not* get $\mathbf{Q}_m[\![A]\!][\![B_2]\!]$.

This takes care of the basic setup. We should now see if we get $\mathbf{Q}_m[\![A]\!][\![A \wedge B_1]\!]$, $\neg\mathbf{Q}_m[\![A]\!][\![A \wedge B_2]\!]$. And indeed we do. To see that the definition satisfies Intensional Conservativity -and therefore gives those results- it is enough to note that (for all $\mathbf{X}, \mathbf{Y}$)

$$|\mathbf{Y}^m \cap \mathbf{X}^m| = |(\mathbf{X}^m \cap \mathbf{Y}^m) \cap \mathbf{X}^m| = |(\mathbf{X} \wedge \mathbf{Y})^m \cap \mathbf{X}^m|.$$

This of course is but a single possible interpretation of a single possible reading of many, but it seems likely that a variety of other options will work equally well, and we will see later that this is indeed the case. Thus, when intensionality is properly accounted for, Conservativity does not need to be given up as a universal property of natural language determiners, not even for many.

3.2.3 On Scandinavians and the Reverse Reading

Taking an intensional approach to *many* not only helps to reclaim Conservativity, it also resolves a different issue: that of the so-called Reverse Reading whereby a quantifier will sometimes seem to take its arguments in the opposite order from what the sentence structure would suggest.

A famous example of this is found in (Westerståhl 1985). Consider the following sentences:

(1) Many winners of the Nobel Prize in Literature are Scandinavian.
(2) Many Scandinavians have won the Nobel Prize in Literature.
(3) Many Scandinavians are Nobel Prize winners in Literature.

As of the year 1984, 14 out of a total of 81 winners of the Nobel Prize in Literature are Scandinavians. This would seem surprisingly many, and it is generally agreed that the sentence (1) is true here. Furthermore, it is generally felt that from an intuitive point at least, sentence (2) should be true.

Sentence (3) would seem to be a slightly different way of phrasing sentence (2). However, Westerståhl argues that (3) is clearly false, on the basis that 14 is a very small number compared to the number of Scandinavians. He goes on to suggest that while (3) certainly corresponds to a possible reading of (2), the preferred reading of (2) is expressed by (1). Thus, the logical form of (2) would have the arguments of the quantifier reversed relative to what the surface form would suggest.

Contrary to this view, I maintain that (2) and (3) should be rendered the same way and can be found to be true without resorting to a reversed reading equivalent to (1). To see how this may be done, let us again take the example reading of *many* we used earlier:

$$\mathbf{Q}_m\mathbf{XY} \Leftrightarrow \left(\frac{|\mathbf{Y}^m \cap \mathbf{X}^m|}{|\mathbf{X}^m|} > \frac{1}{|W|}\sum_{n\in W}\frac{|\mathbf{Y}^n \cap \mathbf{X}^n|}{|\mathbf{X}^n|}\right)$$

Using **S** for "Scandinavian" and **N** for "Nobel Prize in Literature winner", sentence (3) would be true iff the following holds.

$$\left(\frac{|\mathbf{N}^m \cap \mathbf{S}^m|}{|\mathbf{S}^m|} > \frac{1}{|W|}\sum_{n\in W}\frac{|\mathbf{N}^n \cap \mathbf{S}^n|}{|\mathbf{S}^n|}\right)$$

On the left-hand side, we have the relative number of Nobel Prize in Literature winners among Scandinavians in this world. This of course is a tiny number. So why is it wrong to say that this reading is clearly false?
The trick is that the important comparison here is not between Prize winners and Scandinavians, nor even between Scandinavian Prize winners and other Scandinavians. Rather, the comparison that matters is between this possible world and others.
As is conventional, let us assume for the sake of argument that the actual world is fairly normal in the sense that other worlds by and large have a similar amount of Scandinavians as the actual world. Thus, the division by $|\mathbf{S}^m|$ for the actual world is by and large comparable to the division by $|\mathbf{S}^n|$ in others. This suggests the comparison will be true so long as $|\mathbf{N}^m \cap \mathbf{S}^m|$ is substantially larger than the average $\frac{1}{|W|}\sum_{n\in W}|\mathbf{N}^n \cap \mathbf{S}^n|$ across all worlds. But the reason we take (1) to be true in the example is exactly that among the possible worlds we consider there are generally substantially less Scandinavian Nobel Prize winners than in the real world. Thus, so long as W is chosen in a way appropriate to the example this reading will predict that (3) is true.

3.2.4 Other key properties

Conservativity is not the only property taken to apply to virtually all natural language determiners. Two important others are Extension (which I will mostly refer to by the abbreviation EXT[5]) and Isomorphism closure. Let us see how well many does on intensionalized versions of those.

Intensional EXT

We start with Extension. Extension roughly states that when a domain M is extended to M', the interpretation relative to that domain remains the same. For

[5]Given how much here revolves around intensions and extensions, to do otherwise could invite confusion.

traditional binary quantifiers, this is defined as follows (paraphrasing (Westerståhl 2007, p279)):

$$\textit{If } A, B \subseteq M \subseteq M'$$
$$\textit{then } Q_M AB \Leftrightarrow Q_{M'} AB$$

The point of EXT is domain restriction; it serves to make everything in $M - (A \cup B)$ irrelevant to the interpretation of $Q_M AB$.
Under the circumstances one might well think that the highly context- dependent many stands a poor chance of satisfying any version of EXT. Yet it is quite possible.
In fact, we shall intensionalize a more broadly defined property EXT*, defined as:

$$\textit{If } A, B \subseteq M, A, B \subseteq M'$$
$$\textit{then } Q_M AB \Leftrightarrow Q_{M'} AB$$

(It's worth pointing out that in the traditional approach the difference is largely irrelevant, as regular EXT gives $Q_M AB = Q_{A \cup B} AB = Q_{M'} AB$. However, EXT* is more convenient to work with when intensionalizing.) We define our Intensional version of EXT as follows:

$$\text{If } \mathbf{X}^m = \mathbf{X}^{m'}, \mathbf{Y}^m = \mathbf{Y}^{m'}, \text{ then } \mathbf{Q}_m \mathbf{XY} \Leftrightarrow \mathbf{Q}_{m'} \mathbf{XY}$$

This amounts to saying that $\mathbf{Q}_m \mathbf{XY}$ depends on m only insofar as it depends on the interpretations of $\mathbf{X}$ and $\mathbf{Y}$ in m: where those stay the same, so does the evaluation.
This sounds like a tall order, but it is satisfied by the interpretation from our earlier example. To see this, it suffices to note that

$$\frac{|\mathbf{Y}^m \cap \mathbf{X}^m|}{|\mathbf{X}^m|} = \frac{|\mathbf{Y}^{m'} \cap \mathbf{X}^{m'}|}{|\mathbf{X}^{m'}|}.$$

There are some important caveats to this result. First of all, Intensional EXT does *not* mean the quantifier only "has access to" the interpretations in the local universe. It still has access to the properties themselves. What it does mean is that insofar as the quantifier has access to more than the local interpretations of $\mathbf{X}$ and $\mathbf{Y}$, it only has such access in a model-independent way.
For example, in the reading for "many" we used in Section 3.2.2, the quantifier used this access to $\mathbf{X}$ and $\mathbf{Y}$ to generate the comparison standard $\frac{1}{|W|} \sum_{n \in W} \frac{|\mathbf{Y}^n \cap \mathbf{X}^n|}{|\mathbf{X}^n|}$. Such behavior is not undesirable, and arguably is part of the point of using an intensionalized definition.

Second, even this intensional version might not be possible or desirable for every reading we want to model. Those who compare things against alternatives (eg (Cohen 2001)(Tanaka 2003)) risk running foul of it. More on this in Section 3.3.

Intensional Isomorphism closure

Next, we consider Isomorphism closure, sometimes abbreviated ISOM. In the traditional version, this can be rendered as follows ((Westerståhl 2007)[p281]):

If f is a bijection from M to M',
then $Q_M AB \Leftrightarrow Q_{M'} f[A]f[B]$

The point of Isomorphism closure is to ensure that quantifiers cannot distinguish between individual elements in a universe, or even across universes.
Since models in our formalism come with interpretation functions, the Intensional version is slightly more complicated:

If there is a bijection $f : D(m) \to D(m')$
with $f[\mathbf{X}^m] = \mathbf{X}^{m'}, f[\mathbf{Y}^m] = \mathbf{Y}^{m'}$
then $\mathbf{Q}_m\mathbf{XY} \Leftrightarrow \mathbf{Q}_{m'}\mathbf{XY}$

We demand not only a bijection f from $D(m)$ to $D(m')$, but also that the interpretations of $\mathbf{X}$ and $\mathbf{Y}$ in the two models are related through this same bijection. This is similar to demanding that f is an isomorphism, except that the demand is more of a local one for each pair.[6]
(Also, note that since f still works on the level of domains rather than involving properties, the conclusion is phrased a bit differently.)
It is straightforward enough to see that our earlier interpretation of many satisfies this property as well. The key part of that interpretation was the following comparison.

$$\frac{|\mathbf{Y}^m \cap \mathbf{X}^m|}{|\mathbf{X}^m|} > \frac{1}{|W|}\sum_{n \in W} \frac{|\mathbf{Y}^n \cap \mathbf{X}^n|}{|\mathbf{X}^n|}$$

Let us focus on the left side first. Since $f[\mathbf{X}^m] = \mathbf{X}^{m'}$ and f is a bijection, it follows that $|\mathbf{X}^m| = |\mathbf{X}^{m'}|$.
To see that $|\mathbf{Y}^m \cap \mathbf{X}^m| = |\mathbf{Y}^{m'} \cap \mathbf{X}^{m'}|$, note that since f is a bijection, the following holds:

$$\begin{aligned} f[\mathbf{Y}^m \cap \mathbf{X}^m] &= f[\mathbf{Y}^m] \cap f[\mathbf{X}^m] \\ &= \mathbf{Y}^{m'} \cap \mathbf{X}^{m'} \end{aligned}$$

Therefore as before $|\mathbf{Y}^m \cap \mathbf{X}^m| = |\mathbf{Y}^{m'} \cap \mathbf{X}^{m'}|$. Thus the left side of the equation is the same for m and m'. This is trivially true for the right side, and therefore $\mathbf{Q}_m\mathbf{XY} \Leftrightarrow \mathbf{Q}_{m'}\mathbf{XY}$.

[6] As a first thought it might look desirable to go much further and that f be an actual isomorphism; i.e. that $f[\mathbf{X}^m] = \mathbf{X}^{m'}$ holds for all properties. However, one can always define, say, a property $\mathbf{X}$ for which $\mathbf{X}^m$ and $\mathbf{X}^{m'}$ do not even have the same number of elements. Thus, making such a broad demand would guarantee that no such f exists for any structure, rendering the whole thing worthless.
Therefore we are forced to work only with those properties which work well with f (for at least one f).

3.2.5 Relation with Extensional Properties

One may wonder whether we are justified in believing that the 'lifted' properties we have come up with in this section represent the most appropriate way of intensionalizing. But they are far from arbitrary. In all three cases they can be naturally related to their original counterpart through a straightforward lifting function.

3.2.3. DEFINITION. For a non-intensional quantifier Q, define its *intensional lift* Q^* as follows:

$$Q^*_m\mathbf{XY} \Leftrightarrow Q_{D(m)}\mathbf{X}^m\mathbf{Y}^m$$

This lifting function leads to the following correspondence theorem.

3.2.4. THEOREM. *Where Q is a non-intensional quantifier and Q^* is its lift:*

- *Q^* satisfies Intensional Conservativity if and only if Q is Conservative*
- *Q^* satisfies Intensional EXT if and only if Q satisfies EXT*, where EXT* is like EXT but applies for any M, M' such that $A, B \subseteq M$, $A, B \subseteq M'$*
- *Q^* satisfies Intensional Isomorphism closure if and only if Q satisfies Isomorphism closure*

Thus, all three of them are natural and true broadenings of their original counterparts. For proof of the above, see Appendix 3.4.1.

The lifting function suggests another matter of some interest: under which conditions can an intensional quantifier (or at least a quantifier expressed in terms of this framework) be interpreted as the lift of a traditional extensional one? This question is answered in Appendix 3.4.2.
(Of course, appropriate readings of 'many' cannot be interpreted as such lifts.)

3.2.6 General form

The intensionalized properties described above obviously apply to far more than the simple example reading of many. We will generalize that reading greatly to obtain a general form of intensional quantifier they also apply to. Besides being interesting in its own right, this will be useful when looking at other approaches in the next section.
Our sample reading was as follows:

$$\mathbf{Q}_m\mathbf{XY} \Leftrightarrow \left(\frac{|\mathbf{Y}^m \cap \mathbf{X}^m|}{|\mathbf{X}^m|} > \frac{1}{|W|} \sum_{n \in W} \frac{|\mathbf{Y}^n \cap \mathbf{X}^n|}{|\mathbf{X}^n|} \right)$$

Here the fraction of $\mathbf{X}$'s in a particular model that are also $\mathbf{Y}$ had to be larger than the average of that same fraction over all models. To generalize this, we replace "fraction of $\mathbf{X}$'s in a particular model that are also $\mathbf{Y}$" with an arbitrary function a (an Actual value of something) depending only on $|\mathbf{X}^m|$ and $|(\mathbf{X}\wedge\mathbf{Y})^m|$, "larger than" with an arbitrary relation $\succ$, and "the average of ..." with an arbitrary function st (an intensionally determined STandard value) depending only on $\mathbf{X}$, $\mathbf{X}\wedge\mathbf{Y}$ and W.
Formally, then, we get the following.

3.2.5. DEFINITION. A quantifier $\mathbf{Q}$ has the *general form* iff the following is true

$$\mathbf{Q}_m\mathbf{XY} \Leftrightarrow a(|\mathbf{X}^m|, |(\mathbf{X}\wedge\mathbf{Y})^m|) \succ st(\mathbf{X}, \mathbf{X}\wedge\mathbf{Y}, W)$$

with a, st, $\succ$ as above.

It is perhaps not immediately obvious that *All* and *Some* have the general form, but this can be shown to be true if the right choices are made. These and other examples are listed below.

Quantifier	$a(\lvert\mathbf{X}^m\rvert, \lvert(\mathbf{X}\wedge\mathbf{Y})^m\rvert)$	$\succ$	$st(\mathbf{X}, \mathbf{X}\wedge\mathbf{Y}, W)$
All	$\frac{\lvert(\mathbf{X}\wedge\mathbf{Y})^m\rvert}{\lvert\mathbf{X}^m\rvert}$	$\geq$	1
Some	$\lvert(\mathbf{X}\wedge\mathbf{Y})^m\rvert$	$\geq$	1
At least n	$\lvert(\mathbf{X}\wedge\mathbf{Y})^m\rvert$	$\geq$	n
Exactly n	$\lvert(\mathbf{X}\wedge\mathbf{Y})^m\rvert$	$=$	n
At most n	$\lvert(\mathbf{X}\wedge\mathbf{Y})^m\rvert$	$\leq$	n
Most	$\frac{\lvert(\mathbf{X}\wedge\mathbf{Y})^m\rvert}{\lvert\mathbf{X}^m\rvert}$	$>$	$\frac{1}{2}$
More than x% of	$\frac{\lvert(\mathbf{X}\wedge\mathbf{Y})^m\rvert}{\lvert\mathbf{X}^m\rvert}$	$>$	$\frac{x}{100}$

3.2.6. THEOREM. *Every quantifier that has the general form (and indeed, every quantifier whose evaluation depends only on* $\boldsymbol{X}, \boldsymbol{X}\wedge\boldsymbol{Y}, |\boldsymbol{X}^m|, |(\boldsymbol{X}\wedge\boldsymbol{Y})^m|$ *and* W*) satisfies Intensional Conservativity, Intensional EXT and Intensional Isomorphism closure.*

PROOF: A straightforward substitution will show that this is true for Intensional Conservativity. Details left to the reader.
Intensional EXT says that $\mathbf{Q}_m\mathbf{XY} \Leftrightarrow \mathbf{Q}_{m'}\mathbf{XY}$ whenever $\mathbf{X}^m = \mathbf{X}^{m'}$, $\mathbf{Y}^m = \mathbf{Y}^{m'}$. Now if $\mathbf{X}^m = \mathbf{X}^{m'}$, $\mathbf{Y}^m = \mathbf{Y}^{m'}$ then trivially $|\mathbf{X}^m| = |\mathbf{X}^{m'}|$, $|(\mathbf{X}\wedge\mathbf{Y})^m| = |(\mathbf{X}\wedge\mathbf{Y})^{m'}|$. Since the only way in which a quantifier $\mathbf{Q}$ of the general form depends on the specific models m, m' is through its dependence on those cardinalities, $\mathbf{Q}_m\mathbf{XY} \Leftrightarrow \mathbf{Q}_{m'}\mathbf{XY}$ follows.
For Intensional Isomorphism Closure, suppose $h : D(m) \to D(m')$ is a bijection

and $\mathbf{X}^{m'} = h[\mathbf{X}^m]$, $\mathbf{Y}^{m'} = h[\mathbf{Y}^m]$. Then clearly $|\mathbf{X}^m| = |\mathbf{X}^{m'}|$. Similarly,

$$\begin{aligned} |(\mathbf{X} \wedge \mathbf{Y})^m| &= |\mathbf{X}^m \cap \mathbf{Y}^m| \\ &= |h[\mathbf{X}^m \cap \mathbf{Y}^m]| \\ &= |h[\mathbf{X}^m] \cap h[\mathbf{Y}^m]| \\ &= |\mathbf{X}^{m'} \cap \mathbf{Y}^{m'}| \\ &= |(\mathbf{X} \wedge \mathbf{Y})^{m'}| \end{aligned}$$

Thus, the arguments of a are invariant under replacing m by m' under these circumstances, which leads to $\mathbf{Q}_m\mathbf{XY} \Leftrightarrow \mathbf{Q}_{m'}\mathbf{XY}$.

3.3 Other readings of Many

As mentioned in the introduction, I am not the first to notice that any proper treatment of many should have at least an intensional component to it. Thus, through the years a number of readings that have such a component have been proposed. However, it has not yet been looked into how these readings fare with regards to Conservativity. In this section we will investigate some of them to find out just that.
To avoid confusion, we will rephrase these treatments in terms of the framework and notational conventions we have been using so far.

Fernando & Kamp

Fernando and Kamp's account (Fernando and Kamp 1996) states that "…the arguments of many …cannot be interpreted simply by their extensions" and uses a probability-based method for the intensional component. The idea is that a given number of $\mathbf{X}$'s that are $\mathbf{Y}$ qualifies as many if one would have expected there to be less. The quantifier is given by

$$\mathbf{Many}_m(\mathbf{X}, \mathbf{Y}) \Leftrightarrow \bigvee_{n \geq 1} (|(\mathbf{X} \wedge \mathbf{Y})^m| \geq n) \wedge n\text{-is-many}(\mathbf{X}, \mathbf{Y})$$

The probability-driven component n-is-many$(\mathbf{X}, \mathbf{Y})$ comes in a simple version and a more complex one. The simple version asserts that the probability of there being less than n $\mathbf{X}$'s that are $\mathbf{Y}$ is sufficiently high. It is of the form $P(\{m' : |(\mathbf{X} \wedge \mathbf{Y})^{m'}| < n\}) > c$, for a world-independent probability function P and constant c.
While it would be fairly easy to express this in our general form (left to the reader), it unfortunately is also symmetrical. Thus we are more interested in the more advanced reading.
In the advanced version, we do not merely use the probability of there being

less than n such objects, but conditionalize this probability against that of having exactly as many $\mathbf{X}$'s are there happen to be. This gives us the following n-is-many$(\mathbf{X}, \mathbf{Y})$:

$$P\left(\{m' : |(\mathbf{X} \wedge \mathbf{Y})^{m'}| < n\} \mid \{m' : |\mathbf{X}^{m'}| = |\mathbf{X}^m|\}\right) > c$$

Because of this actual world-dependent component, this reading does not have the general form. However, since it depends only on $\mathbf{X}, \mathbf{X} \wedge \mathbf{Y}, |\mathbf{X}^m|, |(\mathbf{X} \wedge \mathbf{Y})^m|$ and the independent c and P it still satisfies Intensional Conservativity, EXT and Isomorphism closure.

Cohen

The Relative Proportional reading introduced by Cohen (Cohen 2001) is based on the notion of *alternatives*. The alternatives of a property are other properties which it is appropriate to compare it to.
For instance, when considering the sentence "Many Scandinavians won a Nobel Prize in Literature" (see also Section 3.2.3), the alternatives to Scandinavian would be various (non-Scandinavian) nationalities. This sentence would be considered true under this reading if the proportion of Scandinavians who have won a Nobel Prize is (significantly) larger than the average proportion of people who have done so from other backgrounds.[7]
Formally, we take $\mathbf{many}_m(\mathbf{X}, \mathbf{Y})$ to be true iff the following holds:

$$\frac{|\mathbf{X}^m \cap \mathbf{Y}^m|}{|\mathbf{X}^m \cap \bigcup\{\mathbf{Z}^m | \mathbf{Z} \in A\}|} > \frac{|\bigcup\{\mathbf{Z}^m | \mathbf{Z} \in A\} \cap \mathbf{Y}^m|}{|\bigcup\{\mathbf{Z}^m | \mathbf{Z} \in A\}|}$$

Here A is a set of pairs of alternatives for $\mathbf{X}$ and Y, given by

$$A = \{\mathbf{X}' \wedge \mathbf{Y}' | \mathbf{X}' \in \text{ALT}(\mathbf{X}), \mathbf{Y}' \in \text{ALT}(\mathbf{Y})\}$$

where ALT$(\mathbf{X})$ gives a set of properties considered to be alternatives to $\mathbf{X}$, including $\mathbf{X}$ itself. It is important to keep in mind that such alternatives are necessarily disjoint everywhere.
The above looks a bit complex because it accounts for the possibility that the alternatives are not exhaustive (that is, that there exist objects that don't fall under any alternative) either for $\mathbf{X}$ or for $\mathbf{Y}$. If they are exhaustive for both it simplifies considerably, leaving

$$\frac{|\mathbf{X}^m \cap \mathbf{Y}^m|}{|\mathbf{X}m|} > \frac{|\mathbf{Y}^m|}{|D(m)|}$$

It is not hard to see that this reading is Symmetric. Cohen admits this much, but does not consider it a significant problem. He also notes in his abstract that

[7] Though note that strictly speaking, 'other' here would include Scandinavian itself.

this reading is not Conservative (in the regular sense), which similarly he does not necessarily consider to be an important issue. It is not a big surprise then that Intensional Conservativity does not necessarily hold either.
To test this, let A remain as before and let A' be the version of A obtained when $\mathbf{X}$ is replaced by $\mathbf{X} \wedge \mathbf{Y}$. This raises the question what kind of alternatives are in $\mathrm{ALT}(\mathbf{X} \wedge \mathbf{Y})$. A straightforward choice for this would be to let $\mathrm{ALT}(\mathbf{X} \wedge \mathbf{Y}) = A$.[8] Hence we get

$$A' = \{\mathbf{X}' \wedge \mathbf{Z}' | \mathbf{X}' \in \mathrm{ALT}(Y), \mathbf{Z}' \in A\}.$$

But because of the nature of $\mathbf{Z}'$, it always either implies or contradicts $\mathbf{X}'$. Therefore what we in fact end up with is $A' = A$. We now obtain

$$\begin{aligned}
\mathbf{many}(\mathbf{X}, \mathbf{X} \wedge \mathbf{Y}) \quad &\Leftrightarrow \quad \frac{|\mathbf{X}^m \cap (\mathbf{X} \wedge \mathbf{Y})^m|}{|\mathbf{X}^m \cap \bigcup\{\mathbf{Z}^m | \mathbf{Z} \in A\}|} > \frac{|\bigcup\{\mathbf{Z}^m | \mathbf{Z} \in A\} \cap (\mathbf{X} \wedge \mathbf{Y})^m|}{|\bigcup\{\mathbf{Z}^m | \mathbf{Z} \in A\}|} \\
&\Leftrightarrow \quad \frac{|(\mathbf{X} \wedge \mathbf{Y})^m|}{|\mathbf{X}^m \cap \bigcup\{\mathbf{Z}^m | \mathbf{Z} \in A\}|} > \frac{|(\mathbf{X} \wedge \mathbf{Y})^m|}{|\bigcup\{\mathbf{Z}^m | \mathbf{Z} \in A\}|} \\
&\Leftrightarrow \quad |\mathbf{X}^m \cap \bigcup\{\mathbf{Z}^m | \mathbf{Z} \in A\}| < |\bigcup\{\mathbf{Z}^m | \mathbf{Z} \in A\}|
\end{aligned}$$

The latter is a tautology, so Intensional Conservativity does not hold.
With the reading depending so much on the extensions of alternatives, we shouldn't expect Intensional EXT to hold either, and it doesn't. Pick $\mathbf{X}, \mathbf{Y}, m, m'$ such that $\mathbf{many}_m(\mathbf{X}, \mathbf{Y})$ is true, $\mathbf{X}^{m'} = \mathbf{X}^m$, $\mathbf{Y}^{m'} = \mathbf{Y}^m$ and every alternative to $\mathbf{X}$ or $\mathbf{Y}$ (except $\mathbf{X}$ and $\mathbf{Y}$ themselves) has empty extension. Then $\mathbf{many}_{m'}(\mathbf{X}, \mathbf{Y})$ reduces to

$$\frac{|\mathbf{X}^m \cap \mathbf{Y}|}{|\mathbf{X}^m \cap \mathbf{Y}|} > \frac{|\mathbf{X}^m \cap \mathbf{Y}|}{|\mathbf{X}^m \cap \mathbf{Y}|},$$

a contradiction.

It is worth pointing out that an important motivation behind the Relative Proportional reading was to provide an alternative to what Cohen calls the Reverse Interpretation view. Thus, as we have seen in Section 3.2.3, the good news is that even if the Relative Proportional reading is not as successful as one may hope, the intensional approach has allowed us to provide an alternative of our own which does satisfy Intensional Conservativity (as well as Intensional EXT and Intensional Isomorphism closure).

Tanaka

Similar to Cohen, Tanaka's account (Tanaka 2003) is based on sets of alternatives, based on taxonomic knowledge. It distinguishes between taking alternatives to

[8] Admittedly this decision is a crucial step, and making a different choice here might potentially lead to a different outcome. Still, the choice seems appropriate enough and no alternative that actually gives a different outcome comes to mind.

the subject or the predicate, and between comparing alternatives of the same level (the Sister-alt reading) or a higher level (the Mother-alt reading).
For instance, in the sentence "Many Scandinavians have won the Nobel Prize in Literature", which Tanaka also discusses, the "sisters" of Scandinavian would be various other nationalities, whereas the "Mother" property would include people of any nationality.
This leads to four possible readings[9] of "Many X's are Y", which can be paraphrased as follows:

	$S - ALT$	$M - ALT$
Subject	$\frac{\|\mathbf{Y}^m \cap \mathbf{X}^m\|}{\|\mathbf{Y}^m\|} > \frac{\|\mathbf{Y}^m \cap \text{sister}(\mathbf{X})^m\|}{\|\mathbf{Y}^m\|}$	$\frac{\|\mathbf{Y}^m \cap \mathbf{X}^m\|}{\|\mathbf{Y}^m\|} > \frac{\|\mathbf{X}^m\|}{\|(\text{mother}(\mathbf{Y}))^m\|}$
Predicate	$\frac{\|\mathbf{X}^m \cap \mathbf{Y}^m\|}{\|\mathbf{X}^m\|} > \frac{\|\mathbf{X}^m \cap \text{sister}(\mathbf{Y})^m\|}{\|\mathbf{X}^m\|}$	$\frac{\|\mathbf{X}^m \cap \mathbf{Y}^m\|}{\|\mathbf{X}^m\|} > \frac{\|\mathbf{Y}^m\|}{\|(\text{mother}(\mathbf{X}))^m\|}$

In the relative M-ALT Subject reading, the relative amount of $\mathbf{Y}$s that are $\mathbf{X}$ is compared to the proportion of $\mathbf{X}$'s among the 'mother' of $\mathbf{Y}$. In the earlier example, this would mean comparing the proportion of Scandinavians who have won the Nobel Prize in Literature to the proportion of Scandinavians among all humans.
The M-ALT readings are not (Intensionally) Conservative: it is easy enough to see that both of them turn into a tautology if $\mathbf{Y}$ is replaced by $\mathbf{X} \wedge \mathbf{Y}$.
In the relative S-ALT Subject reading, the relative amount of $\mathbf{Y}$s that are $\mathbf{X}$ is compared to the same value for some sister of $\mathbf{X}$. It is admittedly not entirely clear to me if this means comparing to a single sister picked arbitrarily, comparing to some constructed 'arbitrary' sister, taking an average among all sisters or something else. Still, it seems unlikely that Intensional Conservativity can be attained.
Since sisters are disjoint, we get $|\text{sister}(\mathbf{X})^m \cap (\mathbf{X} \wedge \mathbf{Y})^m| = |\emptyset| = 0$, and similarly $|\mathbf{X}^m \cap \text{sister}(\mathbf{X} \wedge \mathbf{Y})^m| = |\emptyset| = 0$. A more charitable interpretation based on some constructed 'arbitrary' sister which may overlap the original sister would not help here either: only the part that does overlap the original would be left, so both readings would still produce a tautology.
Another possible interpretation could be to take an average over all sisters, writing the Subject-focused reading as

$$\frac{|\mathbf{Y}^m \cap \mathbf{X}^m|}{|\mathbf{Y}^m|} > \sum_{\mathbf{Z} \in \text{sisters}(\mathbf{X})} \frac{|\mathbf{Y}^m \cap \mathbf{Z}^m|}{|\mathbf{Y}^m|}$$

But even if we do this, replacing $\mathbf{Y}$ with $\mathbf{X} \wedge \mathbf{Y}$ will make the reading either trivially false (if $\mathbf{X}$ itself is counted among the sisters) or trivially true (if it is not).
To make matters particularly odd, Tanaka makes it a point to propose a revised

[9]In addition to two absolute readings which we are not interested in here.

notion of Conservativity, wherein focal mapping determines which element is conservative. This could mean that for some or all of the readings above, he would have us replace not $\mathbf{Y}$ but $\mathbf{X}$ by $\mathbf{X} \wedge \mathbf{Y}$ to test for Conservativity. But the fact of the matter is that this changes nothing. Replacing $\mathbf{X}$ by $\mathbf{X} \wedge \mathbf{Y}$ above turns all four readings into tautologies in essentially the same ways. As it stands I fail to see how his readings could satisfy the notion he introduces.
As for Intensional EXT, it fails for much the same reason it fails for Cohen's reading. The proof for this is left as an exercise for the reader.

Lappin

Lappin provides the only thoroughly intensional treatment I am aware of (Lappin 2000), and it might hold up well. It works by constructing a set S of normative possible situations, then comparing the amount of $\mathbf{X}$'s that are $\mathbf{Y}$ in the actual situation sa with the amounts in the normative ones.[10] Thus it is broadly defined as follows:

$$\mathbf{many}_{sa}(\mathbf{X},\mathbf{Y}) \Leftrightarrow S \neq \emptyset, \text{ and for every } sn \in S, \quad |\mathbf{X}^{sa} \cap \mathbf{Y}^{sa}| \geq |\mathbf{X}^{sn} \cap \mathbf{Y}^{sn}|$$

This account looks good and simple, but is held back by a highly underdefined S. One of the choices for S Lappin discusses is based on historical averages; another aims to be similar to the Fernando & Kamp approach. Some of his less useful suggestions involve using the following, where C is "a comparison set determined in sa":

$$\begin{aligned} S =& \{sn | sn = sa \ \& \ |\mathbf{X}^{sa} \cap \mathbf{Y}^{sa}| \geq |\mathbf{X}^{sn} \cap C|\} \\ S =& \{sn | sn = sa \ \& \ |\mathbf{X}^{sa} \cap \mathbf{Y}^{sa}| \geq |\mathbf{Y}^{sn} \cap C|\} \\ S =& \{sn | sn = sa \ \& \ |\mathbf{X}^{sa} \cap \mathbf{Y}^{sa}| \geq |C|\} \end{aligned}$$

The first conjunct in each of these ensures that only sa is considered for S. Since C is also determined using only sa, the readings generated by these choices for S have $\mathbf{Q}_{sa}\mathbf{XY}$ depend only on $\mathbf{X}^{sa}, \mathbf{Y}^{sa}$ and sa. Thus, these readings are in fact non-intensional ones, and therefore will not be able to overcome Westerståhl's problematic example as discussed in the introduction. Any way to get around this would involve taking the intensions of $\mathbf{X}$ and $\mathbf{Y}$ into account when choosing C.
The examples above show that some extra conditions on S are needed to separate useful readings from less useful ones. To find them, we look to our general form. In line with this, we can make things easier for ourselves by rephrasing the broad definition of many as

$$\mathbf{many}_{sa}(\mathbf{X},\mathbf{Y}) \Leftrightarrow \left(S \neq \emptyset \ \& \ |\mathbf{X}^{sa} \cap \mathbf{Y}^{sa}| \geq \max_{sn \in S} |\mathbf{X}^{sn} \cap \mathbf{Y}^{sn}| \right)$$

[10]Lappin uses "situation" where we would use "world" or "model".

This comes close to matching our general form, provided the right-hand side does not require too much. Specifically, we get the restriction that one must be able to determine S using only $\mathbf{X}$, $\mathbf{X} \wedge \mathbf{Y}$ and W.

Solt

Like Lappin, Solt provides a broad account (Solt 2009) which can cover a lot of possible readings of many by varying a somewhat underspecified parameter. In this case, the readings are built based on a 'neutral range' N_S of amounts that are not considered either many or few. As Solt puts it, "the full range of readings available to *many* and *few* can be derived via manipulation of two elements: the structure of the scale (whether or not an upper bound is assumed) and the choice of the neutral range on that scale". (Solt 2009, p177)
The structure of the scale corresponds to the difference between cardinal and proportional readings. In both cases, the general reading ultimately amounts to

$$|(\mathbf{X} \wedge \mathbf{Y})^m| \geq \sup N_S$$

When it comes to determining N_S, Solt finds that there is sometimes merit to involving possible worlds as Fernando & Kamp and Lappin do, but argues that this is often inappropriate. Instead, she favors constructing N_S as a range around an (implicit) comparison point p_c. A general recipe to determine p_c (in the absence of cues like 'compared to' and 'for a') is not provided.
Still, the general reading above easily fits our general form from Section 3.2.6, allowing us to say that when a possible world-based approach is taken, Intensional Conservativity can be guaranteed simply by demanding N_S depend only on $\mathbf{X}$, $\mathbf{X} \wedge \mathbf{Y}$ and W.

3.4 Appendix A: Reductions

3.4.1 Lifting Theorem

3.4.1. Definition. A *non-intensional quantifier* Q is a function which when given a domain M and two sets $U, V \subseteq M$ gives an evaluation of true or false. We will write $Q_M UV$ to denote that this evaluation is true.
For a non-intensional quantifier Q, define its *intensional lift* $\mathbf{Q}^*$ as follows:

$$\mathbf{Q}^*_m \mathbf{XY} \Leftrightarrow Q_{D(m)} \mathbf{X}^m \mathbf{Y}^m$$

Also, for any set U in domain $D(m)$, the lift $l_m(U)$ is the set of properties $\mathbf{X}$ for which $\mathbf{X}^m = U$.

3.4.2. Theorem. *Where Q is a non-intensional quantifier and $\mathbf{Q}^*$ is its lift:*

- $\boldsymbol{Q}^*$ *satisfies Intensional Conservativity if and only if* Q *is Conservative*
- $\boldsymbol{Q}^*$ *satisfies Intensional EXT if and only if* Q *satisfies EXT*, where EXT* is like EXT but applies for any* M, M' *such that* $A, B \subseteq M$, $A, B \subseteq M'$
- $\boldsymbol{Q}^*$ *satisfies Intensional Isomorphism closure if and only if* Q *satisfies Isomorphism closure*

PROOF: Conservativity is the easiest. First assume $\mathbf{Q}^*$ satisfies Intensional Conservativity. For a given set M, let m be a model with $D(m) = M$. Then

$$\begin{array}{lll} Q_M UV & \Leftrightarrow \exists \mathbf{X} \in l_m(U), \mathbf{Y} \in l_m(V) : \mathbf{Q}^*_m \mathbf{XY} & \textit{by construction} \\ & \Leftrightarrow \exists \mathbf{X} \in l_m(U), \mathbf{Y} \in l_m(V) : \mathbf{Q}^*_m(\mathbf{X})(\mathbf{X} \wedge \mathbf{Y}) & \textit{by Intensional Conservativity} \\ & \Leftrightarrow \exists \mathbf{X} \in l_m(U), \mathbf{Z} \in l_m(U \cap V) : \mathbf{Q}^*_m \mathbf{XZ} & \textit{see below} \\ & \Leftrightarrow Q_M U(U \cap V) & \textit{by definition} \end{array}$$

For the third step , note that

$$(\mathbf{X} \wedge \mathbf{Y})^m = \mathbf{X}^m \cap \mathbf{Y}^m = U \cap V$$

Therefore $(\mathbf{X} \wedge \mathbf{Y}) \in l_m(U \cap V)$.

Next, assume that Q is (regularly) Conservative, m is some model with $D(m) = M$ and $\mathbf{X}$ and $\mathbf{Y}$ are properties. Then

$$\begin{array}{lll} \mathbf{Q}^*_m \mathbf{XY} & \Leftrightarrow Q_M \mathbf{X}^m \mathbf{Y}^m & \textit{by definition} \\ & \Leftrightarrow Q_M \mathbf{X}^m (\mathbf{X}^m \cap \mathbf{Y}^m) & \textit{by Conservativity} \\ & \Leftrightarrow Q_M \mathbf{X}^m (\mathbf{X} \wedge \mathbf{Y})^m & \textit{by definition} \\ & \Leftrightarrow \mathbf{Q}^*_m \mathbf{X}(\mathbf{X} \wedge \mathbf{Y}) & \textit{by definition} \end{array}$$

For EXT*, let $U, V \subseteq M, M'$, $D(m) = M$, $D(m') = M'$ and let $\mathbf{X}, \mathbf{Y}$ be such that $\mathbf{X}^m = U = \mathbf{X}^{m'}$, $\mathbf{Y}^m = V = \mathbf{Y}^{m'}$.
First assume $\mathbf{Q}^*$ satisfies Intensional EXT. Then

$$\begin{array}{lll} Q_M UV & \Leftrightarrow \mathbf{Q}^*_m \mathbf{XY} & \textit{by definition} \\ & \Leftrightarrow \mathbf{Q}^*_{m'} \mathbf{XY} & \textit{by Intensional EXT} \\ & \Leftrightarrow Q_{M'} UV & \textit{by definition} \end{array}$$

For the other direction, assume Q satisfies EXT*. Then

$$\begin{array}{lll} \mathbf{Q}^*_m \mathbf{XY} & \Leftrightarrow Q_M \mathbf{X}^m \mathbf{Y}^m & \textit{by definition} \\ & \Leftrightarrow Q_{M'} \mathbf{X}^{m'} \mathbf{Y}^{m'} & \textit{by EXT}^* \\ & \Leftrightarrow \mathbf{Q}^*_{m'} \mathbf{XY} & \textit{by definition} \end{array}$$

For Isomorphism closure, let f be a bijection from $D(m)$ to $D(m')$ and let $\mathbf{X}^{m'} = f[\mathbf{X}^m], \mathbf{Y}^{m'} = f[\mathbf{Y}^m]$.
First assume that Q satisfies Isomorphism closure. This yields

$$\begin{aligned} \mathbf{Q}^*_m\mathbf{XY} &\Leftrightarrow Q_{D(m)}\mathbf{X}^m\mathbf{Y}^m && \textit{by definition} \\ &\Leftrightarrow Q_{D(m')}f[\mathbf{X}^m]f[\mathbf{Y}^m] && \textit{by Isomorphism Closure} \\ &\Leftrightarrow Q_{D(m')}\mathbf{X}^{m'}\mathbf{Y}^{m'} && \textit{by condition} \\ &\Leftrightarrow \mathbf{Q}^*_{m'}AB && \textit{by definition} \end{aligned}$$

The other direction is almost trivial: where $U, V \subseteq D$, pick a structure with $D(m) = M$, $\mathbf{X}^m = U, \mathbf{Y}^m = V$ and assume $\mathbf{Q}^*$ satisfies Intensional Isomorphism closure to obtain

$$\begin{aligned} Q_{D(m)}UV &\Leftrightarrow \mathbf{Q}^*_m\mathbf{XY} && \textit{by definition} \\ &\Leftrightarrow \mathbf{Q}^*_{m'}\mathbf{XY} && \textit{by Intensional ISOM} \\ &\Leftrightarrow Q_{D(m')}UV && \textit{by definition} \end{aligned}$$

3.4.2 Extensional Intensional Quantifiers

It is a matter of some interest to see under which conditions a given intensional quantifier can be interpreted as a lift of a non-intensional one. As one might expect, the answer is that this is so iff the truth value in a given model depends only on that model and the local extensions there. The following two propositions demonstrate this.

3.4.3. PROPOSITION. *If an intensional quantifier* $\mathbf{Q}$ *is such that* $\mathbf{Q}_m\boldsymbol{XY}$ *is a function of* $\boldsymbol{X}^m$, $\boldsymbol{Y}^m$ *and* $D(m)$, *then there is a non-intensional quantifier* Q^2 *such that* $\mathbf{Q}_m\boldsymbol{XY} \Leftrightarrow (Q^2)^*_m\boldsymbol{XY}$.

PROOF: For the proof, define

$$Q^2_M UV \Leftrightarrow \forall m' \textit{ with domain } M : \forall \mathbf{X} \in l_{m'}(U), \mathbf{Y} \in l_{m'}(V) : \mathbf{Q}_{m'}\mathbf{XY}$$

This gives

$$\begin{aligned} (Q^2)^*_m\mathbf{XY} &\Leftrightarrow Q^2_M\mathbf{X}^m\mathbf{Y}^m \\ &\Leftrightarrow \forall m' \textit{ with domain } M \\ &\qquad \forall \mathbf{X}' \in l_{m'}(\mathbf{X}^m), \mathbf{Y}' \in l_{m'}(\mathbf{Y}^m) : \mathbf{Q}_{m'}\mathbf{X}'\mathbf{Y}' \\ &\Leftrightarrow \forall \mathbf{X}' \in l_m(\mathbf{X}^m), \mathbf{Y}' \in l_m(\mathbf{Y}^m) : \mathbf{Q}_m\mathbf{X}'\mathbf{Y}' \\ &\Leftrightarrow \mathbf{Q}_m\mathbf{XY} \end{aligned}$$

(In the most important step, we may eliminate "$\forall m'$ *with domain* M" because $\mathbf{Q}'_m\mathbf{X}'\mathbf{Y}'$ depends only on the domain and the extensions there and the latter

have already been fixed by quantifying over $l_{m'}(\mathbf{X}^m)$, $l_{m'}(\mathbf{Y}^m)$.
Similarly, the next universal quantification may be eliminated because by definition all $\mathbf{X}' \in l_m(\mathbf{X}^m)$ have the same extension in m as $\mathbf{X}$ (and the same for $\mathbf{Y}$).)

This covers one direction. The other direction is covered by the proposition below, which is trivial enough to require no further proof.

3.4.4. PROPOSITION. *For any lift* $\boldsymbol{Q}^*$ *of a non-intensional quantifier* Q*,* $\boldsymbol{Q}^*_m \boldsymbol{X}\boldsymbol{Y}$ *is a function of* $\boldsymbol{X}^m$, $\boldsymbol{Y}^m$ *and* $D(m)$*.*

As mentioned before, good readings of 'many' (certainly any reading that avoids the problem mentioned in the introduction while still being Intensionally Conservative) will not be interpretable as a lift of this kind. Such readings will necessarily depend on information beyond what can be drawn from the local extensions and domain, and hence will not be interpretable as a function of only these.

3.5 Appendix B: Characterizing the General Form

We have seen in Section 3.2.6 that there is a convenient general form for intensional quantifiers which guarantees Intensional Conservativity, Intensional EXT and Intensional ISOM while also being broad enough to cover various common natural language determiners. For the sake of convenience the definition is briefly restated below.

3.5.1. Definition. An intensional quantifier $\mathbf{Q}$ is of the *general form* if it can be written as

$$\mathbf{Q}_m\mathbf{XY} \Leftrightarrow a(|\mathbf{X}^m|, |(\mathbf{X}\wedge\mathbf{Y})^m|) \succ st(\mathbf{X}, \mathbf{X}\wedge\mathbf{Y}).$$

Having already proven one direction, it would be highly desirable if we could also show that Intensional Conservativity, Intensional EXT and Intensional ISOM taken together imply that an intensional quantifier is of the general form. Unfortunately, the standard way of defining ISOM is not appropriate for this. Instead, we will be using a differently phrased and slightly stronger property called Numericality or NUM. It will turn out that this NUM is true of every quantifier of the general form, and that NUM combined with Intensional Conservativity comes very close to guaranteeing that a quantifier is of that form (a slight extra assumption is needed).

3.5.2. Definition. An intensional quantifier $\mathbf{Q}$ satisfies *Numericality*, abbreviated *NUM*, iff the following is true for it.

$$\text{If } |\mathbf{X}^m| = |\mathbf{X}^{m'}|, |(\mathbf{X}\wedge\mathbf{Y})^m| = |(\mathbf{X}\wedge\mathbf{Y})^{m'}|, \text{ then } \mathbf{Q}_m\mathbf{XY} \Leftrightarrow \mathbf{Q}_{m'}\mathbf{XY}$$

NUM is slightly stronger than both Intensional EXT and Intensional ISOM. Examples of quantifiers with NUM include *All* and *Some* as well as *At least n*, *Exactly n* and *At most n*. It is also an appropriate requirement for *Many* and *Few*.

It may be an unusual move to consider the conjunction $\mathbf{X}\wedge\mathbf{Y}$ in the above definition. However, this move is quite necessary for things to work properly. To see this, define NUM* as the following property:

$$\text{If } |\mathbf{X}^m| = |\mathbf{X}^{m'}|, |\mathbf{Y}^m| = |\mathbf{Y}^{m'}|, \text{ then } \mathbf{Q}_m\mathbf{XY} \Leftrightarrow \mathbf{Q}_{m'}\mathbf{XY}$$

It is easy enough to see that NUM* and Intensional Conservativity together imply NUM. (Use $\mathbf{X}\wedge\mathbf{Y}$ for $\mathbf{Y}$ as a special case, then switch the consequent back.) However, NUM does not similarly imply NUM*, because even with Conservativity one cannot from $|\mathbf{X}^m| = |\mathbf{X}^{m'}|, |(\mathbf{X}\wedge\mathbf{Y})^m|$ infer that $|\mathbf{Y}^m| = |\mathbf{Y}^{m'}|$.

Thus, NUM* is stronger than NUM, the difference being that it also pays attention to exactly that part of $\mathbf{Y}$ which Conservativity tells us should be irrelevant. This extra strength is not harmless; it prevents the normal form from implying

NUM*. For this reason, NUM is the superior choice for our purposes.

Note that the $\mathbf{X} \wedge \mathbf{Y}$ term does not mean NUM guarantees Intensional Conservativity. For example, consider a quantifier expressing "There are more X's than one would expect there to be Y's", which could be phrased as

$$\mathbf{Q}_m\mathbf{XY} \Leftrightarrow |\mathbf{X}^m| > \frac{1}{|W|} \sum_{m' \in W} |\mathbf{Y}^{m'}|$$

This satisfies Intensional NUM, but clearly not Intensional Conservativity.

With these notes out of the way, it is time to move on to the next step. We already know that the general form implies Intensional Conservativity, but we need to show that it also implies NUM. This much is fairly straightforward.

3.5.3. THEOREM. *Every intensional quantifier of the general form satisfies NUM.*

PROOF: Suppose $\mathbf{Q}$ is of the general form. It follows that $\mathbf{Q}_m\mathbf{XY}$ iff $a(|\mathbf{X}^m|, |(\mathbf{X} \wedge \mathbf{Y})^m|) \succ st(\mathbf{X}, (\mathbf{X} \wedge \mathbf{Y}))$.
Now suppose $|\mathbf{X}^m| = |\mathbf{X}^{m'}|, |(\mathbf{X} \wedge \mathbf{Y})^m| = |(\mathbf{X} \wedge \mathbf{Y})^{m'}|$. Then obviously $a(|\mathbf{X}^m|, |(\mathbf{X} \wedge \mathbf{Y})^m|) = a(|\mathbf{X}^{m'}|, |(\mathbf{X} \wedge \mathbf{Y})^{m'}|)$. Since $st(\mathbf{X}, \mathbf{Y})$ stays the same, it immediately follows that $\mathbf{Q}_m\mathbf{XY} \Leftrightarrow \mathbf{Q}_{m'}\mathbf{XY}$.

While the proof that the general form implies NUM is straightforward, the proof that NUM and Intensional Conservativity jointly imply the general form is far from it. Stretching the definition of the general form far, the "standard value" *st* is defined in a way that includes all information about when $\mathbf{Q}_m\mathbf{XY}$ is the case. Because of NUM, the relevant part of this information (whether it is the case in the actual world) can then be extracted using only $|\mathbf{X}^m|$ and $|(\mathbf{X} \wedge \mathbf{Y})^m|$. In the last steps, Conservativity compensates for the unusual form chosen for NUM.
The proof below requires the extra assumption that there is a limit to how large the domains of models can be. However, since this limit can be not just infinity but a particularly large type of infinity this should not be a big issue, especially when analyzing natural language.

3.5.4. THEOREM. *If there is an infinite cardinal α such that $|D(m)| \leq \alpha$ for all models m, then if an intensional quantifier $\mathbf{Q}$ satisfies Intensional Conservativity and NUM, it is of the general form.*

PROOF: Define *st* as a function which takes pairs of properties to sets of pairs of elements of $\alpha + 1$ such that $(u, v) \in st(\mathbf{X}, \mathbf{Y})$ if and only if there is some $w \in W$ such that $|\mathbf{X}^w| = u, |(\mathbf{X} \wedge \mathbf{Y})^w| = v$ and $\mathbf{Q}_w\mathbf{XY}$.
Note that $\mathbf{Q}_m\mathbf{XY}$ iff $(|\mathbf{X}^m|, |(\mathbf{X} \wedge \mathbf{Y})^m|) \in st(\mathbf{X}, \mathbf{Y})$.
Now let $st(\mathbf{X}, \mathbf{Y}) = f(\mathbf{X}, \mathbf{Y})$, $a(x, y) = (x, y)$ and let $\succ$ be the inclusion relation

$\in$.
In this way, $a(|\mathbf{X}^m|, |(\mathbf{X} \wedge \mathbf{Y})^m|) \succ st(\mathbf{X}, \mathbf{Y})$ if and only if

$$(|\mathbf{X}^m|, |(\mathbf{X} \wedge \mathbf{Y})^m|) \in st(\mathbf{X}, \mathbf{Y})$$

which is if and only if there is some $m' \in W$ such that $|\mathbf{X}^{m'}| = |\mathbf{X}^m|, |(\mathbf{X} \wedge \mathbf{Y})^{m'}| = |(\mathbf{X} \wedge \mathbf{Y})^m|$ and $\mathbf{Q}_{m'}\mathbf{XY}$. Because of NUM, the latter is just in case $\mathbf{Q}_m\mathbf{XY}$.
As a specific case, $\mathbf{Q}_m\mathbf{X}(\mathbf{X} \wedge \mathbf{Y})$ if and only if

$$a(|\mathbf{X}^m|, |(\mathbf{X} \wedge (\mathbf{X} \wedge \mathbf{Y}))^m|) \succ st(\mathbf{X}, \mathbf{X} \wedge \mathbf{Y})$$

By Intensional Conservativity of $\mathbf{Q}$ and since $|(\mathbf{X} \wedge (\mathbf{X} \wedge \mathbf{Y}))^m| = |(\mathbf{X} \wedge \mathbf{Y})^m|$, this means $\mathbf{Q}_m\mathbf{XY}$ iff $a(|\mathbf{X}^m|, |(\mathbf{X} \wedge \mathbf{Y})^m|) \succ st(\mathbf{X}, \mathbf{X} \wedge \mathbf{Y})$, completing the proof.

For those who feel the above proof stretches things to the point of cheating, a different version is given below. This version guarantees that $\succ$ is a relation on the real numbers, at the small extra cost of requiring that all domains are at most countably infinite.

3.5.5. Theorem. *If all models have a finite or countably infinite domain, then if an intensional quantifier* $\boldsymbol{Q}$ *satisfies Intensional Conservativity and NUM, it is of the general form with* $\succ$ *a relation on the real numbers.*

Proof: Define f as a function which takes pairs of properties to sets of pairs of elements of $\omega + 1$ such that $(u, v) \in f(\mathbf{X}, \mathbf{Y})$ if and only if there is some $w \in W$ such that $|\mathbf{X}^w| = u, |(\mathbf{X} \wedge \mathbf{Y})^w| = v$ and $\mathbf{Q}_w\mathbf{XY}$.
Note that $\mathbf{Q}_m\mathbf{XY}$ iff $(|\mathbf{X}^m|, |(\mathbf{X} \wedge \mathbf{Y})^m|) \in f(\mathbf{X}, \mathbf{Y})$.
Let g be a bijection from ω to $(\omega + 1)^2$. Let $h(n)$ be 1 iff $g(n) \in f(\mathbf{X}, \mathbf{Y})$, 0 otherwise. Let

$$st(\mathbf{X}, \mathbf{Y}) = \sum_{n=1}^{\infty} \frac{1}{2^n} h(n)$$

This amounts to saying that in binary notation, the n-th digit is a 1 if and only if $g(n) \in f(\mathbf{X}, \mathbf{Y})$.
Now let $a = g^{-1}$ and let $x \succ y$ iff $x = a(u, v)$ and the x-th digit in the binary notation of y is a 1.
In this way, $a(|\mathbf{X}^m|, |(\mathbf{X} \wedge \mathbf{Y})^m|) \succ st(\mathbf{X}, \mathbf{Y})$ if and only if

$$h(g^{-1}(|\mathbf{X}^m|, |(\mathbf{X} \wedge \mathbf{Y})^m|)) = 1$$

which is if and only if

$$(|\mathbf{X}^m|, |(\mathbf{X} \wedge \mathbf{Y})^m|) \in f(\mathbf{X}, \mathbf{Y})$$

The rest of the proof is as above.

3.5.1 NUM and Intensional ISOM

It is worth pointing out that NUM can be rephrased to be much like Intensional ISOM. This leads to a property that might be called Conservative ISOM, as defined below.

3.5.6. DEFINITION. An intensional quantifier $\mathbf{Q}$ satisfies Conservative ISOM if $\mathbf{Q}_m\mathbf{XY} \Leftrightarrow \mathbf{Q}_{m'}\mathbf{XY}$ holds whenever there is a bijection $f : A \to B$ with

$$A \subseteq D(m), B \subseteq D(m'), f[\mathbf{X}^m] = \mathbf{X}^{m'}, [(\mathbf{X} \wedge \mathbf{Y})^m] = (\mathbf{X} \wedge \mathbf{Y})^{m'}$$

As stated above, the two definitions are equivalent.

3.5.7. THEOREM. *Conservative ISOM is equivalent to NUM.*

PROOF: Suppose $\mathbf{Q}$ has Conservative ISOM. Suppose $\mathbf{X}, \mathbf{Y}, m, m'$ are such that

$$|\mathbf{X}^m| = |\mathbf{X}^{m'}|, |(\mathbf{X} \wedge \mathbf{Y})^m| = |(\mathbf{X} \wedge \mathbf{Y})^{m'}|$$

It follows that $|\mathbf{X}^m - \mathbf{Y}^m| = |\mathbf{X}^{m'} - \mathbf{Y}^{m'}|$. Therefore there are bijections $f_1 : (\mathbf{X} \wedge \mathbf{Y})^m \to (\mathbf{X} \wedge \mathbf{Y})^{m'}, f_2 : \mathbf{X}^m - \mathbf{Y}^m \to \mathbf{X}^{m'} - \mathbf{Y}^{m'}$.
We can combine f_1 and f_2 into a bijection $f : \mathbf{X}^m \to \mathbf{X}^{m'}$. Trivially $\mathbf{X}^m \subseteq D(m)$, $\mathbf{X}^{m'} \subseteq D(m')$. By construction f has $f[\mathbf{X}^m] = \mathbf{X}^{m'}, f[(\mathbf{X} \wedge \mathbf{Y})^m] = (\mathbf{X} \wedge \mathbf{Y})^{m'}$. Therefore by Conservative ISOM $\mathbf{Q}_m\mathbf{XY} \Leftrightarrow \mathbf{Q}_{m'}\mathbf{XY}$.
The other direction is left to the reader.

3.6 Appendix C: Logical Properties of Many - A Brief Glimpse

With intensionalization making *many* less problematic as a generalized quantifier, it becomes interesting to look at its logical behaviour. In order to get very far with this, one must first determine when $\boldsymbol{Many}_m\mathbf{XY}$ is true. It seems natural enough to require that any good reading consists of a comparison between an amount in the actual world and an intensionally determined standard; that is, that the reading is of the general form.
Furthermore, we will require that this comparison is through the relation $>$, rather than some arbitrary $\succ$. This is less limiting than it may seem, since cases like "a is significantly larger than st" can be handled by multiplying st with a factor close to 1. Beyond this, though, not much can really be assumed.
The strategy we will use in this section is to consider some constraints on the interpretation of many which seem plausible for at least some readings and see which logical rules they imply. After that we will also look at a number of things these constraints do *not* imply.
A concept we will use repeatedly in this section is that of a pair of properties being intensionally disjoint. This amounts to their being disjoint everywhere, and is defined as follows.

3.6.1. DEFINITION. The properties $\mathbf{X}$ and $\mathbf{Y}$ are *intensionally disjoint* (relative to W) if $\forall m \in W \mathbf{X}^m \cap \mathbf{Y}^m = \emptyset$.

Notation

For the sake of convenience, we will often write $a_m(\mathbf{X}, \mathbf{Y})$ to refer to $a(|\mathbf{X}^m|, |(\mathbf{X} \wedge \mathbf{Y})^m|)$ and $st(\mathbf{X}, \mathbf{Y})$ to refer to $st(\mathbf{X}, \mathbf{X} \wedge \mathbf{Y}, W)$.

3.6.1 Additivity

For actual and standard values based on simple counting, averages, probability functions and the like, a reasonable extra constraint is that of Additivity, where the values can be interpreted as the sum of their parts. This is the first constraint we look at, at it will imply several convenient theorems about disjunctions.
We look first at additivity in the second argument, as this is a property possessed by our earlier example reading. A quantifier is right-Additive if the actual and standard values for the (relative) amount of $\mathbf{X}$ which are $\mathbf{Y} \vee \mathbf{Z}$ are the sums of the actual and standard values for the (relative) amounts of $\mathbf{X}$ which are $\mathbf{Y}$ and $\mathbf{Z}$. Specifically:

3.6.2. DEFINITION. **(Right-Additivity)** An intensional quantifier $\mathbf{Q}$ of the form

$$\mathbf{Q}_m\mathbf{X}\mathbf{Y} \Leftrightarrow a(|\mathbf{X}^m|, |(\mathbf{X} \wedge \mathbf{Y})^m|) \succ st(\mathbf{X}, \mathbf{X} \wedge \mathbf{Y})$$

is *right-Additive* if for all $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ with $\mathbf{Y}, \mathbf{Z}$ intensionally disjoint,

- $a_m(\mathbf{X}, \mathbf{Y} \vee \mathbf{Z}) = a_m(\mathbf{X}, \mathbf{Y}) + a_m(\mathbf{X}, \mathbf{Z})$
- $st(\mathbf{X}, \mathbf{Y} \vee \mathbf{Z}) = st(\mathbf{X}, \mathbf{Y}) + st(\mathbf{X}, \mathbf{Z})$

Note that generally speaking (for natural language examples) the first condition is roughly the same as requiring that $a(x, y+z) = a(x, y) + a(x, z)$ for all numbers x, y, z.
This property holds for our earlier example reading, which was of the form

$$\frac{|(\mathbf{X} \wedge \mathbf{Y})^m|}{|\mathbf{X}^m|} > \frac{1}{|W|} \sum_{m \in W} \frac{|(\mathbf{X} \wedge \mathbf{Y})^m|}{|\mathbf{X}^m|}$$

3.6.3. THEOREM. *The above reading of Many satisfies Constraint 1a.*

Proof: This is left to the reader for the actual value. As for the standard value:

$$\begin{aligned} st(\mathbf{X}, \mathbf{Y} \vee \mathbf{Z}) &= \frac{1}{|W|} \sum_{m \in W} \frac{|\mathbf{X}^m \cap (\mathbf{Y} \vee \mathbf{Z})^m|}{|\mathbf{X}^m|} \\ &= \frac{1}{|W|} \sum_{m \in W} \left(\frac{|\mathbf{X}^m \cap \mathbf{Y}^m|}{|\mathbf{X}^m|} + \frac{|\mathbf{X}^m \cap \mathbf{Z}^m|}{|\mathbf{X}^m|} \right) \\ &= \left(\frac{1}{|W|} \sum_{m \in W} \frac{|\mathbf{X}^m \cap \mathbf{Y}^m|}{|\mathbf{X}^m|} \right) + \left(\frac{1}{|W|} \sum_{m \in W} \frac{|\mathbf{X}^m \cap \mathbf{Z}^m|}{|\mathbf{X}^m|} \right) \\ &= st(\mathbf{X}, \mathbf{Y}) + st(\mathbf{X}, \mathbf{Z}) \end{aligned}$$

Right-Additivity implies the following two convenient properties concerning the (intensional) disjunction.

3.6.4. Theorem. *If $\boldsymbol{Q}$ is right-Additive and $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are intensionally disjoint, then $\boldsymbol{Q}(\boldsymbol{X}, \boldsymbol{Y} \vee \boldsymbol{Z})$ implies $\boldsymbol{Q}(\boldsymbol{X}, \boldsymbol{Y})$ or $\boldsymbol{Q}(\boldsymbol{X}, \boldsymbol{Z})$.*

Proof:

$$\begin{aligned} \mathbf{Q}_m(\mathbf{X}, \mathbf{Y} \vee \mathbf{Z}) &\Leftrightarrow a_m(\mathbf{X}, \mathbf{Y} \vee \mathbf{Z}) > st(\mathbf{X}, \mathbf{Y} \vee \mathbf{Z}) \\ &\Leftrightarrow a_m(\mathbf{X}, \mathbf{Y}) + a_m(\mathbf{X}, \mathbf{Z}) > st(\mathbf{X}, \mathbf{Y}) + st(\mathbf{X}, \mathbf{Z}) \\ &\Rightarrow (a_m(\mathbf{X}, \mathbf{Y}) > st(\mathbf{X}, \mathbf{Y})) \;\; or \;\; (a_m(\mathbf{X}, \mathbf{Z}) > st(\mathbf{X}, \mathbf{Z})) \end{aligned}$$

3.6.5. Theorem. *If $\boldsymbol{Q}$ is right-Additive and $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are intensionally disjoint, then if $\boldsymbol{Q}(\boldsymbol{X}, \boldsymbol{Y})$ and $\boldsymbol{Q}(\boldsymbol{X}, \boldsymbol{Z})$ both hold, so does $\boldsymbol{Q}(\boldsymbol{X}, \boldsymbol{Y} \vee \boldsymbol{Z})$.*

Proof:

$$\begin{aligned} a_m(\mathbf{X}, \mathbf{Y} \vee \mathbf{Z}) &= a_m(\mathbf{X}, \mathbf{Y}) + a_m(\mathbf{X}, \mathbf{Z}) \\ &> st(\mathbf{X}, \mathbf{Y}) + st(\mathbf{X}, \mathbf{Z}) = st(\mathbf{X}, \mathbf{Y} \vee \mathbf{Z}) \end{aligned}$$

The natural other side of right-Additivity is left-Additivity, which of course is defined as follows.

3.6.6. Definition. **(Left-Additivity)** An intensional quantifier $\mathbf{Q}$ of the form

$$\mathbf{Q}_m\mathbf{X}\mathbf{Y} \Leftrightarrow a(|\mathbf{X}^m|, |(\mathbf{X} \wedge \mathbf{Y})^m|) \succ st(\mathbf{X}, \mathbf{X} \wedge \mathbf{Y})$$

is *left-Additive* if for all $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ with $\mathbf{X}, \mathbf{Y}$ disjoint,

- $a_m(\mathbf{X} \vee \mathbf{Y}, \mathbf{Z}) = a_m(\mathbf{X}, \mathbf{Z}) + a_m(\mathbf{Y}, \mathbf{Z})$
- $st(\mathbf{X} \vee \mathbf{Y}, \mathbf{Z}) = st(\mathbf{X}, \mathbf{Z}) + st(\mathbf{Y}, \mathbf{Z})$

As a constraint, left-Additivity is not as appropriate as right-Additivity. It certainly does not hold for proportional readings like our earlier example. It does hold for some absolute readings like the following.

$$\boldsymbol{Many}_m\mathbf{XY} \Leftrightarrow |(\mathbf{X}\wedge\mathbf{Y})^m| > \frac{1}{|W|}\sum_{w\in W}|(\mathbf{X}\wedge\mathbf{Y})^w|$$

However, it has undesirable consequences when combined with Reflexivity or anti-Reflexivity, which we shall see soon.
Naturally, the more immediate consequences of left-Additivity are the same as those of right-Additivity except that they are for the left argument.

3.6.7. THEOREM. *If $\boldsymbol{Q}$ is left-Additive and $\boldsymbol{X}$ and $\boldsymbol{Y}$ are intensionally disjoint, then*

- $\boldsymbol{Q}(\boldsymbol{X}\vee\boldsymbol{Y},\boldsymbol{Z})$ *implies* $\boldsymbol{Q}(\boldsymbol{X},\boldsymbol{Z})$ *or* $\boldsymbol{Q}(\boldsymbol{Y},\boldsymbol{Z})$
- *if* $\boldsymbol{Q}(\boldsymbol{X},\boldsymbol{Z})$ *and* $\boldsymbol{Q}(\boldsymbol{Y},\boldsymbol{Z})$ *both hold, so does* $\boldsymbol{Q}(\boldsymbol{X}\vee\boldsymbol{Y},\boldsymbol{Z})$

PROOF:Analogous to proofs for right-Additivity.

3.6.2 Reflexivity and anti-Reflexivity

Reflexivity and anti-Reflexivity are rather straightforward properties with consequences that are all similar and not too hard to prove, so we will dive right in and discuss their relevance and desirability afterwards.

3.6.8. DEFINITION. **(Reflexivity)** An intensional quantifier $\mathbf{Q}$ is *Reflexive* if for all properties $\mathbf{X}$ and for all m, $\mathbf{Q}_m\mathbf{XX}$.

3.6.9. DEFINITION. **(anti-Reflexivity)** An intensional quantifier $\mathbf{Q}$ is *anti-Reflexive* if for all properties $\mathbf{X}$ and for all m, *Not* $\mathbf{Q}_m\mathbf{XX}$.

3.6.10. THEOREM. ***(Contradiction)****If $\boldsymbol{Q}$ is right-Additive and anti-Reflexive, then $\boldsymbol{Q}_m(\boldsymbol{X},\boldsymbol{Y})$ cannot be true at the same time as $\boldsymbol{Q}_m(\boldsymbol{X},\neg\boldsymbol{Y})$.*
Analogously, if it is left-Additive and anti-Reflexive, $\boldsymbol{Q}_m(\boldsymbol{X},\boldsymbol{Y})$ cannot be true at the same time as $\boldsymbol{Q}_m(\neg\boldsymbol{X},\boldsymbol{Y})$.

PROOF: First off, note that $(\mathbf{Y}\vee\neg\mathbf{Y})^m$ is always $D(m)$ and therefore $\mathbf{X}\wedge(\mathbf{Y}\vee\neg\mathbf{Y})$ is the same property as $\mathbf{X}$. Also note that $\mathbf{X}\wedge\mathbf{Y}$ is intensionally disjoint from $\mathbf{X}\wedge\neg\mathbf{Y}$. By right-Additivity (with some intermediate steps left to the reader), it follows that

$$a_m(\mathbf{X},\mathbf{X}) = a_m(\mathbf{X},\mathbf{X}\wedge\mathbf{Y}) + a_m(\mathbf{X},\mathbf{X}\wedge\neg\mathbf{Y})$$

and the same for $st(\mathbf{X},\mathbf{X})$.
If $\mathbf{Q}_m(\mathbf{X},\mathbf{Y})$ and $\mathbf{Q}_m(\mathbf{X},\neg\mathbf{Y})$ are both true, it follows that $a_m(\mathbf{X},\mathbf{X}) > st(\mathbf{XX})$, which is to say $\mathbf{Q}_m\mathbf{XX}$, contradicting anti-Reflexivity. (The proof for left-Additivity is analogous.)

3.6.11. THEOREM. ***(Excluded Middle)*** *If* $\boldsymbol{Q}$ *is right-Additive and Reflexive, then for all* $\boldsymbol{X}$, $\boldsymbol{Y}$ *and* m*, at least one of* $\boldsymbol{Q}_m(\boldsymbol{X},\boldsymbol{Y})$ *and* $\boldsymbol{Q}_m(\boldsymbol{X},\neg\boldsymbol{Y})$ *is true. Analogously, if it is left-Additive and Reflexive, then for all* $\boldsymbol{X}$, $\boldsymbol{Y}$ *and* m*, at least one of* $\boldsymbol{Q}_m(\boldsymbol{X},\boldsymbol{Y})$ *and* $\boldsymbol{Q}_m(\neg\boldsymbol{X},\boldsymbol{Y})$ *is true.*

PROOF: As in the previous proof, we have

$$a_m(\mathbf{X},\mathbf{X}) = a_m(\mathbf{X},\mathbf{X}\wedge\mathbf{Y}) + a_m(\mathbf{X},\mathbf{X}\wedge\neg\mathbf{Y})$$

and the same for st. Thus if $\mathbf{Q}_m(\mathbf{X},\mathbf{Y})$ and $\mathbf{Q}_m(\mathbf{X},\neg\mathbf{Y})$ are both false, it follows that $a_m(\mathbf{X},\mathbf{X}) \leq st(\mathbf{XX})$, which is to say $\mathbf{Q}_m\mathbf{XX}$, contradicting anti-Reflexivity.

3.6.3 Desirability

Contradiction in the right argument is sometimes desirable and sometimes not. It is appropriate for readings that can be interpreted as “unexpectedly many”, “unusually many” and the like, but not so for readings that amount to “a substantial proportion (but not necessarily 50%)”. Thus, anti-Reflexivity is appropriate for the former but not the latter. The converse is true for Excluded Middle in the right argument: it is not at all appropriate for “unexpectedly many” but fits the latter interpretation nicely.
This makes it tempting to believe any reading will be either Reflexive or anti-Reflexive, but it is worth pointing out that adding a minimal standard will defeat both, as the quantifier “At least n” is neither Reflexive nor anti-Reflexive. On the other hand, a minimal standard will also go against right-Additivity, so a weakened conjecture like the following might still hold.

3.6.12. CONJECTURE. *Every right-Additive natural reading of many is either Reflexive or anti-Reflexive.*

In the left argument, both Contradiction and Excluded Middle are bizarre results for any interpretation of many. Given the conjecture above, this means we should reject left-Additivity as a constraint on many, and indeed should be very skeptical about any reading for it which is left-Additive.

3.6.4 Negative theorems

Disjunctive

Disjointness is necessary when it comes to the consequences of right-Additivity.

3.6.13. THEOREM. *If B and C are not disjoint, then* $\mathrm{Many}(A, B)$ *and* $\mathrm{Many}(A, C)$ *do not jointly imply* $\mathrm{Many}(A, B \vee C)$.

EXAMPLE: Using our example reading, let each world correspond to the outcomes of an experiment where a six-sided die is rolled 40 times, with W covering every possibility once.
Let A be the set of all rolls, B the set of all rolls where the result is 2 or 6 and C the set of all rolls where the result is 3 or 6.
Suppose in the actual world 1 was rolled 4 times and 5 and 6 were each rolled 18 times. Then *Many*(A, B) and *Many*(A, C), as in both cases the actual value $\frac{18}{40}$ is higher than the standard/expected value of $\frac{2}{6} = \frac{1}{3}$. However, it is not the case that *Many*$(A, B \vee C)$, since $\frac{18}{60}$ is lower than the expected value of $\frac{3}{6} = \frac{1}{2}$.

Right-Additivity does not guarantee the consequences of left-Additivity.

3.6.14. THEOREM. *From* $\mathrm{Many}(A \vee B, C)$ *it does not follow that* $\mathrm{Many}(A, C)$ *or* $\mathrm{Many}(B, C)$*, even if A, B disjoint.*

EXAMPLE: Use the same W as above. Let A be the rolls resulting in 3 or 6, B be the rolls resulting in 2, 4 or 5, and C the rolls resulting in 5 or 6.
Suppose in the actual world 6 was rolled 17 times, 5 once, 4 three times and 3 was rolled 19 times. Then *Many*$(A \vee B, C)$, as the actual value of $\frac{18}{40}$ is higher than the standard/expected value $\frac{2}{5} = \frac{16}{40}$.
However, neither *Many*(A, C) nor *Many*(B, C) hold, since $a(A, C) = \frac{17}{36} < \frac{1}{2} = st(A, C)$ and $a(B, C) = \frac{1}{4} < \frac{1}{3} = st(B, C)$.

3.6.15. THEOREM. *Even if A, B disjoint,* $\mathrm{Many}(A, C)$ *and* $\mathrm{Many}(B, C)$ *do not jointly imply* $\mathrm{Many}(A \vee B, C)$.

EXAMPLE: Use the same W as above, and the same A, B and C as above.
Suppose in the actual world the results were as follows:

- 1 was rolled 17 times
- 2 was rolled 6 times
- 3 was rolled 1 times
- 4 was rolled 7 times
- 5 was rolled 7 times
- 6 was rolled 2 times

Of the rolls that were 3 or 6, two-thirds were 6, giving $a(A, C) = \frac{2}{3} > \frac{1}{2} = st(A, C)$. We also get $a(B, C) = \frac{7}{20} > \frac{7}{21} = \frac{1}{3} = st(B, C)$, so *Many*$(B, C)$.
However, of the 23 rolls that were not 1, only nine were 5 or 6, a fraction which falls just slightly short of $\frac{2}{5}$ (left to the reader). Thus, $a(A \vee B, C) = \frac{9}{23} < \frac{2}{5} = st(A \vee B, C)$, so it is no true that *Many*$(A \vee B, C)$.

Conjunctive

3.6.16. THEOREM. Many(A, B) *and* Many(A, C) *do not jointly imply* Many$(A, B \wedge C)$.

EXAMPLE: Let W be as above, let A be the rolls resulting in 1, 2 or 3, let B be the rolls resulting in 1 or 2, and C the rolls resulting in 1 or 3.
Suppose in the actual world 2 and 3 were each rolled 20 times. Then *Many*(A, B) and *Many*(A, C), but not *Many*$(A, B \wedge C)$.

3.6.17. THEOREM. *From* Many$(A, B \wedge C)$ *it does not follow that either* Many(A, B) *or* Many(A, C).

EXAMPLE: Let W be as above, let A be the rolls resulting in 1, 2, 3 or 4, let B be the rolls resulting in 1 or 2, and C the rolls resulting in 1 or 3.
Suppose in the actual world 1 was rolled 16 times and 4 was rolled 24 times. Then *Many*$(A, B \wedge C)$ $(a(A, B \wedge C) = \frac{16}{40} > \frac{1}{4} = st(A, B \wedge C))$, but neither *Many*$(A, B)$ nor *Many*(A, C) $(a(A, B) = \frac{16}{40} < \frac{1}{2} = st(A, B))$.

3.6.18. THEOREM. Many(A, C) *and* Many(B, C) *do not jointly imply* Many$(A \wedge B, C)$.

EXAMPLE: Let W be as above. Let A be the rolls resulting in 1, 2 or 3, let B be the rolls resulting in 1, 2 or 4, and let C be the rolls resulting in 3 or 4.
Suppose in the actual world 1 and 2 were each rolled 8 times and 3 and 4 were each rolled 12 times. Then *Many*(A, C) and *Many*(B, C) but not *Many*$(A \wedge B, C)$.

3.6.19. THEOREM. *From* Many$(A \wedge B, C)$ *it does not follow that either* Many(A, C) *or* Many(B, C).

EXAMPLE: Let W be as above. Let A be the rolls resulting in 1, 2 or 3, let B be the rolls resulting in 1, 2 or 4, and let C be the rolls resulting in 1.
Suppose in the actual world 1 was rolled 4 times and 3 and 4 were each rolled 18 times. Then *Many*$(A \wedge B, C)$ but neither *Many*(A, C) nor *Many*(B, C) $(a(A, C) = \frac{4}{22} < \frac{1}{3} = st(A, C))$.

3.6.20. THEOREM. **(Non-Monotonicity)**Many(A, B) *and* Many(B, C) *do not jointly imply* Many(A, C)

EXAMPLE: Let W be as above. Let A be the rolls resulting in 1, 2 or 3, let B be the rolls resulting in 2, 3 or 4, and let C be the rolls resulting in 3.
Suppose in the actual world the results were as follows:

- 1 was rolled 10 times
- 2 was rolled 13 times

- 3 was rolled 11 times
- 4 was rolled 6 times

Then *Many*(A, B) and *Many*(B, C) but not *Many*(A, C).

Negation

Right-Additivity does not guarantee Excluded Middle. (Related theorems will go analogously.)

3.6.21. THEOREM. **(No Excluded Middle)** *It is possible that neither* Many(A, B) *nor* Many$(A, \neg B)$ *hold.*

EXAMPLE: Let W be as above. Let A be the rolls resulting in 3 or 6 and B the rolls resulting in 5 or 6.
Suppose that in the actual world both 3 and 6 are rolled 20 times. Then *Many*(A, B) is not true, but neither is *Many*$(A, \neg B)$.
(This specific reading requires all examples need to be this close, but not all do. Increasing the standard a bit can make for more examples without violating the constraints.)

Chapter 4

A Syllogistic for Subsective Adjectives

4.1 Introduction

Subsective adjectives are the category of adjectives such that the extension of a noun modified by such an adjective is a part of the extension of the noun proper. To put it another way: the fact that *small* is a subsective adjective allows us to immediately conclude that "Every small elephant is an elephant", or indeed that "Every small pfargtl is a pfargtl".
The natural logic program, in its most ambitious form, pursues a full logic of natural language. In such a 'natural logic', the logical form of a sentence would either mirror its linguistic structure or simply be the same as its surface form. The semantics accompanying this logic would either be a proof-theoretic one based on the logic itself or simply be obsolete.
More modest contributions to this program aim to capture various interesting parts of natural language in fragments of syllogistic logic, using conventional model-theoretic semantics and largely sidestepping issues of logical form by considering entire sentences at once. This chapter fits into this modest approach. (For a better look at natural logic, its main contributors and its various subdivisions, the introduction of (Moss 2008) is a good start.)
In line with this approach, this chapter introduces a syllogistic fragment for dealing with subsective adjectives, and proves it to be complete relative to an appropriate model-theoretic semantics.

In recent years similar work has been done for intersective adjectives as well as comparatives.(Moss 2010)(Moss 2011) But while every intersective adjective is subsective, the converse does not hold. Specifically, intersective adjectives are adjectives which have an extension of their own (using the term loosely) such that the extension of the modified noun is the intersection of the extension of the adjective and that of the noun proper. For example, given that *red* is an intersective adjective, we know right away that a red apple is exactly something

which is both red and an apple. This however does not work for *small*: a small skyscraper is not always both small and a skyscraper.
Perhaps more interestingly, non-intersective subsective adjectives like *small* can show behaviour similar to context-dependence, and therefore do not have the same monotonicity properties as intersective ones: a red apple is a red piece of fruit, but a small skyscraper is not always a small building. These differences should make this endeavour a worthwhile expansion, moreso since the class of subsective adjectives is rather broad, to the point that some argue it covers all natural language adjectives that aren't more appropriately interpreted as modalities.(Partee 2010)

4.2 The Syllogistics

Our first step in creating an appropriate syllogistics is to create an appropriate type of models and semantics for it to work with. We will include nouns and adjectives as distinct entities, and require that each model offers an interpretation both for each noun in isolation and for each combination PA of an adjective and a noun. In fact, we also require a second interpretation $\overline{P}A$ for each such combination, which we will take to be the interpretation of the anti-extension. Later on we will demand that these interpretations are consistent with the adjectives being subsective, as well as with $\overline{P}A$ being such an anti-extension, but we will first prove completeness of a simpler logic in the broader case.
Our semantics will provide interpretations not only for standard existential and universal sentences, but also for *conditional universals*; that is, sentences of the form "*All X that are Y are Z*". While these are generally perhaps little more than a curiosity, in the context of suitable models (which we will define later) they provide us with great expressive power. They will allow us not only to express "*No X are Y*" and "*Some X are not Y*" (see Section 4.2.5), but also to essentially handle arbitrarily large conjunctions through use of the Definition Rule and Extended Definition Rule (Section 4.2.4).

A model consists of a domain M, a subset $[\![A]\!] \subseteq M$ for each noun A and subsets $[\![PA]\!], [\![\overline{P}A]\!] \subseteq M$ for each pair of an adjective P and a noun A. The semantics is then straightforward enough: use the following, where X and Y can be of the form A, PA or $\overline{P}A$:

$$\mathfrak{M} \models \textit{All } X \textit{ are } Y \;\Leftrightarrow\; [\![X]\!] \subseteq [\![Y]\!] \tag{4.1}$$

$$\mathfrak{M} \models \textit{Some } X \textit{ are } Y \;\Leftrightarrow\; [\![X]\!] \cap [\![Y]\!] \neq \emptyset \tag{4.2}$$

$$\mathfrak{M} \models \textit{All } X \textit{ that are } Y \textit{ are } Z \;\Leftrightarrow\; [\![X]\!] \cap [\![Y]\!] \subseteq [\![Z]\!] \tag{4.3}$$

We will generally write (X, Y, Z) to abbreviate *All* X *that are* Y *are* Z. It may be worth noting that as in most modern approaches and contrary to what the ancient syllogistics prescribed, existential import is not valid for this semantics and will not be an acceptable step in the accompanying logic.

$$\frac{}{\text{All X are X}} \qquad \frac{\text{All X are Z} \quad \text{All Z are Y}}{\text{All X are Y}}$$

$$\frac{\text{Some X are Y}}{\text{Some Y are X}} \qquad \frac{\text{All Y are Z} \quad \text{Some X are Y}}{\text{Some X are Z}}$$

$$\frac{}{\text{(X, Y, X)}} \qquad \frac{}{\text{(X, Y, Y)}} \qquad \frac{\text{(X, Y, U)} \quad \text{(X, Y, V)} \quad \text{(U, V, Z)}}{\text{(X, Y, Z)}}$$

$$\frac{\text{All X are Y}}{\text{(X, X, Y)}} \qquad \frac{\text{(X, X, Y)}}{\text{All X are Y}} \qquad \frac{\text{Some U are V} \quad \text{(U, V, X)} \quad \text{(U, V, Y)}}{\text{Some X are Y}}$$

Figure 4.1: The basic logic. The first four rules are kept for convenience; interested readers may prove them from the rest.

4.2.1 Logic

The logic we will start out with is based heavily on work by Moss (Moss 2008). Specifically, we expand his logic for *All* and *Some* with his logic for (X, Y, Z) and a couple of extra rules. We will use $\vdash$ to refer to this logic, the rules of which are shown in Figure 4.2.1. Since I am unaware of any previous effort to combine these particular two, we will need to do some legwork to prove $\vdash$ is complete.

4.2.2 An aside: Predicative use

As it is based around subsective adjectives, the above system allows adjectives to appear only in constructions where they modify nouns. This could potentially limit its usefulness in dealing with natural language, where they can sometimes be used in a directly predicative way such as in the sentence "All men are mortal". Unfortunately, incorporating this use would be far from trivial.

As the interpretation of subsective adjectives can depend on the noun being modified, it would not be an appropriate strategy to interpret all sentences of the form "All A are P" as one would sentences of the form "All A are PA". To do so would mean that "All ants are small" implies "All ants are small ants", which is inappropriate. (Beyond this, there would also be a technical problem when trying to interpret sentences of the form "All PA are Q".)

Furthermore, it is not even a given that there is a single noun one should take the directly predicative Q to modify. For example, consider the sentence "All members of our family are tall". When "members of our family" includes both children and adults, the appropriate reading would be that the children are tall for children of their age and the adults are tall for adults. The creation of an approach that allows us to acquire such interpretation in a non-arbitrary systematic way would be well beyond the scope of this simple piece. Thus, we shall not take the option of directly predicative use into consideration in the rest of this chapter.

4.2.3 General Models

The first part of our completeness proof is to show that $\vdash$ is complete relative to models as defined above. Like the rules, the proofs in this section are based on (Moss 2008).

4.2.1. DEFINITION. For a set of sentences Γ, the existential part Γ_{some} is the set of all sentences of the form *Some X are Y* in Γ.

4.2.2. DEFINITION. For a set of sentences Γ, define the model $\mathfrak{M}_E$ as follows: Let $N = |\Gamma_{some}|$. We think of N as the ordinal number $\{0, 1, \ldots, N-1\}$. For $i \in N$, label the predicates used in Γ_{some} as V_i and W_i such that

$$\Gamma_{some} = \{Some\ V_i\ are\ W_i | i \in N\}$$

Note that for $i \neq j$, we might well have $V_i = V_j$ or $W_i = W_j$.
Now for the domain of $\mathfrak{M}_E$ we take the set N and for the interpretations we define

$$[\![X]\!] = \{i \in N | \Gamma \vdash (V_i, W_i, X)\}$$

4.2.3. LEMMA. $\mathfrak{M}_E \models \Gamma$

PROOF: Let $S \in \Gamma$. If S is existential, then S is of the form "Some V_i are W_i" for some i. Since $\Gamma \vdash (V_i, W_i, V_i), (V_i, W_i, W_i)$, it follows that $i \in [\![V_i]\!] \cap [\![W_i]\!]$. Thus, $\mathfrak{M}_E \models S$.
If S is of the form (X, Y, Z), then suppose $i \in [\![X]\!] \cap [\![Y]\!]$. Then $\Gamma \vdash (V_i, W_i, X)$, (V_i, W_i, Y). Since we trivially have $\Gamma \vdash (X, Y, Z)$, we have

$$\cfrac{\cfrac{\Gamma}{(V_i, W_i, X)} \quad \cfrac{\Gamma}{(V_i, W_i, Y)} \quad \cfrac{\Gamma}{(\text{X, Y, Z})}}{(V_i,\ W_i,\ \text{Z})}$$

Since $\Gamma \vdash (V_i, W_i, Z)$, therefore $i \in [\![Z]\!]$. This shows that $[\![X]\!] \cap [\![Y]\!] \subseteq [\![Z]\!]$, which means $\mathfrak{M}_E \models S$.
If S is of the form "All X are Y", then $\Gamma \vdash (X, X, Y)$. This gives $\mathfrak{M}_E \models (X, X, Y)$ as above, and thus $\mathfrak{M}_E \models S$.

4.2.4. LEMMA. *For existential S, if* $\mathfrak{M}_E \models S$ *then* $\Gamma \vdash S$.

PROOF: Let S be "Some X are Y". Suppose $i \in [\![X]\!] \cap [\![Y]\!]$. Therefore $\Gamma \vdash (V_i, W_i, X), (V_i, W_i, Y)$, and so we have

$$\dfrac{\dfrac{\Gamma}{\text{Some } V_i \text{ are } W_i} \quad \dfrac{\Gamma}{(V_i,\, W_i,\, \text{X})} \quad \dfrac{\Gamma}{(V_i,\, W_i,\, \text{Y})}}{\text{Some X are Y}}$$

4.2.5. THEOREM. *Let* Γ *contain only universal sentences, and let* $S = (X, Y, Z)$. *Then if* $\Gamma \models S$, $\Gamma \vdash S$.

PROOF: Let $\mathfrak{M}$ be the model whose universe is $\{*\}$ and whose interpretation is such that $[\![W]\!] = \{*\}$ if and only if $\Gamma \vdash (X, Y, W)$.
To show that $\mathfrak{M} \models \Gamma$, suppose $(U, V, W) \in \Gamma$. (For sentences of the form "All U are W", simply substitute (U, U, W), which is made true iff "All U are W".) We need to show that $[\![U]\!] \cap [\![V]\!] \subseteq [\![W]\!]$. This is trivial if the left side is empty, so we may assume $[\![U]\!] \cap [\![V]\!] = \{*\}$.
This gives $\Gamma \vdash (X, Y, U), (X, Y, V)$. Therefore we have the following proof of (X, Y, W) from Γ.

$$\dfrac{\dfrac{\Gamma}{(X, Y, U)} \quad \dfrac{\Gamma}{(X, Y, V)} \quad \dfrac{\Gamma}{(U, V, W)}}{(X, Y, W)}$$

Thus $[\![W]\!] = \{*\} \supseteq [\![U]\!] \cap [\![V]\!]$.
Now since $\mathfrak{M} \models \Gamma$, it follows that $\mathfrak{M} \models S$. As a matter of direct rules, $\Gamma \vdash (X, Y, X), (X, Y, Y)$. As such, $[\![X]\!] = \{*\} = [\![Y]\!]$. Therefore $[\![Z]\!] = \{*\}$, and therefore by construction $\Gamma \vdash S$.

4.2.6. THEOREM. $\vdash$ *is complete. (For models that aren't necessarily acceptable in the sense we will define later.)*

PROOF: Let $\Gamma \models S$. If S is existential, then $\Gamma \vdash S$ follows since $\mathfrak{M}_E \models S$.

If it is of the form (X, Y, Z), let $\Gamma' = \Gamma - \Gamma_{Some}$. We claim that $\Gamma' \models S$. To see this, let $\mathfrak{M} \models \Gamma'$. We get a new model $\mathfrak{M}' = \mathfrak{M} \cup \{*\}$ by letting $[\![X]\!]' = [\![X]\!] \cup \{*\}$ for all X. The model $\mathfrak{M}'$ so obtained satisfies Γ' and all *Some* sentences whatsoever in the fragment. Hence $\mathfrak{M}' \models \Gamma$.
So $\mathfrak{M}' \models S$. And since S is a universal sentence $\mathfrak{M} \models S$ as well. This proves our claim that $\Gamma' \models S$. By Theorem 4.2.5, it follows that $\Gamma' \vdash S$.

If S is instead of the form "All X are Y", it can be proven by first proving (X, X, Y) and then using a translation rule for the final step.

4.2.4 Weakly Acceptable Models

To actually get an interpretation where adjectives function as subsective adjectives, we add some more conditions to the models. A model which satisfies all the conditions we are after is deemed acceptable. To help streamline the extensive

$$\frac{}{\text{All PA are A}} \qquad \frac{}{\text{All } \overline{P}\text{A are A}} \qquad \text{Subsectivity 1 \& 2}$$

$$\frac{\text{Some PA are } \overline{P}\text{A}}{\text{S}} \qquad \frac{(X, Y, PA) \qquad (X, Y, \overline{P}A)}{(X, Y, Z)} \qquad \text{Ex Falso 1 \& 2}$$

Figure 4.2: The rules for $\vdash_W$ on top of the rules for $\vdash$.

completeness proof we will also make use of a weaker notion: the weakly acceptable model. Every acceptable model is also weakly acceptable, but not vice versa.

4.2.7. DEFINITION. A model is considered *acceptable* if the following conditions hold:

- $[\![PA]\!] \subseteq [\![A]\!]$, $[\![\overline{P}A]\!] \subseteq [\![A]\!]$
- $[\![\overline{P}A]\!] = [\![A]\!] - [\![PA]\!]$

It is instead considered *weakly acceptable* if the last condition is replaced by:

- $[\![PA]\!] \cap [\![\overline{P}A]\!] = \emptyset$

4.2.8. DEFINITION. Define the expanded logic $\vdash_W$ by adding to $\vdash$ the rules in Figure 4.2.

4.2.9. LEMMA. *Let Γ contain only universal sentences. Then $\Gamma \models (X, Y, Z)$ on weakly acceptable models iff $\Gamma \vdash_W (X, Y, Z)$.*

PROOF: For the interesting direction, suppose $\Gamma \models (X, Y, Z)$. Consider the model $\mathfrak{M}$ with $M = \{*\}$ and for each W, $[\![W]\!] = \{*\}$ iff $\Gamma \vdash_W (X, Y, W)$. As we have previously shown, $\mathfrak{M} \models \Gamma$.
Suppose $[\![PA]\!] = \{*\}$. Then $\Gamma \vdash_W (X, Y, PA)$. By adding subsectivity to the proof, $\Gamma \vdash_W (X, Y, A)$. Thus $[\![A]\!] = \{*\} \supseteq [\![PA]\!]$.
There are two cases to consider. If there are P and A such that $[\![PA]\!] = \{*\} = [\![\overline{P}A]\!]$ then we have:

$$\frac{\dfrac{\Gamma}{(X, Y, PA)} \qquad \dfrac{\Gamma}{(X, Y, \overline{P}A)}}{(X, Y, Z)}$$

If there are no such P and A, then this means $\mathfrak{M}$ is weakly acceptable. Therefore $\mathfrak{M} \models (X, Y, Z)$. It is straightforwardly seen that $* \in [\![X]\!] \cap [\![Y]\!]$. Hence $* \in [\![Z]\!]$. Thus, by construction, $\Gamma \vdash_W (X, Y, Z)$.

4.2.10. THEOREM. *If Γ is consistent, then there is a weakly acceptable model $\mathfrak{M}$ such that:*

- $\mathfrak{M} \models \Gamma$
- *If $\mathfrak{M} \models$Some X is Y, then $\Gamma \vdash_W$Some X is Y.*

PROOF: Define $\mathfrak{M}$ in the same way as $\mathfrak{M}_E$ (Definition 4.2.2, except with $\vdash_W$ instead of $\vdash$). Thus, it suffices to show that $\mathfrak{M}$ is weakly acceptable.
For the first property, suppose $i \in [\![PA]\!]$. Then $\Gamma \vdash_W (V_i, W_i, PA)$, and so we have the following:

$$\dfrac{(V_i, W_i, PA) \qquad (V_i, W_i, PA) \qquad \dfrac{\dfrac{}{\text{All PA are A}}}{\text{(PA, PA, A)}}}{(V_i, W_i, A)}$$

Therefore $\Gamma \vdash_W (V_i, W_i, A)$, and therefore $i \in [\![A]\!]$.
This proves that $[\![PA]\!] \subseteq [\![A]\!]$, and $[\![\overline{P}A]\!] \subseteq [\![A]\!]$ can be proven in the same way.
For the last property, suppose $i \in [\![PA]\!] \cap [\![\overline{P}A]\!]$.
Therefore $\Gamma \vdash_W (V_i, W_i, PA), (V_i, W_i, \overline{P}A)$, giving the following:

$$\dfrac{\dfrac{\dfrac{\Gamma}{\text{Some } V_i \text{ are } W_i} \qquad \dfrac{\Gamma}{(V_i, W_i, PA)} \qquad \dfrac{\Gamma}{(V_i, W_i, \overline{P}A)}}{\text{Some PA are } \overline{P}A)}}{\text{S}}$$

This contradicts the assumption that Γ is consistent.

4.2.11. THEOREM. *$\vdash_W$ is complete on weakly acceptable models.*

PROOF: Let $\Gamma \models S$. If S is existential, then $\Gamma \vdash_W S$ by the previous theorem. If it is universal, then we claim that $\Gamma_U \models S$, where Γ_U contains only the universal sentences in Γ. This claim follows from Lemma 4.2.17, which is proven in the next section.
We can assume wlog that S is of the form (X, Y, Z). Thus by Lemma 4.2.9 $\Gamma_U \vdash_W S$, and therefore $\Gamma \vdash_W S$.

The Universal Model $\mathfrak{M}_U$

For a consistent set of sentences Γ, we define a special model with some interesting properties.

4.2.12. DEFINITION. Let Γ be a ($\vdash_W$-)consistent set of sentences. Let Γ_U consist of all universal sentences in Γ. We construct $\mathfrak{M}_U$ as follows:

- Let I be the set of all nouns plus all PA and $\overline{P}A$ where A is a noun and P an adjective.
- Let Γ' be the set of all universal sentences S for which $\Gamma_U \vdash_W S$. Enumerate it such that $\Gamma' = \{S_1, S_2, \ldots\}$.
- Let $M_0 = \{y \subseteq I | \textit{if } PA \in y \textit{ then } \overline{P}A \notin y\}$.
- $[\![X]\!]_i = \{y \in M_i | X \in y\}$
- $M_{i+1} = \begin{cases} M_i - [\![X]\!]_i \cap (M_i - [\![Y]\!]_i) & \textit{If } S_{i+1} = \textit{All X are Y} \\ M_i - [\![X]\!]_i \cap [\![Y]\!]_i \cap (M_i - [\![Z]\!]_i) & \textit{If } S_{i+1} = (X, Y, Z) \end{cases}$
- Let $\mathfrak{M}_U = (\bigcap M_i, \bigcap [\![.]\!]_i)$

4.2.13. LEMMA. $\mathfrak{M}_U \models \Gamma_U$

PROOF: Suppose $S \in \Gamma_U$. Then trivially $\Gamma_U \vdash_W S$. Therefore $S \in \Gamma'$. Given the enumeration of Γ', let j be such that $S = S_{j+1}$.
By construction, M_{j+1} does not contain any counterexample to S. Since the domain of $\mathfrak{M}_U$ is contained is $(\bigcap M_i) \subseteq M_{j+1}$ and the interpretation function is the same except restricted to its domain, this means the domain of $\mathfrak{M}_U$ does not contain any such counterexample either. Therefore $\mathfrak{M}_U \models S$.

4.2.14. LEMMA. *For universal sentences S, if* $\mathfrak{M}_U \models S$ *then* $\Gamma_U \models S$.

PROOF: Suppose $\mathfrak{M}_U \models S$ and $\mathfrak{M} \models \Gamma_U$. For u, v in the universe of $\mathfrak{M}$, let $u \equiv v$ iff $\forall X : u \in [\![X]\!]_{\mathfrak{M}} \Leftrightarrow v \in [\![X]\!]_{\mathfrak{M}}$.
Now define $\mathfrak{M}' = (M', [\![.]\!]_{\mathfrak{M}'})$ as follows.

$$M' = \{\{X | \forall u \in [u], u \in [\![X]\!]_{\mathfrak{M}}\} | \textit{for } [u] \textit{ in the universe of } \mathfrak{M}/\equiv\}$$

$$[\![X]\!]_{\mathfrak{M}'} = \{y \in M' | X \in y\}$$

Clearly, $\mathfrak{M}'$ is essentially the same model as $\mathfrak{M}/\equiv$. But $\mathfrak{M}'$ has the same valuation function as $\mathfrak{M}_U$ and its universe is a part of the universe of $\mathfrak{M}_U$. Therefore, if S is a universal sentence, then $\mathfrak{M}_U \models S$ implies $(\mathfrak{M}/\equiv) \models S$, which in turn implies $\mathfrak{M} \models S$.

4.2.15. THEOREM. *For existential sentences S, if* $\Gamma \cup \{S\}$ *is consistent then* $\mathfrak{M}_U \models S$.
(As a corollary, if Γ *is consistent then* $\mathfrak{M}_U \models \Gamma$.*)*

PROOF: Suppose that $\mathfrak{M}_U \not\models S$, where S is "Some X is Y". Thus, $[\![X]\!] \cap [\![Y]\!] = \emptyset$. This gives $\mathfrak{M}_U \models (X, Y, PA), (X, Y, \overline{P}A)$. Therefore, by Lemma 4.2.14, $\Gamma_U \models (X, Y, PA), (X, Y, \overline{P}A)$, and therefore (by Lemma 4.2.9) $\Gamma_U \vdash_W (X, Y, PA), (X, Y, \overline{P}A)$.
But this gives

$$\dfrac{\dfrac{\dfrac{\Gamma}{(X,Y,PA)} \quad \dfrac{\Gamma}{(X,Y,\overline{P}A)} \quad \text{Some X is Y}}{\text{Some PA is }\overline{P}\text{A}}}{\text{R}}$$

This shows that $\Gamma \cup \{S\}$ is inconsistent.
(For the corollary, suppose Γ is consistent and $S \in \Gamma$. If S is existential, then $\mathfrak{M}_U \models S$ by the above. If it is universal, then $\mathfrak{M}_U \models S$ by Lemma 4.2.13.)

4.2.16. Lemma. *$\mathfrak{M}_U$ is a weakly acceptable model.*

Proof:

- Let S = "*All PA are A*". Then $\vdash_W S$. Thus, $\Gamma_U \vdash_W S$ and therefore $\mathfrak{M}_U \models S$ by construction. This gives $[\![PA]\!] \subseteq [\![A]\!]$.
- Any element of $[\![PA]\!] \cap [\![\overline{P}A]\!]$ would contain both PA and $\overline{P}A$. But all potential elements satisfying this were removed from M_0.

4.2.17. Lemma. *Let Γ be consistent. Then for any universal sentence S, $\Gamma \models S$ iff $\Gamma_U \models S$.*
(As a corollary, $\Gamma \vdash_W S$ iff $\Gamma_U \vdash_W S$.)

Proof: The right-to-left direction is trivial. Now suppose $\Gamma \models S$. Then $\mathfrak{M}_U \models S$ (since $\mathfrak{M}_U$ is a weakly acceptable model of Γ). Therefore $\Gamma_U \models S$ by Lemma 4.2.14.

Some interesting pseudo-rules

Weakly acceptable models make valid some inference schemes beyond the rules we've used thus far. The following three in particular will be useful to us:

$$\dfrac{\dfrac{[\Delta((X_1,X_2),Z)]}{\vdots \atop \text{S}}}{\text{S}}\ \text{definition}$$

In the Definition rule, $\Delta((X_1, X_2), Z)$ is defined as follows:

$$\Delta((X_1, X_2), Z) := \{(X_1, X_2, Z), \textit{All Z are } X_1, \textit{All Z are } X_2\}$$

Furthermore, $\Delta((X_1, X_2), Z)$ may be withdrawn only if Z occurs neither in S nor in any assumption that has not already been withdrawn.
Essentially, Δ *defines* Z as being the conjunction of X_1 and X_2. The rule then says that this definition is not productive -ie anything concluded with the help of the definition is already true without it.

$$\dfrac{\dfrac{[(X, Y, PA)] \qquad [(X, Y, \overline{P}A)]}{\vdots}}{\dfrac{\text{S}}{\text{Some X are Y}}}\ \text{RAA minor}$$

(Here, the bracketed premises may be withdrawn if the proof works for arbitrary S.) This partial form of Reductio Ad Absurdum further establishes that the combination of (X, Y, PA) and $(X, Y, \overline{P}A)$ amounts to saying "No X are Y". Given this, its validity should not come as a big surprise.

$$\dfrac{\dfrac{[\text{Some X are Z}]}{\vdots}}{\dfrac{\text{Some Y are Z}}{\text{All X are Y}}}\ \text{All-abstraction}$$

Here, "Some X are Z" can be withdrawn if Z does not occur in any assumption that has not already been withdrawn. Thus, the rule says that if any property belonging to some X provably also belongs to some Y, then that must be because all X are Y.

4.2.18. Lemma. *The three rules above are sound on weakly acceptable models.*

Proof: For Definition, suppose $\mathfrak{M} \models \Gamma$ and S is provable from Γ using Definition (with an otherwise sound proof). Then this proof can be done with a Z (plus extra variables) which does not occur in Γ and which does not yet have an interpretation in $\mathfrak{M}$.
Then let $\mathfrak{M}'$ be the model which differs from $\mathfrak{M}$ only in that $[\![Z]\!] = [\![X_1]\!] \cap [\![X_2]\!]$, and $[\![PZ]\!] = [\![Z]\!]$ for all P.
Since Z does not occur in Γ, it follows that $\mathfrak{M}' \models \Gamma$. By assumption there is a sound proof of S from the combination of Γ and $\Delta((X_1, X_2), Z)$. Thus, $\mathfrak{M}' \models S$. But Z doesn't occur in S and $\mathfrak{M}'$ differs only in the interpretation of Z. Hence, $\mathfrak{M} \models S$.
For All-abstraction, suppose $\mathfrak{M} \models \Gamma$ and there is a sound proof of "Some Y are Z" from the combination of Γ and "Some X are Z", where Z does not occur in Γ. This proof is still valid if we choose Z such that it does not have an interpretation on $\mathfrak{M}$. Now for an arbitrary $x \in [\![X]\!]$, let $\mathfrak{M}_x$ be the model which differs from $\mathfrak{M}$ only in that $[\![Z]\!] = \{x\}$, and $[\![PZ]\!] = \{x\}$ for all P.
As before, it is easily seen that $\mathfrak{M}_x \models \Gamma, \textit{Some X are Z}$. Thus, by the assumed soundness, $\mathfrak{M}_x \models \textit{Some Y are Z}$. It follows that $x \in [\![Y]\!]$, and that this is also true for $\mathfrak{M}$. Since we did this for arbitary $x \in [\![X]\!]$, this proves that $[\![X]\!] \subseteq [\![Y]\!]$, and hence $\mathfrak{M} \models \textit{All X are Y}$.
(It is left to the reader to complete the above into a proper induction proof. RAA minor is also left to the reader.)

While the Definition Rule in its above form has its uses already, we will later

need an extended version that allows for any number of X_i to be declared jointly equivalent to Z.
Luckily this can be obtained from the simple form with minimal extra effort by repeated application. For example, to get a suitable $\Delta((X_1, X_2, X_3), Z)$ we introduce a new Y and combine $\Delta((X_1, X_2), Y)$ with $\Delta((Y, X_3), Z)$. Proof is then immediate from using the rule twice. The next lemma generalizes this strategy.

4.2.19. Lemma. ***(Extended Definition Rule)*** *Recursively define*

$$\Delta((X_1, \dots, X_k), Z) = \Delta((X_1, \dots, X_{k-1}), Y_{k-2}) \cup \Delta((Y_{k-2}, X_k), Z)$$

Then, assuming Z and all Y_i do not occur in any assumption that has not already been withdrawn, the following rule is sound on weakly acceptable models.

$$\frac{\begin{array}{c}\underline{[\Delta((X_1, \dots, X_n), Z)]}\\ \vdots \\ \mathrm{S}\end{array}}{\mathrm{S}}\ \text{definition}$$

Proof: Proof is by induction using soundness of the simple case.

Since we already have a complete system for weakly acceptable models, the following follows immediately:

4.2.20. Lemma. *In any correct proof, any occurrence of the above rules can be rewritten using only rules from $\vdash_W$.*

Sadly, to actually provide a specific rewriting method is hard enough to go beyond the scope of this fragment. I suspect that in all these cases the key is that the "inside" proof can only work if a $\vdash_W$-proof of the end conclusion already exists. But of course this insight is not yet particularly constructive.
Regardless, these rules are useful enough to be needed in some proofs in the later sections. We will take this lemma as legitimizing treating these pseudo-rules as if they were actual $\vdash_W$-rules at those places.

4.2.5 Acceptable Models

This rounds out our completeness proof for weakly acceptable models. Relative to weakly acceptable models, acceptable models are characterized by their forcing every element which is A to be either PA or $\overline{P}A$ (where A is a noun and P a predicate).
Thus, to get our logic for acceptable models $\vdash_{AS}$, we simply add the following variation of the Disjunctive Syllogism.

$$\dfrac{\begin{array}{c}\underline{[\text{Some X are PA}]}\\ \vdots \\ \text{S}\end{array} \qquad \begin{array}{c}\underline{[\text{Some X are }\overline{P}\text{A}]}\\ \vdots \\ \text{S}\end{array} \qquad \text{Some X are A}}{\text{S}}\ \text{Disjunction Elimination}$$

Note: Unless specified otherwise, the word *consistent* in this section refers to consistency with regards to $\vdash_{AS}$.

4.2.21. LEMMA. *Disjunction Elimination is sound on acceptable models. (Hence, $\vdash_{AS}$ is sound on acceptable models.)*

PROOF: Left to the reader.

The following similar way of looking at things will be useful later.

4.2.22. LEMMA. *Let S="Some X are A", S_1="Some X are PA", S_2="Some X are $\overline{P}A$". Then if $\Gamma \cup \{S\}$ is consistent, so is at least one of $\Gamma \cup \{S_1\}$ and $\Gamma \cup \{S_2\}$.*

PROOF: Suppose both of them are not. Then an arbitrary R can be proven from $\Gamma \cup \{S\}$ as follows.

$$\dfrac{\dfrac{\Gamma \quad [\text{Some X are PA}]}{\text{R}} \qquad \dfrac{\Gamma \quad [\text{Some X are }\overline{P}\text{A}]}{\text{R}} \qquad \text{Some X are A}}{\text{R}}$$

The Restricted Universal Model $\mathfrak{M}_U^-$

To show completeness for $\vdash_{AS}$, we begin by showing that every Γ that is consistent on it has an acceptable model. We do this by explicitly constructing such a model, similar to $\mathfrak{M}_U$ before.

4.2.23. DEFINITION. Let $\mathfrak{M}_U^-$ be like $\mathfrak{M}_U$ (see Section 4.2.4), except that in the construction we let

$$M_0 = \{X \subseteq I | \textit{if } A \in X \textit{ then } PA \in X \textit{ or } \overline{P}A \in X\textit{, but not both}\}$$

Through this definition we remove any superfluous elements from $\mathfrak{M}_U$, guaranteeing its acceptability. It is easy enough to see that $\mathfrak{M}_U^-$ is an acceptable model and makes the universal parts of Γ true. Thus, our main interest becomes the existential claim.
Since $\mathfrak{M}_U^-$ does not allow elements to be underdetermined in the way $\mathfrak{M}_U$ does, we will prove that each element of $\mathfrak{M}_U$ has at least one completely determined version in $\mathfrak{M}_U$, which is therefore also in $\mathfrak{M}_U^-$. This will guarantee that any existential sentence that has a witness in $\mathfrak{M}_U$ (which includes any existential sentence consistent with Γ) will also have one in $\mathfrak{M}_U^-$.

Disjunction Elimination is essential for this proof, which naturally requires that Γ is consistent. It also makes heavy use of the conditional universal, both directly and through the Extended Definition Rule. Note also that most of it necessarily takes place in $\mathfrak{M}_U$, not $\mathfrak{M}_U^-$.

4.2.24. Lemma. *If* $\Gamma \cup \{S\}$ *is consistent, then if* $\mathfrak{M}_U \models$ "Some X is A"*, it follows that either* $\mathfrak{M}_U \models$ "Some X is PA" *or* $\mathfrak{M}_U \models$ "Some X is $\overline{P}A$"*.*

Proof: Let $z = \{Z|\Gamma \vdash_{AS} (X, A, Z)\}$, $z_1 = \{Z|\Gamma \vdash_{AS} (X, PA, Z)\}$, $z_2 = \{Z|\Gamma \vdash_{AS} (X, \overline{P}A, Z)\}$. Note that none of these can have been removed as a result of the rules. This means that any that are not in M_U are not in M_0.
Therefore, if $z_1 \notin M_U$ then there are Q, B such that $QB, \overline{Q}B \in z_1$. But that means $\Gamma \vdash_{AS} (X, PA, QB), (X, PA, \overline{Q}B)$. So then $\Gamma \cup \{$"*Some X is PA*"$\}$ is inconsistent. Similarly, if $z_2 \notin M_U$ then $\Gamma \cup \{$"*Some X is* $\overline{P}A$"$\}$ is inconsistent.
Because $\Gamma \cup \{S\}$ is consistent and because of the Disjunction Elimination rule, it follows that $z_1 \in M_U$ or $z_2 \in M_U$.

4.2.25. Theorem. *Let* Γ *be consistent. Let* $z = \{X_1, \ldots, X_n, A\}$*,* $z_1 = z \cup \{PA\}$*,* $z_2 = z \cup \{\overline{P}A\}$*. Then if* $z \in M_U$*, there is some* $z' \in M_U$ *with* $z_1 \subseteq z'$ *or* $z_2 \subseteq z'$*.*

Proof: For a Z which doesn't occur in Γ, let $\Delta = \Delta((X_1, \ldots, X_n), Z)$ (see Lemma 4.2.19). Let $\Gamma' = \Gamma \cup \Delta$. Then Γ' is consistent (otherwise Γ would be inconsistent by the Extended Definition Rule).
Let the model $\mathfrak{M}'_U$ be as for the $\mathfrak{M}_U$ construction, but based on Γ'. Then $\mathfrak{M}'_U \models$ "*Some Z is A*". By the previous lemma, it follows that either $\mathfrak{M}'_U \models$ "*Some Z is PA*" or $\mathfrak{M}'_U \models$ "*Some Z is* $\overline{P}A$".
Let z' be a witness to this. Then it follows that $z_1 \subseteq z'$ or $z_2 \subseteq z'$. Now let $\mathfrak{M}_U^Z$ be $\mathfrak{M}_U$ expanded with a noun Z such that

$$[\![Z]\!] = \bigcup_{i=1}^{n} [\![X_i]\!]$$

Since Δ forces Z to have this particular interpretation, it follows that $\mathfrak{M}'_U \subseteq \mathfrak{M}_U^Z$. Thus $z' \in \mathfrak{M}_U^Z$. But then $(z' - \{Z\}) \in \mathfrak{M}_U$.

4.2.26. Theorem. *Let* Γ *be consistent. Then if* $x \in M_U$*, then there is an* $x' \in M_U^-$ *with* $x \subseteq x'$*.*

Proof: By induction using the above theorem.

4.2.27. Lemma. *For existential sentences* S*, if* $\Gamma \cup \{S\}$ *is consistent, then* $\mathfrak{M}_U^- \models S$*.*

Proof: Suppose $\Gamma \cup \{S\}$ is consistent, with S="Some X is Y". By Theorem 4.2.15, we know $\mathfrak{M}_U \models S$. Let x witness this. Then by the above theorem there is an $x' \in M_U^-$ with $x \subseteq x'$. This x' witnesses $\mathfrak{M}_U^- \models S$.

4.2.28. Theorem. *Every consistent set has an acceptable model.*

Proof: Let Γ be such a set. By construction, $\mathfrak{M}_U^-$ makes the universal part of Γ true. By the lemma above, it also makes the existential part of Γ true.
Hence $\mathfrak{M}_U^- \models \Gamma$. The proof that $\mathfrak{M}_U^-$ is an acceptable model is left to the reader.

Finishing the completeness proof

The completeness proof now is a matter of combining the model-existence result we have just proved with the expressive power afforded to us by the conditional universal. For the existential part, we show that if S is not provable, then its negation (which we can express as noted earlier) is consistent, and therefore has a model.

4.2.29. Theorem. *Let Γ be a consistent set of sentences. Let $\Gamma \models S$ on acceptable models for S an existential sentence. Then $\Gamma \vdash_{AS} S$.*

Proof: Let S="Some X is Y". Suppose $\Gamma \not\vdash_{AS} S$. Let $\Gamma' = \Gamma \cup \{(X, Y, PA), (X, Y, \overline{P}A)\}$. Then Γ' is consistent, because otherwise $\Gamma \vdash_{AS} S$ by RAA minor.
Since Γ' is consistent, there is an acceptable model $\mathfrak{M}$ with $\mathfrak{M} \models \Gamma'$. Clearly, $\mathfrak{M} \not\models S$, yet $\mathfrak{M} \models \Gamma$. Thus, $\Gamma \not\models S$.

For the simple universal case, we again show that if S is not provable, its negation (which involves using a newly introduced Z to stand in for the property of not being Y) is consistent.

4.2.30. Theorem. *Let Γ be a consistent set. Let S="All X are Y". If $\Gamma \models S$ on acceptable models, then $\Gamma \vdash_{AS} S$.*

Proof: Suppose $\Gamma \not\vdash_{AS} S$. For a Z which does not occur in Γ, let

$$\Gamma' = \Gamma \cup \{(Y, Z, PA), (Y, Z, \overline{P}A), \textit{Some X are Z}\}$$

Then Γ' is consistent, because if it were not we would get the following proof of S from Γ.

$$\cfrac{\cfrac{\cfrac{\Gamma \qquad [(Y,Z,PA)]^1[(Y,Z,\overline{P}A)]^1 \qquad [\text{Some X are Z}]^2}{\vdots \atop \text{S}}}{\text{Some Y are Z}}\,{-1}}{\text{All X are Y}}\,{-2}$$

Since Γ' is consistent, it has an acceptable model. This model has an element $x \in [\![X]\!] \cap [\![Z]\!]$. It is easily seen that $x \notin [\![Y]\!]$. Thus this model does not make S true. Since it does make Γ true, $\Gamma \not\models S$.

The added expressive power we have just used comes at the cost of having to prove completeness for conditional universals themselves. The Definition Rule has already been convenient to us several times, but this place in particular is one where we cannot do without it.

4.2.31. THEOREM. *Let Γ be a consistent set. Let $S = (X_1, X_2, Y)$. If $\Gamma \models S$ on acceptable models, then $\Gamma \vdash_{AS} S$.*

PROOF: Suppose $\Gamma \models S$. For a Z which does not occur in Γ, let $\Gamma' = \Gamma \cup \Delta((X_1, X_2), Z)$. Obviously, $\Gamma' \models$ "*All Z are Y*". Therefore by the previous theorem $\Gamma' \vdash_{AS}$ "*All Z are Y*".[1]
This gives us the following proof of S from Γ.

$$\cfrac{\cfrac{\cfrac{[\Delta((X_1,X_2),Z)]^1}{(X_1,X_2,Z)} \qquad \cfrac{[\Delta((X_1,X_2),Z)]^1}{(X_1,X_2,Z)} \qquad \cfrac{\cfrac{\Gamma \qquad [\Delta((X_1,X_2),Z)]^1}{\begin{array}{c}\vdots \\ \text{All Z are Y}\end{array}}}{(Z,Z,Y)}}{(X_1,X_2,Y)}}{(X_1,X_2,Y)}\ \text{-1}$$

The following now follows straightforwardly.

4.2.32. THEOREM. *$\vdash_{AS}$ is complete on acceptable models.*

4.3 Conclusion

With the proof we have completed in the previous section, we now know that the syllogistic logic defined by the combination of Figure 4.2.1, Figure 4.2 and the Disjunction Elimination Rule (Section 4.2.5) is sound and complete relative to the class of acceptable models, which are just those models where adjectives are subsective. Thus, it is an appropriate logic to use for subsective adjectives. Since subsective adjectives are perhaps the broadest class of natural language adjectives, this in and of itself is a result of some interest.
Beyond this, having characterized this broad class allows us to characterize more specific classes of adjectives in terms of the same system. For a simple example, consider intersective adjectives. Semantically we may characterize intersectives

[1]This does not work if Γ' is inconsistent, but besides that not being possible, the next step would still work even so.

adjectives as those adjectives P for which there is a $[\![P]\!]$ such that $[\![PA]\!] = [\![P]\!] \cap [\![A]\!]$ for every noun A. On the syllogistic side, this may be characterized by the following rule:[2]

$$\frac{}{(PA, B, PB)}\text{ Intersectiveness}$$

More complex cases involve syllogistic characterizations of adjectives based on underlying relations with particular properties. We will see more of this in the next chapter.

[2]To show completeness, define $[\![P]\!]$ as the union of all $[\![PA]\!]$.

Chapter 5

A Syllogistic Characterization of Gradable Adjectives

5.1 Introduction

Gradable adjectives are adjectives which may hold to a lesser or greater extent, with key examples being words like 'old', 'expensive' and 'tall'. Since gradable adjectives allow for comparative forms, they may be interpreted as being based on an underlying order. Some debate is possible as to whether such an order must come with explicit numerical degrees. We shall not assume that this is necessary, as we can get quite far without doing so.

Rather, we shall assume that gradable adjectives are those subsective adjectives which are based on a Weak Order; that is, a relation which is Transitive, Irreflexive and Almost-Connected. The main goal of this chapter will be to show that this notion can be characterized inside a natural logic where only the extension of the adjective is known beforehand. In Section 5.2 we briefly review the formal system we will be using for this, the syllogistics for subsective adjectives introduced in the previous chapter. In Section 5.3 we work towards the characterization we are after, ultimately capturing it in a single inference rule.

In Section 5.4 we characterize and explore the notion of *commensurability*, which here is defined as holding between groups of gradable adjectives which are based on the same underlying order. This will prove a useful notion in dealing with antonyms and degree modifiers. Finally, in Section 5.5 we discuss a way to extend the system to deal with vagueness.

5.2 The Syllogistics

As mentioned previously, we shall make use of the syllogistics introduced in the previous chapter. The first component of the syllogistics is the class of models and semantics it will work with. Here nouns and (subsective) adjectives are included

as distinct entities, and each model is required to offer an interpretation both for each noun in isolation and for each combination PA of an adjective and a noun. In fact, we also require a second interpretation $\overline{P}A$ for each such combination, which we will take to be the interpretation of the anti-extension.
The semantics provides interpretations not only for standard existential and universal sentences, but also for *conditional universals*; that is, sentences of the form "*All X that are Y are Z*". While these are generally perhaps little more than a curiosity, in this particular context they provide us with great expressive power, which we will come to rely on later.

The comparison-class interpretation

It is important to note that we will not be interpreting nouns as arbitrary comparison classes: while every noun does arguably constitute a comparison class, we do not assume that for every conceivable comparison class a noun is available which matches it. This choice drastically reduces how much we will be able to assume about the formal behaviour of gradable adjectives, making our task rather more challenging.
Of course, the comparison-class interpretation is still interesting in its own right, both because it is the appropriate approach to analyze context-dependence and because it leads to a cleaner underlying weak order. The work for this has essentially been done by van Benthem in (van Benthem 1991). For the sake of completeness and convenience, we redo it our own terms in Appendix 5.6.

5.2.1 Formalism

A model consists of a domain M, a subset $[\![A]\!] \subseteq M$ for each noun A and subsets $[\![PA]\!], [\![\overline{P}A]\!] \subseteq M$ for each pair of an adjective P and a noun A. The semantics is straightforward enough: use the following, where X and Y can be of the form A, PA or $\overline{P}A$:

$$\mathfrak{M} \models \textit{All } X \textit{ are } Y \;\Leftrightarrow\; [\![X]\!] \subseteq [\![Y]\!] \tag{5.1}$$

$$\mathfrak{M} \models \textit{Some } X \textit{ are } Y \;\Leftrightarrow\; [\![X]\!] \cap [\![Y]\!] \neq \emptyset \tag{5.2}$$

$$\mathfrak{M} \models \textit{All } X \textit{ that are } Y \textit{ are } Z \;\Leftrightarrow\; [\![X]\!] \cap [\![Y]\!] \subseteq [\![Z]\!] \tag{5.3}$$

(We will generally write (X, Y, Z) for the latter.)
In order to properly represent subsective adjectives, it is necessary to restrict ourselves to a specific class of models. This class, which we will refer to as *acceptable* models, consists of those models where all adjectives are subsective and $\overline{P}A$ is in fact the anti-extension it should be. Formally then, acceptable models are those models which have the following properties (for all P, A):

- $[\![PA]\!] \subseteq [\![A]\!]$, $[\![\overline{P}A]\!] \subseteq [\![A]\!]$

$$\frac{}{\text{All X are X}} \qquad \frac{\text{All X are Z} \quad \text{All Z are Y}}{\text{All X are Y}}$$

$$\frac{\text{Some X are Y}}{\text{Some Y are X}} \qquad \frac{\text{All Y are Z} \quad \text{Some X are Y}}{\text{Some X are Z}}$$

$$\frac{}{\text{(X, Y, X)}} \qquad \frac{}{\text{(X, Y, Y)}} \qquad \frac{\text{(X, Y, U)} \quad \text{(X, Y, V)} \quad \text{(U, V, Z)}}{\text{(X, Y, Z)}}$$

$$\frac{\text{All X are Y}}{\text{(X, X, Y)}} \qquad \frac{\text{(X, X, Y)}}{\text{All X are Y}} \qquad \frac{\text{Some U are V} \quad \text{(U, V, X)} \quad \text{(U, V, Y)}}{\text{Some X are Y}}$$

$$\frac{}{\text{All PA are A}} \qquad \frac{}{\text{All } \overline{P}\text{A are A}} \qquad \frac{\text{Some PA are } \overline{P}\text{A}}{\text{S}}$$

$$\frac{(X, Y, PA) \quad (X, Y, \overline{P}A)}{(X, Y, Z)}$$

$$\frac{\begin{matrix}\dfrac{\text{[Some X are PA]}}{} \\ \vdots \\ \text{S}\end{matrix} \quad \begin{matrix}\dfrac{\text{[Some X are } \overline{P}\text{A]}}{} \\ \vdots \\ \text{S}\end{matrix} \quad \text{Some X are A}}{\text{S}}$$

Figure 5.1: The syllogistic. The first two rows are a basic syllogistic for *All* and *Some*. The third row is a syllogistic for (X, Y, Z). The fourth row contains rules to properly combine the two. The fifth and sixth row provide for subsectivity and *Ex Falso Quodlibet*. The last rule can be seen as a form of Disjunction Elimination.

- $[\![\overline{P}A]\!] = [\![A]\!] - [\![PA]\!]$

The rules of the syllogistic logic that is used for this are shown in Figure 5.2.1. For the proof that this logic is sound and complete relative to the above semantics on the class of acceptable models, see Chapter 4.

5.3 Characterizing Gradable Adjectives

Using the system defined in the previous section, we can now start to investigate the natural logic properties of gradable adjectives. We will in particular want to consider the consequences of the fact that gradable adjectives are based on an underlying order with certain properties. To this end, our first step should be to create a more formal notion of "based on".

5.3.1. Definition. Let P be an adjective defined on the acceptable model $\mathfrak{M}$.

- *The relation induced by* P, denoted $\prec_P$, is given by

$$x \prec_P y \Leftrightarrow \exists A : x \in [\![\overline{P}A]\!], y \in [\![PA]\!]$$

- A relation $\prec$ on the domain of $\mathfrak{M}$ is *compatible with* P iff it is an extension of $\prec_P$ (that is, iff $x \prec_P y$ implies $x \prec y$). In this case we say that P *is based on* $\prec$.

The induced relation $\prec_P$ will likely look familiar; it is a traditional starting point when one wants to define a "P-er than" ordering. However, keep in mind that unlike some other treatments we are using A to refer to the extension of some noun rather than to an arbitrary context of comparison.
For us, the statement $x \prec_P y$ cannot be simplified along the lines of "$[\![PA]\!] = \{y\}$ for $[\![A]\!] = \{x, y\}$", because there is no guarantee that such an A exists. Thus, while $\prec_P$ will be a key ingredient of any acceptable order for P, we will never expect it to be one in and of itself and it is only the start of our investigations.

5.3.1 No Reversal

The first property we will characterize is the particularly desirable No Reversal property.[1] For our purposes, we define this as follows.

5.3.2. DEFINITION. $\prec$ has the No Reversal property for P if and only if

$$\forall x, y\, (x \prec y \Rightarrow \textit{not}\ y \prec_P x)$$

This of course is to say that if $x \prec y$ then there is no A for which $x \in [\![PA]\!]$, $y \in [\![\overline{P}A]\!]$. A typical example is to consider size. Some elephants who are considered small elephants may at the same time be considered large animals. From this, No Reversal lets us conclude that any large elephants (who are also animals, but this much is clear in the example) will also be large animals. The idea of course is that these large elephants would be larger than those small elephants who are already large animals, and being larger than some other large animals would therefore necessarily be large animals themselves. This relates to the property of monotonicity, which we will get to soon.

Given that they are based on an order, there are a number of reasons why we would want a gradable adjective to be based on one with the No Reversal property. Importantly, if this is the case then two key properties are guaranteed, those of *convexity* and *monotonicity*.
Convexity states that the extension of the adjective has no 'holes' in it, as seen along the dimension of the order. Specifically, if two objects both have PA, then so does any object (with A) located 'between' them.

[1] See also (van Benthem 1991). The interested reader may prove that if we interpret nouns as contexts, P has the No Reversal property defined there if and only if $\prec_P$ has the No Reversal property under our definition.

5.3.3. Theorem. *Let $\prec$ have No Reversal for P. Then P is convex relative to $\prec$.*
That is, if $x \prec y \prec z$ and $x, z \in [\![PA]\!]$, $y \in [\![A]\!]$, then $y \in [\![PA]\!]$.

Proof: Suppose towards contradiction that $y \in [\![\overline{P}A]\!]$. Then $y \prec_P x$. Since $x \prec y$, this contradicts $\prec$ having No Reversal for P.

Monotonicity states that PA is always a top segment of A, covering everything 'above' a certain cutoff point. Specifically, if a given object has PA, then so does any larger object in A.

5.3.4. Theorem. *Let $\prec$ have No Reversal for P. Then P is monotone relative to $\prec$.*
That is, if $x \prec y$ and $x, y \in [\![A]\!]$, $x \in [\![PA]\!]$, then $y \in [\![PA]\!]$.

Proof: Suppose towards contradiction that $y \in [\![\overline{P}A]\!]$. Then $y \prec_P x$. Since $x \prec y$, this contradicts $\prec$ having No Reversal for P.

The No Reversal property can be adequately characterized by the following inference rule.

$$\frac{\text{Some } \overline{P}\text{A are PB}}{\text{All PA that are B are PB}}\ \text{Directedness}$$

(It's easiest to see this is the same property when you keep in mind that "All PA that are B are PB" amounts to saying "No PA are $\overline{P}B$".)

5.3.5. Theorem. *An acceptable model makes Directedness valid for P if and only if there is a relation $\prec$ on its domain which has the No Reversal property for P and is compatible with P.*

Proof: Let $\mathfrak{M}$ be an acceptable model with $\mathfrak{M} \models$ *Some $\overline{P}A$ are PB*. Let $x \in [\![\overline{P}A]\!] \cap [\![PB]\!]$, $y \in [\![PA]\!] \cap [\![B]\!]$. Then $x \prec y$. (Since $x \prec_P y$.)
Therefore, since $x \in [\![PB]\!]$ and $y \in [\![B]\!]$, the No Reversal property implies $y \in [\![PB]\!]$. Since x must exist by assumption and we only assumed that $y \in [\![PA]\!] \cap [\![B]\!]$, it follows that $\mathfrak{M} \models$ *All PA that are B are PB.*

For the more interesting direction, suppose Directedness is valid for P on $\mathfrak{M}$. Then define

$$x \prec y \Leftrightarrow \textit{not } y \prec_P x$$

By construction $\prec$ has the No Reversal property for P. Thus we need only show it is compatible with P.
For this, suppose $x \prec_P y$, $x \not\prec y$. Then by construction $y \prec_P x$, and as such

there are A and B such that $x \in [\![\overline{P}A]\!], y \in [\![PA]\!]$ and $x \in [\![PB]\!], y \in [\![\overline{P}B]\!]$. Because of x, we have $\mathfrak{M} \models$ *Some* $\overline{P}A$ *are* PB. Since $\mathfrak{M}$ makes the No Reversal rule valid, this implies $\mathfrak{M} \models (PA, B, PB)$. But because of y, this does not hold. Contradiction, therefore $\prec$ is compatible with P.

5.3.2 No Cycles

While one could argue that Directedness ensures an adjective P corresponds to a certain direction, this in and of itself does not guarantee that it will correspond to an order with low and high sides. The direction can also be along a cycle: for example, consider a model with domain $\{x, y, z\}$ and the following interpretations

- $[\![A]\!] = \{x, y\}, [\![PA]\!] = \{y\}$
- $[\![B]\!] = \{y, z\}, [\![PB]\!] = \{z\}$
- $[\![C]\!] = \{z, x\}, [\![PC]\!] = \{x\}$

Here we have $x \prec_P y \prec_P z$, but also $z \prec_P x$. Nonetheless, $\prec_P$ has the No Reversal property for P.
We can guarantee acyclicity and hence prevent situations like this from arising by generalizing Directedness into the following rule.

$$\frac{\text{Some } \overline{P}A_1 \text{ are } PA_2 \qquad \ldots \qquad \text{Some } \overline{P}A_{n-1} \text{ are } PA_n}{(PA_1, A_n, PA_n)} \text{ No Cycles}$$

For another size-based example, consider a fantasy world containing giants, humans, dwarves and various hybrids of the three. Then if there exist a half-giant who is considered an average-sized giant but a tall human and a half-dwarf who is considered an average-sized human but a tall dwarf, the No Cycles rule lets us conclude that any half-giant-half-dwarf who is a tall giant is also a tall dwarf. While this appeals to similar underlying intuitions, it is nonetheless a conclusion which we would not be able to draw using No Reversal alone.
The aptness of calling this rule No Cycles is reinforced by the fact that it is valid exactly in those cases where $\prec_P$ is acyclic, which we see below.

5.3.6. THEOREM. *$\mathfrak{M}$ makes the No Cycles rule valid for P if and only if $\prec_P$ is acyclic (ie there are no elements $x_1, \ldots, x_n$ such that $x_1 \prec_P x_2 \prec_P \ldots \prec_P x_n \prec_P x_1$).*

PROOF: Suppose there are such elements. Then there are $A_1, \ldots, A_n$ such that $x_i \in [\![\overline{P}A_i]\!]$, $x_{i+1} \in [\![PA_i]\!]$ [and $x_1 \in [\![PA_n]\!]$].
Thus, $\mathfrak{M}$ makes the following true:

- Some $\overline{P}A_1$ are PA_n (witnessed by x_1)
- Some $\overline{P}A_2$ are PA_1 (witnessed by x_2)
- ...
- Some $\overline{P}A_n$ are PA_{n-1} (witnessed by x_n)

Let $B_1 = A_1, B_2 = A_n, B_3 = A_{n-1}, \ldots, B_n = A_2$. Then the above can be restated as

- Some $\overline{P}B_1$ are PB_2 (witnessed by x_1)
- Some $\overline{P}B_2$ are PB_3 (witnessed by x_n)
- ...
- Some $\overline{P}B_n$ are PB_1 (witnessed by x_2)

If No Cycles were valid for P on $\mathfrak{M}$, then $\mathfrak{M} \models (PB_1, B_n, PB_n)$. The latter has a counterexample in x_2, so No Cycles is not valid here.

For the other direction, suppose the rule is not valid on $\mathfrak{M}$. Then there are $A_1, \ldots, A_n$ such that "Some $\overline{P}A_1$ are PA_2", ..., "Some $\overline{P}A_{n-1}$ are PA_n" are all true on $\mathfrak{M}$ and (PA_1, A_n, PA_n) is not.
Thus, there are $x_1 \in [\![\overline{P}A_1]\!] \cap [\![PA_2]\!]$, $x_2 \in [\![\overline{P}A_2]\!] \cap [\![PA_3]\!]$, ..., $x_{n-1} \in [\![\overline{P}A_{n-1}]\!] \cap [\![PA_n]\!]$, $y \in [\![PA_1]\!] \cap [\![\overline{P}A_n]\!]$. By definition this gives

$$y \prec_P x_{n-1} \prec_P x_{n-2} \prec_P \ldots \prec_P x_2 \prec_P x_1 \prec_P y$$

5.3.3 The Consequences of No Cycles

While it is good to think of Acyclicity as the main thing characterized by the No Cycles rule, we shall see that in so doing it also characterizes some far more significant things, up to and including our main objective, the existence of an order with desirable properties.
The first step towards this is to obtain Transitivity. A lack of cycles means the relation $\prec_P$ can be extended into a transitive one. We show below that the No Cycles rule in fact characterizes the ability to do so while retaining No Reversal.

5.3.7. THEOREM. *Let $\mathfrak{M}$ be an acceptable model. Then it makes the No Cycles rule valid for P if and only if P is based on a transitive relation with the No Reversal property.*

PROOF: Suppose P is based on such a relation $\prec$. Suppose towards contradiction that $\mathfrak{M}$ does not make No Cycles valid for P. Then there are $x_1, \dots, x_n$ with $x_1 \prec_P x_2 \prec_P \dots \prec_P x_n \prec_P x_1$. Since $\prec$ is compatible with P and transitive, it follows that $x_2 \prec x_1$. However, since $\prec$ has the No Reversal property this implies $x_1 \not\prec_P x_2$. Contradiction.

For the other direction it suffices to note that when No Cycles is valid the transitive closure of $\prec_P$ has the No Reversal property. (This is an immediate corollary of Theorem 5.3.6: if $x \prec_P y \prec_P z$ then by Theorem 5.3.6 $z \not\prec_P x$.)

Note that any transitive relation on which P is based is an extension of the transitive closure of $\prec_P$. Also, P is always based on the transitive closure of $\prec_P$. Thus Theorem 5.3.7 could also be phrased as saying that No Cycles is valid iff this transitive closure has No Reversal for P. This is an important detail, which leads into the following conveniently phrased Lemma.

5.3.8. LEMMA. *Let $\mathfrak{M}$ be an acceptable model. Let $\prec_P^T$ be the transitive closure of $\prec_P$. Then $\mathfrak{M}$ makes No Cycles valid for P if and only if $\prec_P^T$ is a strict partial order (that is, a transitive, irreflexive relation) with the No Reversal property.*

As a mere partial order, $\prec_P^T$ is one sense greatly falls short of the weak order we would like to have. But in another sense it comes very close indeed. For we can close the distance by using the well-known Order-extension Principle, detailed below.

5.3.9. THEOREM. **(Order-extension Principle, Marczewski)** *Every strict partial order can be extended into a strict linear order.*

Here a strict linear order is a strict partial order which is also connected (meaning for all $x \neq y$ either $x < y$ or $y < x$). For a proof, see (Szpilrajn 1930).

Using the Order-extension Principle, it is straightforward to prove that the No Cycles rule corresponds to P being based on a strict linear order with No Reversal.

5.3.10. THEOREM. *Let $\mathfrak{M}$ be an acceptable model. Then it makes the No Cycles rule valid for P if and only if P is based on a strict linear order with the No Reversal property.*

PROOF: Suppose $\mathfrak{M}$ makes No Cycles valid for P. By Lemma 5.3.8 and the Order-extension Principle, $\prec_P^T$ can be extended into a strict linear order. Let $<$ be such an extension. Since $<$ is an extension of $x \prec_P y$, P is based on $<$.
As a strict linear order, $<$ is antisymmetric (left to the reader). Since $x \prec_P y$ implies $x \neq y$, this antisymmetry means it implies $y \not< x$. By contraposition, $y < x$ implies $x \not\prec_P y$, proving that $<$ has the No Reversal property for P.

The other direction is an immediate consequence of Theorem 5.3.7.

Showing that No Cycles corresponds to P being based on a weak order with No Reversal (our main goal) is now a matter of putting one and one together.

5.3.11. THEOREM. *Let $\mathfrak{M}$ be an acceptable model. Then it makes the No Cycles rule valid for P if and only if P is based on a weak order with No Reversal.*

PROOF: If P is based on a weak order with No Reversal, then No Cycles is valid for P by Theorem 5.3.7. If No Cycles is valid for P, then by Theorem 5.3.10 P is based on a strict linear order with No Reversal, which is also a weak order with No Reversal.

5.4 The Commensurability of Adjective Groups

Sometimes groups of gradable adjectives can be said to act on the same underlying order. For instance, this is true of the group 'big', 'large', 'immense', 'huge', 'giant', 'vast', 'enormous', 'tremendous', 'colossal', 'gigantic', and becomes false if we add 'expensive' to this group. When considering a group of adjectives, it is of interest to see if they are related in this way or not, and in particular to see if this can be determined based on their interpretations alone.
When a group of gradable adjectives is based on the same order in the same way, we will refer to this as *commensurability*, which for our purposes we define as follows:

5.4.1. DEFINITION. A set Δ of adjectives is *commensurable* if there is a single weak order $\prec$ such that for every $P \in \Delta$, $\prec$ satisfies No Reversal for P and P is based on $\prec$.

So long as the adjectives in Δ are individually based on weak orders, this notion has a number of potential uses. When vagueness is modelled in terms of potentially acceptable interpretations, one would want the group of such interpretations of a single adjective to be commensurate. When looking at adjectival modification, it would be interesting to look at those modifiers where the modified adjective is commensurate with the unmodified one, something we explore in Section 5.4.3.
Concerning evaluative adjectives, commensurability implies a great deal of interpersonal compatibility, appropriate for some and inappropriate for others. This is discussed in Section 5.4.2. Finally, commensurability can be generalized to a form of inverted commensurability to handle polar antonyms, which we shall do in Section 5.4.1.

So how can we characterize commensurability? In the previous section we have seen that the No Cycles rule characterizes the possibility of a weak order. To arrive at a weak order that fits all adjectives at the same time, we generalize the No Cycles rule to a version which does the same, creating the No Δ-Cycles rule.

5.4.2. Definition. Let Δ be a set of adjectives. The *No* $\Delta-$*Cycles* rule allows inferences of the following form

$$\frac{\varphi_1(A_1, A_2) \qquad \dots \qquad \varphi_{n-1}(A_{n-1}, A_n)}{\psi(A_1, A_n)} \text{ No } \Delta\text{-Cycles}$$

where each $\varphi(A_i, A_{i+1})$ is of the form "Some $\overline{P}A_i$ are QA_{i+1}" for some $P, Q \in \Delta$ and where $\psi(A_1, A_n)$ is of the form (PA_1, A_n, QA_n) for some $P, Q \in \Delta$.

The No $\Delta-$Cycles rule generalizes the earlier No Cycles rule by forbidding cycles using any combination of adjectives in Δ. That is, it deals with the relation $\prec_\Delta$, detailed below.

5.4.3. Definition. Let Δ be a set of adjectives. Then $x \prec_\Delta y$ iff $\exists P \in \Delta$: $x \prec_P y$.

If it is unclear why the allowable forms of $\varphi(A_i, A_{i+1})$ in the definition are the right way to go, the next lemma should provide some insight.

5.4.4. Lemma. *If $x_3 \prec_\Delta x_2 \prec_\Delta x_1$, then there are A_2, A_3 such that one of the allowable forms of $\varphi_2(A_2, A_3)$ is true.*

Proof: Let $x_3 \prec_P x_2 \prec_Q x_1$, with $P, Q \in \Delta$. Then there are A_2, A_3 such that $x_3 \in [\![\overline{P}A_3]\!], x_2 \in [\![PA_3]\!], x_2 \in [\![\overline{Q}A_2]\!], x_1 \in [\![QA_2]\!]$. Thus, x_2 shows that "Some $\overline{P}A_2$ are QA_3" is true.

The effect of the No $\Delta-$Cycles rule, then, is to forbid cycles in the combined induced relation $\prec_\Delta$.

5.4.5. Theorem. *An acceptable model $\mathfrak{M}$ makes No $\Delta-$Cycles valid if and only if $\prec_\Delta$ is acyclic on its domain.*

Proof: Analogous to Theorem 5.3.6.

Given the above, we can now pursue a strategy analogous to the one we used in Section 5.3.3. Thus, we first look at the transitive closure of $\prec_\Delta$.

5.4.6. Lemma. *Let $\mathfrak{M}$ be an acceptable model. Let $\prec_\Delta^T$ be the transitive closure of $\prec_\Delta$. Then $\mathfrak{M}$ makes No $\Delta-$Cycles valid if and only if $\prec_\Delta^T$ is a strict partial order (that is, a transitive, irreflexive relation) which has the No Reversal property for each $P \in \Delta$.*

PROOF: If $\prec_{\Delta}^{T}$ is a strict partial order, then $\prec_{\Delta}$ is acyclic. (For if $\prec_{\Delta}$ contains a cycle then taking the transitive closure closes that cycle so that some x has $x \prec_{\Delta}^{T} x$.) By Theorem 5.4.5, $\mathfrak{M}$ makes No $\Delta-$Cycles valid.
For the other direction, suppose $\mathfrak{M}$ makes No $\Delta-$Cycles valid. By Theorem 5.4.5, $\prec_{\Delta}$ is acyclic, and therefore irreflexive. Thus if there is some $x \prec_{\Delta}^{T} x$ then this was added by the transitive closure operation. But that is only possible if there are some $y_1, \ldots, y_n$ such that $x \prec_{\Delta} y_1 \prec_{\Delta} \ldots \prec_{\Delta} y_n$, which is exactly what is impossible given that $\prec_{\Delta}$ is acyclic. Therefore, $x \prec_{\Delta}^{T}$ is irreflexive.
It is by construction also transitive, and thus is a strict partial order. Having established that it is irreflexive, it is easily seen that $\prec_{\Delta}^{T}$ also has the No Reversal property for $P \in \Delta$. For if $x \prec_{\Delta}^{T} y \prec_{P} x$ for $P \in \Delta$ then $x \prec_{\Delta}^{T} x$, which would contradict irreflexivity.

As in Section 5.3.3, combining this with the Order-extension Principle leads to the desired result.

5.4.7. THEOREM. *Let $\mathfrak{M}$ be an acceptable model. Then it makes the No $\Delta-$Cycles rule valid if and only if there is a strict linear order $<$ which is compatible with all $P \in \Delta$ and which has the No Reversal property for all $P \in \Delta$.*

PROOF: Analogous to Theorem 5.3.10.

5.4.1 Polar antonyms

Where commensurable adjectives take up the same side of a spectrum, antonyms take up opposite sides of one. One key property of antonym pairs is a lack of overlap, i.e. $[\![PA]\!] \cap [\![QA]\!] = \emptyset$. Obviously this can be characterized by the following inference rule (proof left to the reader):

$$\frac{}{\text{All } PA \text{ are } \overline{Q}A}\ \text{No Overlap}$$

This however is not enough to get a good notion of antonymy, even when P and Q individually satisfy No Cycles. Even if they only occur outside P, instances of Q may still be restricted by a separate direction having nothing to do with P. To illustrate, consider the following example.
Let P be an interpretation of 'tall' where a person is in $[\![PA]\!]$ if his length is greater than the average lengths of people in $[\![A]\!]$. Opposed to this, let Q be the notion of 'odd-short', defined so that $[\![QA]\!]$ contains those people who are not in $[\![PA]\!]$ and also have a length which expressed in centimeters is an odd number. Here P and Q exclude each other and each can be based on a weak order, but Q is not an antonym of P in a sufficiently strong sense of the word.[2]

[2]It may be hard to see at first how Q can be based on a weak order, when it is obviously not convex relative to the standard length-based ordering. The answer of course is that Q is based

To ensure that antonyms are based on the same underlying order, we will use essentially the same method as we used to guarantee commensurability in general. Thus, we start by defining an analogue of the No Δ-Cycles rule for antonym pairs.
Note that the approach we take in the rest of this section does not imply a lack of overlap. Such a lack of overlap needs to be characterized separately per above.

5.4.8. Definition. The No $P, \overline{Q}$-Cycles rule allows all deductions of the following form

$$\frac{\varphi_1(A_1, A_2) \qquad \dots \qquad \varphi_{n-1}(A_{n-1}, A_n)}{\psi(A_1, A_n)} \text{No P,Q-Cycles}$$

where each $\varphi_i(A_i, A_{i+1})$ is of one of the following forms

- "Some $\overline{P}A_i$ are PA_{i+1}"
- "Some $\overline{P}A_i$ are $\overline{Q}A_{i+1}$"
- "Some QA_i are PA_{i+1}"
- "Some QA_i are $\overline{Q}A_{i+1}$"

and where $\psi(A_1, A_n)$ is of one of the following forms

- (PA_1, A_n, PA_n)
- $(PA_1, A_n, \overline{Q}A_n)$
- $(\overline{Q}A_1, A_n, PA_n)$
- $(\overline{Q}A_1, A_n, \overline{Q}A_n)$

Again, we add a lemma to help see why the definition above is the appropriate one.

5.4.9. Lemma. *Let $x \prec_{(P,Q)} y$ iff $x \prec_P y$ or $y \prec_Q x$. If $x_3 \prec_{(P,Q)} x_2 \prec_{(P,Q)} x_1$, then there are A_2, A_3 such that one of the acceptable forms of $\varphi_2(A_2, A_3)$ is true.*

Proof:

- If $x_3 \prec_P x_2 \prec_P x_1$, then there are A_2, A_3 such that $x_3 \in [\![\overline{P}A_3]\!]$, $x_2 \in [\![PA_3]\!]$, $x_2 \in [\![\overline{P}A_2]\!]$, $x_1 \in [\![PA_2]\!]$. Thus, x_2 shows that "Some $\overline{P}A_2$ are PA_3" is true.

on a different weak order, specifically the one where $x \prec y$ iff either x has an even length (when expressed in centimeters) and y an odd one or both have an odd one and y is taller than x. Relative to this ordering, it is Q which is convex while P is not, which can be seen as a reason to suspect incommensurability.

- If $x_2 \prec_Q x_3, x_2 \prec_P x_1$, then there are A_2, A_3 such that $x_3 \in [\![QA_3]\!]$, $x_2 \in [\![\overline{Q}A_3]\!]$, $x_2 \in [\![\overline{P}A_2]\!]$, $x_1 \in [\![PA_2]\!]$. Thus, x_2 shows that "Some $\overline{P}A_2$ are $\overline{Q}A_3$" is true.

- If $x_3 \prec_P x_2, x_1 \prec_Q x_2$, then there are A_2, A_3 such that $x_3 \in [\![\overline{P}A_3]\!]$, $x_2 \in [\![PA_3]\!]$, $x_2 \in [\![QA_2]\!]$, $x_1 \in [\![\overline{P}A_2]\!]$. Thus, x_2 shows that "Some QA_2 are PA_3" is true.

- If $x_2 \prec_Q x_3, x_1 \prec_Q x_2$, then there are A_2, A_3 such that $x_3 \in [\![QA_3]\!]$, $x_2 \in [\![\overline{Q}A_3]\!]$, $x_2 \in [\![QA_2]\!]$, $x_1 \in [\![\overline{P}A_2]\!]$. Thus, x_2 shows that "Some QA_2 are $\overline{Q}A_3$" is true.

The next step is to show that No $P, \overline{Q}-$Cycles corresponds with the acyclicity of a particular relation, in this case the relation $\prec_{(P,Q)}$ which is defined by letting $x \prec_{(P,Q)} y$ iff $x \prec_P y$ or $y \prec_Q x$.
Given the similar theorems we have done previously (Theorem 5.3.6 and Theorem 5.4.5), this should be straightforward enough to see.

5.4.10. Theorem. *Let $x \prec_{(P,Q)} y$ iff $x \prec_P y$ or $y \prec_Q x$. Then $\mathfrak{M}$ makes No $P, \overline{Q}$-Cycles valid if and only if there are no $x_1, \ldots, x_n$ such that $x_1 \prec_{(P,Q)} x_2 \prec_{(P,Q)} \cdots \prec_{(P,Q)} x_n \prec_{(P,Q)} x_1$.*

Proof: Left to the reader.

Next we argue that this implies that the transitive closure is a partial order, analogous to Lemma's 5.3.8 and 5.4.6.

5.4.11. Lemma. *Let $\mathfrak{M}$ be an acceptable model. Let $\prec^T_{(P,Q)}$ be the transitive closure of $\prec_{(P,Q)}$. Then $\mathfrak{M}$ makes No $P, \overline{Q}$-Cycles valid if and only if $\prec^T_{(P,Q)}$ is a strict partial order (that is, a transitive, irreflexive relation) which has the No Reversal property for P and for which the reverse has No Reversal for Q (that is, if $y \succ^T_{(P,Q)} x$ then not $x \prec_Q y$).*

Proof: Left to the reader.

The last step is again little more than an application of the Order-extension Principle.

5.4.12. Theorem. *Let $\mathfrak{M}$ be an acceptable model. Then it makes the No $P, \overline{Q}-$Cycles rule valid if and only if there is a strict linear order $<$ which is compatible with and has No Reversal for P and of which the reverse is compatible with and has No Reversal for Q.*

Proof: Analogous to Theorem 5.3.10.

5.4.2 Evaluative Adjectives and Personal Taste

As hinted at earlier, the notion of commensurability is strong enough so as to prohibit modeling certain forms of evaluative adjectives. For example, consider a situation where Alice feels movie x is a good movie and y isn't, while Bob feels just the opposite -a plausible enough scenario.
When we use P to represent Alice's notion of 'good', Q to represent Bob's and A to mean 'movie', this generalizes into the notion of P, Q*-Reversal*.

5.4.13. Definition. The model $\mathfrak{M}$ has P, Q*-Reversal* iff there are a noun A and elements x, y in the domain of $\mathfrak{M}$ such that

- $x \in [\![PA]\!]$, $y \in [\![\overline{P}A]\!]$
- $y \in [\![QA]\!]$, $x \in [\![\overline{Q}A]\!]$

This notion of reversal concerns multiple adjectives and a single noun, and thus is distinct from the earlier notion of reversal which concerned a single adjective and multiple nouns. Still, it is easy to see that the No Δ-Cycles rule prohibits this kind of reversal from taking place (where $P, Q \in \Delta$). For in the case of P, Q-Reversal we have $x \prec_P y \prec_Q x$ and thus $x \prec_\Delta y \prec_\Delta x$, which is prevented by Theorem 5.4.5.
Since No Δ-Cycles is a necessary condition for commensurability (Theorem 5.4.7), it follows that there can be no P, Q-Reversal when P and Q are commensurable. If this outcome seems to go too far, consider the way we defined commensurability. To be commensurable, P and Q must be based on the same weak order $\prec$. Being based on it here means that any distinction drawn by either adjective must be justifiable based solely on referring to $\prec$. Thus, a the situation above would necessarily imply $x \prec y \prec x$, which of course prevents $\prec$ from being a weak order. It may seem tempting to allow for P, Q-reversal by introducing a weaker notion of commensurability, one where the commensurable adjectives don't have to be based on the weak order $\prec$, but that relation merely has to have No Reversal (in the sense of Definition 5.3.2) relative to all of them. This however will not do, since under such a definition the trivial empty relation (where $x \not\prec y$ for all x, y) would always work, rendering any combination of adjectives commensurable.

The above does not mean that all forms of interpersonal disagreement are excluded. For there are situations where we would not want P, Q-Reversal to be a possibility. For example, consider differing notions of 'tall'. If Alice considered person x to be a tall basketball player and y not to be while Bob would consider things to be just the opposite, this would be odd indeed. Regardless of further details and despite some personal differences being acceptable, we would feel that at least one of them must have a defective notion of 'tall'.
The reason for this is exactly that we expect judgements of tallness to be based on the objective underlying 'taller than' relation. Thus, commensurability is still an

appropriate notion in situations where personal judgements are restricted in such a way, and serves as a useful way to distinguish such situations from 'anything goes' taste issues.

5.4.3 Strictenings and Loosenings

When we look at adjectival modifiers like 'very', 'really', 'extremely', 'perfectly', etcetera, we see a group that has in common that the modified meaning is a strictening of the original meaning, in at least the following way:

5.4.14. Definition. Let P, Q be adjectives interpreted by the acceptable model $\mathfrak{M}$. Then P is a *strictening* of Q (and Q a *loosening* of P) iff $[\![PA]\!] \subseteq [\![QA]\!]$ for all A.

This of course is characterizable by $\mathfrak{M}$ making the following rule valid for all A (proof left to the reader).

$$\frac{}{\text{All PA are QA}}\ \text{P,Q-Strictening}$$

Despite what we have seen thus far, two adjectives being commensurable does not guarantee that one is a strictening of the other. The simplest possible formal example of this is the following:

- $[\![A]\!] = [\![B]\!] = \{x, y\}$
- $[\![PA]\!] = \{x, y\}$, $[\![QA]\!] = \{y\}$
- $[\![PB]\!] = \{y\}$, $[\![QB]\!] = \{x, y\}$

Here the weak order where $x \prec y$ is compatible with P and Q and has No Reversal for both, proving they are commensurable. Still, neither one is a strictening of the other.

Conversely, being a strictening in this sense does not imply commensurability either, because it does not imply that P covers a particular side of Q. For example, while one possible strictening of 'not cold' is 'hot', another possible strictening is 'lukewarm' (in a sense that excludes hot)
For a formal example of this, consider the following situation:

- $[\![A]\!] = [\![B]\!] = \{x, y, z\}$
- $[\![QA]\!] = \{y, z\}$, $[\![QB]\!] = \{z\}$
- $[\![PA]\!] = \{y\}$, $[\![PB]\!] = \emptyset$

Here P is a strictening of Q, and P and Q individually make No Cycles valid. However, while we have $x \prec_Q y \prec_Q z$, we also have $z \prec_P x$. Thus, No Δ-Cycles is not valid, so P and Q are not commensurable.

5.4.4 The Logic of Commensurability

It is of some interest to note that commensurability is preserved under standard combination operators. The proof for this is almost trivial.

5.4.15. THEOREM. *Let Δ be a commensurable set of adjectives, with $P, Q \in \Delta$. Then the following sets are also commensurable:*

- $\Delta \cup \{P \wedge Q\}$, *where* $[\![(P \wedge Q)A]\!] := [\![PA]\!] \cap [\![QA]\!]$
- $\Delta \cup \{P \vee Q\}$, *where* $[\![(P \vee Q)A]\!] := [\![PA]\!] \cup [\![QA]\!]$

PROOF: Let Δ be commensurable. Let $\prec$ be a weak order which is compatible with every $P \in \Delta$ and has No Reversal for every such P. If $x \prec_{P \wedge Q} y$, then either $x \prec_P y$ or $x \prec_Q y$ (possibly both). It follows immediately that $\prec$ is compatible with and has No Reversal for $P \wedge Q$.
The same argument holds for $P \vee Q$.

Similarly, it is interesting to look at negation. For our purposes we will use an internal negation which corresponds to swapping P and $\overline{P}$. That is, we let $[\![(\neg P)A]\!] = [\![\overline{P}A]\!]$.
This is convenient because it creates a more workable way to deal with antonyms than the No P, Q-Cycles rule from Section 5.4.1. (Proof left to the reader.)

5.4.16. THEOREM. *No P, Q-Cycles is valid if and only if No Δ-Cycles is valid for $\Delta = \{P, \neg Q\}$.*

These notions and results let us express a couple of interesting conjectures about natural language.

5.4.17. CONJECTURE. *If No Cycles is true for P and f is a natural language modifier, then $f(P)$ is a boolean combination of commensurable strengthenings /loosenings of P.*

For a simple example, for the modifier 'very', *very*(P) is itself a commensurable strengthening of P.
This conjecture does not imply that $f(P)$ itself is commensurable with P. The most obvious example of this is 'not', which can be rendered as *not*$(P) = \neg P$.
Boolean combinations are allowed to handle more complex cases like a non-inclusive 'somewhat' (that is, a notion where 'somewhat warm' does not include things that are very warm). For that, something like *somewhat*$(P) = P \wedge \neg$*very*(P) could be used.

5.4.18. CONJECTURE. *If $\prec$ is a weak order, then any natural concept based on $\prec$ can be expressed as a boolean combination of elements of some Δ whose commensurability is proven by $\prec$.*

An easy example would be ‘lukewarm’, which could either be rendered as *somewhat*(*warm*), above, or as, say, $\neg cold \wedge \neg hot$. For an example of something that would be excluded by this conjecture, suppose we were to call someone ‘of prime length’ iff his length in centimetres (rounded up) is a prime number. In this case, being of prime length is not a natural concept, or more accurately is not a natural concept based on the “taller than” relation.
(It could still be considered a natural concept based on the relation where all people of prime length are considered greater than all other people. Thus, to completely distinguish natural and unnatural concepts one would have to distinguish natural and unnatural relations, which goes beyond the scope of this chapter.)

5.5 Vagueness, Distinguishability and The Sorites

Gradable adjectives such as ‘pretty’, ‘bright’ and ’expensive’ are key examples of vagueness. This makes it interesting to see what we can say about the vagueness of such gradable adjectives given what we have established so far. When looking at a specific gradable adjective P, an important notion to consider is when two objects are similar enough in terms of the underlying order associated with P that they cannot be distinguished between on that basis.
Traditionally, one approach to this is to simply check if they are already being distinguished by one of them being considered PA and the other $\overline{P}A$ for some A. Under this method, x and y would be $P-$indistinguishable if and only if neither $x \prec_P y$ nor $y \prec_P x$.
Another option would be to let y count as larger than x if and only if it is so in every linear order compatible with P, letting the two be indistinguishable if $x < y$ for some such orders and $y < x$ for others.
However these two approaches are not as distinct as they may appear, and rather are separated by nothing more than a transitive closure.

5.5.1. Theorem. *Let $\mathfrak{M}$ be an acceptable model making No Cycles valid for the adjective P. Let $x \prec y$ if and only if $x < y$ for every strict linear order $<$ which is an extension of the transitive closure of $\prec_P$.*
Then $\prec$ is itself the transitive closure of $\prec_P$.

Proof: It suffices to show that $x \prec y$ implies $x \prec_P^T y$, which we shall do by contraposition. Suppose $x \not\prec_P^T y$. Now if $y \prec_P^T x$ then every linear extension $<$ has $y < x$ and by irreflexivity $x \not< y$, which would give $x \not\prec y$ immediately. Thus we may assume $y \not\prec_P^T x$.
Construct $\prec^*$ by extending $\prec_P^T$ with $y \prec^* x$ and taking the transitive closure. Suppose $z \prec^* z$ for some z. As $\prec_P^T$ is irreflexive, this must be introduced by the

construction. Thus we have $z \prec_P^T y$, $x \prec_P^T z$.[3] But that would imply $x \prec_P^T y$, contradicting our original assumption. Hence there can be no such z, proving $\prec^*$ is irreflexive.
Thus $\prec^*$ is a strict partial order. Therefore by the Order-extension Theorem there is some strict linear order $<$ extending $\prec^*$. This $<$ is also an extension of $\prec_P^T$, yet has $y < x$ and therefore $x \not< y$. Thus, $x \not\prec y$.

Given this result, the following notion of $P-$indistinguishability seems as good as any.

5.5.2. Definition. For an adjective P, x and y are *$P-$indistinguishable*, denoted as $x \sim_P y$, iff $x \not\prec_P^T y$ and $y \not\prec_P^T x$.

This definition gives $\sim_P$ certain desirable if perhaps rather obvious properties.

5.5.3. Theorem. *If No Cycles is valid for P, then $\sim_P$ is reflexive and symmetric.*

Proof: *Symmetry*: By construction. *Reflexivity*: If $x \not\sim_P x$ then $x \prec_P^T x$, but by Theorem 5.3.6 this is exactly what is made impossible by No Cycles.

Note that transitivity is not generally a property of $\sim_P$, even if No Cycles is valid. This fact is related to the traditional issue known as the Sorites Paradox, which we discuss soon.

5.5.1 The Sorites Paradox and Incomplete Judgement

While $\sim_P$ is a good start and certainly a necessary part of any good indistinguishability relation, it cannot truly be enough. Since it is based on distinctions between P and $\overline{P}$, it will always lead to these two being sharply distinguishable. This immediately validates the inductive step of the Sorites paradox.

5.5.4. Lemma. **(Inductive Step)** *If $x \sim_P y$, then for all A $(x \in [\![PA]\!], y \in [\![A]\!]) \Rightarrow y \in [\![PA]\!]$.*

Proof: Suppose there is some A for which this is not the case. Then $x \in [\![PA]\!], y \in [\![\overline{P}A]\!]$. But then $y \prec_P x$ and therefore $x \not\sim_P y$.

As a result, a line of pairwise indistinguishable objects cannot have a P end and a $\overline{P}$ end.

5.5.5. Theorem. **(Sorites Paradox)** *There cannot be A, $x_1, x_2, \ldots, x_n$ such that*

[3] A longer path from z to y or x to z may exist, but $z \prec_P^T y$, $x \prec_P^T z$ then follows by transitivity of $\prec_P^T$.

- $x_1, \ldots, x_n \in [\![A]\!]$
- $x_1 \sim_P x_2 \sim_P \ldots \sim_P x_n$
- $x_1 \in [\![\overline{P}A]\!]$
- $x_n \in [\![PA]\!]$

PROOF: Contradiction by induction using the Inductive Step.

We can get around this problem by extending the indistinguishability relation. If $\sim$ is known only to be an extension of $\sim_P$, then $x \prec_P y$ does not imply $x \not\sim y$, so the inductive step does not necessarily hold.
This move leads to a couple of questions. One is how this extended $\sim$ should be obtained. Another is what makes it attractive to believe the Inductive Step is true. However, while these questions are natural and traditional enough, both of them turn out to in fact be poorly phrased.
On the first question we are wondering how to blur the overly sharp lines we possess, when the truth is we should not possess these sharp lines in the first place. The lack of vagueness here is due to the system's implicit assumption that for every object in A we possess a specific judgement of it being either PA or $\overline{P}A$. This of course is not generally appropriate when dealing with vagueness. We can loosen this requirement by moving to a different kind of model, the *weakly acceptable model*.

5.5.6. DEFINITION. A *weakly acceptable model* is a model where for every adjective P and noun A, $[\![PA]\!] \subseteq [\![A]\!]$, $[\![\overline{P}A]\!] \subseteq [\![A]\!]$.

The semantics for weakly acceptable models is the same as the semantics for acceptable models. The logic of weakly acceptable models is also known to us.

5.5.7. THEOREM. *Let $\vdash_W$ be the logic consisting of the rules of Figure 5.2.1 except for the last one. Then $\vdash_W$ is sound and complete on weakly acceptable models.*

For the proof, see Theorem 3.2.11.
Beyond the bare logic, we must also look at Directedness and No Cycles, as the latter in particular is a driving force behind most of our analysis. Here we can reproduce our key theorems by making only slight changes to our definitions. Note that when working with the acceptable models we have been working with so far there is no difference between these notions and the original ones.

5.5.8. DEFINITION. $\prec$ has *Strong No Reversal for P* if and only if

$$\forall x, y(x \prec y \Rightarrow (PAx \& Ay \Rightarrow PAy))$$

$x \precsim_P y$ iff one of the following holds:

- $x \prec_P y$
- For some A, $x \in [\![\overline{P}A]\!]$, $y \in [\![A]\!], y \notin [\![\overline{P}A]\!], y \notin [\![PA]\!]$.
- For some A, $y \in [\![PA]\!]$, $x \in [\![A]\!], x \notin [\![\overline{P}A]\!], x \notin [\![PA]\!]$.

$\prec_P$ is Strongly Acyclic iff there are no $x_1, \ldots, x_n$ such that

$$x_1 \prec_P x_2 \prec_P x_3 \prec_P \ldots \prec_P x_n \precsim_P x_1$$

5.5.9. Theorem. *For weakly acceptable models $\mathfrak{M}$, the following hold:*

1. *If there is a relation $\prec$ on the domain of $\mathfrak{M}$ which has Strong No Reversal for P and is compatible with P, then Directedness is valid on $\mathfrak{M}$.*
2. *If Directedness is valid on $\mathfrak{M}$ then there is a relation $\prec$ on the domain of $\mathfrak{M}$ which has No Reversal for P as is compatible with P.*
3. *No Cycles on $\mathfrak{M}$ is valid if and only if $\prec_P$ is strongly acyclic.*

Proof: For the second point, the relevant part of the proof of Theorem 4.3.5 continues to work for weakly acceptable models. For the first and third point, the proof can be obtained by making minimal changes to the proofs of Theorems 4.3.5 and 4.3.6, left to the reader.

Thus, we have dropped the assumption of complete judgement at a minimal cost. So how does the Sorites Paradox behave under Incomplete Judgement? We previously defined $\sim_P$ by letting $x \sim_P y$ iff $x \not\prec_P y$ and $y \not\prec_P x$. Under this definition, the Inductive Step is not valid and so the Paradox is avoided. However, this definition is perhaps too broad when working with weakly acceptable models. A tempting stricter alternative is to let $x \sim y$ iff $x \not\precsim_P y$ and $y \not\precsim_P x$, but that would be exactly the wrong thing to do, as it would amount to treating non-judgement as a third truth value treated exactly as sharply as the values we started out with, recreating that which we were trying to move away from.
Instead, we must take an approach where such unjudged cases are relevant, but it is still possible for a judged and an unjudged case to be observationally indistinguishable. Given an existing set of judgements, the broadest good way to do so would be by the $\sim^*_P$ defined below.

5.5.10. Definition. $x \sim^*_P y$ if and only if one of the following is true.

- $x \sim_P y$ and $x \precsim_P y$
- $x \not\precsim_P y$ and $y \not\precsim_P x$

To see that this is broad enough, let $[\![A]\!] = \{x_1, x_2, x_3\}$, $[\![\overline{P}A]\!] = \{x_1\}$, $[\![PA]\!] = \{x_3\}$. Then (if there are no further relevant extensions) we get $x_1 \sim_P^* x_2 \sim_P^* x_3$. Clearly, then, no good rephrasing of the Inductive Step will be valid for $\sim_P^*$ are another sufficiently broad indistinguishability relation.
So why does the Inductive Step seem true? As noted earlier, the answer is that this question too is poorly phrased. In contexts where a series of x_i is equidistantly separated using a minute measure, such as the traditional individual grains of sand in a heap or individual hairs on a head, the matter of whether an adjective applies must be seen as one of judgement rather than absolute truth. Consequently, the Inductive Step is not something which is true or false of the world, but rather a standard against which ones judgements are held. On this view, the paradox is that sometimes there is no way for a series of judgements to both meet this standard and judge the obvious endpoints appropriately.

5.6 Appendix: Nouns as Contexts

Contrary to how we use it in the rest of this chapter, the formal system we work with can support an interpretation where every context has a matching noun A where $[\![PA]\!]$ is the extension of P relative to that context for any P. Important existing work in this area, specifically (van Benthem 1991), leads to the conclusion that under the right set of circumstances this makes $\prec_P$ itself a weak order. In this section we shall briefly repeat the key steps of this work in our own terms.

To enforce that a noun is available to match any context, we create the notion of the fully detailed model, which is a model where each finite context (subset of the domain) can be referred to with an existing noun.

5.6.1. Definition. The model $\mathfrak{M}$ is *fully detailed* if for every finite subset X of the domain of $\mathfrak{M}$, there is some noun A such that $[\![A]\!] = X$.

For the sake of convenience, we combine the constraints of Upward Difference and Downward Difference from (van Benthem 1991) into a single rule, that of Conservation of Significance.[4]

$$\frac{\text{Some } \overline{P}A \text{ are B} \qquad \text{Some PA are B}}{\text{Some B are } \overline{P}B\text{, Some B are PB}}\ \text{Conservation of Significance}$$

The Conservation of Significance rule can be understood as appealing to the notion of significant difference: if x is $\overline{P}A$ and y is PA, then y is significantly more P than x. Now if both of these are also in B, then B contains covers a

[4]On fully detailed models, it is easy enough to see that Conservation of Significance does indeed correspond exactly to this combination: where x, y is the difference pair in question, simply go 'Downward' from A to $\{x, y\}$ and then 'Upward' from $\{x, y\}$ to B.

significant range. The idea now is that while P might divide B based on some other distinction which is even more important than the difference between x and y, it is not acceptable for it to not divide B at all, as that would wrongly suggest B does not contain any significant differences in $P-$ness within it.
The astute reader may have noticed that our definition of $\prec_P$ is not precisely the same as the one used in (van Benthem 1991). However, under the circumstances we are working with the two do generate exactly the same relation

5.6.2. LEMMA. *Let $\mathfrak{M}$ be a fully detailed model which makes No Cycles and Conservation of Significance valid for P. If $x \prec_P y$ and $[\![A]\!] = \{x, y\}$, then $[\![PA]\!] = \{y\}$.*

PROOF: Because of Conservation of Significance, $[\![PA]\!]$ is either $\{x\}$ or $\{y\}$. Because of No Cycles, it cannot be $\{x\}$.

We can now prove the theorem we are after, which is largely a matter of straightforward case-checking.

5.6.3. THEOREM. *Let $\mathfrak{M}$ be a fully detailed model which makes No Cycles and Conservation of Significance valid for P. Then $\prec_P$ is a weak order.*

PROOF:
Anti-Symmetry: Trivial.
Transitivity: Suppose $x \prec_P y \prec_P z$. Let $[\![A]\!] = \{x, y\}$, $[\![B]\!] = \{y, z\}$, $[\![C]\!] = \{x, y, z\}$. Now consider the eight a priori possibilities for $[\![PC]\!]$.

- $[\![PC]\!] = \{x, y, z\}$ and $[\![PC]\!] = \emptyset$ are excluded because of Conservation of Significance. (The lemma gives us *Some $\overline{P}A$ are C*, *Some PA are C*.)
- $[\![PC]\!] = \{x\}$ and $[\![PC]\!] = \{x, y\}$ are excluded because of No Cycles. (They would imply $z \prec_P x$.)
- $[\![PC]\!] = \{y\}$ and $[\![PC]\!] = \{x, z\}$ are excluded because of No Cycles. (They would respectively imply $z \prec_P y$ and $y \prec_P x$.)
- $[\![PC]\!] = \{z\}$ and $[\![PC]\!] = \{y, z\}$ are left as the only possibilities, and both imply $x \prec_P z$.

Almost-Connectedness: Suppose $x \prec_P y$. Let $[\![A]\!] = \{x, y\}$, $[\![B]\!] = \{x, y, z\}$. Again, consider the eight a priori possibilities for $[\![PB]\!]$.

- $[\![PB]\!] = \{x, z\}$, $[\![PB]\!] = \{x, y, z\}$ and $[\![PB]\!] = \emptyset$ are excluded for reasons seen above.
- $[\![PB]\!] = \{x\}$ and $[\![PB]\!] = \{x, y\}$ imply $z \prec_P x$ and therefore by transitivity $z \prec_P y$.
- $[\![PB]\!] = \{y\}$ implies $z \prec_P y$.
- $[\![PB]\!] = \{z\}$ and $[\![PB]\!] = \{y, z\}$ imply $x \prec_P z$.

Chapter 6

Making the Right Exceptions

6.1 Introduction

Discussions often end before the issues that started them have been resolved. In the eighties and nineties of the previous century default reasoning was a hot topic in the field of logic & AI. The result of this discussion was not one single theory that met with general agreement, but a collection of alternative theories, each with its merits, but none entirely satisfactory. This paper aims to give a new impetus to this discussion.

The issue is the logical behavior of sentences of the form

S's are normally P

Such sentences function as default rules: when you are confronted with an object with property S, and you have no evidence to the contrary, you are legitimized to assume that this object has property P.

The 'evidence to the contrary' can vary. Sometimes it simply consists in the empirical observation that the object concerned is in fact an exception to the rule. On other occasions the evidence may be more indirect. Consider:

premise 1	*A's are normally E*
premise 2	*S's are normally not E*
premise 3	*S's are normally A*
premise 4	*c is A and c is S*
by default	*c is not E*

This is a case of conflicting defaults.[1] At first sight one might be tempted to draw both the conclusion that *c is E* (from premises 1 and 4) and that *c is not E* (from premises 2 and 4) , and maybe on second thought to draw neither. But the

[1]If a concrete example is wanted, substitute '*adult*' for A, '*employed*' for E, and '*student*' for S.

third premise states that objects with the property S normally have the property A as well. So, apparently, *normal* S's are *exceptional* A's, as the rule that *A's are normally E* does not hold for them. In other words, only the S-defaults apply to c. So, presumably, c is not E.

Default reasoning has been formalized in various ways, and within each of the existing theoretical frameworks a number of strategies have been proposed to deal with conflicting defaults — many of them rather *ad hoc*. In the following we will focus on two of these frameworks, Circumscription, and Inheritance Nets[2], and implement a new, principled strategy to deal with conflicting rules in each of these.

6.2 Naive Circumscription

Within the circumscriptive approach a sentence of the form *S's are normally P* is represented by a formula of the form

$$\forall x((Sx \wedge \neg Ab_{SxPx}\, x) \rightarrow Px)$$

Here $Ab_{SxPx}\, x$ is a one place predicate. The subscript '$SxPx$' serves as an index, indicating the rule concerned. If an object a satisfies the formula $Ab_{SxPx}\, x$, this means that a is an *abnormal* object with respect to this rule.

More generally, let $\mathcal{L}_0$ be a language of *monadic* first order logic. With each pair $\langle \varphi(x), \psi(x) \rangle$[3], we associate a new one-place predicate $Ab_{\varphi(x)\psi(x)}$, thus obtaining the first order language $\mathcal{L}$.

A *default rule* is a formula of $\mathcal{L}$ of the form

$$\forall x((\varphi(x) \wedge \neg Ab_{\varphi(x)\psi(x)}\, x) \rightarrow \psi(x))$$

Here, $\varphi(x)$ and $\psi(x)$ must be formulas of $\mathcal{L}_0$ that are quantifier-free and in which no individual constant occurs. The formula $\varphi(x)$ is called the *antecedent* of the rule, $Ab_{\varphi(x)\psi(x)}\, x$ is its *abnormality clause*, and $\psi(x)$ its *consequent*. Again, the index $\varphi(x)\psi(x)$ is there just to indicate that it concerns the abnormality predicate of the rule with antecedent $\varphi(x)$ and consequent $\psi(x)$. When it is clear which variable is at stake we will write $Ab_{\varphi\psi}$ rather than $Ab_{\varphi(x)\psi(x)}$. And often we will shorten '$\forall x((\varphi(x) \wedge \neg Ab_{\varphi\psi}\, x) \rightarrow \psi(x))$' further to

$$\forall x(\varphi(x) \rightsquigarrow \psi(x))$$

Since it is clear from the antecedent and the consequent of a default rule what the abnormality clause is, this should not cause confusion.[4]

[2] See (McCarthy 1987),(McCarthy 1990),(Horty, Thomason and Touretzy 1990)

[3] Notation: we write $\varphi(x)$ to denote a formula φ of $\mathcal{L}_0$ in which (at most) the variable x occurs freely

[4] Some readers may not like the fact that in this set up the formulas $\forall x(Sx \rightsquigarrow Px)$ and

In ordinary logic, for an argument to be valid, the conclusion must be true in *all* models in which the premises are true. The basic idea underlying circumscription is that not all models of the premises matter but only the most normal ones — only the ones in which the extension of the abnormality predicates is minimal given the information at hand. Formally:

6.2.1. DEFINITION.

(i) Let $\mathfrak{A} = \langle \mathcal{A}, \mathcal{I} \rangle$ and $\mathfrak{A}' = \langle \mathcal{A}', \mathcal{I}' \rangle$ be two models with the following properties:

  (a) $\mathcal{A} = \mathcal{A}'$

  (b) for all individual constants c, $\mathcal{I}(c) = \mathcal{I}'(c)$

  (c) for all predicates $Ab_{\varphi\psi}$, $\mathcal{I}(Ab_{\varphi\psi}) \subseteq \mathcal{I}'(Ab_{\varphi\psi})$

  Then $\mathfrak{A}$ *is at least as normal as* $\mathfrak{A}'$. If $\mathfrak{A}$ is at least as normal as $\mathfrak{A}'$, but $\mathfrak{A}'$ is not at least as normal as $\mathfrak{A}$, then $\mathfrak{A}$ *is more normal than* $\mathfrak{A}'$.

(ii) Let $\mathfrak{C}$ be a class of models. Then $\mathfrak{A} = \langle \mathcal{A}, \mathcal{I} \rangle$ is an *optimal* model in $\mathfrak{C}$ iff $\mathfrak{A} \in \mathfrak{C}$ and there is no model in $\mathfrak{C}$ that is more normal than $\mathfrak{A}$.

(iii) Let Δ be a set of sentences. Then $\Delta \models_c \varphi$ iff φ is true in all optimal models of Δ.

If $\Delta \models_c \varphi$, we say that φ follows by circumscription from Δ. Here is an example of an argument for which this is so.

premise 1	*Adults normally have a bank account*
premise 2	*Adults normally have a driver's license*
premise 3	*John is an adult*
premise 4	*John does not have a driver's license*
by default	*John is an adult with a bank account*

This can be formalized as

$\forall y(Sy \rightsquigarrow Py)$ are not logically equivalent, because they contain different abnormality predicates. We could remedy this defect by introducing the same abnormality predicate $Ab_{\varphi(\cdot)\psi(\cdot)}$ for all pairs $\langle \varphi(x), \psi(x) \rangle$, independent of the free variable x occurring in $\varphi(x)$ and $\psi(x)$. Here '$\cdot$' refers to a symbol that does not belong to the vocabulary of $\mathcal{L}_0$, and by $\varphi(\cdot)$, we mean the expression that one obtains from $\varphi(x)$ by replacing each free occurrence of x by an occurrence of $\cdot$.

Some readers may insist that on top of this we should enforce that whenever $\varphi(x)$ is logical equivalent to $\chi(x)$, and $\psi(x)$ to $\theta(x)$, $\forall x(\varphi(x) \rightsquigarrow \psi(x))$ gets equivalent to $\forall x(\chi(x) \rightsquigarrow \theta(x))$. This can be done by stipulating that we are only interested in models that assign the same extension to $Ab_{\varphi(\cdot)\psi(\cdot)}$ and $Ab_{\chi(\cdot)\theta(\cdot)}$ if $\varphi(x)$ is logical equivalent to $\chi(x)$ and $\psi(x)$ to $\theta(x)$. However, for our purposes, we can keep things simple.

premise 1	$\forall x((Ax \land \neg Ab_{AB}\, x) \rightarrow Bx)$
premise 2	$\forall x((Ax \land \neg Ab_{AD}\, x) \rightarrow Dx)$
premise 3	Aj
premise 4	$\neg Dj$
by circumscription	Bj

This example illustrates why the abnormality predicates have a double index referring to both the antecedent and the consequent of the rule, rather than a single one referring to just the antecedent. It is not sufficient to distinguish between normal and abnormal A's, and formalize a sentence like *Adults normally have a bank account* as $\forall x((Ax \land \neg Ab_A\, x) \rightarrow Bx)$. The distinction has to be more fine grained. An object with the property A can be a normal A in some respects and an abnormal A in other. Even though John is an abnormal adult in not having a driver's license, he is a normal adult in having a bank account, or at least we want to be able to conclude by default that he is. If we had formalized the argument in the following way, we would not have gotten very far.

premise 1	$\forall x((Ax \land \neg Ab_A\, x) \rightarrow Bx)$
premise 2	$\forall x((Ax \land \neg Ab_A\, x) \rightarrow Dx)$
premise 3	Aj
premise 4	$\neg Dj$

Let us now look at the case of conflicting defaults introduced at the end of section 1. The formalized version looks like this:

premise 1	$\forall x(Ax \rightsquigarrow Ex)$
premise 2	$\forall x(Sx \rightsquigarrow \neg Ex)$
premise 3	$\forall x(Sx \rightsquigarrow Ax)$
premise 4	$Ac \land Sc$
by circumscription	$\neg Ea$

Unfortunately, in this simple set up the conclusion $\neg Ea$ does not follow from the premises. We find two kinds of optimal models: in some the sentences $\neg Ab_{SA}\, c$, $\neg Ab_{S\neg E}\, c$, and $Ab_{AE}\, c$ hold, which is fine, but in the other the sentences $\neg Ab_{SA}\, c$, $Ab_{S\neg E}\, c$, and $\neg Ab_{AE}\, c$ are true.

Recall that in the informal discussion of this example it was suggested that the three default rules involved together imply that objects with property S are *exceptional* A's; normal A's have the property E, but normal S's don't, even though normal S's do have property A.

In the next section we will see how one can enforce that in all models in which these three defaults hold, also the formula $\forall x(Sx \rightarrow Ab_{AE}\, x)$ wil be true. Once we have this, the only optimal models will be models in which $\neg Ab_{SA}\, c$, $\neg Ab_{S\neg E}\, c$, and $\neg Ab_{AE}\, c$ are true. Which means that the conclusion follows.

6.3 Exemption and Inheritance

In the following, we will distinguish two kinds of rules, rules that allow for exceptions and rules that do not allow for exceptions. So far we only talked about the first kind, but we also want to discuss the second kind. In order to do so, sentences of the form $\forall x(\varphi(x) \rightarrow \psi(x))$ can get a special status as *strict rules*. These strict rules are to be distinguished from universal sentences that are only accidentally true, and they will be treated differently.[5]

The general set up will be this: Let Σ be a set of default and strict rules and Π be a set of sentences. Think of $I = \langle \Sigma, \Pi \rangle$ as the *information* of some agent at some time, where Σ is the set of rules the agent is acquainted with, and Π the agent's factual information. We correlate with I a pair $\langle \mathcal{U}_I, \mathcal{F}_I \rangle$, and call this the (information) *state generated by I*. $\mathcal{U}_I$ is called the *universe* of the state. The elements of $\mathcal{U}_I$ are models of Σ, but not all models of Σ are allowed. The universe $\mathcal{U}_I$ must satisfy some additional *constraint* that will be discussed below. $\mathcal{F}_I$ consists of all models in $\mathcal{U}_I$ that are models of Π.

In this set up we can define validity as follows:

$$\Sigma, \Pi \models_d \varphi \text{ iff for all optimal models } \mathfrak{A} \in \mathcal{F}_I, \mathfrak{A} \models \varphi$$

Read '$\Sigma, \Pi \models_d \varphi$' as '$\varphi$ follows by default from Σ and Π'.

Before we can turn to a discussion of the constraint, we need to introduce some technical notions .

6.3.1. Definition. (i) Suppose $\mathfrak{A} \models \forall x(\varphi(x) \rightsquigarrow \psi(x))$, and let d be an element of the domain of $\mathfrak{A}$. Then *d complies with* $\forall x(\varphi(x) \rightsquigarrow \psi(x))$ (in $\mathfrak{A}$) iff d does not satisfy $Ab_{\varphi\psi}x$.

(ii) Let Σ be a set of rules, and let d be some element of the domain of some model $\mathfrak{A}$ of Σ. Then *d complies with* Σ (in $\mathfrak{A}$) iff d complies with all the default rules in Σ.

So, if an object satisfying $\varphi(x)$ complies with $\forall x(\varphi(x) \rightsquigarrow \psi(x))$, it will also satisfy $\psi(x)$. But notice that the definition allows for the following situations:

- The object d complies with $\forall x(\varphi(x) \rightsquigarrow \psi(x))$, but d does not satisfy $\varphi(x)$.
- The object d satisfies $\varphi(x)$ and $\psi(x)$, but d does not comply with $\forall x(\varphi(x) \rightsquigarrow \psi(x))$.

We will present examples later on. For now, just take 'comply' as a technical term.

[5]It is tempting to introduce a necessity operator in the object language to distinguish rules form accidental statements, but we resist this temptation, and only make the distinction at a meta-level.

6.3.1 The Exemption Principle

The constraint we will impose on $\mathcal{U}_I$ is motivated by the following minimal requirement.

> If the *only* factual information about some object is that it has property P, it must be valid to infer by default that this object complies with all the default rules for objects with property P.

What would be the use of these rules if they would not at least allow this inference?

It may seem easy to satisfy this requirement, but it is not.

6.3.2. DEFINITION.

(i) An exemption clause is a formula of the form $\forall x(\varphi(x) \to \bigvee_{\delta\in\Delta} Ab_\delta\, x)$, for Δ a set of default rules.[6]

(ii) Let Σ be a set of rules. $\Sigma^{\varphi(x)}$ is the set of rules in Σ with antecedent $\varphi(x)$.

(iii) The exemption clause $\forall x(\varphi(x) \to \bigvee_{\delta\in\Delta} Ab_\delta\, x)$ is an exemption clause *for* Σ iff $\Sigma \models \forall x(\varphi(x) \to \bigvee_{\delta\in\Delta\cup\Sigma^{\varphi(x)}} Ab_\delta\, x)$.

To see how these definitions work, consider again

$$\Sigma = \{\forall x(Ax \rightsquigarrow Ex), \forall x(Sx \rightsquigarrow \neg Ex), \forall x(Sx \rightsquigarrow Ax)\}$$

Here $\Sigma^{Sx} = \{\forall x(Sx \rightsquigarrow Ax), \forall x(Sx \rightsquigarrow \neg Ex)\}$. Let $\Delta = \{\forall x(Ax \rightsquigarrow Ex)\}$. Clearly, there is no model such that some object in its domain satisfies Sx and complies with $\Delta \cup \Sigma^{Sx}$. So,

$$\Sigma \models \forall x(Sx \to \bigvee_{\delta\in\Delta\cup\Sigma^{Sx}} Ab_\delta\, x)$$

By *(iii)* above this means that $\forall x(Sx \to \bigvee_{\delta\in\Delta} Ab_\delta\, x)$, i.e. $\forall x(Sx \to Ab_{AE}x)$, is an exemption clause for Σ, the idea being that objects with property S are, so to speak, *exempted* from the rule that A's are normally E.

The word 'exempted' suggests that default rules are some kind of normative rules. Indeed, often it is helpful to think of them that way. The use of the word 'normally', already suggests that we are dealing with a kind of norms here. To count as a normal S, S's must be A, and to count as a normal A, A's must be E, but here an exception is made for the S's. S's must be A, but they do not have to be E, they are not subjected to this rule. Actually, they must be not E.

In the following definition it is made explicit for any set of rules Σ which kinds of objects are exempted from which rules in Σ.

[6]Where δ is a default rule, Ab_δ is the abnormality clause of δ. By definition, if $\Delta = \emptyset$, $\bigvee_{\delta\in\Delta} Ab_\delta\, x = \bot$.

6.3.3. Definition. Let Σ be a set of rules, and let Π be an arbitrary set of formulas.

(i) The *exemption extension* $\Sigma^{\in}$ of Σ is given by

$$\Sigma^{\in} = \bigcup_{n\in\omega} \Sigma^{\in}_n$$

where $\Sigma^{\in}_0 = \Sigma$ and $\Sigma^{\in}_{n+1} = \Sigma^{\in}_n \cup \{\varphi \mid \varphi \text{ is an exemption clause for } \Sigma^{\in}_n\}$

(ii) The state generated by $I = \langle \Sigma, \Pi \rangle$ is the state $\langle \mathcal{U}_I, \mathcal{F}_I \rangle$ given by

(a) $\mathfrak{A} \in \mathcal{U}_I$ iff $\mathfrak{A}$ is a model of $\Sigma^{\in}$;

(b) $\mathcal{F}_I$ consists of all models in $\mathcal{U}_I$ that are models of Π.

Notice that $\Sigma^{\in}$ has the following property, which we will call the *Exemption Principle*.

$$\text{If } \Sigma^{\in} \models \forall x(\varphi(x) \rightarrow \bigvee_{\delta\in\Delta\cup\Sigma^{\varphi(x)}} Ab_\delta\, x), \text{ then } \Sigma^{\in} \models \forall x(\varphi(x) \rightarrow \bigvee_{\delta\in\Delta} Ab_\delta\, x)$$

In fact $\Sigma^{\in}$ is the weakest extension of Σ with this property.

6.3.4. Proposition (Minimal requirement).
Suppose $\forall x(\varphi(x) \rightsquigarrow \psi(x)) \in \Sigma$. *Then* $\Sigma, \{\varphi(c)\} \models_d \psi(c)$.

Proof: Let $\langle \mathcal{U}_I, \mathcal{F}_I \rangle$ be the state generated by $I = \langle \Sigma, \{\varphi(c)\} \rangle$. It suffices to show that every optimal model in $\mathcal{F}_I$ has the property that the object named c complies with $\Sigma^{\varphi(x)}$. If $\mathcal{F}_I = \emptyset$, this holds trivially. Suppose $\mathcal{F}_I \neq \emptyset$. Consider any model $\mathfrak{A} = \langle \mathcal{A}, \mathcal{I} \rangle$ in $\mathcal{F}_I$ in which the object $\mathcal{I}(c)$ does not comply with $\Sigma^{\varphi(x)}$. We will show that $\mathfrak{A}$ is not optimal.

Let Δ be the set of defaults in $\Sigma^{\in}$ with which $\mathcal{I}(c)$ complies. Apparently, $\Sigma^{\in} \not\models \forall x(\varphi(x) \rightarrow \bigvee_{\delta\in\Delta} Ab_\delta\, x)$. By the exemption principle this means that $\Sigma^{\in} \not\models \forall x(\varphi(x) \rightarrow \bigvee_{\delta\in\Delta\cup\Sigma^{\varphi(x)}} Ab_\delta\, x)$. Hence, there exists a model $\mathfrak{A}' = \langle \mathcal{A}', \mathcal{I}' \rangle$ in $\mathcal{U}_I$ such that some element d_0 in $\mathcal{A}'$ satisfies $(\varphi(x) \wedge \neg \bigvee_{\delta\in\Delta\cup\Sigma^{\varphi(x)}} Ab_\delta\, x)$.
Now, let $\mathfrak{A}'' = \langle \mathcal{A}'', \mathcal{I}'' \rangle$ be defined as follows:
– $\mathcal{A}'' = \mathcal{A}$;
– For individual constants a, $\mathcal{I}''(a) = \mathcal{I}(a)$;
– For P an ordinary predicate or an abnormality predicate,
if $d \neq \mathcal{I}(c)$, then $d \in \mathcal{I}''(P)$ iff $d \in \mathcal{I}(P)$, and
if $d = \mathcal{I}(c)$, then $d \in \mathcal{I}''(P)$ iff $d_0 \in \mathcal{I}'(P)$.
Consider any quantifier-free formula $\theta(x)$ in which no individual constant occurs. Clearly, if $d \neq \mathcal{I}''(c)$, then d satisfies $\theta(x)$ in $\mathfrak{A}''$ iff d satisfies $\theta(x)$ in $\mathfrak{A}$, while $\mathcal{I}''(c)$ satisfies $\theta(x)$ in $\mathfrak{A}''$ iff $I'(c)$ satisfies $\theta(x)$ in $\mathfrak{A}'$.

Given that all sentences of $\Sigma^{\in}$ are of the form $\forall x\theta(x)$ with θ as described, $\mathfrak{A}''$ will be a model of $\Sigma^{\in}$. And clearly, $\mathfrak{A}''$ is more normal than $\mathfrak{A}$. Therefore $\mathfrak{A}$ is not optimal.

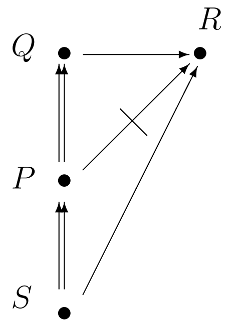
6.3.2 The Inheritance Property

On the face of it the exemption principle is not very strong. But it is amazing to see its consequences. One is the *inheritance property*, which in its simplest form runs as follows:

Let Σ be a set of rules. Suppose that $\Sigma^{\in} \models \forall x(\varphi(x) \rightsquigarrow \psi(x))$ and $\Sigma^{\in} \models \forall x(\psi(x) \to Ab_{\chi\theta}\, x)$. Then $\Sigma^{\in} \models \forall x(\varphi(x) \to Ab_{\chi\theta}\, x)$

To see how this works, consider the theory Σ consisting of the following five rules

$$\begin{array}{l} \forall x(Qx \rightsquigarrow Rx) \\ \forall x(Px \to Qx) \\ \forall x(Px \rightsquigarrow \neg Rx) \\ \forall x(Sx \to Px) \\ \forall x(Sx \rightsquigarrow Rx) \end{array}$$



Consider the first three rules, and notice that the exemption principle enforces that $\forall x(Px \to Ab_{QR}\, x) \in \Sigma^{\in}$.[7] Now, the inheritance principle yields that $\forall x(Sx \to Ab_{QR}\, x) \in \Sigma^{\in}$. By applying the the exemption principle to the last three rules we also have $\forall x(Sx \to Ab_{P\neg R}\, x) \in \Sigma^{\in}$.

So, all S's are P's but all S's are exceptional P's because they normally have the property R whereas P's normally do not have property R. The P's are exceptional Q's because Q's normally do have the property R. Now, does this make the S's normal Q 's? No! The S's neither count as normal P's nor as normal Q's. Exceptions to exceptions are not normal. The S's are doubly exceptional Q's rather than normal Q's. (Would you call a flying penguin a normal bird?)

The example illustrates the fact that it is possible for an object not to comply with a rule whereas both the antecedent and the consequent of the rule hold for it. Objects with the property S do not comply with the rule $\forall x(Qx \rightsquigarrow Rx)$, but in optimal circumstances they will have both the properties Q and R.

We will now prove a general form of the inheritance property.

6.3.5. PROPOSITION (INHERITANCE PROPERTY).
Let $\langle \mathcal{U}_I, \mathcal{F}_I \rangle$ be the state correlated with with the information $I = \langle \Sigma, \Pi \rangle$. Let $\Delta \subseteq \Sigma$ be a set of default rules.

[7] If a proof is wanted: Take $\Sigma = \{\forall x(Qx \rightsquigarrow Rx), \forall x(Px \to Qx), \forall x(Px \rightsquigarrow \neg Rx)\}$ and $\Delta = \{\forall x(Qx \rightsquigarrow Rx)\}$. Note that $\Sigma \models \forall x(Px \to \bigvee_{\delta \in \Delta \cup \Sigma^{Px}} Ab_\delta\, x)$, or simply put $\Sigma \models \forall x(Px \to (Ab_{QR} \vee Ab_{P\neg R}\, x))$; apply the exemption principle to find that $\Sigma^{\in} \models \forall x(Px \to Ab_{QR} x)$.

Suppose

$$(a)\ \Sigma^{\in} \models \forall x(\varphi(x) \rightsquigarrow \psi(x))\ \ and\ (b)\ \Sigma^{\in} \models \forall x(\psi(x) \to \bigvee_{\delta\in\Delta} Ab_\delta\, x)$$

Then

$$\Sigma^{\in} \models \forall x(\varphi(x) \to \bigvee_{\delta\in\Delta} Ab_\delta\, x)$$

Proof: By first-order logic alone, it is trivially true that

$$\Sigma^{\in} \models \forall x(\varphi(x) \to (\psi(x) \vee \neg\psi(x)))$$

Given (a), objects that satisfy $\varphi(x)$ and $\neg\psi(x)$ will also satisfy $Ab_{\varphi\psi}\, x$. Thus, the above statement remains true when we replace $\neg\psi(x)$ with $Ab_{\varphi\psi}\, x$. Similarly, given (b) we can replace $\psi(x)$ in the formula above with $\bigvee_{\delta\in\Delta} Ab_\delta\, x$ while keeping the statement true. This gives us

$$\Sigma^{\in} \models \forall x(\varphi(x) \to (\bigvee_{\delta\in\Delta} Ab_\delta\, x \vee Ab_{\varphi\psi}\, x))$$

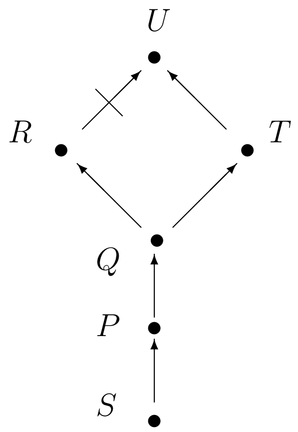
Given the the exemption principle this means

$$\Sigma^{\in} \models \forall x(\varphi(x) \to \bigvee_{\delta\in\Delta} Ab_\delta\, x)$$

6.3.3 Some more examples

(i) Using the inheritance principle it is easy to see why the following argument is valid.

$$\begin{array}{ll} & \forall x(Rx \rightsquigarrow \neg Ux) \\ & \forall x(Tx \rightsquigarrow Ux) \\ & \forall x(Qx \rightsquigarrow Tx) \\ & \forall x(Qx \rightsquigarrow Rx) \\ & \forall x(Px \rightsquigarrow Qx) \\ & \forall x(Sx \rightsquigarrow Px) \\ & Sc \\ \hline \therefore & Rc \wedge Tc \end{array}$$



Looking at the first four rules, we see that the exemption principle enforces that $\forall x(Qx \to (Ab_{R\neg U}\, x\ \vee\ Ab_{TU}\, x)) \in \Sigma^{\in}$. By applying the Inheritance Principle twice we see that $\forall x(Sx \rightarrow (Ab_{R\neg U}\, x \ \vee \ Ab_{TU}\, x)) \in \Sigma^{\in}$. So in all relevant models either $Ab_{R\neg U}\, c$ or $Ab_{TU}\, c$ is true. From this it follows that in all optimal models $\neg Ab_{S\neg P}$, $\neg Ab_{PQ}$, $\neg Ab_{QR}$, and $\neg Ab_{QT}$ are true, which enables us to conclude by default that Pc, Qc, Rc and Tc.

Notice that on the naive account from section 2 none of these can be concluded. It would not even be possible to make the first step upwards from Sc to Pc. Here we can not only make this first step but also a second to Qc and further up to Rc and Tc. Only when we hit a direct conflict do we need to stop. By having the upper abnormalities propagate downward, we do not have to take into account potential abnormalities at the lower levels.

(ii) Both *Defeasible Modus Ponens* and *Defeasible Modus Tollens* are valid.[8]

$$\begin{array}{ll} & \forall x(Sx \rightsquigarrow Px) \\ & Sc \\ \hline \therefore & Pc \end{array} \qquad\qquad \begin{array}{ll} & \forall x(Sx \rightsquigarrow Px) \\ & \neg Pc \\ \hline \therefore & \neg Sc \end{array}$$

The latter shows that an object need not have property S to count as an object that complies with the rule $\forall x(Sx \rightsquigarrow Px)$. Intuitively, if the object c had property S, it would be an abnormal S. So, assuming that the object c is normal and *complies* with the rule, it will not have property S.

Now, consider the following premises

premise 1	$\forall x(Sx \rightsquigarrow Px)$
premise 2	$\forall x(Px \rightsquigarrow \neg Sx)$
premise 3	Sc

At first sight one might be tempted to conclude Pc by *Defeasible Modus Ponens* and $\neg Pc$ by *Defeasible Modus Tollens*, but in fact the exemption principle enforces that $\forall x(Sx \rightarrow Ab_{P\neg S}\, x) \in \Sigma^{\in}$. This means that the only default conclusion to be drawn is Pc.

The reason we bring this up is because several authors have questioned the validity of *Defeasible Modus Tollens* with putative counterexamples like the following:

premise 1	*Men normally don't have a beard*
premise 2	*John has a beard*
by default	*John is not a man*

[8]There is a huge difference between this kind of *Modus Tollens* (From $\forall x(Sx \rightsquigarrow Px)$ and $\neg Pa$ it follows (by default) that $\neg Sa$) and *Contraposition* (From $\forall x(Sx \rightsquigarrow Px)$ it follows that $\forall x(\neg Px \rightsquigarrow \neg Sx)$). For a discussion, see (Caminada 2008).

However, all this example shows is that one has to be very careful in providing 'intuitive' counterexamples when dealing with default arguments. One must be sure that the premises faithfully represent *all* one knows about the matter at issue.

In this case we know in fact more than the premises state. For instance, people with a beard normally are men. (This is why the conclusion sounds weird in the first place.)

Now, if we state this explicitly as a third premise we get:

premise 1	*People with a beard normally are men*
premise 2	*Men normally don't have a beard*
premise 3	*John has a beard*

And as we saw, Defeasible modus ponens beats defeasible modus tollens, so the only conclusion to be drawn is that *John is a man.*

6.3.4 Coherence

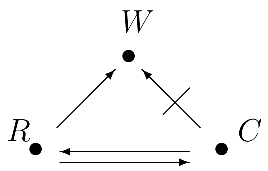
Every set Σ of default rules is consistent.[9] This does not mean that from a logical point of view every such set is okay. Here are some examples.

Consider

$$\Sigma = \{\forall x(Sx \rightsquigarrow Px), \forall x(Sx \rightsquigarrow \neg Px)\}$$

Clearly, a theory of this form is of no use. Note that $\Sigma \models \forall x(Sx \to (Ab_{SP} \vee Ab_{S\neg P}))$. We can apply the exemption principle (take $\Delta = \emptyset$ and $\varphi(x) = Sx$) to find that $\forall x(Sx \to \bot)$ is an exemption clause for Σ. So, $\Sigma^{\in} \models \neg\exists x Sx$.

A more complicated example is this one:



$$\forall x(Rx \rightsquigarrow Cx)$$
$$\forall x(Cx \rightsquigarrow Rx)$$
$$\forall x(Rx \rightsquigarrow Wx)$$
$$\forall x(Cx \rightsquigarrow \neg Wx)$$

'Rainy days normally are cold', 'Cold days normally are rainy', 'On rainy days the wind is normally west', 'On cold days the wind is normally not west'. Something is wrong with this theory. The exemption principle does not allow such days: $\Sigma^{\in} \models \neg\exists x Rx$. Proof: note first that $\forall x(Cx \to Ab_{RW}x) \in \Sigma^{\in}$. By the inheritance property it follows that $\forall x(Rx \to Ab_{RW}) \in \Sigma^{\in}$. Applying the exemption principle once more yields $\forall x(Rx \to \bot) \in \Sigma^{\in}$.

[9]To see why this is so, consider a model in which all objects are abnormal in all respects.

A third example is given by

$$\Sigma = \{\forall x(Sx \rightsquigarrow Px), \forall x((Sx \wedge Qx) \rightsquigarrow \neg Px), \forall x((Sx \wedge \neg Qx) \rightsquigarrow \neg Px)\}$$

Again, this does not sound like an acceptable theory. Too many exceptions are being made. $\Sigma^{\in}$ does not allow this. Note that $\Sigma \models \forall x((Sx \wedge Qx) \rightarrow (Ab_{(Sx \wedge Qx)\neg Px}x) \vee Ab_{SxPx}x))$. Hence, by the exemption principle $\forall x((Sx \wedge Qx) \rightarrow Ab_{SxPx}x) \in \Sigma^{\in}$; similarly, $\forall x((Sx \wedge \neg Qx) \rightarrow Ab_{SP}x) \in \Sigma^{\in}$. Hence, $\Sigma^{\in} \models \forall x(Sx \rightarrow Ab_{SP}x)$. But then $\forall x(Sx \rightarrow \bot) \in \Sigma^{\in}$.

The above leads to the following definition:

6.3.6. Definition. A set of rules Σ is coherent iff for every $\varphi(x)$ which is an antecedent of a rule in Σ, $\Sigma^{\in} \cup \{\exists x \varphi(x)\}$ is consistent.

A set of rules is incoherent if it is logically impossible to satisfy the minimal requirement. In such a case there is some property such that no object with this property can comply with all the rules for objects with this property. Given the exemption principle, no such objects are allowed.

As will become clear in the next section, for inheritance nets we can give an exact syntactic characterization of the sets of rules that are incoherent.

6.4 Networks

Inheritance networks are, simply put, the kind of directed graphs we have used to illustrate some of the examples in the previous sections. Thus, an inheritance network is a directed graph where the arrows represent default rules, nodes represent properties and specifically marked arrows are used for negative rules and for strict rules. More formally:

6.4.1. Definition. An *Inheritance Network* is a pair $\langle V, \Sigma \rangle$, where each element of Σ is a combination of an ordered pair of elements of V and a *polarity* which may be positive, negative, strict positive or strict negative.
Elements of Σ are referred to as arrows going from the first element of the ordered pair to the second. We will generally refer to an arrow from u to v as uv if positive, uv^- if negative, uv^* if strict positive and uv^{*-} if strict negative.

Nodes will generally represent properties, but may also represent objects or individuals, provided they are only connected to other nodes by strict arrows. In many examples, there is a single node representing an individual, and strict arrows from it to nodes representing properties indicate that that individual has those properties. The definition we use does not distinguish between nodes representing individuals and nodes representing properties: the difference is purely a matter of interpretation.
For making inferences in these networks, the notion of a *path* is crucial.

6.4.2. Definition. Let $\langle V, \Sigma \rangle$ be an Inheritance Network, with $a, b \in V$.

(i) A *positive path* from a to b is a subset $\{\alpha_1, \ldots, \alpha_n\} \subseteq \Sigma$ such that there exist $v_1, \ldots, v_{n-1} \in V$ such that:

- α_1 is a positive (or strict positive) arrow from a to v_1
- α_i is a positive (or strict positive) arrow from v_{i-1} to v_i, where $1 < i < n$
- α_n is a positive (or strict positive) arrow from v_{n-1} to b

Furthermore, the empty set is considered a positive path from any $v \in V$ to itself.

(ii) $X \subseteq \Sigma$ is a *negative path* from a to b if there are X_1, X_2, a', b', α such that

- $X = X_1 \cup \{\alpha\} \cup X_2$
- X_1 is a positive path from a to a'
- X_2 is a positive path from b to b'
- α is a negative (or strict negative) arrow from a' to b', or from b' to a'[10]

If there exists a positive (negative) path from a to b, this serves as *prima facie* evidence that objects with property a have (do not have) property b. Of course, in interesting examples we have prima facie evidence for both b and not b, which brings us to the next key notion: the *conflicting set*.

6.4.3. Definition. Where $\langle V, \Sigma \rangle$ is an inheritance network and $a \in V$, a subset $X \subseteq \Sigma$ is a *conflicting set relative to* a iff there is some $b \in V$ such that X contains both a positive and a negative path from a to b.
Such an X is a *minimal conflicting set* if every proper subset of X is not a conflicting set relative to a.

In the above definition, note that 'minimal' does mean having the least possible number of elements. Rather, it simply means that nothing more can be taken out without losing the property.

6.4.1 Making inferences in inheritance nets

Let $\langle V, \Sigma \rangle$ be an inheritance network, and $u, v \in V$. In the following we will write $u \asymp v$ to indicate that there is a positive path from u to v and a positive path from v to u.

[10]Note that it's possible that $a = a', b = b'$ and X_1 and X_2 are empty.

6.4.4. Definition. Where $\langle V, \Sigma \rangle$ is an inheritance network and $a \in V$, let

$$Ess_{\Sigma}(a) = \{uv \in \Sigma \mid u \asymp a\} \cup \{uv^{-} \in \Sigma \mid u \asymp a\} \cup \{\alpha \in \Sigma \mid \alpha \text{ is a strict arrow}\}$$

For a given property a, the set $Ess_{\Sigma}(a)$ contains the rules that are essential for a, i.e. all rules from which the objects with property a cannot be exempted. No object can be exempted from any strict rule; the objects with property a cannot be exempted from any rule for objects with property a, and more generally, the objects with property a cannot be exempted from any rule for objects with a property b that is "default equivalent" to a.

6.4.5. Definition. Where $\langle V, \Sigma \rangle$ is an inheritance network and $a \in V$, let

$$d(a) = \{X - Ess_{\Sigma}(a) \mid X \textit{ is a minimal conflicting set relative to } a\}$$

Note that $d(a)$ is not a set of arrows but rather a set of sets of arrows. The intuition is that the objects with property a are exempted from at least one rule in every set in $d(a)$.

The inheritance property comes in by letting the d function propagate backwards along positive paths, collecting d-sets in the D function defined below.

6.4.6. Definition. Where $\langle V, \Sigma \rangle$ is an inheritance network and $a \in V$, let

$$D(a) = \bigcup \{d(b) \mid \textit{there is a positive path from } a \textit{ to } b\}$$

Thus, $D(a)$ is the union of all $d(b)$ for b's to which there is a positive path from a. Its elements are sets of arrows, just like the elements of each $d(b)$ are.

We are now close to defining the consequence relation for networks. This will be done in terms of *exception sets*, potential sets of default rules (that is, arrows) to which an exception must be made.

6.4.7. Definition. Where $\langle V, \Sigma \rangle$ is an inheritance network and $a \in V$, $X \subseteq \Sigma$ is an *acceptable exception set of* a iff for all $Y \in D(a)$ there is some $\alpha \in X$ such that $\alpha \in Y$.

Such an X is a *minimal exception set* if every proper subset of X is not an acceptable exception set of a.

Each minimal exception set represents a way to make as few exceptions as possible. A given conclusion b now follows from a in a network if b can be reached from a under each of these ways.

6.4.8. Definition. Let $\langle V, \Sigma \rangle$ be an inheritance network. Let $a, b \in V$.

- $a \vdash_{\Sigma} b$ iff for every minimal exception set X of a there is a positive path Y from a to b such that $X \cap Y = \emptyset$.

- $a \vdash_\Sigma \neg b$ iff at least one of the following is true:

  (i) For every minimal exception set X of a there is a negative path Y from a to b such that $X \cap Y = \emptyset$.

  (ii) No minimal exception set X of a is also an acceptable exception set of b.

We did not prepare the reader for the second clause of negative entailment. It is there for the special case in which there is no path from a to b.[11] In such a case it may happen that objects with property b are so abnormal that one can safely assume that the object under consideration does not have property b.
When we know nothing about an object, we like to assume that it is normal in all respects. Thus if objects with property b are never normal in all respects, like a penguin which is either a non-flying bird or an even more abnormal flying penguin, we assume that objects we do not know anything about do not have property b. This is certainly how it works in the circumscription semantics. (Note that if b does not force exceptions to be made then any set is an acceptable exception set of b. Thus only 'exceptional' b's are affected.)
We do not cover arguments from complete ignorance here, but the above also holds if we do know something about the object but what we know (in this case, that it has property a) is completely unrelated to b. So for example, if we combine a Nixon Diamond and a Tweety Triangle into a single inheritance network (without adding any extra arrows), this clause lets us conclude that Nixon is presumably not a penguin and vice versa.
Determining exactly when a and b are not sufficiently related is non-trivial. There are situations where a and b are both connected to some third node c, yet still distinct enough that we should allow $a \vdash_\Sigma \neg b$ to follow. The key here is that if a necessarily creates the same abnormalities b does, then someone who already accepts a cannot reject b on the basis of those abnormalities. This is what is stated by the condition that the minimal exception sets for a are all acceptable exception sets for b. We will see in the Appendix that this condition is the correct one for the sake of making the completeness proof work.

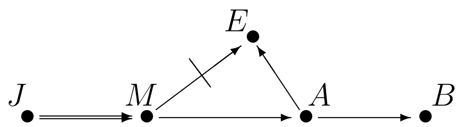
6.4.2 Examples

As a first example, we consider the following desirable inference.

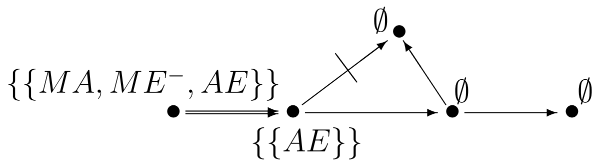
[11] If there is a path from a to b every minimal exception set of a is an acceptable exception set for b.

premise 1	*Adults normally have a bank account*
premise 2	*Master students are normally adults*
premise 3	*Master students are normally not employed*
premise 4	*Adults are normally employed*
premise 5	*John is a master student*
by default	*John is an adult with a bank account, but he is not employed*

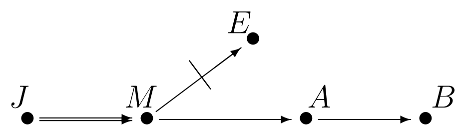
Rendered as an inheritance network, this looks as follows.



Our first step is to determine the d function. Since there are no conflicting sets relative to A, B, and E, we have $d(A) = d(B) = d(E) = \emptyset$. The conflicting sets relative to master student are $\{MA, ME^-, AE\}$ and $\{AB, MA, ME^-, AE\}$. Only the first of these is minimal. Since $Ess_\Sigma(M) = \{MA, ME^-, JM\}$, we obtain $d(M) = \{\{AE\}\}$.
Similarly, there is a single minimal conflicting set relative to J: the set $\{MA, ME^-, AE, JM\}$. We have $Ess_\Sigma(J) = \{JM\}$, so $d(J) = \{\{MA, ME^-, AE\}\}$.



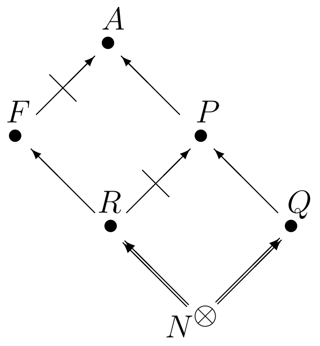
We can now determine $D(J)$. Since there is a positive path from J to every other node, $D(J)$ is the union of all the d's. Only two are non-trivial, so $D(J) = \{\{MA, ME^-, AE\}, \{AE\}\}$.
Since $\{AE\} \in D(J)$, every acceptable exception set for John will contain arrow AE. Since $\{AE\}$ is itself an acceptable exception set, this makes it the only minimal exception set. Thus, a conclusion is acceptable iff there is a path from J to it that does not use arrow AE. That is, if there is a path in the following network.



Therefore as desired we obtain $J \vdash_\Sigma \neg E$, $J \vdash_\Sigma A$, $J \vdash_\Sigma B$.

The Double Diamond

The following network is a well-known extension of the Nixon Diamond, generally referred to as the Double Diamond.



premise 1	*Nixon is a Republican and a Quaker*
premise 2	*Quakers are normally Pacifist*
premise 3	*Republicans are normally not Pacifist*
premise 4	*Republicans are normally Football fans*
premise 5	*Pacifists are normally Anti-military*
premise 6	*Football fans are normally not Anti-military*

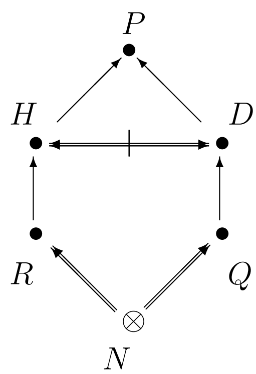
The question is whether Nixon is Anti-military. In traditional pre-emption based approaches (notably (Horty et al. 1990)), the positive path from N to A is disabled by the negative path from N to P, so that $\neg A$ may be concluded. This outcome is considered counterintuitive since the negative path to A is itself disabled by its positive counterpart. This has led to paths like that being referred to as *zombie paths.*(Makinson and Schlechta 1991) Since our own approach is not based on this kind of pre-emption, we can do a bit better here.

The first thing to notice is that there are no pairs of conflicting paths starting at P, F, A, R, or Q. Therefore all of them have empty d, and $D(N) = d(N)$. We subsequently find that $D(N) = \{\{QP, RP^-\}, \{QP, RF, PA, FA^-\}\}$. (Details left to the reader.) It is important to keep in mind that "minimal exception set" does not mean "exception set with the smallest amount of elements", meaning that $\{QP\}$ is not the only minimal exception set (relative to N) here. The others are $\{RP^-, RF\}, \{RP^-, PA\}$ and $\{RP^-, FA^-\}$.
We trivially obtain $N \vdash_\Sigma R$, $N \vdash_\Sigma Q$. But as to the other proprties, nothing can be concluded. While this seems natural enough for P and A, some people might see it as counterintuitive for F. However, it should be noted that there is both a positive and a negative path from N to F.

A floating conclusion

The next example is much discussed in the literature on inheritance nets.

premise 1	*Nixon is a Republican and a Quaker*
premise 2	*Quakers are normally Doves*
premise 3	*Republicans are normally Hawks*
premise 4	*Nobody is both a hawk and a dove*
premise 5	*Hawks normally are politically motivated*
premise 6	*Doves normally are politically motivated*

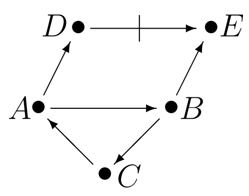


Does it follow that Nixon is politically motivated? According to the theory presented here, the answer to this question is 'Yes'.[12] It is easy to see that $D(N) = d(N)$. Furthermore, $d(N) = \{\{RH, QD\}\}$ (left to the reader). This means there are two minimal exception sets for N, namely $\{RH\}$ and $\{QD\}$.

The exception set $\{RH\}$ does not contain any element of the rightmost path from N to P, and the exception set $\{QD\}$ does not contain any element of the leftmost path from N to P. Thus, for each minimal exception set there is a positive path from N to P which does not contain any element of that set. Therefore $N \vdash_\Sigma P$.

Closed loops

The algorithm we will present below can also handle inheritance nets with cyclic paths. For an example of how this works, consider the following



premise 1	A's are normally B
premise 2	B's are normally C
premise 3	C's are normally A
premise 4	A's are normally D
premise 5	D's are normally not E
premise 6	B's are normally E
premise 7	x is A
by default	x is E

This example overlaps a small loop with part of a Nixon diamond. At first glance then, one might expect $d(A) = \{\{DE^-, BE\}\}$. However, this is not the case. Since all points of the loop must be taken into account, we have $Ess_\Sigma(A) = \{AD, AB, BE, BC, CA\}$. Therefore the conflicting set $X = \{AB, AD, DE^-, BE\}$ leads to the inclusion of not $\{DE^-, BE\}$ but rather $X - Ess_\Sigma(A) = \{DE^-\}$ in $d(A)$. Thus, E may be validly concluded when starting at A, B or C.

6.4.3 An algorithm

The way inheritance works in this system makes a backward-induction approach ideal. Consider the following pseudo-code algorithm to determine d and D across a network.

for $i = 1$ to n **do**
  for each positive path X starting at x_i **do**
    for each negative path Y starting at x_i **do**
      if X and Y have the same endpoint **then**
        $d(x_i) := d(x_i) \cup \{X \cup Y - Ess_\Sigma(x_i)\}$

[12] This is what most people working in this field want. Horty ((Horty 2002)) provides a counterexample, but it concerns normative rules rather than defaults. See also Prakken((Prakken 2002)) for an insightful discussion.

```
            end if
        end for
    end for
end for
for i = 1 to n do
    for X ∈ d(x_n) do
        D(x_i) := D(x_i) ∪ {X}
    end for
    for j = i + 1 to n do
        if x_j x_i ∈ Σ or x_j x_i^* ∈ Σ then
            D(x_j) := D(x_j) ∪ D(x_i)
        end if
    end for
end for
for i = 1 to n do
    for X ∈ D(x_i) do
        if ∃Y ∈ D(x_i) : Y ⊂ X then
            D(x_i) := D(x_i) − {X}
        end if
    end for
end for
```

The above will work so long as the nodes have have already been put in backward-induction order; that is, so long as for $i < j$ there is never a positive arrow from i to j. In cases where such an ordering is impossible (ie when the network contains positive loops), the correct results can still be obtained by simply rerunning the parts for D until the results stop changing.[13]

The algorithm is polynomial-time relative to n, Σ and P, where P is the number of paths there are. While we know $|\Sigma| \leq n^2$, P of course cannot be guaranteed to be less than exponential in Σ.

Since it is based on pairs of paths, we know that $|d(x_i)| < 0.5P^2$ for any i. The inheritance thereby puts $|D(x_i)|$ in the order of nP^2. In the absence of a way to reduce this figure, this means the most intensive part of the algorithm is the part where non-minimal elements are removed from D. Indeed, this is why it is generally more efficient to do this at the end (as we do here), rather than on-the-fly inside another loop.

While determining D is the bulk of the work when trying to determine whether $a \vdash_\Sigma b$ or whether the first option for $a \vdash_\Sigma \neg b$ holds, more is needed to check for the second option for $a \vdash_\Sigma \neg b$. Recall that under this item, $a \vdash_\Sigma \neg b$ is true if no minimal exception set X of a is an acceptable exception set of b. Instead of constructing every such X, we will check this for every *choice set*

[13] Specifically constructed perverse examples can necessitate any amount of runs up to n, but generally only a couple should be needed.

of $D(a)$. A choice set of $D(a)$ is a set $X \subseteq D(a)$ constructed by choosing for each $Y \subseteq D(a)$ one element $y \in Y$ to put in X. Each minimal exception set is contained in such a choice set (left to the reader) and each such choice set contains a minimal exception set (since it is itself an acceptable exception set), so it follows that this has the same result as checking all minimal exception sets.

```
for each choice set X of D(a) do
  Acceptable:=true
  for all Y ∈ D(b) do
    if X ∩ Y = ∅ then
      Acceptable:=false
    end if
  end for
  if Acceptable=true then
    return a ⊬_Σ ¬b
  end if
end for
return a ⊢_Σ ¬b
```

In this algorithm, for each choice set X of $D(a)$, we first assume that X is an acceptable exception set of b and then check if there is a reason to revise this. If it is indeed an acceptable exception set of b then we conclude that $a \nvdash_\Sigma \neg b$ (we assume the other option for $a \vdash_\Sigma \neg b$ has already been ruled out) and halt the algorithm. Otherwise we move on to the next choice set X. If no choice set X is an acceptable exception set of b, then we conclude that $a \vdash_\Sigma \neg b$.
Of course, the time it takes to create all choice sets is exponential in $|D(a)|$, so one may wish to be careful about when to choose to use this second algorithm.

6.4.4 Completeness

Networks are a natural way to illustrate most examples even when working in a circumscriptive framework, so it will come as no surprise that the inheritance networks from this chapter can be interpreted in terms of the system we introduced before. However, what is far from trivial is the interpretation can be done in such a way that all the results coincide; that is, that the network-based approach is sound and complete (as to what it can express) relative to the other framework.

We provide the (rather straightforward) translation and the formal statement here. For the extensive proof, see Appendix 6.6.

6.4.9. Definition. Let $N = \langle V, \Sigma \rangle$ be an inheritance network, with $V = \{v_1, \ldots, v_n\}$. We associate with every $v_i \in V$ a predicate P_i, and with every arrow α a rule α given by

$$\begin{aligned} v_i v_j^{\uparrow} &= \forall x(P_i x \rightsquigarrow P_j x) \\ v_i v_j^{-\uparrow} &= \forall x(P_i x \rightsquigarrow \neg P_j x) \\ v_i v_j^{*\uparrow} &= \forall x(P_i x \rightarrow P_j x) \\ v_i v_j^{*-\uparrow} &= \forall x(P_i x \rightarrow \neg P_j x) \end{aligned}$$

We will call $\Sigma^{\uparrow} = \{\alpha \mid \alpha \in \Sigma\}$ the *lift* of N.

Note that since networks do not distinguish between individuals and properties, the lift will convert to predicates any individuals used in example networks. A premise like "John is an Adult", which in the circumscription framework could be represented as Aj, would be represented in an inheritance network as a strict arrow from J to A, the lift of which would be $\forall x(Jx \rightarrow Ax)$.

6.4.10. THEOREM. ***(Soundness-Completeness Theorem)*** *Let* $N = \langle V, \Sigma \rangle$ *and* $\Sigma^{\uparrow}$ *be as in the definition. Suppose* $\Sigma^{\uparrow}$ *is coherent. Then* $v_i \vdash_{\Sigma} v_j$ *if and only if* $\Sigma^{\uparrow}, \{P_i c\} \models_d P_j c$, *and* $v_i \vdash_{\Sigma} \neg v_j$ *if and only if* $\Sigma^{\uparrow}, \{P_i c\} \models_d \neg P_j c$.

Since the above theorem only works in the case of coherence, it is desirable to have a comparable network-based notion. This is where the following theorem comes in. Again, the proof can be found in Appendix 6.6.

6.4.11. THEOREM. *Let* $\Sigma^{\uparrow}$ *be the lift of the inheritance network* $\langle V, \Sigma \rangle$. *Then* $\Sigma^{\uparrow}$ *is coherent if and only if there is no* $v \in V$ *with* $\emptyset \in d(v)$.

Broadly speaking, the latter is the case if two equivalent points yield unresolvably different conclusions about a third. This is made explicit by the following definition and proposition, also proven in the Appendix.

6.4.12. DEFINITION. The vertex x *semi-strictly implies* (*semi-strictly refutes*) y if there is a positive (negative) path from x to y where every arrow after the first is strict.

6.4.13. PROPOSITION. *Let* $\langle V, \Sigma \rangle$ *be an inheritance network. If* $\emptyset \in d(x)$, *then there are some* z *and some* $y \approx x, y' \approx x$ *such that* y *semi-strictly implies* z *and* y' *semi-strictly refutes* z.

6.5 Conclusion

In the above we have studied the logical properties of defaults, or more particularly of sentences of the form *S's are normally P*. We have shown that their capricious logical behavior can be wholly explained on the basis of one simple underlying principle that determines, in cases of conflicting defaults, which objects are *exempted* from which rules. We have developed the theory both semantically (within a circumscriptive theory) and syntactically (using inheritance nets). In the appendix we will prove a completeness theorem showing that arguments that can be expressed in both systems are valid on the one account iff they are valid on the other.

Despite the length of this chapter, we have only taken the first steps developing these systems. Undoubtedly, a more systematic *model theoretic* study of the circumscriptive part will result in a more elegant proof of the completeness theorem. We also think that on the *algorithmic* side further investigations may yield simplifications. For example, things get a lot less complicated (and complex) if the nets do not have cycles. Finally, a study like this should be complemented by a study which answers the question under which conditions a set of default rules can be safely adopted as a guiding line for taking decisions. Maybe this is a question for methodologists rather than for logicians, but the answer is important to everybody interested in common sense reasoning.

6.6 Appendix: Completeness of Networks relative to the Semantics

When defining the d and D functions we already suggested that they amount to implementing the inheritance property and a weak version of the exemption principle. Before starting with the completeness proof proper, we will first make this explicit and prove that when working with inheritance networks the combination of this weak version of the exemption principle and the inheritance property is equivalent to the regular exemption principle.

6.6.1 New constraints, same consequences

6.6.1. Definition. Let Σ be a set of rules. The formulas φ and ψ are *equivalent in* Σ iff there are $\varphi_1, \ldots, \varphi_m$, $\psi_1, \ldots, \psi_n$ such that $\varphi_m = \psi = \psi_1$, $\psi_n = \varphi = \varphi_1$ and for all $1 \leq i < m$, $1 \leq j < n$

$$\begin{array}{rcl} \Sigma & \models & \forall x(\varphi_i(x) \rightsquigarrow \varphi_{i+1}(x)) \\ \Sigma & \models & \forall x(\psi_j(x) \rightsquigarrow \psi_{j+1}(x)) \end{array}$$

We denote this as $\varphi \approx_\Sigma \psi$, or simply $\varphi \approx \psi$ is no confusion is possible.

6.6.2. Definition.

(i) The clause $\forall x(\varphi(x) \to \bigvee_{\delta\in\Delta} Ab_\delta x)$ is an expanded exemption clause for Σ iff there are $\psi_1 \approx \psi_2 \approx \ldots \approx \psi_n \approx \varphi$ such that

$$\Sigma \models \forall x(\varphi(x) \to \bigvee_{\delta\in\Delta\cup\Sigma^{\varphi(x)}\cup\Sigma^{\psi_1(x)}\cup\ldots\cup\Sigma^{\psi_n(x)}} Ab_\delta x)$$

.

(ii) The expanded weak exemption extension Σ^W of Σ is given by

$$\Sigma^W = \Sigma \cup \{\varphi \mid \varphi \text{ is an expanded exemption clause for } \Sigma\}$$

6.6.3. Definition.

(i) The clause $\forall x(\varphi(x) \to \bigvee_{\delta\in\Delta} Ab_\delta x)$ is an inherited clause for Σ iff there is some ψ such that $\forall x(\varphi(x) \rightsquigarrow \psi(x)) \in \Sigma$ and $\Sigma \models \forall x(\psi(x) \to \bigvee_{\delta\in\Delta} Ab_\delta x)$.

(ii) The inheritance extension Σ^I of Σ is given by

$$\Sigma^I = \bigcup_{n\in\omega} \Sigma^I_n$$

where $\Sigma^I_0 = \Sigma$ and $\Sigma^I_{n+1} = \Sigma^I_n \cup \{\varphi \mid \varphi \text{ is an inherited clause for } \Sigma^I_n a\}$

6.6.4. Theorem. $\Sigma^\in \models \Sigma^{WI}$

Proof: We first prove that $\Sigma^\in \models \Sigma^W$. Let $\theta \in \Sigma^W$. We may assume that $\Sigma \not\models \theta$ (otherwise $\Sigma^\in \models \theta$ follows immediately). Therefore θ is of the form

$$\theta = \forall x(\varphi(x) \to \bigvee_{\delta\in\Delta} Ab_\delta x)$$

with

$$\Sigma \models \forall x(\varphi(x) \to \bigvee_{\delta\in\Delta\cup\Sigma^{\varphi(x)}\cup\Sigma^{\psi_1(x)}\cup\ldots\cup\Sigma^{\psi_n(x)}} Ab_\delta x)$$

for some $\psi_1 \approx \psi_2 \approx \ldots \approx \psi_n \approx \varphi$.
Since $\psi_1 \approx \varphi$, (repeated) use of the inheritance property lets us conclude

$$\Sigma^\in \models \forall x(\psi_1(x) \to \bigvee_{\delta\in\Delta\cup\Sigma^{\varphi(x)}\cup\Sigma^{\psi_1(x)}\cup\ldots\cup\Sigma^{\psi_n(x)}} Ab_\delta x)$$

By taking $\Delta' = \Delta \cup \Sigma^{\varphi(x)}$, we may use the exemption principle to conclude

$$\Sigma^\in \models \forall x(\psi_1(x) \to \bigvee_{\delta\in\Delta\cup\Sigma^{\varphi(x)}\cup\Sigma^{\psi_2(x)}\cup\ldots\cup\Sigma^{\psi_n(x)}} Ab_\delta x)$$

Now by (repeatedly) using the inheritance property again we arrive at

$$\Sigma^{\in} \models \forall x(\varphi(x) \to \bigvee_{\delta \in \Delta \cup \Sigma^{\varphi(x)} \cup \Sigma^{\psi_2(x)} \cup \ldots \cup \Sigma^{\psi_n(x)}} Ab_\delta x)$$

The same process can be repeated for all ψ_i, leaving us with

$$\Sigma^{\in} \models \forall x(\varphi(x) \to \bigvee_{\delta \in \Delta \cup \Sigma^{\varphi(x)}} Ab_\delta x)$$

from which it follows through the exemption principle that

$$\Sigma^{\in} \models \forall x(\varphi(x) \to \bigvee_{\delta \in \Delta} Ab_\delta x)$$

This proves that $\Sigma^{\in} \models \Sigma^{W}$. Therefore $\Sigma^{\in I} \models \Sigma^{WI}$. Since the exemption principle implies the inheritance property, $\Sigma^{\in} \models \Sigma^{\in I}$, and thus $\Sigma^{\in} \models \Sigma^{WI}$.

How about $\Sigma^{WI} \models \Sigma^{\in}$? We doubt this holds for every Σ, but it does hold for the special case that Σ is the lift of an inheritance network. Before we turn to the proof of this statement some more observations are needed.

The rules and exemption clauses figuring in the sets $(\Sigma^{\uparrow})^{WI}$ have a very specific syntactic form, which gives us a lot of freedom when we construct models of such sets. For example, all the sentences concerned are universal, so every $(\Sigma^{\uparrow})^{WI}$ is preserved under submodels. Note also that if the only difference between two models $\mathcal{A}$ and $\mathcal{A}'$ is that $\mathcal{A}'$ has more abnormalities than $\mathcal{A}$, then $\mathcal{A}'$ will be a model of $(\Sigma^{\uparrow})^{WI}$ if $\mathcal{A}$ is. This also holds if for some some predicates P_i that do not occur in the consequent of any rule in $(\Sigma^{\uparrow})^{WI}$, the extension of P_i in $\mathcal{A}'$ is a subset of the extension of P_i in $\mathcal{A}$. More precisely:

6.6.5. LEMMA. *Let $\Sigma^{\uparrow}$ be the lift of an inheritance network $\langle V, \Sigma\rangle$, with $V = \{v_1, \ldots, v_m\}$. Let Γ consist of sentences of the form $\forall x(Q_j x \to \bigvee_{\delta \in \Delta_j} Ab_\delta x)$. Let $\mathfrak{A} = \langle \mathcal{A}, \mathcal{I}\rangle$ and $\mathfrak{A}' = \langle \mathcal{A}', \mathcal{I}'\rangle$ be two models with the following properties:*

(a) $\mathfrak{A} \models \Sigma^{\uparrow} \cup \Gamma$;

(b) $\mathcal{A} = \mathcal{A}'$;

(c) for all individual constants c, $\mathcal{I}(c) = \mathcal{I}'(c)$;

(d) for all predicates P_i, the following holds:

  (da) If P_i does not occur in the consequent of any rule in $\Sigma^{\uparrow}$, then $\mathcal{I}'(P_i) \subseteq \mathcal{I}(P_i)$;

  (db) Otherwise, $\mathcal{I}'(P_i) = \mathcal{I}(P_i)$;

(e) for all predicates $Ab_{P_iP_j}$, $\mathcal{I}(Ab_{P_iP_j}) \subseteq \mathcal{I}'(Ab_{P_iP_j})$;

Then $\mathfrak{A}' \models \Sigma^\uparrow \cup \Gamma$.

Proof: Left to the reader.

On the way to the completeness theorem, we are often looking for correspondences between notions that play a role in inheritance nets on the one hand and notions in the circumscription framework on the other. One such notion is the notion of a *path.*

Note that if in a network $\langle V, \Sigma\rangle$ there is a positive path from v_i to v_j, then $\Sigma^\uparrow \models \forall x((P_i x \wedge \bigwedge_{\alpha\in X} \neg Ab_\alpha x) \rightarrow P_j x)$.[14] For coherent theories the converse is also true. This follows immediately from the following more general proposition.

6.6.6. Lemma. *Let $\Sigma^\uparrow$ be the lift of an inheritance network $\langle V, \Sigma\rangle$, with $V = \{v_1, \dots, v_m\}$. Let Γ consist of sentences of the form $\forall x(Q_j x \rightarrow \bigvee_{\delta\in\Delta_j} Ab_\delta x)$. Let $\varphi(x)$ be a quantifier-free formula in which all predicates are abnormality predicates, and such that $\Sigma^\uparrow \cup \Gamma \cup \{\exists x(P_i x \wedge \varphi(x))\}$ is consistent. If $\Sigma^\uparrow \cup \Gamma \models \forall x((P_i x \wedge \varphi(x)) \rightarrow P_j x)$, then there is a positive path from v_i to v_j.*

Proof: We proceed with an unusual induction, one on the number of distinct consequents of rules in $\Sigma^\uparrow$.
Case n=0: If $\emptyset \cup \Gamma \models \forall x((P_i x \wedge \varphi(x)) \rightarrow P_j x)$ and $\emptyset \cup \Gamma \cup \{\exists x(P_i x \wedge \varphi(x))\}$ is consistent, then $i = j$. So, there is a path from v_i to v_j.
Induction Hypothesis: The theorem is true if the number of distinct consequents occurring in the rules of $\Sigma^\uparrow$ is at most n.
Case n+1: Let $\Sigma^\uparrow$ have $n+1$ such consequents.

We first show that there is at least one l such that $\Sigma^\uparrow$ contains the rule $\forall x((P_l x \wedge \neg Ab_{P_lP_j} x) \rightarrow P_j x)$.

Suppose there is no such l. Given the fact that both $\Sigma^\uparrow$ and Γ consist of universal sentences we can construct a model $\mathfrak{A}$ of $\Sigma^\uparrow \cup \Gamma \cup \{\forall x(P_i x \wedge \varphi(x))\}$. Since $\Sigma^\uparrow \cup \Gamma \models \forall x((P_i x \wedge \varphi(x)) \rightarrow P_j x)$, $\forall x P_j x$ is true in $\mathfrak{A}$. Now, notice that if we change the interpretation of P_j in $\mathfrak{A}$, while leaving the interpretation of all other predicates the same, the resulting model $\mathfrak{A}'$ will still be a model of Γ, $\forall x(P_i x \wedge \varphi(x))$, and also $\Sigma^\uparrow$, the latter by lemma 6.6.5. However, $\mathfrak{A}'$ is not a model of $\forall x((P_i x \wedge \varphi(x)) \rightarrow P_j x)$ any more. This contradicts the fact that $\Sigma^\uparrow \cup \Gamma \models \forall x((P_i x \wedge \varphi(x)) \rightarrow P_j x)$.

Now, let L be the set of l for which $\Sigma^\uparrow$ contains the rule $\forall x(P_l x \wedge \neg Ab_{P_lP_j} x \rightarrow P_j x)$. Let $\Sigma^\uparrow_{-j}$ be $\Sigma^\uparrow$ with all rules in which P_j is the consequent removed. The next claim is that for at least one $l \in L$, $\Sigma^\uparrow_{-j} \cup \Gamma \models \forall x(P_i x \wedge \varphi(x) \rightarrow P_l x)$.

To prove this let $\mathfrak{A}_{-j}$ be a model of $\Sigma^\uparrow_{-j} \cup \Gamma \cup \{\forall x(P_i x \wedge \varphi(x))\}$. Suppose the claim does not hold. Then we can change the interpretation of P_l for all $l \in L$

[14]We are a bit sloppy here. We should have written '$Ab_{\alpha^\uparrow}$' instead of 'Ab_α', because it concerns the abnormality predicate of the lift $\alpha^\uparrow$ of the arrow α.

in such a manner that $\forall x \neg P_l x$ gets true for all $l \in L$ and such that $\forall x \neg P_j x$ gets true, while leaving the interpretation of all other predicates the same, without affecting the truth of the sentences in $\Sigma^{\uparrow}_{-j} \cup \Gamma \cup \{\forall x(P_i x \wedge \varphi(x))\}$. (Again by Lemma 6.6.5.)

The model would then trivially make true all default rules with $P_l x$ in the antecedent for any $l \in L$, and therefore be a model of $\Sigma^{\uparrow} \cup \Gamma$. However, it would make $\forall x(P_i x \wedge \varphi(x) \rightarrow P_j x)$ false, contradicting the fact that $\Sigma^{\uparrow} \cup \Gamma \models \forall x(P_i x \wedge \varphi(x) \rightarrow P_j x)$.

So we find some l such that $\Sigma^{\uparrow}$ contains a rule of the form $\forall x(P_l x \wedge \neg Ab_{P_{l'} P_j} x \rightarrow P_j x)$ and $\Sigma^{\uparrow}_{-j} \cup \Gamma \models \forall x(P_i x \wedge \varphi(x) \rightarrow P_l x)$. Note that $\Sigma^{\uparrow}_{-j}$ has n distinct consequents of default rules in it. Thus by the induction hypothesis there is a positive path from v_i to v_l. The rule $\forall x(P_{l'} x \wedge \neg Ab_{P_{l'} P_j} x \rightarrow P_j x)$ corresponds to an arrow from $v_{l'}$ to v_j, extending the path to one from v_i to v_j.

6.6.7. Theorem. *If $\Sigma^{\uparrow}$ is the lift of an inheritance network $\langle V, \Sigma \rangle$ and is coherent, then $(\Sigma^{\uparrow})^{WI} \models (\Sigma^{\uparrow})^{WI\in}$*

Proof: It suffices to show that $(\Sigma^{\uparrow})^{WI}$ satisfies the exemption principle. So, let θ, θ' be any clauses of the form below:

$$\theta = \forall x(P_i x \rightarrow \bigvee_{\delta \in \Delta} Ab_\delta x)$$

$$\theta' = \forall x(P_i x \rightarrow \bigvee_{\delta \in \Delta \cup (\Sigma^{\uparrow})^{P_i x}} Ab_\delta x)$$

We have to prove that whenever such a θ' is implied by $(\Sigma^{\uparrow})^{WI}$, so is θ.

Suppose $(\Sigma^{\uparrow})^{WI} \models \theta'$. Note first that if $\Sigma^{\uparrow} \models \theta'$, then $(\Sigma^{\uparrow})^{W} \models \theta$ by construction (because $\varphi \approx \varphi$), and we're done. So, the interesting case is when $\Sigma^{\uparrow} \not\models \theta'$.

Set $(\Sigma^{\uparrow})^{WI} = \Sigma^{\uparrow} \cup \{\phi_1, \phi_2, \ldots\}$, where

$$\phi_j = \forall x(Q_j x \rightarrow \bigvee_{\delta \in \Delta_j} Ab_\delta x)$$

For any k, let $(\Sigma^{\uparrow})^{WI}_k = \Sigma^{\uparrow} \cup \{\phi_1, \ldots, \phi_k\}$.

Now, there is some n such that $(\Sigma^{\uparrow})^{WI}_{n-1} \not\models \theta'$, while $(\Sigma^{\uparrow})^{WI}_n \models \theta'$. Two important things are true about ϕ_n.

Claim 1: There is a path from the node corresponding to P_i to the node corresponding to Q_n.

To show this, consider a model $\mathfrak{A}_1$ of $(\Sigma^{\uparrow})^{WI}_{n-1}$ where θ' is not true. Define $\mu(x) = \bigwedge_{\delta \in \Delta \cup (\Sigma^{\uparrow})^{P_i x}} \neg Ab_\delta x$. Thus, $\mathfrak{A}_1 \models \exists x(P_i x \wedge \mu(x))$, demonstrating that $(\Sigma^{\uparrow})^{WI}_{n-1} \cup \{\exists x(P_i x \wedge \mu(x))\}$ is consistent.

If the claimed path does not exist, contraposition of Lemma 6.6.6 tells us it cannot be the case that $(\Sigma^{\uparrow})^{WI}_{n-1} \models \forall x((P_i x \wedge \mu(x)) \rightarrow Q_n x)$. Thus there is a model

$\mathfrak{A}_2$ of $(\Sigma^\uparrow)^{WI}_{n-1}$ with some element d satisfying $P_i x \wedge \neg Q_n x \wedge \mu(x)$. Restrict $\mathfrak{A}_2$ to d to get $\mathfrak{A}_3$. Since $\mathfrak{A}_3 \models \forall x \neg Q_n x$, trivially $\mathfrak{A}_3 \models \phi_n$. Therefore $\mathfrak{A}_3 \models (\Sigma^\uparrow)^{WI}_n$. However, $\mathfrak{A}_3 \not\models \theta'$, contradicting the choice of n. This contradiction proves the claimed path must exist.

Claim 2: $\Delta_n \subseteq \Delta \cup (\Sigma^\uparrow)^{P_i x}$.

For this, let $\mathfrak{A}_4$ be the restriction of $\mathfrak{A}_1$, above, to elements satisfying $P_i x \wedge \mu(x)$. Construct $\mathfrak{A}_5$ from $\mathfrak{A}_4$ by making $\forall x Ab_\delta x$ true for all $\delta \in \Delta_n - (\Delta \cup (\Sigma^\uparrow)^{P_i x})$. Note that $\mathfrak{A}_5$ is still a model of $(\Sigma^\uparrow)^{WI}_{n-1}$. Also, $\mathfrak{A}_5 \models \neg\theta'$. However, if the claim is false then trivially $\mathfrak{A}_5 \models \phi_n$ and hence $\mathfrak{A}_5 \models (\Sigma^\uparrow)^{WI}_n$. By contradiction, the claim must be true.

Having proven these claims, we now distinguish two cases, depending on where ϕ_n was added.

Case I: $\phi_n \in (\Sigma^\uparrow)^W$.

If $\phi_n \in (\Sigma^\uparrow)^W$, then there are $Q'_1 \approx \ldots \approx Q'_u \approx Q_n$ such that

$$\Sigma^\uparrow \models \forall x (Q_n x \to \bigvee_{\delta \in \Delta_n \cup (\Sigma^\uparrow)^{Q_n} \cup (\Sigma^\uparrow)^{Q'_1} \cup \ldots \cup (\Sigma^\uparrow)^{Q'_u}} Ab_\delta x).$$

Given that $\Delta_n \subseteq \Delta \cup (\Sigma^\uparrow)^{P_i x}$, this implies

$$\Sigma^\uparrow \models \forall x (Q_n x \to \bigvee_{\delta \in \Delta \cup (\Sigma^\uparrow)^{P_i} \cup (\Sigma^\uparrow)^{Q_n} \cup (\Sigma^\uparrow)^{Q'_1} \cup \ldots \cup (\Sigma^\uparrow)^{Q'_u}} Ab_\delta x).$$

This leaves two possibilities. If $Q_n \approx P_i$ then $(\Sigma^\uparrow)^W \models \forall x (Q_n x \to \bigvee_{\delta \in \Delta} Ab_\delta x)$, which implies $\theta \in (\Sigma^\uparrow)^{WI}$.

If it is not the case that $Q_n \approx P_i$, the above can be simplified to

$$\Sigma^\uparrow \models \forall x (Q_n x \to \bigvee_{\delta \in \Delta \cup (\Sigma^\uparrow)^{Q_n} \cup (\Sigma^\uparrow)^{Q'_1} \cup \ldots \cup (\Sigma^\uparrow)^{Q'_u}} Ab_\delta x).$$

To prove this, let χ be the simplified formula and χ' the unsimplified one. Suppose there is no path from the node corresponding to Q_n to the node corresponding to P_i and yet $\Sigma^\uparrow \not\models \chi$. Define $\mu(x)$ as follows:

$$\mu(x) = \left(\bigwedge_{\delta \in \Delta \cup (\Sigma^\uparrow)^{Q_n} \cup (\Sigma^\uparrow)^{Q'_1} \cup \ldots \cup (\Sigma^\uparrow)^{Q'_u}} \neg Ab_\delta x) \right) \wedge \left(\bigvee_{\delta \in (\Sigma^\uparrow)^{P_i}} Ab_\delta x \right).$$

Then $\Sigma^\uparrow \cup \{\exists x (Q_n x \wedge \mu(x))\}$ is consistent. (Since $\Sigma^\uparrow$ is coherent, $\Sigma^\uparrow \cup \{\exists x Q_n x\}$ is consistent. Therefore this follows directly from $\Sigma^\uparrow \models \chi'$, $\Sigma^\uparrow \not\models \chi$.) Since there is no path from the node corresponding to Q_n to the node corresponding to P_i, contraposition of Lemma 6.6.6 tells us it cannot be the case that $\Sigma^\uparrow \models$

$\forall x((Q_n x \wedge \mu(x)) \to P_i x)$. Therefore there is a model of $\Sigma^\uparrow$ with some element d satisfying $Q_n x \wedge \neg P_i x \wedge \mu(x)$.
Adjust this model such that for no δ in $(\Sigma^\uparrow)^{P_i}$ d satisfies $Ab_\delta x$. Since d does not satisfy $P_i x$, this adjusted model is still a model of $\Sigma^\uparrow$ (if there is no path as above). However, this model does not make χ' true. This contradiction proves that if such a path does not exist then $\Sigma^\uparrow \models \chi$.

Given that $\Sigma^\uparrow \models \chi$, it follows that $\forall x(Q_n x \to \bigvee_{\delta\in\Delta} Ab_\delta x) \in (\Sigma^\uparrow)^W$. Since there is a path from the node corresponding to P_i to the one corresponding to Q_n, this in turn leads to $\theta \in (\Sigma^\uparrow)^{WI}$.
Case II: $\phi_n \in (\Sigma^\uparrow)^{WI} - (\Sigma^\uparrow)^W$.
In this case there is some Q' such that there is a positive path from the node corresponding to Q_n to the node corresponding to Q' and

$$(\Sigma^\uparrow)^W \models \forall x(Q'x \to \bigvee_{\delta\in\Delta_n} Ab_\delta x).$$

Recall that we have established $\Delta_n \subseteq \Delta \cup (\Sigma^\uparrow)^{P_i}$. Thus the above implies

$$(\Sigma^\uparrow)^W \models \forall x(Q'x \to \bigvee_{\delta\in\Delta\cup(\Sigma^\uparrow)^{P_i}} Ab_\delta x).$$

Now pick m such that $(\Sigma^\uparrow)^{WI}_m$ implies the above formula and $(\Sigma^\uparrow)^{WI}_{m-1}$ does not. Since $(\Sigma^\uparrow)^W$ does so, we may assume that $\phi_m \in (\Sigma^\uparrow)^W$. Therefore by the same arguments as above (the ones used for the case that $\phi_n \in (\Sigma^\uparrow)^W$) it follows that $\theta \in (\Sigma^\uparrow)^{WI}$.

The above theorems give us $\Sigma^\in \models \Sigma^{WI}$ and $\Sigma^{WI} \models \Sigma^{WI\in}$. Since it is trivially true that $\Sigma^{WI\in} \models \Sigma^\in$, this means $\Sigma^\in$ and Σ^{WI} have the same models.

What is perhaps easier to see but still important to prove is that the alternative constraints leading to Σ^{WI} correctly model what happens in constructing the D function. The following Lemma and Proposition cover this part.

6.6.8. LEMMA. *Let $\Sigma^\uparrow$ be the lift of some network $\langle V, \Sigma\rangle$, with $V = \{v_1, \ldots, v_n\}$. Then $X \subseteq \Sigma$ is a conflicting set relative to v_i if and only if*

$$\Sigma^\uparrow \models \forall x \left(P_i x \to \bigvee_{\alpha\in X} Ab_\alpha x \right)$$

(Note that the above means the formula is true on every model of $\Sigma^\uparrow$, even those which are not models of $(\Sigma^\uparrow)^{WI}$.)

PROOF: Suppose $X \subseteq \Sigma$ is a conflicting set relative to v_i. Suppose towards contradiction that there is a model $\mathfrak{A}$ of $\Sigma^\uparrow$ such that

$$\mathfrak{A} \models \exists x \left(P_i x \wedge \bigwedge_{\alpha\in X} \neg Ab_\alpha x \right)$$

Since X is a conflicting set relative to v_i, there is some v_j such that X contains both a positive and a negative path to v_j. Therefore by repeated modus ponens (as well as modus tollens, possibly) it follows that both $P_j x$ and $\neg P_j x$. Contradiction.

For the other direction, suppose X is not a conflicting set relative to v_i. Let $\mathfrak{A}$ be a model where $\forall x P_i x$ and $\forall x \neg Ab_\alpha x$ for all $\alpha \in X$ hold, with the rest of the predicates having their truth-value determined by applying the rules in $\Sigma^\uparrow$. Since there are no logical relations between the predicates other than those provided by $\Sigma^\uparrow$, this can be done while letting $\mathfrak{A}$ be a consistent model of $\Sigma^\uparrow$. But the relevant formula is now false on $\mathfrak{A}$. Hence, $\Sigma^\uparrow$ does not entail it.

6.6.9. PROPOSITION. *Let $\Sigma^\uparrow$ be the lift of an inheritance network $\langle V, \Sigma \rangle$, with $V = \{v_1, \ldots, v_n\}$. Let*

$$\phi = \forall x \left(P_i x \to \bigvee_{\alpha \in X} Ab_\alpha x \right)$$

If $(\Sigma^\uparrow)^{WI} \models \phi$, then $Y \in D(v_i)$ for some $Y \subseteq X$. Conversely, if $X \in D(v_i)$ then $(\Sigma^\uparrow)^{WI} \models \phi$.

PROOF: Suppose $(\Sigma^\uparrow)^{WI} \models \phi$. By the construction of $(\Sigma^\uparrow)^{WI}$, there must be some k such that there is a positive path from v_i to v_k and

$$(\Sigma^\uparrow)^{W} \models \forall x \left(P_k x \to \bigvee_{\alpha \in X} Ab_\alpha x \right)$$

By the construction of $(\Sigma^\uparrow)^{W}$, it follows that

$$\Sigma^\uparrow \models \forall x \left(P_k x \to \bigvee_{\alpha \in X \cup Ess_\Sigma(v_k)} Ab_\alpha x \right)$$

By Lemma 6.6.8, this means that $X \cup Ess_\Sigma(v_k)$ is a conflicting set relative to v_k. Therefore $Y \in d(v_k)$ and hence $Y \in D(v_i)$, where $Y = X - Ess_\Sigma(v_k) \subseteq X$.

For the converse, suppose $X \in D(v_i)$. Then there is some v_j such that there is a positive path from v_i to v_j and $X \in d(v_j)$. Therefore $X \cup Ess_\Sigma(v_j)$ is a conflicting set relative to v_j. By Lemma 6.6.8,

$$\Sigma^\uparrow \models \forall x \left(P_j x \to \bigvee_{\alpha \in (X \cup Min_\Sigma(v_j))} Ab_\alpha x \right)$$

By construction of $(\Sigma^\uparrow)^{W}$,

$$(\Sigma^\uparrow)^{W} \models \forall x \left(P_j x \to \bigvee_{\alpha \in X} Ab_\alpha x \right)$$

And therefore by construction $(\Sigma^\uparrow)^{WI} \models \phi$.

6.6.2 Completeness Proof

Knowing (via Σ^{WI}) how the D function and $\Sigma^{\in}$ are related is an important step on our way to completeness, but we are far from done. At this point it may not be entirely clear to what thing on the inheritance network side the models in the sets $\mathcal{F}$ of the states $\langle \mathcal{U}, \mathcal{F} \rangle$ correspond. The bulk of the completeness proof lies in showing that they correspond to acceptable exception sets, with optimal models corresponding to minimal exception sets. The correspondence can be interpreted through the notion defined below.

6.6.10. DEFINITION. Let $N = \langle V, \Sigma \rangle$ be an inheritance network. Let $\langle \mathcal{U}, \mathcal{F} \rangle$ be the state generated by the lift of N. Let $\mathfrak{A}$ be a model in $\mathcal{F}$ whose domain contains an element referred to by the constant c. Let $X \subseteq \Sigma$ be an exception set.

We say that $\mathfrak{A}$ *models* the exception set X for c if $\mathfrak{A} \models Ab_{\alpha^\uparrow} c$ if and only if $\alpha \in X$.

Now we first show that $\mathcal{F}$ consists of those models in $\mathcal{U}$ which correspond to an acceptable exception set of v_i.

6.6.11. PROPOSITION. *Let $\Sigma^\uparrow$ be the lift of the inheritance network $\langle V, \Sigma \rangle$, with $V = \{v_1, \ldots, v_n\}$. Let $I = \langle \Sigma^\uparrow, \{P_i c\} \rangle$. Let $\langle \mathcal{U}, \mathcal{F} \rangle$ be the information state generated by I.*
For $\mathfrak{A} \in \mathcal{U}$, we have $\mathfrak{A} \in \mathcal{F}$ if and only if $\mathfrak{A}$ models an acceptable exception set of v_i for c and makes $P_i c$ true.

PROOF: Let $\mathfrak{A} \in \mathcal{U}$. Then $\mathfrak{A} \in \mathcal{F}$ if and only if $\mathfrak{A} \models P_i c$. This makes the right-to-left direction trivial, so now assume $\mathfrak{A} \models P_i c$.
Choose $X \subseteq \Sigma$ such that $\mathfrak{A}$ models X for c. (By definition there is exactly one way to do this.) The only thing left to show is that X is an acceptable exception set of v_i. Let $Y \in D(v_i)$. We must show that $\exists \delta \in Y : \delta \in X$.

By Proposition 6.6.9,

$$(\Sigma^\uparrow)^{WI} \models \forall x \left(P_i x \to \bigvee_{\alpha \in Y} Ab_{\alpha^\uparrow} x \right)$$

Therefore $\mathfrak{A} \models \bigvee_{\alpha \in Y} Ab_{\alpha^\uparrow} c$. Hence, there is some $\alpha \in Y$ such that $\mathfrak{A} \models Ab_\alpha c$. Since $\mathfrak{A}$ models X for c, this implies implies $\delta \in X$.

Next we show that every minimal exception set is in fact represented by at least one model.

6.6.12. THEOREM. *Let $\Sigma^\uparrow$ be the lift of the inheritance network $\langle V, \Sigma \rangle$, with $V = \{v_1, \ldots, v_n\}$. Let $I = \langle \Sigma^\uparrow, \{P_i c\} \rangle$. Let $\langle \mathcal{U}, \mathcal{F} \rangle$ be the information state generated by I.*
For every minimal exception set X relative to v_i there is a model in $\mathcal{F}$ which models X for c.

PROOF: Let X be a minimal exception set relative to v_i. Construct $\mathfrak{A}$ as follows:

- For the domain, take the same domain as that of some other model in $\mathcal{F}$.
- Let $\mathfrak{A} \models P_i c$ and let $\mathfrak{A}$ model X for c.
- For all P_j, let $\mathfrak{A} \models P_j c$ if and only if there is a positive path from v_i to v_j that does not contain an element of X.
- For all y other than c in its domain and for all P_j, let $\mathfrak{A} \models \neg P_j y$.

We need to show that $\mathfrak{A} \in \mathcal{U}$. (The previous proposition then implies $\mathfrak{A} \in \mathcal{F}$.) For this it suffices to show that $\mathfrak{A} \models (\Sigma^{\uparrow})^{WI}$. (Since $(\Sigma^{\uparrow})^{\in}$ and $(\Sigma^{\uparrow})^{WI}$ have the same models, $\mathcal{U}$ consists exactly of all models of $(\Sigma^{\uparrow})^{WI}$.)
For elements other than c, the predicate assignments are trivially consistent with all rules and exemption clauses in $(\Sigma^{\uparrow})^{WI}$. For c, we first look at the rules in $\Sigma^{\uparrow}$.
Rules in $\Sigma^{\uparrow}$: So let $\phi \in \Sigma^{\uparrow}$, where

$$\phi = \forall x((P_j x \wedge \neg Ab_{P_j P_k} x) \rightarrow P_k x)$$

We may assume that $\mathfrak{A} \models P_j c \wedge \neg Ab_{P_j P_k} c$. (Otherwise c is trivially consistent with the rule.) Thus there is a positive path from v_i to v_j that does not contain an element of X, and the arrow from v_j to v_k is not in X. Therefore there is also such a path from v_i to v_k, and thus $P_k c$.
For negative rules, again take $\phi \in \Sigma^{\uparrow}$ but now with

$$\phi = \forall x((P_j x \wedge \neg Ab_{P_j P_k} x) \rightarrow \neg P_k x).$$

Again we may assume that $\mathfrak{A} \models P_j c \wedge \neg Ab_{P_j P_k} c$. Thus there is a negative path from v_i to v_k containing no element of X. Suppose there is also a positive path from v_i to v_k, and let Y be the union of these two paths. Then Y is a conflicting set relative to v_i. Since X is a minimal exception set relative to v_i, some $\alpha \in Y$ must be in X. Since the negative path had no such overlap, this α must be part of the positive path.
As we've shown that every such positive path contains an element of X, it follows by construction that $\mathfrak{A} \models \neg P_k c$. Therefore the valuation for c is consistent with this rule.
Exemption clauses in $(\Sigma^{\uparrow})^{WI}$: Suppose $\theta \in (\Sigma^{\uparrow})^{WI}$, where

$$\theta = \forall x(P_j x \rightarrow \bigvee_{\alpha \in \Delta} Ab_{\alpha^{\uparrow}} x)$$

By Proposition 6.6.9, $Y \in D(v_j)$ for some $Y \subseteq \Delta$. We may assume that $P_j c$. Therefore there is a positive path from v_i to v_j, and thus $Y \in D(v_i)$. Since X is a minimal exception set relative to v_i, it follows that there is some $\alpha' \in Y$ for which $\alpha' \in X$. By construction, $\mathfrak{A} \models \neg Ab_{\alpha'} c$, and therefore c is consistent with θ.

Finally, we show that minimal exception sets correspond to optimal models.

6.6.13. THEOREM. *Let $\Sigma^{\uparrow}$ be the lift of the inheritance network $\langle V, \Sigma\rangle$, with $V = \{v_1, \ldots, v_n\}$. Let $I = \langle \Sigma^{\uparrow}, \{P_i c\}\rangle$. Let $\langle \mathcal{U}, \mathcal{F}\rangle$ be the information state generated by I.*
Every optimal model of $\mathcal{F}$ models an minimal exception set of v_i for c, and every minimal exception set of v_i has a model (for c) which is optimal in $\mathcal{F}$.

PROOF: For the first part, let $\mathfrak{A}$ be optimal in $\mathcal{F}$. Per Proposition 6.6.11, $\mathfrak{A}$ models some acceptable exception set X of v_i for c. Assume towards contradiction that X is not a minimal exception set of v_i, and that $X' \subset X$ is. Per Theorem 6.6.12, there is a model $\mathfrak{A}' \in \mathcal{F}$ which models X'.
Now construct model $\mathfrak{A}''$ to be exactly like $\mathfrak{A}$ except that when evaluating predicates (including abnormality predicates) applied to c, it uses the same evaluation as $\mathfrak{A}'$.[15] Now the abnormality predicates made true by $\mathfrak{A}''$ are a strict subset of those made true by $\mathfrak{A}$. Thus it is strictly more normal than $\mathfrak{A}$, which is therefore not optimal.

For the second part, let X be a minimal exception set of v_i. By Theorem 6.6.12, there are models in $\mathcal{F}$ which model X for c. Pick $\mathfrak{A}$ to be a model which is optimal amongst those models. Suppose $\mathfrak{B} \in \mathcal{F}$ is at least as normal as $\mathfrak{A}$.
By Proposition 6.6.11, $\mathfrak{B}$ models some acceptable exception set Y of v_i. Since $\mathfrak{B}$ is at least as normal as $\mathfrak{A}$, we have $Y \subseteq X$. Since X is minimal, this means $Y = X$. As we picked $\mathfrak{A}$ to be optimal amongst those that model X, this means $\mathfrak{A}$ is at least as normal as $\mathfrak{B}$.
Thus, $\mathfrak{A}$ is an optimal model.

Having proven the correspondence between optimal models and minimal exception sets, the last step in the completeness proof is to go from these models to the allowable inferences as defined in Definition 6.4.8. After doing this in the next theorem, the result we are after follows almost as a corollary.

6.6.14. THEOREM. *Let $\Sigma^{\uparrow}$ be the lift of the inheritance network $\langle V, \Sigma\rangle$, with $V = \{v_1, \ldots, v_n\}$. Let $I = \langle \Sigma^{\uparrow}, \{P_i c\}\rangle$. Let $\langle \mathcal{U}, \mathcal{F}\rangle$ be the information state generated by I. Let X be a minimal exception set of v_i.*
Then:

1. *If there is a positive path from v_i to v_j which doesn't contain any element of X, then every $\mathfrak{A} \in \mathcal{F}$ which models X makes $P_j c$ true.*
2. *If there is a negative path from v_i to v_j which doesn't contain any element of X, then every $\mathfrak{A} \in \mathcal{F}$ which models X makes $\neg P_j c$ true.*
3. *If X is not an acceptable exception set of v_j, then every $\mathfrak{A} \in \mathfrak{F}$ which models X makes $\neg P_j c$ true.*

[15] Showing that $\mathfrak{A}'' \in \mathcal{F}$ is fairly trivial and left to the reader.

4. *If every $\mathfrak{A} \in \mathcal{F}$ which models X makes $P_j c$ true, then there is a positive path from v_i to v_j which doesn't contain any element of X.*

5. *If every $\mathfrak{A} \in \mathcal{F}$ which models X makes $\neg P_j c$ true, then either there is a negative path from v_i to v_j which doesn't contain any element of X or X is not an acceptable exception set of v_j.*

PROOF: Point 1 and 2 are trivial by repeated modus ponens/tollens. Point 3 is almost as easy: If X is not an acceptable exception set of v_j, then there is some $Y \in D(v_j)$ such that $X \cap Y = \emptyset$. Since $Y \in D(v_j)$, $(\Sigma^{\uparrow})^{WI} \models \forall x(P_j x \rightarrow \bigvee_{\alpha \in Y} Ab_\alpha x)$ (Proposition 6.6.9). Suppose $\mathfrak{A} \in \mathfrak{F}$ models X. Since $X \cap Y = \emptyset$, $\mathfrak{A}$ does not make $\bigvee_{\alpha \in Y} Ab_\alpha c$ true. Therefore $\mathfrak{A} \models \neg P_j c$.
For point 4, suppose every $\mathfrak{A} \in \mathcal{F}$ which models X makes $P_j c$ true. Construct $\mathfrak{B}$ as follows:

- For the domain, take the same domain as that of some other model in $\mathcal{F}$.
- Let $\mathfrak{B} \models P_i c$ and let $\mathfrak{B}$ model X for c.
- For all P_j, let $\mathfrak{B} \models P_j c$ if and only if there is a positive path from v_i to v_j that does not contain an element of X.
- For all y other than c in its domain and for all P_j, let $\mathfrak{B} \models \neg P_j y$.

We have shown in the proof of Theorem 6.6.12 that $\mathfrak{B} \in \mathcal{F}$. Thus, by construction there is a positive path from v_i to v_j that does not contain an element of X.
For point 5, suppose every $\mathfrak{A} \in \mathcal{F}$ which models X makes $\neg P_j c$ true. Now construct $\mathfrak{B}'$ to be as $\mathfrak{B}$ except that $\mathfrak{B}' \models P_j c$. Then $\mathfrak{B}'$ is not in $\mathcal{F}$, and more specifically $\mathfrak{B}' \not\models (\Sigma^{\uparrow})^{WI}$. Pick $\phi \in (\Sigma^{\uparrow})^{WI}$ such that $\mathfrak{B}' \models \neg\phi$. A number of cases arise, depending on ϕ.

a $\phi = \forall x(P_k x \wedge \neg Ab_\phi x \rightarrow \neg P_j x)$ for some k, with $\mathfrak{B}' \models P_k c \wedge \neg Ab_\phi c$. In this case, there is a negative path from v_i to v_j (via v_k) that does not contain an element of X.

b $\phi = \forall x(P_j x \wedge \neg Ab_\phi x \rightarrow \neg P_k x)$ for some k, with $\mathfrak{B}' \models P_k c \wedge \neg Ab_\phi c$. In this case too, there is a negative path from v_i to v_j (via v_k using modus tollens at the end) that does not contain an element of X.

c $\phi = \forall x(P_j x \rightarrow \bigvee_{\delta \in \Delta} Ab_\delta x)$ for some Δ, with $\mathfrak{B}' \models \neg \bigvee_{\delta \in \Delta} Ab_\delta c$. Then it follows that $X \cap \Delta = \emptyset$, and therefore by Proposition A.11, $Y \in D(v_j)$ for some $Y \subseteq \Delta$. Since X contains no element of Δ, it contains no element of this Y. Therefore X is not an acceptable exception set of v_j.

d $\phi = \forall x(P_j x \wedge \neg Ab_\phi \rightarrow P_k x)$ for some k, with $\mathfrak{B}' \models \neg P_k c \wedge \neg Ab_\phi c$. In this case, change the model one step further, making $P_k c$ true. As the new model still

cannot be in $\mathfrak{F}$, find a new ϕ' it now contradicts.
If this ϕ' is like in case a or b, then there is still a negative path, which is just one step longer. (Recall that a negative path can go through any amount of positive arrows 'in the wrong direction' at the end.) If it is like case c, then the Y which is found is also part of v_j. If it is itself like case d, then we continue to proceed in the same way.
Since no amount of making predicates true will make the model part of $\mathfrak{F}$, going on long enough will lead to a ϕ' of one of the first three forms. The only potential complication in this induction is the possibility that we are led to a formula like type a or b where P_k is true merely because of a change we made to the model. In this case there is a negative path from v_j to itself of which no element is in X. Since this path is a contradicting set relative to v_j, it follows that X is not an acceptable exception set of v_j.

6.6.15. Theorem. ***(Soundness-Completeness Theorem)*** *Suppose $\Sigma^{\uparrow}$ is coherent. Then $v_i \vdash_{\Sigma} v_j$ if and only if $\Sigma^{\uparrow}, \{P_i c\} \models_d P_j c$, and $v_i \vdash_{\Sigma} \neg v_j$ if and only if $\Sigma^{\uparrow}, \{P_i c\} \models_d \neg P_j c$.*

Proof: Let $\langle \mathcal{U}, \mathcal{F} \rangle$ correspond to $\langle \Sigma^{\uparrow}, \{P_i c\} \rangle$.

- By definition $v_i \vdash_{\Sigma} v_j$ holds if and only if for every minimal exception set X of v_i, there is a positive path Y from v_i to v_j with $X \cap Y = \emptyset$. Likewise, $v_i \vdash_{\Sigma} \neg v_j$ holds iff either for every such X there is a negative path Y like that, or no such X is an acceptable exception set of v_j.
- By Theorem 6.6.14, this is true iff each $\mathfrak{A} \in \mathcal{F}$ which models a minimal exception set of v_i makes $P_j c$ true ($\neg P_j c$ for the negative case).
- By Theorem 6.6.13, this is true iff each optimal model in $\mathcal{F}$ makes $P_j c$ ($\neg P_j c$) true.
- By definition this is true iff $\Sigma^{\uparrow}, \{P_i c\} \models_d P_j c$ ($\neg P_j c$).

6.6.3 Coherence

6.6.16. Theorem. *Let $\langle V, \Sigma \rangle$ be an inheritance network with $V = \{v_1, \ldots, v_n\}$. Then $\Sigma^{\uparrow}$ is incoherent if and only if there is some v_i such that $\emptyset \in D(v_i)$.*

Proof: $\Sigma^{\uparrow}$ is incoherent if and only if there is some P_i such that $\Sigma^{\uparrow WI} \cup \{\exists x P_i x\}$ is inconsistent. This is if and only if $(\Sigma^{\uparrow})^{WI} \models \forall x \neg P_i x$ for some P_i. By the convention on empty disjunctions, $\forall x \neg P_i x$ is equivalent to $\forall x (P_i x \to \bigvee_{\emptyset} Ab_{\alpha} x)$. Therefore the last step follows from Proposition 6.6.9.

6.6.17. Proposition. *Let $\langle V, \Sigma \rangle$ be an inheritance network without strict arrows. If $\emptyset \in d(x)$, then there are some z and some $y \approx x, y' \approx x$ such that Σ contains a positive arrow from y to z and a negative arrow from y' to z.*

PROOF: Suppose $\emptyset \in d(x)$. Then there is some minimal conflicting set $X \subseteq Ess_\Sigma(x)$. We may assume without loss of generality that X is the union of a positive path $\{xy_1, y_1y_2, \ldots, y_mz\}$ and a negative path $\{xy'_1, y'_1y'_2, \ldots, y'_nz^-\}$. Since $y_mz \in X$, it follows that $y_mz \in Ess_\Sigma(x)$. Therefore $x \approx y_m$. Analogously, $x \approx y'_n$.

6.6.18. DEFINITION. The vertex x *semi-strictly implies* (*semi-strictly refutes*) y if there is a positive (negative) path from x to y where every arrow after the first is strict.

6.6.19. PROPOSITION. *Let $\langle V, \Sigma\rangle$ be an inheritance network. If $\emptyset \in d(x)$, then there are some z and some $y \approx x, y' \approx x$ such that y semi-strictly implies z and y' semi-strictly refutes z.*

PROOF: Suppose $\emptyset \in d(x)$. Then there is some minimal conflicting set $X \subseteq Ess_\Sigma(x)$. We may assume without loss of generality that X is the union of a positive path $\{xy_1, y_1y_2, \ldots, y_nz\}$ and a negative path $\{xy'_1, y'_1y'_2, \ldots, y'_nz^-\}$ (where some of these may actually be strict).
Pick the smallest i for which y_i strictly implies z.[16] Since $y_{i-1}y_i \in X$, it follows that $y_{i-1}y_i \in Ess_\Sigma(x)$. But by construction $y_{i-1}y_i$ is not strict. Therefore $y_{i-1} \approx x$.
Analogously, $y'_{j-1} \approx x$ when we pick the smallest j for which y'_j strictly refutes z. (If no y'_j does so, pick $j = n + 1$ instead.) Now let $y = y_{i-1}$, $y' = y'_{j-1}$. By construction, y semi-strictly implies z and y' semi-strictly refutes it.

[16] For y_{i-1} to exist we must assume x does not semi-strictly imply z, but this is safe because if it does then we can pick $y = x$ and skip the next couple of steps in the proof.

Bibliography

Barwise, J. and Cooper, R.: 1981, Generalized quantifiers and natural language, *Linguistics and Philosophy* **4**, 159–219. 10.1007/BF00350139.
URL: `http://dx.doi.org/10.1007/BF00350139`

Bastiaanse, H.: 2011, The rationality of round interpretation, *in* R. Nouwen, R. van Rooij, U. Sauerland and H.-C. Schmitz (eds), *Vagueness in Communication*, Vol. 6517 of *Lecture Notes in Computer Science*, Springer, pp. 37–50.

Bastiaanse, H. A.: 2013, The intensional many - conservativity reclaimed, *Journal of Philosophical Logic* pp. 1–19.
URL: `http://dx.doi.org/10.1007/s10992-013-9301-7`

Caminada, M.: 2008, On the issue of contraposition of defeasible rules, *Frontiers in Artificial Intelligence and Applications* **172**, 109.

Cohen, A.: 2001, Relative readings of many, often, and generics, *Natural Language Semantics* **9**, 41–67. 10.1023/A:1017913406219.
URL: `http://dx.doi.org/10.1023/A:1017913406219`

Cuzzolin, P. and Lehmann, C.: 2004, Comparison and gradation, *Booïj, G., Mugdan, J., SS, Lehmann, C.(eds.) Morphologie. Ein internationales Handbuch zur Flexion und Wortbiltung* **2**, 1212–1220.

Dehaene, S. and Mehler, J.: 1992, Cross-linguistic regularities in the frequency of number words, *Cognition* **43**(1), 1–29.

Fernando, T. and Kamp, H.: 1996, Expecting many, *Proceedings of SALT 6*, pp. 53–68.

Hofstadter, D. R.: 1979, *Gödel, Escher, Bach: An Eternal Golden Braid*, Basic Books, New York.

Horty, J. F.: 2002, Skepticism and floating conclusions, *Artificial Intelligence* **135**(12), 55 – 72.

Horty, J. F., Thomason, R. M. and Touretzy, D. S.: 1990, A skeptical theory of inheritance in nonmonotonic semantic networks, *Artificial Intelligence* **42**, 311–348.

Ito, R. and Tagliamonte, S.: 2003, Well weird, right dodgy, very strange, really cool: Layering and recycling in English intensifiers, *Language in Society* **32**(02), 257–279.

Jansen, C. J. M. and Pollmann, M. M. W.: 2001, On round numbers: Pragmatic aspects of numerical expressions, *Journal of Quantitative Linguistics* **8**(3), 187–201.

Keenan, E. L. and Stavi, J.: 1986, A semantic characterization of natural language determiners, *Linguistics and Philosophy* **9**, 253–326. 10.1007/BF00630273.
URL: `http://dx.doi.org/10.1007/BF00630273`

Kennedy, C.: 2007, Vagueness and grammar: the semantics of relative and absolute gradable adjectives, *Linguistics and Philosophy* **30**(1), 1–45.

Krifka, M.: 2007, Approximate interpretation of number words: A case for strategic communication, *Cognitive foundations of interpretation* pp. 111–126.

Lappin, S.: 2000, An intensional parametric semantics for vague quantifiers, *Linguistics and Philosophy* **23**, 599–620. 10.1023/A:1005638918877.
URL: `http://dx.doi.org/10.1023/A:1005638918877`

Lorenz, G.: 2002, Really worthwhile or not really significant? a corpus-based approach to the delexicalization, *New reflections on grammaticalization* **49**, 143.

Makinson, D. and Schlechta, K.: 1991, Floating conclusions and zombie paths: two deep difficulties in the "directly skeptical" approach to defeasible inheritance nets, *Artificial Intelligence* **48**, 199–209.

McCarthy, J.: 1987, Circumscription: A form of non-monotonic reasoning, *in* M. L. Ginsberg (ed.), *Readings in Nonmonotonic Reasoning*, Kaufmann, Los Altos, CA, pp. 145–151.

McCarthy, J.: 1990, Applications of circumscription to formalizing common sense knowledge, *in* V. Lifschitz (ed.), *Formalizing Common Sense: Papers by John McCarthy*, Ablex Publishing Corporation, Norwood, New Jersey, pp. 198–225.
URL: `citeseer.ist.psu.edu/mccarthy86applications.html`

Méndez-Naya, B.: 2003, On intensifiers and grammaticalization: The case of swathe, *English Studies* **84**(4), 372–391.

Moss, L.: 2010, Intersecting adjectives in syllogistic logic, *in* C. Ebert, G. Jger and J. Michaelis (eds), *The Mathematics of Language*, Vol. 6149 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 223–237.

Moss, L.: 2011, Syllogistic logic with comparative adjectives, *Journal of Logic, Language and Information* **20**, 397–417. 10.1007/s10849-011-9137-x.
URL: `http://dx.doi.org/10.1007/s10849-011-9137-x`

Moss, L. S.: 2008, Completeness theorems for syllogistic fragments, *in* F. Hamm and S. Kepser (eds), *Logics for Linguistic Structures*, Vol. 29 of *Trends in Linguistics*, De Gruyter Mouton, pp. 143–173.

Partee, B.: 2010, Privative adjectives: Subsective plus coercion, *in* U. R. Rainer Buerle and T. E. Zimmermann (eds), *Presuppositions and Discourse: Essays Offered to Hans Kamp*, Emerald Group Publishing Limited, pp. 273–283.

Peters, H.: 1994, Degree adverbs in early modern english, *Studies in Early Modern English* pp. 269–288.

Potts, C.: 2008, Interpretive Economy, Schelling Points, and evolutionary stability, *Ms., University of Massachusetts Amherst* .

Prakken, H.: 2002, Intuitions and the modelling of defeasible reasoning: some case studies, *arXiv preprint cs/0207031* .

Solt, S.: 2009, *The Semantics of Adjectives of Quantity*, PhD thesis, The City University of New York.

Stoffel, C.: 1901, *Intensives and down-toners: A study in English adverbs*, number 1 in *Anglistische Forschungen*, C. Winter's universitätsbuchhandlung.

Szpilrajn, E.: 1930, Sur l'extension de l'ordre partiel, *Fundamenta Mathematicae* **16**(1), 386–389.

Tanaka, T.: 2003, Semantic interpretation of many, *JELS 20: Papers from the Twentieth National Conference of the English Linguistic Society of Japan*, pp. 188–197.

Tribushinina, E.: 2008, *Cognitive Reference Points: semantics beyond the prototypes in adjectives of space and colour*, PhD thesis, Leiden University.

van Benthem, J.: 1984, Questions about quantifiers, *Journal of Symbolic Logic* pp. 443–466.

van Benthem, J.: 1991, *The Logic of Time: A Model-Theoretic Investigation into the Varieties of Temporal Ontology and Temporal Discourse*, Vol. 156 of *Studies in Epistemology, Logic, Methodology, and Philosophy of Science*, Kluwer Academic Publishers.

Westerståhl, D.: 1985, Logical constants in quantifier languages, *Linguistics and Philosophy* **8**, 387–413. 10.1007/BF00637410.
URL: `http://dx.doi.org/10.1007/BF00637410`

Westerståhl, D.: 2007, Quantifiers in formal and natural languages, *in* D. Gabbay and F. Guenthner (eds), *Handbook of Philosophical Logic*, Vol. 14 of *Handbook of Philosophical Logic*, Springer Netherlands, pp. 223–338.
URL: `http://dx.doi.org/10.1007/978-1-4020-6324-4_4`

Wierzbicka, A.: 1990, The meaning of color terms: semantics, culture, and cognition, *Cognitive Linguistics (includes Cognitive Linguistic Bibliography)* **1**(1), 99–150.

Wilson, D. and Sperber, D.: 2002, Truthfulness and relevance, *Mind* **111**(443), 583–632.

Index

$a \vdash_\Sigma b$, 102
$d(a)$, 102
$D(a)$, 102
$Ess_\Sigma(a)$, 101
$\Delta \models_c \varphi$, 91
$\varphi \approx_\Sigma \psi$, 110
$\Sigma, \Pi \models_d \varphi$, 93
$\Sigma^\in$, 95
$\Sigma^\uparrow$, 108
Σ^W, 110
Σ^I, 110

abnormality predicate, 90
acceptable exception set, 102
additivity, 43–45
  left-additivity, 44
  right-additivity, 43
adjectives
  evaluative, 79
  gradable, 67–88
  subsective, 51–66
anti-reflexivity, 45

circumscription, 90–100
coherence (circumscription), 99–100
commensurability, 75–82
compliance (circumscription), 93
conflicting set, 101
Conservative ISOM, 42
Conservativity, 19, 22
convexity, 71

default rules, 89–123
Directedness, 71
Double Diamond, 104–105

excluded middle, 46, 49
exemption clause, 94
Exemption Principle, 94–95
EXT, 25
Extension (property), *see* EXT

floating conclusions, 105–106

Game Theory, 8–11
  Evolutionary Game Theory, 14–16
Generalized Quantifiers, 19

inheritance networks, 100–109
Inheritance Property, 96–97
Intensional Conservativity, 22
Intensional EXT, 25
Intensional ISOM, 27
intensionality, 19–49
ISOM, 27

language evolution, 12–16

many, 19–49
minimal conflicting set, *see* conflicting set

minimal exception set, *see* acceptable exception set
Minimal Requirement (circumscription), 94–95
monotonicity, 71

No Cycles, 72
  consequences, 73
  No $\Delta-$Cycles, 75
  No $P, \overline{Q}$-Cycles, 77
No Overlap, 77
No Reversal, 70
NUM, 39
Numericality, *see* NUM

optimal model (circumscription), 91

P, Q-Reversal, 79
path (inheritance networks), 100
probability, 23, 30, 43
  conditional probability, 6–8, 16–17

rationality, *see* Game Theory
reflexivity, 45
Reverse Reading, 24
round numbers, 5–17

semi-strict implication (inheritance networks), 109
syllogistics, 51–88

very, 13

Samenvatting

Dit proefschrift is (na de introductie) verdeeld in vijf afzonderlijke hoofdstukken, die elk betrekking hebben op een onderwerp dat gerelateerd is aan vaagheid. Deze hoofdstukken zijn aangepaste versies van manuscripten die uiteindelijk zullen verschijnen in diverse vakbladen.

Het tweede hoofdstuk gaat over de neiging het gebruik van bepaalde getallen te interpreteren als 'rond', dat wil zeggen als een uitspraak die niet alleen dat exacte getal omvat maar ook andere getallen die daar zo dichtbij liggen dat ze bij het afronden daarnaar afgerond zouden worden. Op basis van speltheorie en Bayesiaanse statistiek laat dit hoofdstuk zien dat de neiging tot ronde interpretatie in voorkomende gevallen verdedigd kan worden als een rationale keuze.
Hetzelfde mechanisme draagt ook bij aan een 'losse' interpretatie van andere woorden. Als een dergelijke losse interpretatie vervolgens standaard wordt -wat aannemelijker is dan bij getallen omdat daar de strikte interpretatie nooit echt uit het oog verloren kan worden- kan vervolgens met die lossere standaard hetzelfde gebeuren. Als dit blijft gebeuren is kan een woord dat eerst niet vaag was dat na verloop van tijd wel worden. Er bestaan voorbeelden van woorden die zich in de loop van enkele eeuwen daadwerkelijk op deze manier ontwikkeld hebben. Derhalve biedt dit mechanisme een gedeeltelijke verklaring voor de oorsprong van vaagheid in natuurlijke taal, waarbij het ook suggereert dat iedere natuurlijke taal mettertijd sporen van vaagheid zal vertonen.

Het derde hoofdstuk heeft betrekking op het Engelse woord *many* (veel/vele), een vaag woord. In de theorie van Generalized Quantifiers was *many* lange tijd een probleemgeval, aangezien er geen geschikte formele interpretatie leek te zijn die voldoet aan de eigenschap Conservativiteit, een eigenschap waar vrijwel alle andere determinators in de natuurlijke taal wel aan voldoen.
Dit hoofdstuk bepleit dat er een probleem is met een van de belangrijkste voorbeelden waaruit lang geconcludeerd is dat *many* een probleemgeval is, namelijk

dat *many* een intensionele aanpak vereist die verder niet of nauwelijks voorkomt in de bestaande theorie. Door het gebruiken van een intensioneel systeem met een intensionele notie van Conservativiteit vormt *many* niet langer een probleem. Verder gaat dit hoofdstuk in op intensionele versies van enkele andere belangrijke eigenschappen, geeft het een algemene vorm waaronder quantoren automatisch aan die eigenschappen voldoen, en besteedt het enige aandacht aan de logische eigenschappen van *many* in het bijzonder en intensionele quantoren in het algemeen.

Het vierde hoofdstuk omvat een syllogistische logica voor subsectieve bijvoeglijk naamwoorden. Het vijfde hoofdstuk gebruikt deze logica om de eigenschappen te onderzoeken van zogenaamde gradeerbare bijvoeglijk naamwoorden, een categorie die veel standaardvoorbeelden van vaagheid omvat. Het laat zien dat, wanneer deze categorie gedefinieerd wordt als bestaande uit die subsectieve bijvoeglijk naamwoorden die gebaseerd zijn op een zwakke ordening, een karakterisatie mogelijk is enkel op basis van de extensies, dus zonder vooraf die ordening te hoeven kennen.
Verder wordt het een formeel concept geintroduceerd van *commensurabiliteit* van groepen gradeerbare bijvoeglijk naamwoorden; losjes uitgedrukt is een dergelijke groep commensurabel als ze allen gebaseerd zijn op dezelfde onderliggen ordening. Met behulp van dit concept worden antoniemen, bijvoeglijk naamwoorden die betrekking hebben op persoonlijke voorkeuren, versterkende en verzwakkende bijvoeglijke bepalingen en booleanse combinaties in het raamwerk ingepast. Voorts wordt er nog besproken hoe het systeem uitgebreid kan worden voor gevallen waar vaagheid een belangrijke rol speelt.

Hoofdstuk zes gaat in op een vorm van vaagheid dit niet direct te zien is, via het volgende vraagstuk: wat wordt er precies bedoeld in algemene constructies zoals "Vogels (kunnen) vliegen"? Dit soort constructies, ook wel aangeduid als *default* regels, kunnen niet geinterpreteerd worden als een eenvoudige universele quantificatie. Pinguins en verscheidene andere vogelsoorten kunnen niet vliegen, maar die tegenvoorbeelden worden niet geacht een probleem te vormen voor de waarheid van de algemene uitspraak dat vogels vliegen.
Evenmin kunnen ze geinterpreteerd worden als gelijk aan *de meeste*. Uit “Het is niet zo dat de meeste Nederlanders blond zijn” volgt “De meeste Nederlanders zijn niet blond”, maar uit “Het is niet zo dat Nederlanders blond zijn” (in algemene zin) volgt niet “Nederlanders zijn niet blond”. Bovendien volgt uit default regels van de vorm “A’s zijn B” en “A’s zijn C” dat “A’s zijn B en C”, terwijl uit “De meeste A zijn B” en “De meeste A zijn C” niet afgeleid kan worden dat “De meeste A zijn B en C”.
Een geschiktere interpretatie van "Vogels vliegen" zou liggen in de trant van "Alle normale vogels vliegen" of "Alle goede voorbeelden van vogels vliegen", zinnen

waarvan de vaagheid stukken evidenter is. De manier waarop we in hoofdstuk zes naar de betekenis van dit soort zinnen kijken is door te kijken hoe iemand die ze voor waar aanneemt zou moeten redeneren. De belangrijkste vraag in deze is wat er geconcludeert mag worden in situaties waar verschillen combinaties van default regels tot tegenstrijdige conclusies kunnen leiden. Een enkel onderliggend principe over de betekenis van zulke regels leidt uiteindelijk tot een systematisch antwoord op deze vraag.

Dit antwoord wordt eerst gegeven in de vorm van een model-theoretische semantiek en daarna in termen van overervingsnetten, een simpelere methode waarbij geen specifieke modellen of domeinen van objecten nodig zijn. Deze tweede vorm kan gevangen worden in een handzaam algoritme om in voorkomende gevallen te bepalen waar uitzonderingen gemaakt dienen te worden. Tot slot wordt bewezen dat de twee methodes, waar beide mogelijk, tot dezelfde resultaten leiden.

Abstract

This thesis is divided into five separate chapters, each of which deals with an issue related to vagueness. These chapters are adaptations of manuscripts to be published as papers in various journals. Their abstracts as they (will) appear in these journals are repeated below (but with 'paper' replaced by 'chapter') so as to comply with standard conventions.

Chapter 1 - The Rationality of Round Interpretation Expanding on a point made by Krifka (Krifka 2007, p.7-8), we show that the fact that a round number has been used significantly increases the posterior probability that that number was intended as an approximation.
This increase should typically be enough to make assuming that an approximation was indeed intended a rational choice, and thereby helps explain why round numbers are often seen as simply having an approximate meaning.
Generalization into non-number words is also discussed, resulting in a possible origin of (some) vagueness.

Chapter 2 - The Intensional Many Following on Westerståhl's argument that many is not Conservative (Westerståhl 1985), I propose an intensional account of Conservativity as well as intensional versions of EXT and Isomorphism closure. I show that an intensional reading of many can easily possess all three of these, and provide a formal statement and proof that they are indeed proper intensionalizations.
It is then discussed to what extent these intensionalized properties apply to various existing readings of many.

Chapter 3 - A Syllogistic for Subsective Adjectives I introduce a syllogistic logic for reasoning about subsective adjectives, and prove that it is complete relative to an appropriate class of models.

Chapter 4 - A Syllogistic Characterization of Gradable Adjectives Building on an existing syllogistics for subsective adjectives (Chapter 3), I show that if gradable adjectives are defined as those subsective adjectives which are based on a weak order, this notion can be characterized in a natural logic without prior access to that weak order.
Furthermore, generalizing this characterization allows for the characterization of a useful notion of commensurability of groups of adjectives into a single scale.

Chapter 5 - Making the Right Exceptions Conflicts among default rules are very common. This chapter provides a principled answer to the question of how to deal with them. It does so in two ways: semantically within a circumscriptive theory, and syntactically by supplying an algorithm for inheritance networks. Arguments that can be expressed in both frameworks are valid on the circumscriptive account if and only if the inheritance algorithm has a positive outcome.

Titles in the ILLC Dissertation Series:

ILLC DS-2009-01: **Jakub Szymanik**
Quantifiers in TIME and SPACE. Computational Complexity of Generalized Quantifiers in Natural Language

ILLC DS-2009-02: **Hartmut Fitz**
Neural Syntax

ILLC DS-2009-03: **Brian Thomas Semmes**
A Game for the Borel Functions

ILLC DS-2009-04: **Sara L. Uckelman**
Modalities in Medieval Logic

ILLC DS-2009-05: **Andreas Witzel**
Knowledge and Games: Theory and Implementation

ILLC DS-2009-06: **Chantal Bax**
Subjectivity after Wittgenstein. Wittgenstein's embodied and embedded subject and the debate about the death of man.

ILLC DS-2009-07: **Kata Balogh**
Theme with Variations. A Context-based Analysis of Focus

ILLC DS-2009-08: **Tomohiro Hoshi**
Epistemic Dynamics and Protocol Information

ILLC DS-2009-09: **Olivia Ladinig**
Temporal expectations and their violations

ILLC DS-2009-10: **Tikitu de Jager**
"Now that you mention it, I wonder...": Awareness, Attention, Assumption

ILLC DS-2009-11: **Michael Franke**
Signal to Act: Game Theory in Pragmatics

ILLC DS-2009-12: **Joel Uckelman**
More Than the Sum of Its Parts: Compact Preference Representation Over Combinatorial Domains

ILLC DS-2009-13: **Stefan Bold**
Cardinals as Ultrapowers. A Canonical Measure Analysis under the Axiom of Determinacy.

ILLC DS-2010-01: **Reut Tsarfaty**
Relational-Realizational Parsing

ILLC DS-2010-02: **Jonathan Zvesper**
Playing with Information

ILLC DS-2010-03: **Cédric Dégremont**
The Temporal Mind. Observations on the logic of belief change in interactive systems

ILLC DS-2010-04: **Daisuke Ikegami**
Games in Set Theory and Logic

ILLC DS-2010-05: **Jarmo Kontinen**
Coherence and Complexity in Fragments of Dependence Logic

ILLC DS-2010-06: **Yanjing Wang**
Epistemic Modelling and Protocol Dynamics

ILLC DS-2010-07: **Marc Staudacher**
Use theories of meaning between conventions and social norms

ILLC DS-2010-08: **Amélie Gheerbrant**
Fixed-Point Logics on Trees

ILLC DS-2010-09: **Gaëlle Fontaine**
Modal Fixpoint Logic: Some Model Theoretic Questions

ILLC DS-2010-10: **Jacob Vosmaer**
Logic, Algebra and Topology. Investigations into canonical extensions, duality theory and point-free topology.

ILLC DS-2010-11: **Nina Gierasimczuk**
Knowing One's Limits. Logical Analysis of Inductive Inference

ILLC DS-2010-12: **Martin Mose Bentzen**
Stit, Iit, and Deontic Logic for Action Types

ILLC DS-2011-01: **Wouter M. Koolen**
Combining Strategies Efficiently: High-Quality Decisions from Conflicting Advice

ILLC DS-2011-02: **Fernando Raymundo Velazquez-Quesada**
Small steps in dynamics of information

ILLC DS-2011-03: **Marijn Koolen**
The Meaning of Structure: the Value of Link Evidence for Information Retrieval

ILLC DS-2011-04: **Junte Zhang**
System Evaluation of Archival Description and Access

ILLC DS-2011-05: **Lauri Keskinen**
Characterizing All Models in Infinite Cardinalities

ILLC DS-2011-06: **Rianne Kaptein**
Effective Focused Retrieval by Exploiting Query Context and Document Structure

ILLC DS-2011-07: **Jop Briët**
Grothendieck Inequalities, Nonlocal Games and Optimization

ILLC DS-2011-08: **Stefan Minica**
Dynamic Logic of Questions

ILLC DS-2011-09: **Raul Andres Leal**
Modalities Through the Looking Glass: A study on coalgebraic modal logic and their applications

ILLC DS-2011-10: **Lena Kurzen**
Complexity in Interaction

ILLC DS-2011-11: **Gideon Borensztajn**
The neural basis of structure in language

ILLC DS-2012-01: **Federico Sangati**
Decomposing and Regenerating Syntactic Trees

ILLC DS-2012-02: **Markos Mylonakis**
Learning the Latent Structure of Translation

ILLC DS-2012-03: **Edgar José Andrade Lotero**
Models of Language: Towards a practice-based account of information in natural language

ILLC DS-2012-04: **Yurii Khomskii**
Regularity Properties and Definability in the Real Number Continuum: idealized forcing, polarized partitions, Hausdorff gaps and mad families in the projective hierarchy.

ILLC DS-2012-05: **David García Soriano**
Query-Efficient Computation in Property Testing and Learning Theory

ILLC DS-2012-06: **Dimitris Gakis**
Contextual Metaphilosophy - The Case of Wittgenstein

ILLC DS-2012-07: **Pietro Galliani**
The Dynamics of Imperfect Information

ILLC DS-2012-08: **Umberto Grandi**
Binary Aggregation with Integrity Constraints

ILLC DS-2012-09: **Wesley Halcrow Holliday**
Knowing What Follows: Epistemic Closure and Epistemic Logic

ILLC DS-2012-10: **Jeremy Meyers**
Locations, Bodies, and Sets: A model theoretic investigation into nominalistic mereologies

ILLC DS-2012-11: **Floor Sietsma**
Logics of Communication and Knowledge

ILLC DS-2012-12: **Joris Dormans**
Engineering emergence: applied theory for game design

ILLC DS-2013-01: **Simon Pauw**
Size Matters: Grounding Quantifiers in Spatial Perception

ILLC DS-2013-02: **Virginie Fiutek**
Playing with Knowledge and Belief

ILLC DS-2013-03: **Giannicola Scarpa**
Quantum entanglement in non-local games, graph parameters and zero-error information theory

ILLC DS-2014-01: **Machiel Keestra**
Sculpting the Space of Actions. Explaining Human Action by Integrating Intentions and Mechanisms

ILLC DS-2014-02: **Thomas Icard**
The Algorithmic Mind: A Study of Inference in Action

ILLC DS-2014-03: **Harald A. Bastiaanse**
Very, Many, Small, Penguins: Vaguely Related Topics