Negation in Natural Language

David Woodford Ripley

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree Doctor of Philosophy in the Department of Philosophy.

Chapel Hill

2009

Approved by:

Keith Simmons

William Lycan

Thomas Hofweber

J. Michael Terry

Dean Pettit

Abstract

David Woodford Ripley: Negation in Natural Language

(Under the direction of Keith Simmons)

Negation is ubiquitous in natural language, and philosophers have developed plenty of different theories of the semantics of negation. Despite this, linguistic theorizing about negation typically assumes that classical logic's semantics for negation—a simple truth-functional toggle—is adequate to negation in natural language, and philosophical discussions of negation typically ignore vital linguistic data. The present document is thus something of an attempt to fill a gap, to show that careful attention to linguistic data actually militates *against* using a classical semantics for negation, and to demonstrate the philosophical payoff that comes from a nonclassical semantics for natural-language negation.

I present a compositional semantics for natural language in which these questions can be posed and addressed, and argue that propositional attitudes fit into this semantics best when we use a nonclassical semantics for negation. I go on to explore several options that have been proposed by logicians of various stripes for the semantics of negation, providing a general framework in which the options can be evaluated. Finally, I show how taking non-classical negations seriously opens new doors in the philosophy of vagueness.

To Paul and Jean Kriksciun

Acknowledgements

A lot of people helped tremendously in producing this document; I think I may have done little more than introduce what errors there must be. A number of the ideas here first coalesced in my mind at the University of Connecticut during conversations with Yael Sharvit, Sigrid Beck, and especially Jc Beall, who has been an unflagging source of support and inspiration (and who deserves most of the credit/blame for getting me excited about philosophy in the first place). At the University of North Carolina, my fellow graduate students provided a great deal of help working out these ideas. Particularly helpful were Jason Bowers, who is always willing to cheerfully chase every last rabbit as far as it will run, and Eric Mandelbaum, who I suspect has had more of an effect on my thinking than either of us yet knows. My committee—Keith Simmons, Bill Lycan, Thomas Hofweber, J. Michael Terry, and Dean Pettit—provided excellent feedback on drafts. Keith's readiness to meet and explore the issues raised here, as well as his detailed feedback, were incredibly helpful. The University of Melbourne—with Conrad Asmus, Toby Meadows, Stephan Kubicki, Tama Coutts, Zach Weber, Allen Hazen, Graham Priest, Greg Restall, Ross Brady, and Lloyd Humberstone—served as my intellectual Promised Land (although I have NO OPINION about who is Joshua, David, Saul, Solomon, &c.). My time there was also vitally motivating; they, jointly, by their example, got me to actually buckle down and write this thing. A number of pieces incorporated here have had independent life at conferences, departmental talks, and the like; for vital feedback, I thank Ole Hjortland, Andi Pietz, Ram Neta, David Chalmers, Patrick Greenough, Diana Raffman, James Hampton, Bart Geurts, Daniel Nolan, Bob Adams, Marc Lange, Geoff Sayre-Mccord, Mark Jago, Dominic Hyde, Mark Colyvan, and Edwin Mares. There's a way of reading this dissertation on which it's carrying Richard Sylvan's project forward; I never met him, but I owe him a great debt just the same.

Financially, the bulk of this project was supported jointly by a William R. Kenan Fellowship and by the UNC Department of Philosophy, and the remainder by a CNRS postdoc job at Institut Jean Nicod. Without them all, I would still be bargain manager at Barnes & Noble. That wasn't a bad gig, but I am still very (very) grateful.

Finally, I'd like to thank Suzy Killmister, for her near-literally-incredible understanding and support, my parents, who believed in me before I believed anything at all, and my grandparents, for teaching me how to aim high.

Table of Contents

List of Figures

Chapter 1

Introduction

Negation is ubiquitous in natural language, and philosophers have developed plenty of different theories of the semantics of negation. Despite this, linguistic theorizing about negation typically assumes that classical logic's semantics for negation—a simple truth-functional toggle—is adequate to negation in natural language, and philosophical discussions of negation typically ignore vital linguistic data. The present document is thus something of an attempt to fill a gap, to show that careful attention to linguistic data actually militates *against* using a classical semantics for negation, and to demonstrate the philosophical payoff that comes from a nonclassical semantics for natural-language negation. Here's a brief overview:

The first two chapters set the stage, providing a framework in which to discuss questions of compositional semantics, and arguing that a classical semantics cannot work for natural-language negation. Chapter 2 presents an intensional (possible-worlds-based) semantics for a fragment of English. It's based on (Heim & Kratzer, 1998) and (von Fintel & Heim, 2007), but it does not follow them exactly. In particular, the treatment of intensionality is slightly different; whereas von Fintel & Heim (2007) opt for a semantics that assigns two distinct semantic values—an extension and an intension—to each expression, the present semantics operates purely at the level of intensions. The two methods are equivalent at a certain level of description, but the present method connects more cleanly to the logical frameworks occurring later (in chapter 5).

Chapter 3 applies this semantics to the case of propositional attitudes. It points out familiar problems that arise in this application of possible-worlds-based intensional semantics, and considers two possible fixes: structuralism (as defended in (Carnap, 1956), (Lewis, 1970), (Cresswell, 1985), (Cresswell & Von Stechow, 1982), and (King, 2007)) and circumstantialism (as defended

in (Edelberg, 1994), (Goddard & Routley, 1973), and (Priest, 2005)). According to structuralism, the difficulties faced by possible-worlds semantics should be fixed by treating the semantic values of clauses as syntactically structured in some way. According to circumstantialism, the difficulties should be fixed by considering more circumstances than just possible worlds. (In particular, the circumstantialist maintains, we need negation-incomplete and negation-inconsistent circumstances.) I argue that structuralism falls down on two fronts: it identifies propositions that should be kept distinct, and it distinguishes propositions that should be identified. I show that circumstantialism does not have to share these shortcomings (although it can: thus (Goddard & Routley, 1973)'s and (Priest, 2005)'s circumstantialisms are rejected for similar reasons). I also respond to two arguments: the motivation for structuralism given in (Cresswell & Von Stechow, 1982) and (Cresswell, 1985), and the argument against circumstantialism in (Soames, 1987).

If chapter 3 is correct, classical semantics for negation cannot be correct, as it would rule out negation-incomplete and negation-inconsistent circumstances. Thus, we need a different semantics for negation in natural language. The next two chapters begin to explore the available alternatives. Chapter 4 gives an overview of an important strain in recent logical literature on negation—the strain that treats negation as a unary modal operator—and provides an overview of certain linguistic research on various negative phrases and contexts. I draw important parallels between the two traditions, and show that the linguistic tradition has found important uses for negations weaker than those that can be treated using modal semantics. I show that the role played by neighborhood semantics in the study of positive modalities adapts to negative modalities as well, providing a framework in which to study these weaker negations.

Chapter 5 is the formal meat of the dissertation—here, I provide a unified approach to the semantics of nonclassical negations. This chapter adapts techniques from neighborhood semantics for intensional logic (as in (Chellas, 1980) and (Hansen, 2003)) and frame semantics for relevant and substructural logics (as in (Routley et al., 1982) and (Restall, 2000b)) to develop the theory of substructural neighborhood semantics for negative unary connectives. (Although substructural neighborhood semantics have been glanced at before—in (Mares & Fuhrmann, 1995) and (Mares, 2004)—both of these deal with conditionals; neither considers negation, or indeed any unary connective.) I give a base system—distributive lattice logic plus a certain sort of contraposition—and several extensions of it, proving soundness, completeness, and correspondence for each extension. I then show how neighborhood semantics collapses to the more familiar modal semantics for negation (discussed in chapter 4) when certain conditions are met. Showing that neighborhood semantics is equivalent to the familiar Routley star semantics for negation (see eg (Routley et al., 1982)) in certain

circumstances requires developing a bit of the algebraic theory of substructural neighborhood semantics. The chapter closes with an expanded "kite of negations" displaying inclusion relationships between many negations, both familiar and new.

The two closing chapters apply nonclassical theories of negation to the problem of vagueness, showing that exploring the semantics of negation can bear philosophical fruit. Chapter 6 considers the motivation behind a supervaluationist theory of vagueness, and argues that the motivation is better satisfied by a dialetheic theory of vagueness, such as that given by the paraconsistent logic LP, typically presented as a three-valued logic. One usual objection to three-valued logics of vagueness (pressed in eg (Williamson, 1994) and (Smith, 2008)) holds that considerations of "higher-order vagueness" cannot be properly addressed on such an approach. I show that, if the objection is taken seriously, the defender of an LP-based treatment of vagueness is forced to abandon their three-valued logic and adopt a particular infinite-valued logic—but this logic is itself LP! (The proof of this fact is in (Priest, 1984), but the application to vagueness and higher-order vagueness is original.) Thus, an LP-based theory of vagueness is immune to the objection.

If an LP-based theory of vague predicates is correct, then some contradictions are true. In particular, contradictions like 'She's nice and she's not nice', when 'she' picks out a borderline case of niceness, are true. Chapter 7 presents empirical research on speakers' responses to borderline cases, showing that they agree (to a quite high degree) with such borderline contradictions. I consider several interpretations of this research, falling into roughly three classes: a) that participants in the experiment didn't interpret the sentences in question as contradictions, b) that participants did interpret the sentences as contradictions, but they were wrong to agree, and c) that participants interpreted the sentences as contradictions, and were right to agree. Multiple interpretations fall into each of these classes. I argue that only two interpretations explain participants' responses in a natural manner: a certain variety of contextualism about vagueness, augmented with hypotheses about participants' understanding of conjunction, or dialetheism of the sort I recommend. I point out that these two interpretations are actually very similar; it is their similarity (in particular, they both allow that an item can both have and not have the very same property—but the contextualist requires a context shift for this to happen, where the dialetheist does not) that allows them both to work here.

Thus, the current document 1) shows that classical theories of negation are inadequate to negation as it occurs in natural language; 2) develops a semantic framework in which a great variety of nonclassical negations can be explored and compared to each other; and 3) demonstrates the philosophical payoff of nonclassical theories of negation. Along the way, new contributions are made to

theories of propositions, neighborhood and substructural semantics for unary connectives, theories of vagueness, and experimental semantics.

Chapter 2

An intensional semantics

There are many frameworks for compositional semantics available; to some extent any theorist's selection of one over the others is arbitrary. Nonetheless, to explore the semantics of negation in natural language it's vital to have some semantic framework in which to work. The task of this chapter is to present such a framework.

The arguments against classical theories of negation that follow will rely on intensional constructions, so it is vital for my purposes to use a framework sufficient to handle such constructions as conditionals, modals, and propositional attitudes. Well-known intensional frameworks have been presented in (Montague, 2002) and (Lewis, 1970), but those pieces rely on a syntax relatively far removed from the best current syntactic research. On the other hand, (Heim & Kratzer, 1998) and (Authier & Reed, 1999) present compositional frameworks sensitive to contemporary syntactic research, but these pieces remain in a purely extensional framework, and so are ill-equipped to deal with constructions I'll be making heavy use of.

The closest relative to the framework I'll present here can be found in (von Fintel & Heim, 2007). This uses the extensional framework of (Heim & Kratzer, 1998), but adds an intensional component, using possible worlds to provide semantics for a wider range of natural language constructions. Despite the similarities, my framework differs from theirs in functioning purely at the level of intension, rather than trafficking heavily in both intension and extension. (The difference is more stylistic than substantive.)

This chapter is structured as follows: §2.1 covers notation and preliminaries necessary for §§2.2–2.4, which present (Heim & Kratzer, 1998)'s treatment of predication, relative clauses, and quantification. Then, §2.5 covers preliminaries necessary for §§2.6–2.7, which present a treatment of modals,

conditionals, and propositional attitudes that hews closely to the treatment in (von Fintel & Heim, 2007).

2.1 Extensional preliminaries

The extensional framework I present here (in §§2.1–2.4) is (almost entirely) the framework of (Heim & Kratzer, 1998). Any differences will be explicitly noted.
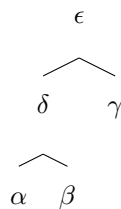
2.1.1 Notation

Boldface will be used to quote object-language expressions; thus Mary is a person but **Mary** is a name.

We use a function $[\![\ ]\!]$ to interpret object-language expressions as semantic entities. For example, for our purposes, the denotation of a name is its bearer.[1] Thus $[\![Mary]\!]$ = Mary. The interpretation function is officially to be read as relative to a model M (and perhaps also a time, context, variable assignment, &c.). When this needs to be made explicit, it'll show up as a superscript on $[\![\ ]\!]$; thus $[\![\mathbf{Mary}]\!]^M$ is the denotation of **Mary** in the model M, $[\![\mathbf{Mary}]\!]^{M,t,g}$ is the denotation of **Mary** relative to a number of different parameters. For most of this chapter, though, we'll be able to get by ignoring this, and so the $[\![\ ]\!]$ function will stand alone much of the time.

2.1.2 Syntax

Here's the plan: a sentence's syntax determines the way we combine lexical denotations to get the denotation of the sentence. For example, if a sentence has the syntax (1a), we calculate its denotation by first combining $[\![\alpha]\!]$ with $[\![\beta]\!]$ to get $[\![\delta]\!]$, then combining $[\![\delta]\!]$ with $[\![\gamma]\!]$ to get $[\![\epsilon]\!]$, the sentence's denotation. On the other hand, if a sentence has the syntax (1b), we'd calculate its denotation by first combining $[\![\beta]\!]$ and $[\![\gamma]\!]$ to get $[\![\delta]\!]$, then combining $[\![\alpha]\!]$ with $[\![\delta]\!]$ to get $[\![\epsilon]\!]$.

(1)   a.

$$
\begin{array}{c}
\epsilon \\
\diagup\diagdown \\
\delta \quad\ \gamma \\
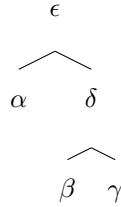\diagup\diagdown \\
\alpha\ \ \beta
\end{array}
$$

---

[1] 'Denote' and 'denotation' are here used in a very thin way; every semantically non-vacuous linguistic expression denotes something (its semantic value), and no claims are implied (nor do I make any anywhere in this document) about reference or anything heavy-duty like that.

b.

$$\epsilon$$
$$\alpha \quad \delta$$
$$\beta \quad \gamma$$

What is a sentence's syntax? The syntacticians will tell us. In particular, I'll pay attention to what the syntacticians call LF: the level of syntactic representation that is interpreted by whatever semantic system there is.[2]

That said, there is considerable controversy surrounding some of the features of LF representations. Much of this controversy won't affect anything I have to say here; I'll use syntactic representations general enough to apply more or less no matter how the controversies are settled. When the controversy matters (as in the case of Quantifier Raising, or QR), I'll be more explicit. Also, the syntactic representations I use are general enough to apply to either a Government & Binding-style syntax (as in (Heim & Kratzer, 1998)) or a Minimalist-style syntax (as in (Authier & Reed, 1999)).

2.1.3 Types

This will be a typed semantics; I'll use several types of semantic value, each with its own domain.[3] For example, members of the set $D_t = \{0, 1\}$ of truth-values are of type t. Entities (things like you, me, tables, bananas, emotions, &c.) are of type e, and they make up the domain of entities D_e. The only two primitive types (for now) are e and t, but there are complex types too, formed from other types. If α and β are types, then $\langle \alpha, \beta \rangle$ is a type too. Every type has its own domain: the domain of semantic values of any type α is D_α. We specify D_e and D_t, and build the remaining domains as follows: $D_{\langle \alpha, \beta \rangle} = D_\beta{}^{D_\alpha}$ (the set of functions from D_α to D_β).[4]

I take the denotation of a proper name to be the name's bearer. For example, $[\![GeorgeWashington]\!]$ = George Washington, $[\![NiagaraFalls]\!]$ = Niagara Falls. Name-bearers are of type e.[5] Sentences have truth-values (1 or 0) as their denotations (for now). These values are of type t, the other primitive type.

---

[2]Some recent phase-based theories of syntax dispense with such a level altogether; see eg (Chomsky, 2004). Although I won't pursue this here, I suspect that the present semantics can also be made to fit these theories, even if the fit has to be described a bit differently.

[3]For details on typed languages and their semantics, see eg (Church, 1940; Henkin, 1950; Rennie, 1974).

[4]Were a complete proof theory going to be a desideratum later, we'd want to be careful here, but it won't be, so I won't be.

[5]But see Partee (1987) for possible exceptions.

2.1.4 λ

Since the other domains (besides D_e and D_t) contain only functions, we'll be talking about functions a lot, and some notation will help. I'll follow (Heim & Kratzer, 1998) in using a version of the λ notation explored in eg (Church, 1941; Routley, 1980).

As an example, consider the following function: $\lambda e[e$'s favorite llama]. Call this function g; for each $x \in D_e$, $g(x) = x$'s favorite llama. Since x's favorite llama is of type e as well (indeed, all llamas are), g is a function of type $\langle e, e \rangle$.

I'll also use functions like the following (of type $\langle e, t \rangle$): $\lambda e(e$ smokes$)$. Call this function f. It looks as though $f(x) = x$ smokes, but that would be weird, since 'x smokes' isn't of type t. (Remember, $D_t = \{1, 0\}$.) In fact I mean something slightly different: that $f(x) = 1$ iff x smokes, and $f(x) = 0$ otherwise. Thus f really is of type $\langle e, t \rangle$. This'll be important to keep in mind as we move forward.

Note as well that I frequently use the type labels e, t, s, p, u, v, a, and their primes, as variables; when that happens, the variables should be interpreted as restricted to the appropriate types.[6] An overriding convention, for when explicitness matters: whenever a variable bears a subscripted type label, it should be interpreted as restricted to that type; whenever a denotation bears a subscripted type label, it is of that type (so $[\![eats]\!]_{\langle e,t \rangle}$ is of type $\langle e, t \rangle$); and whenever an object-language expression bears a subscripted type label, its denotation is of that type ($[\![\mathbf{eats}_{\langle e,t \rangle}]\!]$ is of type $\langle e, t \rangle$ too).

2.2 Predication

This section explores the extensional semantics of predication. Here (and throughout) I'll ignore tense and aspect.

2.2.1 Intransitive verbs

The simplest case of predication involves a subject and a simple intransitive predicate. As an example, consider the sentence **Sam flies**. It has the following (immediately relevant) structure:[7]

---

[6]Don't worry about s, p, u, v, or a yet; they'll come up later.

[7]Although I use labels like **vP**, I do so mostly for convenience in referring to various constituents. Nothing here is incompatible with a so-called 'bare phrase structure' theory of syntax. See eg (Chomsky, 1995).

(2)   **vP**

$$\text{Sam} \quad \text{flies}$$
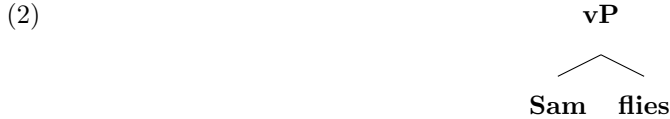
To interpret this sentence, we need axioms interpreting the words in it. As above, we'll treat the denotations of names as of type e. When it comes to intransitive verbs, we'll take them to have denotations of type $\langle e, t \rangle$. Assume:

(3)   a. $[\![Sam]\!] = \text{Sam}$

          b. $[\![flies]\!] = \lambda e[e \text{ flies}]$

We also need some story about how these denotations combine. The following rule of Functional Application (FA) is the linchpin of the framework in (Heim & Kratzer, 1998):

(4)   a.   When α, β, and γ are nodes in configuration (4b), and when $[\![\alpha]\!]$ can take $[\![\beta]\!]$ as an argument, then $[\![\gamma]\!] = [\![\alpha]\!]([\![\beta]\!])$:[8]

          b. γ

$$\alpha \quad \beta$$

Putting our lexical denotations together following the rule FA, we get:

(5)   a. $[\![vP]\!] = [\![flies]\!]([\![Sam]\!]) = [\![flies]\!](\text{Sam}) =$

          b. $\lambda e[e \text{ flies}](\text{Sam})$

          c. $[\![vP]\!] = 1$ iff Sam flies

We can see that this is the desired result; getting this result thus offers some support for the system of hypotheses (at least: the lexical entries, the syntactic analysis of **Sam flies**, and the rule FA) that entailed it.

2.2.2   Transitive verbs

Now that we see how the rough idea works, we can expand it beyond the simplest subject-predicate sentences. Consider a sentence involving a transitive verb like **kick**, and the appropriate number of proper names:

---

[8]We want to be flexible here w/r/t linear order: maybe the function comes first, and maybe its argument does. Heim & Kratzer (1998) achieve this with a disjunctive version of FA, but it's equally achievable by not taking linear order into account in the first place, paying attention only to hierarchical structure, and that's the route we'll take here. Thus, in the structure (4b), no assumptions are made about which of α or β is pronounced first; all that's relevant are dominance relations. This is consonant with the hypothesis (as in eg (Chomsky, 2004)) that linear order is a phonological phenomenon only, unrelated to interpretation. NB: It can't happen (since our types are well-founded) that $[\![\alpha]\!]$ and $[\![\beta]\!]$ are of appropriate types to take each other as arguments, so we don't need linear order to disambiguate: FA gives unique results.

(6) a. **Mary kicked Barry**

b.

$$
\begin{array}{c}
\textbf{vP} \\
\diagup \qquad \diagdown \\
\textbf{Mary} \qquad \textbf{v}' \\
\diagup \quad \diagdown \\
\textbf{kicked} \quad \textbf{Barry}
\end{array}
$$
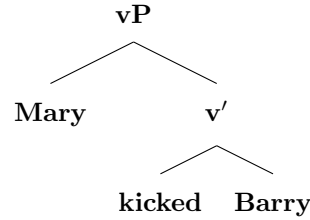
The lexical entries for the names work as before. The entry for the verb is slightly more complex; we suppose that (mono)transitive verbs denote functions of type $\langle e, \langle e, t \rangle \rangle$.

(7) a. $[\![Mary]\!] = \mathrm{Mary}$

b. $[\![Barry]\!] = \mathrm{Barry}$

c. $[\![kicked]\!] = \lambda e \lambda e'[e' \text{ kicked } e]$

Note the order of the arguments in $[\![kicked]\!]$: the 'kickee' is the first argument taken, the 'kicker' the second. This corresponds to the structure in (6b): the object of the verb, **Barry**, is merged before the subject, **Mary**. This makes the calculation come out as follows:

(8) a. $[\![v']\!] = [\![kicked]\!]([\![Barry]\!]) = [\![kicked]\!](\mathrm{Barry}) =$

b. $\lambda e \lambda e'[e' \text{ kicked } e](\mathrm{Barry}) =$

c. $\lambda e'[e' \text{ kicked Barry}]^9$

(9) a. $[\![vP]\!] = [\![v']\!]([\![Mary]\!]) = [\![v']\!](\mathrm{Mary}) =$

b. $\lambda e'[e' \text{ kicked Barry}](\mathrm{Mary})$

c. $[\![vP]\!] = 1$ iff Mary kicked Barry

Again, this is clearly the right answer.

2.2.3 Common nouns

The simplest way to think of common nouns is as of type $\langle e, t \rangle$, the same as intransitive verbs. If that's the way we go, we'll probably ignore the copula and article typically involved: we'll take $[\![bear]\!]$ to be the same as $[\![isabear]\!]$.[10] Then the calculation for **Sam is a bear** looks much like the calculation for **Sam flies** above.

---

[9] Note that this is an appropriate (extensional) denotation for **kicked Barry**. That **kicked Barry** has some denotation of its own worth exploring can be seen in subsentential uses of **kicked Barry**, to answer questions like **What did Mary do yesterday?**. For more, see eg (Barton & Progovac, 2004; Stainton, 2005).

[10] For more on the contribution made by 'is' and 'a' here, see eg (Partee, 1987).

2.2.4 Predicate Modification

Predicate nouns don't always occur on their own: they come modified by adjectives as well. If $[\![bear]\!]$ is of type $\langle e, t \rangle$, so too it seems should be $[\![bluebear]\!]$. That would seem to suggest that $[\![blue]\!]$ should be of type $\langle\langle e, t\rangle, \langle e, t\rangle\rangle$. But wait! Look at **Jake is blue**. The appearance of **blue** as a predicate adjective would seem to suggest that $[\![blue]\!]$ is of type $\langle e, t \rangle$.

Three potential solutions:
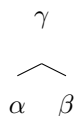
(10)   a. Leave $[\![blue]\!]$ as type $\langle\langle e, t\rangle, \langle e, t\rangle\rangle$, and postulate an invisible nominal of type u in **Jake is blue**.

         b. Allow $[\![blue]\!]$ to type-shift: it's of type $\langle e, t \rangle$ in **Jake is blue** and $\langle\langle e, t\rangle, \langle e, t\rangle\rangle$ in **Jake is a bear**.

         c. Add a new rule of predicate modification: $[\![blue]\!]$ is always of type $\langle e, t \rangle$, but there's a rule to allow a direct combination of two nodes with type $\langle e, t \rangle$.

Heim & Kratzer (1998) go with (10c) for adjectives like **blue**, and the rule of Predicate Modification (PM) to be found there is simple and straightforward:[11]

(11)   a. When α, β, and γ are nodes in configuration (11b), and when $[\![\alpha]\!]$ and $[\![\beta]\!]$ are both of type $\langle e, t \rangle$, then $[\![\gamma]\!]_{\langle e,t\rangle} = \lambda e[\,[\![\alpha]\!](e) = [\![\beta]\!](e) = 1]$

         b.

$$\gamma$$
$$\overset{\frown}{}$$
$$\alpha \quad \beta$$

This works for so-called *intersective* adjectives, but note that not every adjective can be accommodated this way. Consider **big**. PM doesn't get us the right result; Fran can be a big mouse without being both big and a mouse.[12] So $[\![big]\!]$ has to be of type $\langle\langle e, t\rangle, \langle e, t\rangle\rangle$, at least sometimes; one of (10a) or (10b) is in order, for some (occurrences of some) adjectives. I won't go further into this issue here.

2.3 Relative clauses

Of course, natural language constructions are not limited to simple predications. Consider the sentence **Mary saw the cat which John ate**. We'll deal with definite descriptions later (§2.4.3);

---

[11]Note that it can never happen that both FA and PM apply to the same tree, since no function of type $\langle e, t \rangle$ can take another such function as an argument.

[12]Fran has to be a mouse to be a big mouse (**big** is subsective, in the terms of (Forbes, 2006)), but Fran doesn't have to be big.

for now, consider just the relative clause **which John ate**. Heim & Kratzer (1998) give the following hypothesis about its denotation: it ought to be a function of type $\langle e, t \rangle$, one that takes x to 1 if John ate x and to 0 otherwise. This section shows how to arrive at this denotation compositionally.

We parse relative clauses as follows:

(12)
$$\begin{array}{c}
\textbf{CP} \\
\diagup \quad \diagdown \\
\textbf{which}_1 \quad \diagdown \\
\textbf{C} \quad \textbf{John ate } t_1
\end{array}$$
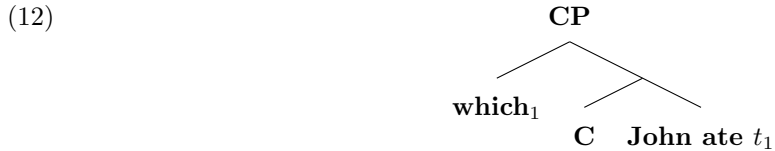
2.3.1 Traces and variable assignments

In order to interpret these trees, it seems we need to assign some value to $[\![t_1]\!]$. But since we're going to end up treating (some) traces as variables, it won't work to assign t_1 an absolute denotation. Instead, we'll assign it a denotation relative to a variable assignment. A variable assignment g is a member of $D_e{}^{\mathbb{N}}$, a function from the natural numbers into D_e.[13] Our relativized function $[\![\ ]\!]^g$ assigns a trace with index i the value $g(i)$: $[\![t_i]\!]^g = g(i)$.

Now that we're paying attention to variable assignments, it'll help to be explicit about the relation between $[\![\ ]\!]^g$ and $[\![\ ]\!]$. Every object-language expression is within the domain of the relativized function $[\![\ ]\!]^g$, but some expressions are not in the domain of $[\![\ ]\!]$. An expression A is in the the domain of $[\![\ ]\!]$ iff for every variable assignment g, $[\![A]\!]^g$ is the same; $[\![A]\!]$ is then that invariant denotation.

2.3.2 Predicate Abstraction

To handle relative clauses, Heim & Kratzer (1998) add a rule of predicate abstraction (PA) triggered by the index on the relative pronoun:

(13) a. When α, β, and γ are nodes in configuration (13b), α an index k,[14] then $[\![\gamma]\!]^g = \lambda e[\,[\![\beta]\!]^{g^{e/k}}]$[15]

b.
$$\begin{array}{c}
\gamma \\
\diagup \diagdown \\
\alpha \quad \beta
\end{array}$$
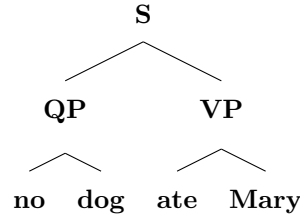
Now consider our tree in a bit more detail:

---

[13] We could use variables of other types as well, if we wanted, but here we won't have the need.

[14] α can be a relative pronoun with index k too.

[15] $g^{e/k}$ is the variable assignment g' such that: i) $g'(i) = g(i)$ for every i other than k and ii) $g'(k) = e$.

(14)

```
                              CP
                          ／      ＼
                    which₁        ／ ＼
                              C      ／ ＼
                                 John    ／ ＼
                                      ate    t₁
```

$[\![\textbf{ate}]\!]^g$ is of type $\langle\langle e,t\rangle, \langle e,t\rangle\rangle$; $[\![t_1]\!]^g$ and $[\![\textbf{John}]\!]^g$ are of type e. We treat **which** and **C** as vacuous, but not the index on **which**; it triggers PA.

(15) a. $[\![\textbf{ate } t_1]\!]^g = [\![\textbf{ate}]\!]^g([\![t_1]\!]^g) = [\![\textbf{ate}]\!]^g(g(1)) =$

  b. $\lambda e \lambda e'[e' \text{ ate } e](g(1)) =$

  c. $\lambda e'[e' \text{ ate } g(1)]$

(16) a. $[\![\textbf{John ate } t_1]\!]^g = [\![\textbf{ate } t_1]\!]^g([\![\textbf{John}]\!]^g) = [\![\textbf{ate } t_1]\!]^g(\text{John}) =$

  b. $\lambda e'[e' \text{ ate } g(1)](\text{John}) = 1$ iff John ate $g(1)$

(17) a. $[\![\textbf{C John ate } t_1]\!]^g = [\![\textbf{John ate } t_1]\!]^g$

(18) a. $[\![\textbf{which}_1 \textbf{ C John ate } t_1]\!]^g = \lambda e[[\![\textbf{John ate } t_1]\!]^{g^{e/1}}] =$

  b. $\lambda e[\text{John ate } e]^{16}$

This is the denotation of type $\langle e,t\rangle$ we were after. Note that in calculating eg $[\![catwhichJohnate]\!]$, we can combine $[\![cat]\!]$ (of type $\langle e,t\rangle$) with $[\![whichJohnate]\!]$ (of type $\langle e,t\rangle$) using the rule of PM: something is a cat which John ate iff it's a cat and John ate it.

2.4 Quantification

This section follows (Heim & Kratzer, 1998) (and, so, given the transitivity of following, (Montague, 2002)) in taking quantified phrases (like **some dogs**, **no bananas**) to be of type $\langle\langle e,t\rangle, t\rangle$, and the quantifiers themselves (eg **some**, **no**) to be of type $\langle\langle e,t\rangle, \langle\langle e,t\rangle, t\rangle\rangle$.

2.4.1 Subject quantification

The simplest cases of sentences involving quantifiers have their QPs in subject position. Consider **no dog ate Mary**, with the following structure:

---

[16]Note that $g^{e/1}(1)$ must be e, whatever g is. This means that $[\![\textbf{which}_1 \textbf{ John ate } t_1]\!]^g$ does not depend on g, and so **which₁ John ate** t_1 is in the domain of $[\![]\!]$. On the other hand, since $[\![\textbf{John ate } t_1]\!]^g$ does depend on g, **John ate** t_1 is not in the domain of $[\![]\!]$.

(19)

```
                    S
                   / \
                  /   \
                QP     VP
               / \     / \
              /   \   /   \
            no   dog ate  Mary
```

Suppose our lexical items denote as follows:

(20) a. $[\![Mary]\!] = $ Mary

b. $[\![ate]\!] = \lambda e \lambda e'[e' \text{ ate } e]$

c. $[\![dog]\!] = \lambda e[e \text{ is a dog}]$

d. $[\![no]\!] = \lambda u_{\langle e,t \rangle} \lambda u'_{\langle e,t \rangle}[\text{There is no } e \text{ such that } u(e) = u'(e) = 1]^{17}$

Then we can calculate the sentence's denotation like so:

(21) a. $[\![ateMary]\!] = [\![ate]\!]([\![Mary]\!]) = [\![ate]\!](\text{Mary}) = $

b. $\lambda e \lambda e'[e' \text{ ate } e](\text{Mary}) = $

c. $\lambda e'[e' \text{ ate Mary}]$

(22) a. $[\![nodog]\!] = [\![no]\!]([\![dog]\!]) = [\![no]\!](\lambda e'[e' \text{ is a dog}])^{18} = $

b. $\lambda u'_{\langle e,t \rangle}[\text{There is no } e \text{ such that } [\![dog]\!](e) = u'(e) = 1] = $

c. $\lambda u'_{\langle e,t \rangle}[\text{There is no } e \text{ such that } e \text{ is a dog and } u'(e) = 1]$

(23) a. $[\![nodogateMary]\!] = [\![nodog]\!]([\![ateMary]\!]) = $

b. 1 iff there is no e such that e is a dog and e ate Mary

Clearly, this is just what we want this sentence to denote.

2.4.2 Object quantification

Subject position, though, is not the only position we find quantifiers in; consider (24a)–(24d):

(24) a. **Jack saw every caterpillar**

b. **Ann gave every carpenter three ducks**
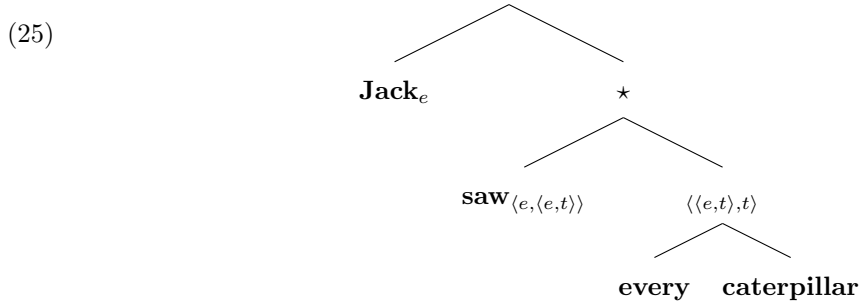
c. **Harry was hated by few orangutans**

---

[17]Or: there is no e in some contextually-determined subset of D_e, if we want to deal with domain restrictions here. See eg (Stanley & Szabó, 2007) for arguments against this approach, though.

[18]Note the change of bound variable from e to e' here. This is harmless and will be engaged in willy-nilly without further comment.

       d. **Tre sat on many spiders**

(24a) gives us an example of a quantifier phrase in direct object position, (24b) two quantifier phrases—one indirect object, one direct object—and (24c) and (24d) a quantifier phrase as the object of a preposition. Our initial type assignment had quantifier phrases as type $\langle\langle e,t\rangle,t\rangle$, but here this results in a type mismatch. For example, consider (24a), which can be treeified as in (25):

(25)

$$
\begin{array}{c}
\text{Jack}_e \qquad \star \\[1em]
\text{saw}_{\langle e,\langle e,t\rangle\rangle} \qquad \langle\langle e,t\rangle,t\rangle \\[1em]
\text{every} \quad \text{caterpillar}
\end{array}
$$

A calculation of this tree's semantic value runs into trouble at the ⋆ed node. Since its daughters are of type $\langle e,\langle e,t\rangle\rangle$ and $\langle\langle e,t\rangle,t\rangle$, neither is of the right type to take the other as a denotation, nor are they of the right type for PM, nor is either one an index at which to apply PA. None of our rules assigns a denotation to this node.

We might try to gerrymander our rules so that $[\![saw]\!]$ can take $[\![Jack]\!]$ as an argument first, yielding a result of type $\langle e,t\rangle$, to which $[\![everycaterpillar]\!]$ could then apply to result in a t for the overall tree. Such a calculation would violate our current rules, but we could presumably come up with some way to permit it. However, this will not solve the problem; this calculation is well-typed, but it yields the wrong meaning. It gives us the truth value 1 iff every caterpillar sees Jack; it gets its subjects and objects the wrong way around.
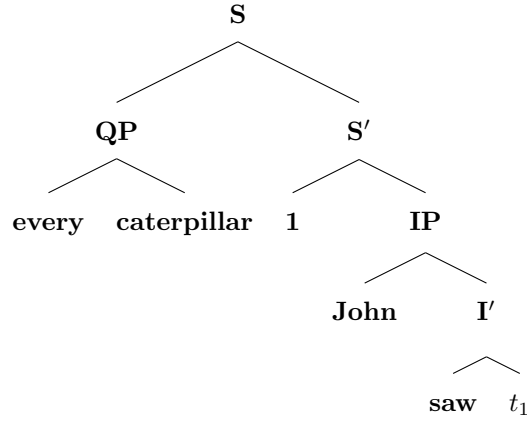
What to do? Heim & Kratzer (1998) consider two possible solutions, both of which are compatible with the present system, as will be shown below.

QR

One way to resolve the issue is to suppose that (24a)'s LF is not (25), but instead something like (26):

(26)

```
                              S
                     ┌────────┴────────┐
                    QP                 S′
                ┌────┴────┐        ┌────┴────┐
             every    caterpillar  1        IP
                                        ┌────┴────┐
                                      John        I′
                                              ┌────┴────┐
                                            saw        t₁
```

This fits with hypotheses in eg (May, 1985), suggesting that quantifier phrases raise covertly on the way to LF. This movement is there called Quantifier Raising, or QR, so we will call this the QR hypothesis. Note the addition of the index just below the QP's landing site.[19] This will trigger our rule of PA, giving us an overall denotation of type $\langle e, t \rangle$ for **1 John saw** t_1, which can then combine with $[\![everycaterpillar]\!]_{\langle\langle e,t\rangle,t\rangle}$ to give us a truth value, of type t. This time, it's even the right truth value. The calculation:

(27) a. $[\![saw]\!] = \lambda e \lambda e'[e' \text{ saw } e]$

b. $[\![John]\!] = \text{John}$

c. $[\![caterpillar]\!] = \lambda e[e \text{ is a caterpillar}]$

d. $[\![every]\!] = \lambda u_{\langle e,t\rangle} \lambda u'_{\langle e,t\rangle}[\{e : u(e) = 1\} \subseteq \{e : u'(e) = 1\}]$[20]

(28) a. $[\![\textbf{saw } t_1]\!]^g = [\![\textbf{saw}]\!]^g([\![t_1]\!]^g) = [\![\textbf{saw}]\!]^g(g(1)) =$

b. $\lambda e \lambda e'[e' \text{ saw } e](g(1)) =$

c. $\lambda e'[e' \text{ saw } g(1)]$

(29) a. $[\![\textbf{John saw } t_1]\!]^g = [\![\textbf{saw } t_1]\!]^g([\![\textbf{John}]\!]^g) = [\![\textbf{saw } t_1]\!]^g(\text{John}) =$

b. $\lambda e'[e' \text{ saw } g(1)](\text{John}) =$

c. 1 iff John saw $g(1)$

(30) a. $[\![\textbf{1 John saw } t_1]\!]^g = \lambda e[\, [\![\textbf{John saw } t_1]\!]^{g^{e/1}}\,] =$

b. $\lambda e[\text{John saw } g^{e/1}(1)] =$

c. $\lambda e[\text{John saw } e]$

---

[19] If all movement/chain-formation is driven by feature-checking, and if features get checked in a Spec-Head relation, then this index might itself be the the head driving the QP to raise. That's purely speculative.

[20] Again, we might want to build in domain restriction here, if we're so inclined. Note that the set-theory talk is really just for space-saving; if this were $[\![most]\!]$, we'd have to use English, which is effectively what we're using anyway.

(31) a. $[\![\mathbf{every\ caterpillar}]\!]^g = [\![\mathbf{every}]\!]^g([\![\mathbf{caterpillar}]\!]^g) =$

b. $\lambda u_{\langle e,t\rangle}\lambda u'_{\langle e,t\rangle}[\{e : u(e) = 1\} \subseteq \{e : u'(e) = 1\}](\lambda e[e \text{ is a 'pillar}]) =$

c. $\lambda u'_{\langle e,t\rangle}[\{e' : e' \text{ is a 'pillar}\} \subseteq \{e' : u'(e') = 1\}]$

(32) a. $[\![\mathbf{every\ 'pillar\ 1\ John\ saw\ } t_1]\!]^g =$

b. $[\![\mathbf{every\ 'pillar}]\!]^g([\![\mathbf{1\ John\ saw\ } t_1]\!]^g) =$

c. $\lambda u'_{\langle e,t\rangle}[\{e' : e' \text{ is a 'pillar}\} \subseteq \{e' : u'(e') = 1\}](\lambda e[\text{John saw } e]) =$

d. 1 iff $\{e' : e' \text{ is a 'pillar}\} \subseteq \{e' : \text{John saw } e'\}$

And this is the right denotation. Similar strategies will work for prepositional objects, indirect objects, multiple quantifiers,[21] &c., as shown in (Heim & Kratzer, 1998).

Type-shifting

Alternately, we might think there is no syntactic operation of QR. If this is the case, quantifiers in object position have to be handled in some different way. One solution is to suppose that quantifiers take a different semantic type when they occur in object position. For example, in **Jack saw every caterpillar**, we'd want $[\![everycaterpillar]\!]$ to be of type $\langle\langle e, \langle e,t\rangle\rangle, \langle e,t\rangle\rangle$, rather than $\langle\langle e,t\rangle, t\rangle$. In particular, rather than the (subject-position-adequate) denotation in (33a), we want the denotation in (33b):

(33) a. $\lambda u_{\langle e,t\rangle}[\{e : e \text{ is a caterpillar}\} \subseteq \{e : u(e) = 1\}]$

b. $\lambda v_{\langle e,\langle e,t\rangle\rangle}\lambda e'[\{e : e \text{ is a 'pillar}\} \subseteq \{e : v(e)(e') = 1\}]$

This denotation will combine just fine with the usual denotations for **saw** and **John**:

(34) a. $[\![saweverycaterpillar]\!] = [\![everycaterpillar]\!]([\![saw]\!]) =$

b. $\lambda v_{\langle e,\langle e,t\rangle\rangle}\lambda e'[\{e : e \text{ is a 'pillar}\} \subseteq \{e : v(e)(e') = 1\}](\lambda e''\lambda e'''[e''' \text{ saw } e'']) =$

c. $= \lambda e'[\{e : e \text{ is a 'pillar}\} \subseteq \{e : e' \text{ saw } e\}]$

(35) a. $[\![Johnsaweverycaterpillar]\!] = [\![saweverycaterpillar]\!]([\![John]\!])$

b. $= \lambda e'[\{e : e \text{ is a 'pillar}\} \subseteq \{e : e' \text{ saw } e\}](\text{John}) =$

c. 1 iff $\{e : e \text{ is a 'pillar}\} \subseteq \{e : \text{John saw } e\}$

This is the denotation we want.

---

[21]Note that scope ambiguities can be accommodated quite easily by this strategy; simply QR the phrases in a different order. (Although note as well that (May, 1985) argues against this way of handling scope.)

A type-shifting explanation must bear two burdens: it must give some way of calculating the post-shift denotation on the basis of the pre-shift denotation, and it must offer some reason for such a shift to occur. I won't address the latter issue here,[22] but note that the former issue can be addressed by specifying an appropriate type-shifting operator. In the case we just considered, it must take (33a) as an argument, and yield value (33b); thus it must be an operator of type $\langle\langle\langle e,t\rangle,t\rangle,\langle\langle e,\langle e,t\rangle\rangle,\langle e,t\rangle\rangle\rangle$. The following function will suffice:

(36) $\lambda q_{\langle\langle e,t\rangle,t\rangle}\lambda v_{\langle e,\langle e,t\rangle\rangle}\lambda e[q(\lambda e'[v(e')(e)])]$

Applying this function to $[\![everycaterpillar]\!]_{\langle\langle e,t\rangle,t\rangle}$ yields $[\![everycaterpillar]\!]_{\langle\langle e,\langle e,t\rangle\rangle,\langle e,t\rangle\rangle}$. The very same function will work to transform any subject quantifier (of type $\langle\langle e,t\rangle,t\rangle$) into its corresponding object quantifier (of type $\langle\langle e,\langle e,t\rangle\rangle,\langle e,t\rangle\rangle$). The type-shifting operators necessary to handle cases of ditransitive verbs or multiple quantifiers will be more complicated, but the idea remains essentially the same.

2.4.3   Definite descriptions

In this work, I'll treat definite descriptions as quantifiers. (Arguments supporting this so-called Russellian treatment can be found in eg (Neale, 1990).) One obvious pleasant consequence of this choice: our treatment of quantifiers above is sufficient for handling definite descriptions as well. A less obvious, but still pleasant, consequence: we can keep our set of truth-values small ($\{1, 0\}$), and not have to worry about adding a third value (or using partial functions) for sentences like **The present King of France is bald**. Our quantifier treatment yields the conclusion that such sentences simply take value 0, a perfectly reasonable conclusion.

2.5   Intensional Preliminaries

Over the next few sections, I'll present a semantics for various intensional constructions. This semantics is adapted from (von Fintel & Heim, 2007). The key semantic ingredient here is provided by *possible worlds*.[23] I won't worry here about what possible worlds are; everything I have to say should be compatible with any theory of possible worlds you favor. The key feature of possible worlds that will be used here: sentences are true at possible worlds, and different sentences are true at different possible worlds.

---

[22]except to suggest that this might provide a semantic role for the syntactician's AgrO head

[23]Although I'll argue later that possible worlds aren't enough, the literature on intensional semantics mostly tries to make do with them, and I'll follow suit for now.

Add to our semantics a domain D_s of possible worlds, and a primitive semantic type s, the type of possible worlds. This gives us many new types; for example, we now have a type $\langle s, t\rangle$, the type of functions from possible worlds to truth-values. This is a type that will concern us greatly; I'll abbreviate it as type p, and refer to it as the type of *propositions*.[24] In this intensional framework, full sentences will no longer denote truth values; instead they will denote propositions.

A few other abbreviations will be used here: the type $\langle e, p\rangle = \langle e, \langle s, t\rangle\rangle$ will be abbreviated as u; it is the type of the denotations of 1-place predicates. The type $\langle e, u\rangle = \langle e, \langle e, \langle s, t\rangle\rangle\rangle$ will be abbreviated as v; it's the type of the denotations of 2-place predicates. And finally, the type $\langle p, \langle e, p\rangle\rangle = \langle\langle s, t\rangle, \langle e, \langle s, t\rangle\rangle\rangle$ will be abbreviated a; it's the type of verbs of propositional attitude.

2.5.1 Individuals

It's worth commenting a bit on the treatment of type e here. Although there is a type $\langle s, e\rangle$ generated by our principles, it's not a type I'll use. (This diverges from the treatment in (von Fintel & Heim, 2007).) I'll continue to treat proper names as simply denoting things of type e. However, I will (as will become apparent in chapter 3) be treating names as somewhat intensional, at least in that, for example, $[\![Hesperus]\!] \neq [\![Phosphorus]\!]$, despite the actual truth of **Hesperus is Phosphorus**. Thus, the entities of type e I'll use here draw distinctions that things of type $\langle s, e\rangle$ are sometimes invoked to draw. That invocation is not necessary, as will become apparent.

Priest (2005) talks of individuals and their *identities* in various worlds—for him, an individual is a transworld continuant, and its identities are the guises it takes on in each world. So, for Priest, Hesperus is not (alworlds) the same thing as Phosphorus, but they do share an identity at the actual world. My treatment of individuals here is similar to Priest's, but does without identities. For Priest, the role played by identities is as objects of predication—predicates denote not things of type $\langle e, p\rangle$, as they do here, but rather things of type $\langle i, p\rangle$, where i is the type of identities.[25] This causes him a bit of awkwardness that's avoided by using fully-intensional predication of the type to be explored here. For example, (Priest, 2005) claims that, since (37a) is clearly true, so too must be (37b)–(37d).[26]

---

[24]Note that the functions that are propositions are the characteristic functions of sets of possible worlds. I'll be harmlessly (but frequently) sloppy about the distinction between the characteristic function and the set.

[25]If i is the type of identities, then Priest's individuals can be treated as of type $\langle s, i\rangle$, although he'd prefer to think of them as *determining* such functions, rather than being them.

[26]Priest gives nonexistent objects an equal footing to existent objects, so the fictional nature of the example doesn't (much) affect the point at issue—although to tighten up all the joints here we should evaluate these sentences at a world that is such that the Superman stories describe it (and so not our own world).

(37)   a. Lois Lane prefers Superman to Clark Kent.

       b. Lois Lane prefers Clark Kent to Superman.

       c. Lois Lane prefers Superman to Superman.

       d. Lois Lane prefers Clark Kent to Clark Kent.

He is forced to this conclusion by treating $[\![prefers]\!]$ as of type $\langle i, \langle i, p \rangle \rangle$, and taking the truth of **Clark Kent is Superman** to guarantee that **Clark Kent** and **Superman** denote the same identity (at the world in question). It is clear, I take it, that this is not a pre-or-even-post-theoretically desirable consequence; it is rather a bullet that Priest must bite, given his treatment of predication.

If predicate-denotations take individuals as arguments, though, and not their identities, then (37a)–(37d) can all be appropriately distinguished, even at worlds where **Clark Kent is Superman** holds. So that's precisely how I'll treat predicate-denotations.[27] This removes the semantic need for identities at all, although, to the extent that they make good metaphysical sense, they may well still be important objects of a fuller theory.

2.6   Modals

We can treat modals (words like **must**, **can**, **should**, &c.) as having denotations of type $\langle p, p \rangle$. That is, a modal denotes a function from propositions to propositions. If a modal is to occur in a sentence, then, and we use our rule of FA to conduct the calculation, the modal's sister must be something that denotes a proposition (or something that can take a modal as an argument; we won't explore that possibility here). If we look directly at a sentence like (38), though, that seems not to be the case:

(38)   John [must [call Mary]]

After all, $[\![callMary]\!]$ is presumably of type u not p. Following (von Fintel & Heim, 2007), I'll assume that the relevant structure for our purposes is not (38) but instead (39):

(39)   must [John [call Mary]]

In this structure, it should be clear, $[\![must]\!]$ can get an argument of type p.

Which function of type $\langle p, p \rangle$ should $[\![must]\!]$ be? The answer is derived from work in modal logic: we first appeal to a function R of type $\langle s, p \rangle$.[28] Then we say that a proposition p must hold at a

---

[27] For a similar argument, based on (and christening) the "Partee puzzle", see (Montague, 2002).

[28] This function plays the role of the relative possibility relation in so-called Kripke semantics for modal logic.

world s iff the proposition p holds at every world in $R(s)$ (that is, iff $R(s) \subseteq p$). Different flavors of modality (alethic, deontic, epistemic, &c.) can be modeled by using different functions in place of R. In effect, we take a modal like **must** to have an incomplete meaning; some function must be provided to play the role of R. In actual conversation, this is done implicitly, but here I'll make it explicit.

Looking at this sentence in more detail will provide an opportunity to see our intensional semantics at work.

(40) a. $[\![John]\!] = \text{John}$

  b. $[\![Mary]\!] = \text{Mary}$

  c. $[\![call]\!] = \lambda e \lambda e' \lambda s [e' \text{ calls } e \text{ in } s]$

  d. $[\![must]\!] = \lambda p \lambda s [R(s) \subseteq p]$

(41) a. $[\![callMary]\!] = [\![call]\!]([\![Mary]\!]) = [\![call]\!](\text{Mary}) =$

  b. $\lambda e \lambda e' \lambda s [e' \text{ calls } e \text{ in } s](\text{Mary}) =$

  c. $\lambda e' \lambda s [e' \text{ calls Mary in } s]$

(42) a. $[\![JohncallMary]\!] = [\![callMary]\!]([\![John]\!]) = [\![callMary]\!](\text{John})$

  b. $= \lambda e' \lambda s [e' \text{ calls Mary in } s](\text{John}) =$

  c. $\lambda s [\text{John calls Mary in } s]$

(43) a. $[\![must[JohncallMary]]\!] = [\![must]\!]([\![JohncallMary]\!]) =$

  b. $\lambda p \lambda s [R(s) \subseteq p](\lambda s [\text{John calls Mary in } s]) =$

  c. $\lambda s [R(s) \subseteq \{s' : \text{John calls Mary in } s'\}]$

This is the right denotation. To see this, we must fill in the right function for R, which will depend on what is meant by **must**. Suppose **must** is meant deontically; we mean it's required by the rules that John calls Mary. Then $R(s)$, for any world s will be the set of worlds that in which everything that happens is compatible with the rules as they are in s. (Note that there's no reason to expect in general that $s \in R(s)$, for this choice of R.) Then the proposition denoted by **John must call Mary** will be the set of worlds s such that every world compatible with s's rules is a John-calls-Mary world. These worlds are just the worlds where **John calls Mary** is required by the rules; that is, the worlds where John must call Mary, if we read **must** deontically.

Other readings of **must** can be supplied by using different functions in the place of R. This same semantics will work for other universal modals, such as **should**, **have to**, **ought to**, &c. A semantics for existential modals such as **can**, **may**, **might**, &c. can be provided as follows:

21

(44) $[\![can]\!] = \lambda p \lambda s [R(s) \cap p \neq \emptyset]$

I leave it as an exercise to show that this gives the right results.

2.7 Attitudes

Propositional attitude expressions function semantically much like modals, except that they leave less to context. They, like modals, take a proposition as an argument, but they require a bit more: someone to hold the attitude. Consider the sentence (45):

(45) Ortcutt thinks Joan is suspicious

$[\![Joan\,is\,suspicious]\!]$ is a proposition (call it p), and $[\![Ortcutt]\!]$ is Ortcutt. What relation has to obtain between Ortcutt and p for (45) to be true? von Fintel & Heim (2007) provide an answer: it must be that the worlds compatible with everything Ortcutt thinks are all in p. We can associate with the verb **think** a function T of type $\langle e, \langle s, p \rangle \rangle$; given a thinker e and a world s, this function gives us the set of worlds compatible with everything e thinks in s.[29] This suggests the following denotation for **think**:

(46) $[\![think]\!] = \lambda p \lambda e \lambda s [T(e)(s) \subseteq p]$

If we use this denotation, it's quick to find the denotation of (45):

(47) a. $[\![Ortcutt]\!] = $ Ortcutt

b. $[\![Joan]\!] = $ Joan

c. $[\![is\,suspicious]\!] = \lambda e \lambda s [e \text{ is suspicious in } s]$

(48) a. $[\![Joan\,is\,suspicious]\!] = [\![is\,suspicious]\!]([\![Joan]\!]) = $

b. $\lambda e \lambda s [e \text{ is suspicious in } s](\text{Joan}) = $

c. $\lambda s [\text{Joan is suspicious in } s]$

(49) a. $[\![thinks\,Joan\,is\,suspicious]\!] = [\![thinks]\!]([\![Joan\,is\,suspicious]\!])$

b. $= \lambda p \lambda e \lambda s [T(e)(s) \subseteq p](\lambda s [\text{Joan is suspicious in } s]) = $

c. $\lambda e \lambda s [T(e)(s) \subseteq \{s' : \text{Joan is suspicious in } s'\}]$

(50) a. $[\![Ortcutt\,thinks\,Joan\,is\,suspicious]\!] = [\![thinks\,Joan\,is\,suspicious]\!]([\![Ortcutt]\!]) = $

---

[29]Although the function plays a similar 'background' role to the relative possibility function R we used to interpret the modals, note that it is much less dependent on context, being determined simply by which attitude verb gets used (modulo effects of context on just what that verb is doing).

b. $\lambda e \lambda s[T(e)(s) \subseteq \{s' : \text{Joan is suspicious in } s'\}](\text{Ortcutt}) =$

c. $\lambda s[T(\text{Ortcutt})(s) \subseteq \{s' : \text{Joan is suspicious in } s'\}]$

This is just what we wanted: the set of worlds s such that every world compatible with what Ortcutt thinks in s is a world where Joan is suspicious.

## 2.7.1 Quantifiers in embedded clauses

Things can get more complicated, though. For example, (51a) is two-ways ambiguous; it has readings that differ in their scope, as indicated in the informal truth-conditions (51b) and (51c).

(51)   a. **Mary believes every panda ate**

       b. [Mary believes the proposition that every panda ate]

       c. [every panda x] [Mary believes that x ate]

(51b) is the reading that would be generated if we simply calculated $[\![everypandaate]\!]$ and plugged it into $[\![believe]\!]$ in the way we did in §2.7. How can we generate the reading (51c)? We can use the same tricks as in §2.4: QR or type-shifting.

QR

If we QR **every panda**, we get the tree in (52):

(52)
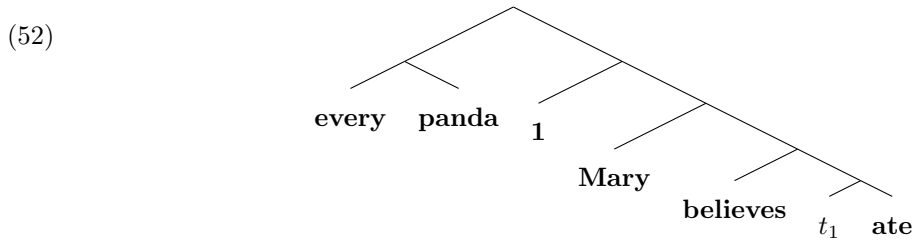
Note that this gets us the right denotation, with the rules we already have in play:

(53)   a. $[\![Mary]\!] = \text{Mary}$

       b. $[\![believe]\!] = \lambda p \lambda e \lambda s[B(e)(s) \subseteq p]$[30]

       c. $[\![every]\!] = \lambda u \lambda u' \lambda s[\{e : u(e)(s)\} \subseteq \{e : u'(e)(s)\}]$

       d. $[\![ate]\!] = \lambda e \lambda s[e \text{ ate in } s]$

       e. $[\![panda]\!] = \lambda e \lambda s[e \text{ is a panda in } s]$

---

[30]I use B here for the function that takes a believer e and a world s to the set of worlds compatible with everything e believes in s.

(54) a. $[\![t_1 \text{ ate}]\!]^g = [\![\text{ate}]\!]^g([\![t_1]\!]^g) = [\![\text{ate}]\!]^g(g(1)) =$

   b. $\lambda s[g(1) \text{ ate in } s]$

(55) a. $[\![\text{believes } t_1 \text{ ate}]\!]^g = [\![\text{believes}]\!]^g([\![t_1 \text{ ate}]\!]^g) =$

   b. $\lambda e \lambda s[B(e)(s) \subseteq \{s' : g(1) \text{ ate in } s'\}] =$

(56) a. $[\![\text{Mary believes } t_1 \text{ ate}]\!]^g = [\![\text{believes } t_1 \text{ ate}]\!]^g(\text{Mary}) =$

   b. $\lambda s[B(\text{Mary})(s) \subseteq \{s' : g(1) \text{ ate in } s'\}]$

(57) a. $[\![\mathbf{1} \text{ Mary believes } t_1 \text{ ate}]\!]^g = \lambda e[\, [\![\text{Mary believes } t_1 \text{ ate}]\!]^{g^{e/1}}]$

   b. $= \lambda e \lambda s[B(\text{Mary})(s) \subseteq \{s' : e \text{ ate in } s'\}]$

(58) a. $[\![\text{every panda}]\!]^g = [\![\text{every}]\!]^g([\![\text{panda}]\!]^g) =$

   b. $\lambda u \lambda u' \lambda s[\{e' : u(e')(s)\} \subseteq \{e' : u'(e')(s)\}](\lambda e \lambda s[e \text{ is a panda in } s]) =$

   c. $\lambda u' \lambda s[\{e' : e' \text{ is a panda in } s\} \subseteq \{e' : u'(e')(s)\}]$

(59) a. $[\![\text{every panda } \mathbf{1} \text{ Mary believes } t_1 \text{ ate}]\!]^g =$

   b. $[\![\text{every panda}]\!]^g([\![\mathbf{1} \text{ Mary believes } t_1 \text{ ate}]\!]^g) =$

   c. $\lambda s[\{e' : e' \text{ is a panda in } s\} \subseteq \{e' : B(\text{Mary})(s) \subseteq \{s' : e' \text{ ate in } s'\}\}]$

It might take a bit of parsing, but it can be seen that this is the right proposition to capture the reading we were after: it's the set of worlds s such that all the things that are pandas in s are also things such that all of Mary's belief-worlds-in-s make them out to have eaten.

Type-shifting

As in §2.4.2, if we are hesitant to postulate a syntactic movement like QR, there is a semantic option: type-shifting. Here we must suppose that $[\![everypanda]\!]$ shifts from its usual type $\langle u, p \rangle$ to the type $\langle u, \langle a, u \rangle \rangle$. Consider the following type-shifting operator:

(60) $\lambda x_{\langle u, p \rangle} \lambda u \lambda a \lambda e \lambda s[x(\lambda e'[a(u(e'))(e)])(s)]$

If we apply this type-shifting operator to $[\![everypanda]\!]_{\langle u, p \rangle}$, we get a denotation of the right type:

(61) $\lambda u \lambda a \lambda e \lambda s[\{e' : e' \text{ is a panda in } s\} \subseteq \{e' : a(u(e'))(e)(s)\}]$

The calculation done this way:

(62) a. $[\![everypandaate]\!] = [\![everypanda]\!]_{\langle u, \langle a, u \rangle \rangle}([\![ate]\!]) =$

   b. $\lambda u \lambda a \lambda e \lambda s[\{e' : e' \text{ is a panda in } s\} \subseteq \{e' : a(u(e'))(e)(s)\}](\lambda e \lambda s[e \text{ ate in } s])$

   c. $= \lambda a \lambda e \lambda s[\{e' : e' \text{ is a panda in } s\} \subseteq \{e' : a(\lambda s'[e' \text{ate in } s'])(e)(s)\}]$

(63) a. $[\![believes\,every\,panda\,ate]\!] = [\![every\,panda\,ate]\!]([\![believes]\!])$

  b. $= \lambda a \lambda e \lambda s[\{e' : e' \text{ is a panda in } s\} \subseteq \{e' : a(\lambda s'[e'\text{ate in } s'])(e)(s)\}](\lambda p \lambda e \lambda s[B(e)(s) \subseteq p]) =$

  $p]) =$

  c. $\lambda e \lambda s[\{e' : e' \text{ is a panda in } s\} \subseteq \{e' : B(e)(s) \subseteq \{s' : e' \text{ ate in } s'\}\}]$

(64) a. $[\![Mary\,believes\,every\,panda\,ate]\!] = [\![believes\,every\,panda\,ate]\!](\text{Mary}) =$

  b. $\lambda s[\{e' : e' \text{ is a panda in } s\} \subseteq \{e' : B(\text{Mary})(s) \subseteq \{s' : e' \text{ ate in } s'\}\}]$

This is the very same proposition we arrived at using the QR strategy. The decision between the QR and type-shifting strategy is a decision that has no semantic effect. It is to be made on the basis of syntax: if there is good evidence for the movement postulated by QR, we ought to suppose that's how it's done, and if there's good evidence against it, we ought to suppose type-shifting is the key here. Either way, though, we see how to provide semantics for quantifiers in embedded clauses, whatever their scope.

2.7.2 De re attitudes

A well-known ambiguity in propositional attitude reports arises in cases like 65a, which can have truth conditions as shown in 65b (the so-called de dicto reading) or 65c (the so-called de re reading):

(65) a. Joan thinks the man upstairs is a spy.

  b. Joan thinks something with the content: the man upstairs is a spy

  c. The man upstairs is such that Joan thinks *he* is a spy

The semantic framework laid out so far is already sufficient to handle this sort of case. Since I'm treating definite descriptions as quantifiers, I see the de dicto/de re ambiguity as an instance of the quantifier scope ambiguity handled above. Either strategy presented there (QR or type-shifting) will work to handle this ambiguity as well.

2.8 Negation

We're now in a position to approach negation. I will treat sentential negation like the modals we saw above: as having a denotation of type $\langle p, p \rangle$, the type of functions from propositions to propositions. A classical semantics for negation assigns it the function (66):

(66) $\lambda p \lambda s[s \notin p]$

Such a semantics has the immediate consequence that every possible world must be both negation-consistent and negation-complete. That is, for any sentence A and any world s, either the sentence or its negation must be true at s, and it cannot be that both A and its negation are true at s.

Thus, every contradiction will denote the same proposition—the empty proposition—and every tautology will denote the same proposition—the set of all possible worlds. In chapter 3, I'll argue that this gives us an unworkable semantics. Thus, the classical semantics for negation given by (66) must be mistaken. In chapters 4 and 5, I'll begin to explore the range of possible alternatives.

Chapter 3

Against structured propositions

This is an essay in compositional semantics: the project of understanding how the meanings of sentences depend systematically on the meanings of their parts, and the way those meanings are combined. One way to model this process is to adapt tools from the study of modal or other intensional logics (see eg (Montague, 2002), (Gamut, 1991b), (von Fintel & Heim, 2007)), and that's the method I'll be pursuing here. My particular task in this essay is to use data about sentences with embedded clauses to provide evidence for theories of what those clauses denote. Call whatever clauses denote, according to a semantic theory, that theory's 'propositions'; then this essay tries to adduce some evidence about what propositions are like.

Here's the plan: in §3.1, I'll discuss a traditional idea—that propositions are sets of possible worlds—and point out some familiar problems with such an approach. In §3.2, I briefly outline two possible improvements on possible-worlds propositions that solve these familiar problems—circumstantialism and structuralism. The remainder of the paper comprises arguments against structuralism and in favor of (a certain form of) circumstantialism: in §3.3 I present arguments against structuralism, and consider some structuralist responses to these arguments, and in §3.4, I answer some influential arguments against circumstantialism.

Two key pieces of notational stuff: I use **boldface type** for quotation (cuts down on quotes everywhere), and ⟦double brackets⟧ to talk about denotations of linguistic items. So, if we think names denote their bearers, then ⟦**Mary**⟧ = Mary. Here we go!

3.1 Problems with the possible-worlds approach

The intensional semantics in (von Fintel & Heim, 2007) uses sets of possible worlds as the denotations[1] of sentences and clauses.[2] For many applications, this works just fine—as (von Fintel & Heim, 2007) show, this picture of propositions allows for a quite natural semantics of modal verbs, conditionals, propositional attitudes, and other intensional constructions—but as has long been realized, the possible-worlds approach results in some trouble around the edges.

3.1.1 Too much equivalence

If propositions are just sets of possible worlds, then any two sentences true in all the same possible worlds will denote the same proposition. Suppose we have two such sentences, A and B. Then the sentences **Jacek believes that** A and **Jacek believes that** B must themselves denote the same proposition, since they are built up in the same way from parts with the same denotations; so it would be impossible for one of these sentences to be true while the other is untrue. This is a bad consequence.

Consider the sentences **Penguins party and penguins don't party** and **Cats are dogs and cats aren't dogs**. These sentences are true in the same possible worlds, viz, none at all.[3] And so, by the above argument, the sentences **Jacek believes that penguins party and penguins don't party** and **Jacek believes that cats are dogs and cats aren't dogs** must themselves denote the same proposition, and so must either be true together or untrue together. But suppose Jacek simply has one contradictory belief—say, that penguins party and penguins don't party—without having the other. Then these sentences should not be true together; rather, the former should be true and the latter false. A possible-world approach cannot provide the correct predictions here.

3.1.2 Logical omniscience

A standard semantics for, say 'believe', has it that a person believes a proposition when the person's belief-worlds (the set of worlds compatible with the person's beliefs) are a subset of the proposition. This has the consequence that if a person believes a proposition p, and p is a subset of a proposition q, then the person must believe q as well—another bad consequence.

---

[1]I use 'denotes' in an unloaded way; it's shorthand for 'takes as a semantic value'. No claims are being made about the capital-R Referential status of any linguistic items.

[2]Or, equivalently, the characteristic functions of such sets. I follow (von Fintel & Heim, 2007) in being harmlessly sloppy about the distinction.

[3]For our purposes here, I assume that no contradiction is true in any possible world.

In particular, this semantics predicts that people always believe all the logical consequences of anything they believe. After all, if A entails B, then the set of possible worlds at which A holds must be a subset of the set of possible worlds where B holds. And since A entails B whenever B is a logical truth, everyone is predicted to believe all logical truths. These are clearly unwanted consequences.

3.1.3 Counterpossibles

Another problem with the possible-worlds approach can be brought out by a pair of counterpossible conditionals:[4]

(67) a. If intuitionists are right, then the law of excluded middle isn't valid

   b. If intuitionists are right, then bees eat cats

Intuitively, (67a) is true and (67b) isn't. After all, one of the claims intuitionists make is that the law of excluded middle isn't valid. On the other hand, I don't know of any intuitionists who think that bees eat cats. But since there are no possible worlds where intuitionists are right, our semantics for conditionals cannot yield these predictions.

In particular, since the possible-worlds approach predicts a conditional to be true just when some contextually-determined set of worlds in which the antecedent holds is a subset of the set of worlds where the consequent holds, if the antecedent of a conditional holds in no possible worlds, this condition is always trivially satisfied; we must predict the conditional to be true. This is the right prediction about (67a) (although note that it's for the wrong reason), but the wrong prediction about (67b).

3.2 Circumstantialism and Structuralism

In this section, I present two possible solutions to these puzzles: I'll call them the circumstantialist and structuralist approaches. Both solutions hold that propositions must be something other than sets of possible worlds, something more fine-grained. Thus, both circumstantialists and structuralists go beyond the semantics for propositions given in (von Fintel & Heim, 2007).[5] Nonetheless,

---

[4]I assume here that these are indeed counterpossible conditionals; that is, that it's not possible that intuitionists are right. If that doesn't work for you, substitute your favorite impossibility.

[5]Some authors use the term 'circumstantialist' in such a way that possible-world approaches count as circumstantialist. Here, it will be clearer if I don't, so I won't; possible-world approaches are *not* circumstantialist, as I use the term.

much of the machinery used in (von Fintel & Heim, 2007) to provide semantics for intensional locutions will carry over to these modifications. (This carryover is often more straightforward for the circumstantialist, but I won't make much of this here.)

3.2.1 Circumstantialism

Circumstantialist propositions are like sets of possible worlds, in that they are sets of circumstances. We merely relax the assumption that all circumstances are possible. Some circumstances are impossible. For example, a circumstance at which water is not H_2O is, following (Kripke, 1980), typically thought to be (metaphysically) impossible. Nonetheless, a circumstantialist semantics will have use for circumstances at which water is not H_2O—the proposition denoted by **Water is not H_2O** will simply be the set of all such circumstances.[6]

Classical possible worlds are circumstances, and the circumstantialist certainly acknowledges them; thus, any propositions that a possible-worlds approach distinguishes will also be distinguished by the circumstantialist. However, there are sets A and B such that A and B contain all the same classical possible worlds, but $A \neq B$, since A and B contain different other circumstances. Thus, the circumstantialist can distinguish propositions that a possible-worlds approach must identify. This is how the circumstantialist handles the problems faced by the possible-worlds approach.

What circumstances are there?

Different circumstantialists invoke different sets of circumstances. Some, notably (Goddard & Routley, 1973) and (Priest, 2005), hold that for every set of sentences, there is a circumstance at which all and only the sentences in that set are true. This view, then, will allow propositions-as-sets-of-circumstances to cut very finely; in particular, every sentence will denote a distinct proposition.[7] Other circumstantialists, notably (Barwise & Perry, 1999), are not so liberal with their circumstances, but still allow for circumstances beyond just possible worlds.

What all these circumstantialists agree about, however, is that hypotheses about what circumstances there are are to be settled by paying attention to the role they play in our theory: which distinctions does a good semantics need to draw? How many distinctions are too many? As will

---

[6]Note that the circumstantialist does *not* (at least not qua circumstantialist) claim that such circumstances are possible. While someone might take issue with the (relatively orthodox) position that water is necessarily H_2O, the circumstantialist need do no such thing. (I use the water/H_2O example merely for illustration; the important part is that insofar as we can believe, doubt, wonder about, assert, question, &c. impossible propositions, the circumstantialist will need to invoke impossible circumstances.)

[7]And there will be more propositions besides; if $\mathcal{L}$ is our set of sentences, then the set of circumstances is $\wp\mathcal{L}$, and so the set of sets of circumstances will be $\wp\wp\mathcal{L}$. Put heredity restrictions on now—how's the cardinality play out?

emerge below, I disagree with the liberal view advanced in (Goddard & Routley, 1973) and (Priest, 2005); in fact, many of the arguments I deploy against structuralism apply to this variety of circumstantialism as well. The circumstantialism in (Barwise & Perry, 1999) also does not hold up well to the sort of argument to be given here.[8]

What *are* impossible circumstances, anyway?

The metaphysics of such ordinary creatures as possible worlds is a contentious and vexed topic. There is no reason to expect the metaphysics of impossible worlds to be any less contentious or vexed. At first blush, it's tempting to think that any of the many positions that has been adopted w/r/t the metaphysics of possible worlds is available as a position about the metaphysics of impossible circumstances. That may not be true; something about impossibility may well interact with certain theories of possible worlds to make those theories less (or more) plausible as theories of impossible circumstances.[9] Anyhow, the whole topic is one deserving of considerable exploration. I don't propose to conduct that exploration here.

Semantics of modals, conditionals, attitudes

The semantics for modals, conditionals, and propositional attitudes, and other intensional constructions can all be carried over unchanged from the possible-worlds approach. Since all that's changed is what things propositions are sets of, not that propositions are sets of truth-supporting states, the possible-world semantics for these constructions continues to work just fine. In particular, the problems for the possible-world approach enumerated in §3.1 can all be solved by making use of impossible circumstances.

3.2.2 Structuralism

The structuralist solution is considerably farther from the possible-worlds approach. Instead of supposing propositions to be unstructured sets, the structuralist takes propositions to have some sort of hierarchical structure. Typically, this structure is inherited from the sentence that denotes

---

[8]In fact, since (Barwise & Perry, 1999) does not acknowledge inconsistent circumstances, it has the same difficulties as a pure possible-worlds semantics handling attitude verbs with inconsistent complements.

[9]It's worth mentioning an argument in (Lewis, 1986) supposedly from modal realism against the idea of an 'impossible world'. This argument crucially assumes a particular semantics for negation, a semantics that is not available to the circumstantialist. To use Lewis's argument against circumstantialism, then, is simply to beg the question; as far as I can see, 'circumstantial realism' is as tenable as Lewis's modal realism.

the proposition. In fact, for all the structuralists whose theories I'll consider here, this is the case.[10]

For the structuralist, any two sentences with different syntax must denote different propositions, since the propositions denoted will inherit their structure from the sentences that denote them, and differently-structured propositions must be different propositions. But although different syntactic structure is sufficient for sentences' denoting different propositions, it is not necessary. To see this, consider the following two sentences:

(68)　a.　Sam saw Pam.

　　　b.　Mark ate George.

These two sentences share the same syntactic structure, but the structuralist will still hold that they denote different propositions, since that one syntactic structure is filled in in different ways. Since structuralists don't want propositions simply to *be* the sentences that denote them, the propositions they invoke won't be structured arrangements of *words*—rather, they will be structured arrangements of *what the words denote*. For (Carnap, 1956), (Lewis, 1970), and (Cresswell, 1985), what words denote is their possible-worlds intensions; thus these authors combine a structuralist approach to propositions with a possible-worlds theory of lexical denotation. In (Carnap, 1956) and (Lewis, 1970), this is done quite explicitly with an eye to the problems presented above; that is, these authors develop a possible-worlds theory of meaning, and then, realizing that it doesn't cut meanings finely enough, patch up the problem by adding syntax to their propositions. ((King, 2007) does not use the machinery of possible worlds; in this work, names denote their bearers and predicates denote the properties you'd expect them to denote, &c.)

Note that it's possible to be both circumstantialist and structuralist; for example, you could take propositions to be structured arrangements of *circumstantialist* intensions. I don't know of anyone who holds this position, and for good reason: it seems to solve the same problem twice. Circumstantialism and structuralism are both invoked to solve the same problem. If either of them works, there is no need for the other. In the arguments below, then, I'll ignore this possibility; I'll deny the circumstantialist any use of structuralist resources, and vice versa. (Structuralists, of course, can and do continue to invoke possible-worlds intensions in their theories.)

---

[10]The views I attribute here to the structuralist are common to (Carnap, 1956), (Lewis, 1970), (Cresswell, 1985), and (King, 2007), unless I specifically note otherwise.

Semantics of modals, conditionals, attitudes

The possible-world semantics for intensional constructions cannot be straightforwardly carried over to a structuralist picture; since propositions are no longer simple sets of truth-supporting circumstances, modifications have to be made. This is a challenge that any structuralist must meet. I don't doubt, however, that with sufficient formal ingenuity, such challenges can be met. So although I propose to criticize structuralism in what follows, I won't do so on these grounds. I'll assume that some adequate semantics for intensional constructions can be arrived at by the structuralist. In particular, I'll assume that the structuralist can solve all the problems for the possible-worlds approach mentioned in §3.1.

3.3   Arguing for circumstantialism—fineness of grain

The structuralist and circumstantialist pictures make different predictions about which sentences will denote identical propositions. In this section, I argue that the structuralist gets things wrong in two ways: first, by identifying propositions that should be kept distinct, and second, by distinguishing propositions that should be identified.

3.3.1   How to individuate propositions

In order to press objections of this form, I first need to say something about which propositions should be identified and which should not. To do this, I'll be appealing to a principle called 'Principle P' in (King, 2007):

Principle P: (p and q express different propositions) iff (for some propositional connective O, Op and Oq differ in truth value in some circumstance)[11]

Note that for P to be fully specified, it would have to take some kind of stand on which circumstances count. Only possible circumstances, or impossible ones too? Which choice we make doesn't affect the arguments that follow, so it seems to me the best move is to explicitly restrict 'some circumstance' in P to possible circumstances. After all, some parties to the debate don't believe in impossible circumstances, and possible circumstances are enough to do all the work we'll need. Now, let's break P down into its two directions.

P (RTL): If for some propositional connective O, Op and Oq differ in truth value in some circumstance, then p and q express distinct propositions.

---

[11]King uses 'express' where I use 'denote'—we mean the same.

The right-to-left direction of P is straightforwardly entailed by our compositional picture of semantics. If the function $[\![O]\!]$ yields different outputs when given inputs $[\![p]\!]$ and $[\![q]\!]$, then it must be that $[\![p]\!] \neq [\![q]\!]$. (It's this direction that caused all the trouble for possible-worlds approaches above.)

P (LTR): If the sentences p and q express different propositions, then for some propositional connective O, Op and Oq differ in truth value in some circumstance.

P (LTR) (contrapositive): If there is no propositional connective O such that Op and Oq differ in truth value in some circumstance, then p and q express the same proposition.

The left-to-right direction of P is more controversial than its right-to-left direction. In presenting Principle P, (King, 2007) accepts the right-to-left direction, but argues against its the left-to-right direction (I'll discuss his argument in §3.3.4). Nonetheless, I think the left-to-right direction is an important guideline in developing a semantic theory; if held to, it helps us avoid drawing needless distinctions.

To see this, a few examples will help. First, imagine a language something like ours, but with no contexts that could discriminate between necessarily equivalent sentences. That is, this language would include no clause-embedding verbs like 'believes', 'says', 'proves', &c., no conditionals, none of the constructions that cause problems for a simple possible-worlds theory.[12] A semantics for this language (ex hypothesi) would have no need to appeal to impossible circumstances, and so it shouldn't appeal to them. It could, and should, make do with just the possible. See if you don't agree. But why would this be? Presumably because of something like the left-to-right direction of Principle P.

A second example makes the point in a different way. Start with your favorite theory of propositions (if you don't have a favorite theory, humor me for a moment and pretend). Now imagine the following theory of propositions: the proposition denoted by a sentence p is the ordered pair whose first member is the proposition assigned to p by your favorite theory and whose second member is the date on which p was first uttered, or, if p has never been uttered, then the number of letters in p. It should be clear that this theory (call it the *silly theory*) is less good than your favorite theory. But why? The silly theory can draw every distinction your favorite theory can draw, and more besides; so we can hardly convict it of a violation of the right-to-left direction of Principle P. The natural thing to say about the silly theory is that it draws *too many* distinctions, or *useless* distinctions;

---

[12]The language of modal logic, for example, might be such a language. Note that the language could still contain clause-embedding verbs that meant something like 'believes something true at all the same possible worlds as', although they'd have to be primitive, since we're supposing that 'believes' is ruled out.

after all, the denotations of complexes involving p will depend only on the first member of $[\![p]\!]$—the second member is idle. It is the left-to-right direction of Principle P that rules out the silly theory, by ruling out distinctions that don't get used.

This is enough intuitive support for P for now. Further on, I'll consider King's objection to it; for now, I want to point out the trouble P, if it is right, causes for the structuralist.

3.3.2 Propositions that should be distinct

Here, I give some examples of propositions that should be distinct, according to Principle P, but that the structuralist cannot distinguish.

Woodchucks and groundhogs

A classic example comes from (Lewis, 1970):[13] consider these sentences:

(69)   a. All woodchucks are groundhogs

      b. All woodchucks are woodchucks

It should be clear that (69a) and (69b) have the same syntax. Assume (what no party to this debate has denied) that 'groundhog' and 'woodchuck' are necessarily coextensive; there is no possible world where something is one but not the other.[14] Then, for the structuralist, $[\![\mathbf{groundhog}]\!] =$ $[\![\mathbf{woodchuck}]\!]$. Since the propositions denoted by these sentences are made of the same denotations arranged in the same way, they are themselves identical for the structuralist: $[\![(69a)]\!] = [\![(69b)]\!]$.[15]

But now note (70a)–(71b):

(70)   a. Alice believes that all woodchucks are groundhogs

      b. Alice believes that all woodchucks are woodchucks

---

[13]In (Lewis, 1970), Lewis calls it an 'oddity' that (69a) comes out logically true on his view. That is indeed an oddity, but the problem I'm pushing on is not about logical truth; it rather comes from the structuralist's failure to distinguish the propositions in play here.

[14]Note that one might resist this conclusion; a descriptivist about natural kinds might take the woodchucks at a world simply to be the things that chuck wood (or whatever) at that world, and the groundhogs at a world the things that hog ground (or whatever) there. Although these happen to be the same things at our world, they might come apart at other worlds. For my purposes here, though, I assume a more Kripkean picture of natural kinds, ignoring this descriptivist alternative. Similar issues arise below, in the case of names, and I'll make a similar assumption there. (This issue is orthogonal to the circumstantialist/structuralist debate, as long as there are at least two necessarily coextensional predicates.)

[15]This objection applies directly to the accounts in (Carnap, 1956), (Lewis, 1970), and (Cresswell, 1985). A structuralist following the model of (King, 2007) might have resources to dodge this; remember, King doesn't take words to denote their possible-worlds intensions, but rather individuals and properties directly. Whether such a structuralist can distinguish these propositions depends on whether the properties groundhoghood and woodchuckhood are distinct. If they are not distinct, the same objection applies, but if they are, there's an escape route for the structuralist here (if a relatively unpalatable one).

(71) a. Billy fears that all woodchucks are groundhogs

b. Billy fears that all woodchucks are woodchucks

It's clear that (70a) and (70b) (and (71a) and (71b)) can differ in truth value in some circumstance; just imagine Alice not realizing that woodchucks and groundhogs are the same critters (or Billy, who knows he's deathly allergic to groundhogs, realizing he's been bitten by a woodchuck). Thus, 'Alice believes that' (and 'Billy fears that') are propositional operators of the sort relevant to Principle P, and so (69a) and (69b) should denote distinct propositions.

The circumstantialist has no problem distinguishing these propositions; a circumstantialist theory can allow that $[\![\mathbf{woodchuck}]\!] \neq [\![\mathbf{groundhog}]\!]$, since it can allow for impossible circumstances at which something is a groundhog but not a woodchuck, and vice versa. Since these are impossible circumstances, there is no need for a circumstantialist to deny the necessary coextensiveness of 'woodchuck' and 'groundhog'.

The problem for the structuralist is caused by the syntactic simplicity of 'groundhog' and 'woodchuck'. Since the structuralist strategy is to appeal to syntax (or divergence in possible-worlds intension) to distinguish what needs to be distinguished, where there is no syntax (or divergence in possible-worlds intension) the structuralist strategy simply won't work.[16]

Frege's puzzle

A similar problem arises in Hesperus/Phosphorus cases. Consider:

(73) a. Hesperus is Phosphorus

b. Hesperus is Hesperus

Again, since (what no party to this debate denies) Hesperus is Phosphorus in every possible world, the structuralist concludes that $[\![\mathbf{Hesperus}]\!] = [\![\mathbf{Phosphorus}]\!]$. Since (73a) and (73b) have the same syntax, they denote the same structured proposition. But again, this results in a violation of Principle P, since (74a) and (74b) can differ in truth value in some circumstance (in fact, they actually differ in truth value):

(74) a. The Babylonians believed that Hesperus is Phosphorus

---

[16]NB: The structuralist has no trouble distinguishing the propositions denoted by (72a) and (72b):

(72) a. All bachelors are unmarried men

b. All bachelors are bachelors

Although there is no possible world where something is a bachelor without being an unmarried man (or vice versa), 'bachelor' and 'unmarried man' are syntactically different, so the structuralist strategy works here.

b. The Babylonians believed that Hesperus is Hesperus

The circumstantialist view, on the other hand, can allow that $\llbracket \textbf{Hesperus} \rrbracket \neq \llbracket \textbf{Phosphorus} \rrbracket$ without having to suppose that it's possible for anything to be one without being the other, simply by allowing that there are *impossible* circumstances where something is Hesperus but not Phosphorus (or vice versa). Thus, for the circumstantialist, $\llbracket (73a) \rrbracket \neq \llbracket (73b) \rrbracket$, the proper conclusion.

But it is not simply that the structuralist view does not cut propositions finely enough. In other cases, it cuts propositions too finely.

3.3.3 Propositions that should be identified

There are a few different sorts of examples to be considered here.[17] In each case, the structuralist will be shown to distinguish propositions that should be identified, according to Principle P, while the circumstantialist (of a certain stripe) need not. In passing, these cases will show that the circumstantialism advanced by eg (Goddard & Routley, 1973) and (Priest, 2005) (according to which there is a circumstance for every set of sentences that makes true all and only the sentences in that set) also draws too many distinctions, and so also violates Principle P.

Passivization

Consider (75a) and (75b):

(75)   a. Bears ate my neighbor

       b. My neighbor was eaten by bears

Claim: there is no propositional operator O such that O(75a) and O(75b) differ in truth value in any circumstance. After all, if someone thinks that bears ate my neighbor, then they think that my neighbor was eaten by bears; if someone said that bears ate my neighbor, they said that my neighbor was eaten by bears; if they're about to prove that bears ate my neighbor, they're about to prove that my neighbor was eaten by bears; &c.[18] But then, by Principle P, (75a) and (75b) should

---

[17]Since Carnap's account structures propositions according to sentential argument structure rather than syntax broadly understood, his account cuts propositions more coarsely than the other structuralists considered here, and he is thus immune to *some* of the objections that follow. Footnotes will point out which objections don't apply to Carnap. The other structuralists in question all line up as targets of each of the following arguments.

[18]There are *sentential operators* that fit the bill; just use quotation marks. But Principle P had better exclude such; otherwise it would require different propositions, eg, for each typeface a sentence might occur in, since (76a) and (76b) can differ in truth value:

(76)   a. She wrote her book in the same typeface as 'this'
       b. She wrote her book in the same typeface as '$\mathcal{THIS}$'

.

denote the same proposition. The structuralist, though, must distinguish the propositions denoted by these sentences. After all, the sentences themselves have different syntactic structure; so too, then, must the propositions they denote, for the structuralist.[19,20]

The circumstantialist of a certain stripe, though, can make the right prediction here; she merely needs to suppose that there are no circumstances in which (75a) is true but (75b) isn't, or vice versa. The circumstantialist theories of (Goddard & Routley, 1973) and (Priest, 2005), though, cannot make the right predictions here, as they do allow such circumstances. This shows how observation of which propositional operators do and do not occur can constrain our theories of circumstance.

Idempotence

Consider (77a)–(77c):

(77) a. Penguins waddle

b. Penguins waddle and penguins waddle

c. Penguins waddle or penguins waddle

Similarly, no propositional operator discriminates between any of these three, and yet, since the sentences have different syntax (or, in the case of (77b) and (77c), differently-denoting lexical items occupying slots in the syntax), the structuralist must say they denote different propositions. This is another violation of Principle P.

Again, the circumstantialist need have no trouble here. Simply supposing that the usual (classical) truth-at-a-circumstance conditions for conjunction and disjunction hold over all circumstances, possible and impossible, yields the desired result. Further, this option is not available to (Goddard & Routley, 1973) and (Priest, 2005); they too must violate Principle P here.

Commutativity, Expletives, Ditransitives

Similar trouble is generated by each of the following pairs of sentences:

(78) a. Carrots are orange and spinach is tasty

b. Spinach is tasty and carrots are orange

(79) a. Carrots are orange or spinach is tasty

---

[19]Is passivization a syntactic transformation? Are these sentences identical at some level of syntactic analysis? No and no!

[20]Depending on how argument-theory interacts with passivization, Carnap might be immune to this charge.

b. Spinach is tasty or carrots are orange

(80) a. Two cats are juggling

    b. There are two cats juggling

(81) a. Pam gave the keys to Sally

    b. Pam gave Sally the keys

By now, the argument is familiar: no propositional operator discriminates between the members of either pair,[21] but since the members differ syntactically, the structuralist must assign them distinct propositions, in violation of Principle P. The circumstantialist has no such obligation.

Translation

Although this objection is structurally much the same as the objections above, it'll bear a bit of extra discussion. Consider (83a)–(83g):

(83) a. Snow is white

    b. Schnee ist weiss (German)
       snow   is white
       'Snow is white'

    c. Snö är vitt   (Swedish)
       snow is white
       'Snow is white'

    d. La nieve es blanca (Spanish)
       the snow is white
       'Snow is white'

    e. Yuki-wa   shiroi-des (Japanese)
       snow-TOP white-is
       'Snow is white'

    f. Ha-shelleg lavan (Hebrew)
       the-snow   white
       'Snow is white'

---

[21]As is well known, conjunct and disjunct order can sometimes matter:

(82) a. I urinated and went to sleep
    b. I went to sleep and urinated
    c. You should eat your vegetables or you're going straight to bed
    d. You're going straight to bed or you should eat your vegetables

In these cases, there are propositional operators that can discriminate. But whatever the explanation is of these phenomena, it should be clear that nothing like this is afoot in (78) and (79).

    g. Nix   nivea est (Latin)
       snow white is

       'Snow is white'

On a structuralist view, (83a)–(83c) denote the same proposition; after all, they have the same syntax and co-denoting lexical items filling that syntax in. But this proposition is different from the propositions the structuralist assigns to (83d) or (83e) or (83f) or (83g), since the syntax differs in each of these cases.

Now, no monolingual speaker speaks a language that allows more than one of the above to be embedded in a propositional operator, so it might at first seem that Principle P doesn't tell us much about (83a)–(83g). But this would be too fast; there are multilingual speakers who are quite comfortable with switching languages mid-sentence. Consider such a speaker; for concreteness say she's a native speaker of both English and Spanish. Both 'Arnold said that snow is white' and 'Arnold said que la nieve es blanca' are in her idiolect; I claim that these will not differ in truth-value in any circumstance, and the same will hold for every propositional operator.[22] If my claim is right, then the structuralist, by assigning different propositions to (83a) and (83d), violates Principle P. The circumstantialist, again, has no difficulty accommodating this data.

On the other hand, suppose my claim is wrong. Suppose, that is, that there is some propositional operator O such that which language a clause p is in can make a difference to the truth value of Op in some circumstance. The structuralist can't deliver this result either; the structuralist must identify the propositions expressed by (83a) and (83c), and so must identify the propositions denoted by eg 'Arnold said that snow is white' and 'Arnold said att snö är vitt' in the mouth of an English/Swedish bilingual speaker, again in violation of Principle P (the other direction this time). The circumstantialist, by allowing for circumstances where snow is white without snö's äring vitt, can accommodate the data.

3.3.4 Structuralist responses

King and too many propositions

In (King, 2007, pp. 95–101), King considers some versions of the too-many-propositions objections and bites the bullet, arguing that the propositions the structuralist must distinguish should in fact be kept distinct. He does not directly consider the idempotence objection of §3.3.3, but his response

---

[22]These sentences *will* differ in assertibility in some circumstances; if our speaker is talking to someone she knows to be a monolingual English speaker, for example, a fully-English sentence would be more cooperative than an English-Spanish sentence. But their truth values will not come apart.

to the commutativity objection in §3.3.3 seems to generalize to cover it. He treats the translation objection of §3.3.3 separately.[23]

King's response to the conjunct-order objection proceeds as follows. First, he attributes principle P to the objector:

P: Two sentences p and q express distinct propositions if and only if for some propositional connective O, Op and Oq differ in truth value in some circumstance.

Of course, in my case, this is a fair attribution; I've used Principle P explicitly throughout my objections, and offered a brief defense of it above in §3.3.1. King, though, means to attack the left-to-right direction of Principle P, by giving a reason for two sentences to denote different propositions despite exhibiting identical embedding behavior. I quote his reasoning in full:

> Suppose that it turns out that (e.g.) '1 = 2' and '2 = 1', when embedded relative to any propositional connective, result in sentences that have the same truth values in all circumstances. P says that they express the same proposition. But there are reasons for holding that they don't. Assume that '=' expresses a relation and that '1' and '2' are names of objects. Now we know that for other expressions that express relations and those very same names, switching the order of the names results in a different proposition expressed. '2 > 1' and '1 > 2' express different propositions. It is quite reasonable to suppose that the propositions expressed by '2 > 1' and '1 > 2' have the same constituents (e.g. two objects and a relation) and differ in the way in which these constituents are put together. Thus '2 > 1' and '1 > 2' suggest that in a sentence consisting of two names flanking a relation sign, the different possible orders of the names encodes [sic] some difference in the way in which the entities named and the relation expressed combine to form a proposition. And this gives us reason to think that '1 = 2' and '2 = 1' express different propositions in virtue of having *their* constituents differently combined (King, 2007, pp. 96–97).

Applied to our (78a) and (78b), his argument runs as follows: it's clear that 'Carrots are orange if spinach is tasty' and 'Spinach is tasty if carrots are orange' express distinct propositions. But since these sentences contain all the same constituents, it must be the different order of the clauses that makes the difference. This suggests that the difference in the order of the clauses encodes some difference in the way the propositions denoted by the clauses and the operation denoted by the clausal connective are combined to form a proposition. That same difference should then be encoded by the difference in the order of the clauses in (78a) and (78b), so the propositions denoted by 'carrots are orange' and 'spinach is tasty', along with the operation denoted by 'and', should be combined in a different way to arrive at the propositions denoted by the two sentences.[24]

---

[23]He does not consider the objections from passives, expletives, or ditransitives, and the responses he offers to the other objections do not seem to generalize to cover these problems.

[24]This response seems to generalize without much difficulty to our cases (77a)–(77c); we need merely consider a non-idempotent operator like 'if'. Clearly 'penguins waddle' and 'penguins waddle if penguins waddle' express distinct propositions, so repetition can matter. The rest of the argument is the same, *mutatis mutandis*.

It's true that examples like '2 > 1' and 'Spinach is tasty if carrots are orange' lead us to conclude that the order in which names or clauses occur in a sentence *can* make a difference to the proposition expressed. But why should we conclude from that that it *must*? King's story: since order can matter, it must make a difference in the way constituents are combined. But then it should make this difference no matter what the constituents are.

We can (and should) accept all this, though. It still does not lead to King's conclusion. Grant that the constituents of the propositions expressed by (78a) and (78b) are combined in different ways; there is still an open question—have those different ways resulted in the same proposition, or have those different ways resulted in different propositions? It's not clear what could settle this question, other than Principle P. King seems to go with the latter option; he thinks the different ways of combining '[give] us reason' to think we've ended up with different propositions, but it's not clear what that reason is.

It's worth noting that King presents this argument in the context of defending his account against other structuralist accounts, not in defending it against circumstantialist accounts. It may be that, if we can assume structuralism (in particular, if we assume that propositions must reflect in their structure how they are arrived at), King's argument becomes more convincing; note, though, that it doesn't address the circumstantialist concerns presented here.

King and translation

King has a separate response to something like the objection from translation presented in §3.3.3. His presentation of the objection differs importantly from the presentation there, and his response to it does not directly address the objection pressed there; nor is it clear how he would respond to the present objection. To see this, let's look at his presentation of a translation objection and his response.

King (p. 98) considers three different claims an objector might make:

A. At least some proposition(s) can be expressed in different natural languages.

B. At least some proposition(s) expressed in one natural language can be expressed in any natural language.

C. All propositions that can be expressed in one natural language can be expressed in any other.

He grants A, and shows that B and C come out as substantive empirical questions on his account, which he claims is as it should be. It should be clear that the discussion in §3.3.3 makes no claims

as strong as B or C; out of these three claims, I'm only willing to advance claim A. And since King is willing to grant A, it might seem that there is no dispute to be had here.

But this is too fast. King grants A on the following grounds:

> An important part of the philosophical motivation for propositions is the intuition that the same piece of information can be encoded by means of different sentences, whether in different languages or in the same language. Hence some might take A to be a sort of constraint on any theory of propositions. To take a clear case, it would be desirable for a theory of propositions to yield the result that 'Schnee ist weiss' and 'Snow is white' express the same proposition. Any theory which doesn't yield this result will have a lot of explaining to do, [punctuation sic] (though I for one don't think this by itself should sink a theory) (King, 2007, p. 99).

So King's clear case is the pair of our (83a) and (83b). To the extent that these clearly denote the same proposition, it should be just as clear that (83d)–(83g) denote that proposition as well; but King's account cannot deliver this consequence, and so has 'a lot of explaining to do'. No such explanation is offered, though.[25]

Note as well that this quote itself comes dangerously close to pressing the above objections against the structuralist account. What are the different sentences that are supposed to encode the same information? We might think of pairs like 'Nobody's in the room' and 'There isn't anybody in the room', but the structuralist can't allow that these denote the same proposition; after all, they differ syntactically. About the only sentences that the structuralist *can* allow denote the same proposition are sentences like (69a) and (69b)—that is, sentences that *shouldn't* be taken to denote the same proposition! If the intuition in question is really an important part of the motivation for propositions, again the circumstantialist fares much better than the structuralist.

Structuralism Plus

This sort of response allows structuralists to draw the right sort of distinctions between woodchuck and Hesperus cases. 2 responses:

- Why structuralism at all, then?

- Can draw enough distinctions now, but still draw too many.

---

[25]Someone might object, say, that 'white' is not a perfect translation of 'blanca' (say, because they denote slightly different properties), and that the propositions expressed by (83a) and (83d) should be distinguished for that reason. If indeed they are not perfect translations of each other, then these propositions should be distinguished, and both the circumstantialist and the structuralist have the resources to do so. My point here is that if they *are* perfect translations of each other, then the propositions should not be distinguished; the circumstantialist can deliver this result, while the structuralist cannot.

3.4 Answering arguments for structuralism

In this section, I consider two influential arguments for structuralism: an argument from de re propositional attitudes (in (Cresswell, 1985)), and an argument from direct reference (in (Soames, 1987)). In both cases, I claim that these arguments miss their target; the arguments, properly understood, do not offer any reason to believe in structuralism. Cresswell's argument amounts to showing that a certain bad semantics for de re attitudes is unavailable to the circumstantialist, when in fact there is a better semantics equally available to the circumstantialist and the structuralist, and Soames's argument amounts to pressing the familiar Hesperus/Phosphorus problem on the circumstantialist, when it is no less a problem for the structuralist. I offer a potential solution to this problem, but this solution is unavailable to the (usual) structuralist. Thus, Soames's argument, far from putting pressure on the circumstantialist, is hoist by its own petard: it provides us further reason to believe in circumstantialism.

3.4.1 De re attitudes

In (Cresswell, 1985) and (Cresswell & Von Stechow, 1982), Cresswell and Von Stechow present and argue for a structuralist view. Their core argument involves de re belief attributions; they think that structuralism is the key to providing semantics for both readings of sentences like (84):

(84)   Mary thinks that the man upstairs is a spy

On one reading of this sentence (the de dicto reading), it is true iff Mary has a thought with the content *the man upstairs is a spy*. On its other reading (its de re reading), it's true iff Mary thinks *of the man upstairs* that he is a spy, whether or not she thinks of him under that description, or even knows that he is the man upstairs.

Cresswell and Von Stechow use structuralism to provide these readings. On their semantics, the complementizer **that** (or its silent counterpart, in eg **Mary thinks the man upstairs is a spy**) is ambiguous. On one of its readings, ⟦**that the man upstairs is a spy**⟧ is a set of possible worlds: the worlds w at which the man upstairs in w is a spy in w. On the other reading of **that**, ⟦**that the man upstairs is a spy**⟧ is a structured proposition: [[⟦**the man upstairs**⟧]] [⟦**is a spy**⟧]]. Note that this requires, as they point out, that ⟦**that**⟧ (on the latter reading) doesn't take ⟦**the man upstairs is a spy**⟧ as a single argument, but rather that it takes the denotations of the constituents of **the man upstairs is a spy** severally as arguments (or at least that it take

⟦**the man upstairs**⟧ and ⟦**is a spy**⟧ as two separate arguments).[26]

On either reading of **that**, ⟦**thinks**⟧ is a binary relation between a thinker and something thought; sometimes thinkers think sets of possible worlds, and sometimes they think structures.[27] When we think the structure [[⟦**the man upstairs**⟧] [⟦**is a spy**⟧]], we are having a de re thought about the man upstairs—that he is a spy. On the other hand, when we think the set of possible worlds in which the man upstairs is a spy, we are having a de dicto thought to the effect that the man upstairs is a spy.

It's worth pointing out that this solution is not available to the (pure) circumstantialist. Since it crucially requires ⟦**that the man upstairs is a spy**⟧ to (sometimes) be a structured proposition, it requires some **that**-clauses to denote structured propositions. But the circumstantialist doesn't think that anything ever denotes a structured proposition. So this solution is resolutely anti-circumstantialist. Since the structuralist has this way to provide semantics for both readings of (84), and the circumstantialist cannot avail herself of it, there seems to be an argument for structuralism here.

My response is twofold: first, I'll argue that this semantics for de re attitudes is problematic. Second, I'll point to a classic strategy for handling de re attitudes, one that's equally available to the structuralist and the circumstantialist, and doesn't share the problems of Cresswell and Von Stechow's semantics. Since the problems in their semantics should lead us to abandon it for a more workable semantics, and since that more workable semantics is available to everyone, I conclude that de re attitudes do not provide any reason to adhere to structuralism over circumstantialism.

Two problems with Cresswell and Von Stechow's approach

The first problem with Cresswell and Von Stechow's semantics was hinted at above; I present it here in the form of a dilemma. Either **that** is both semantically and syntactically ambiguous, or it's semantically ambiguous without being syntactically ambiguous.[28] Suppose first that **that** is

---

[26]More on this presently.

[27]Cresswell (Cresswell, 1985, p. 29) takes propositional attitude verbs to be of the same category as other binary relation verbs, like 'kick' or 'love'. He concludes from this that **that**-clauses are noun phrases or nominals. If this is right, thinkers actually think sets of possible worlds, or we think structures (which sounds at best awkward to my ear). There's decent syntactic evidence against this view of **that**-clauses—see eg (King, 2007, p. 107)—and it seems to be idiosyncratic to Cresswell among the structuralists I consider. My response to Cresswell's argument, accordingly, doesn't depend on this idiosyncrasy; a more orthodox structuralist presumably could press a similar argument from de re belief, and the response I'll offer presently would answer that similar argument as well.

[28]It's not clear to me which of these best captures Cresswell and Von Stechow's view. The Appendix to Cresswell & Von Stechow (1982), where this issue comes closest to being resolved, uses ' "logical form" ' (with scare-quotes) to name the level at which surface ambiguity is to be resolved. If this is the syntactician's LF, then their theory holds that **that** is syntactically ambiguous, since it can result in many different LFs; but if their ' "logical form" ' is not LF, then perhaps they mean to posit only a semantic ambiguity.

both semantically and syntactically ambiguous. Then (84) (renumbered here (85a)) has two distinct syntactic parsings: (85b) and (85c).

(85)   a.   Mary thinks that the man upstairs is a spy

b.

Mary

thinks

that   X

the man upstairs is a spy

c.

Mary

thinks

that   Y   Z

the man upstairs   is a spy
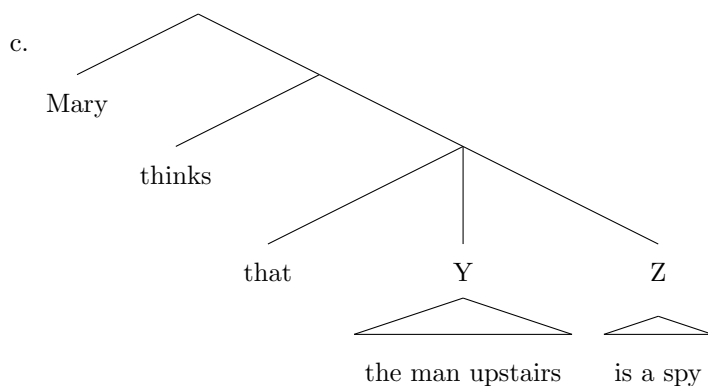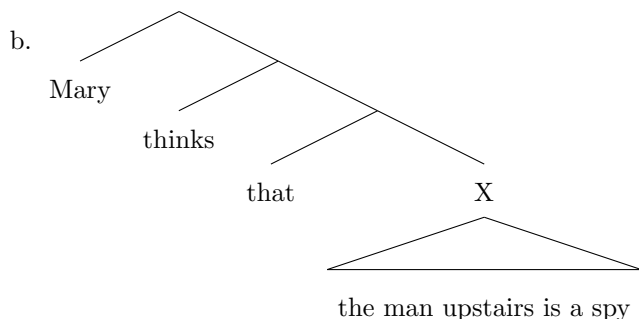
In (85b), **that** has as its only sister the node X, which contains the entire embedded clause (I ignore the structure within the clause here; it's not relevant). On the other hand, in (85c), **that** has two sisters: the node Y, containing the subject of the embedded clause, and Z, containing its predicate. The problem with this hypothesis is simple: there is no syntactic evidence of this supposed structural ambiguity (and they point this out, in fact). In fact, supposing that there is more-than-binary branching at all goes against much well-motivated syntactic theory.[29]

On the other hand, suppose that Cresswell and Von Stechow do not mean to posit a syntactic ambiguity, but rather only a semantic ambiguity. Note that the semantic ambiguity in question is not just ambiguity of the 'bank'/'bank' sort, where both readings are still one-place predicates. Rather, the ambiguity in **that** is (among other things) an ambiguity in how many arguments ⟦**that**⟧ takes. On the de dicto reading of (85a), it can get away with taking a single argument: ⟦**the man upstairs is a spy**⟧.[30] On the de re reading, though, ⟦**that**⟧ takes two arguments: ⟦**the man upstairs**⟧ and ⟦**is a spy**⟧. Unlike the syntactic-ambiguity proposal above, though, this

---

[29]See eg Kayne (1984) or Haegeman (1994). Note as well that Cresswell and Von Stechow's account, interpreted as positing a syntactic ambiguity, is going to force not just trinary branching, but n-ary branching for any finite n.

[30]Actually, this isn't quite true; see Cresswell & Von Stechow (1982) for details. But nothing here hangs on the difference.

proposal has it that on both readings, (85a) has the syntactic structure given in (85b).

But this breaks the tight link between syntax and semantics that is a core feature of compositional semantics. Of course, we might find in the course of our exploration that we need to violate this feature at some point, but we should be very hesitant—this guiding assumption has proven fruitful in the past, and it constrains our theory in useful ways. Besides, there are perfectly good ways (see below) of providing semantics for de re attitudes that allow semantics to follow syntax.

The second problem comes into sharp relief if we look at a fuller range of data. (84) was only two-ways ambiguous, depending on whether the description was read de dicto or de re. But (86) is four-ways ambiguous:

(86)   Billy thinks that my favorite goat is the smallest goat in Tennessee

Each definite description in the **that**-clause can be read either de dicto or de re, resulting in four possible readings. Cresswell and Von Stechow provide these readings by positing more ambiguity in **that**; it's at least four-ways ambiguous, actually. But the phenomenon iterates:

(87)   a. John thinks that my doctor gave Mary's cat to the Postmaster General

       b. John thinks that my doctor gave the toy Mary's cat used to play with to the tallest goat that lives at Lucy's farm

Cresswell and Von Stechow should conclude from (87a) that **that** is at least eight-ways ambiguous, and from (87b) that **that** is at least eighteen-ways ambiguous,[31] and it doesn't stop there; more (and more-ways ambiguous) examples can be generated ad infinitum.

I don't think that positing more and more ambiguity in **that** (whether purely semantic, or both semantic and syntactic) is a helpful way to understand this data. In particular, it misses the following generalization: the same de dicto/de re ambiguities occur in the following sentences:

(88)   a. Billy wonders whether my favorite goat is the smallest goat in Tennessee

       b. Mary asked who my doctor gave Mary's cat to

In (88a) and (88b), the ambiguity cannot be an ambiguity in **that**, but instead must be an ambiguity either in **whether** and **who**. Positing an ambiguity in **that**, then, has missed the core of the phenomenon. Cresswell and Von Stechow acknowledge this generalization, and that their account misses it.

---

[31]Note that when a description is embedded in another description (as **Mary's cat** in **the toy Mary's cat used to play with**, or **Lucy's farm** in **the tallest goat that lives at Lucy's farm**, there are only three possible readings for the larger description: de re, de dicto/inner-description-de-re, and de dicto/inner-description-de-dicto. If the larger description is read de re, the embedded description cannot be read de dicto; it's not even clear what such a reading would amount to.

It would be better, then, if we had some other way to understand the de re attitudes. Fortunately, we do; for example, (von Fintel & Heim, 2007) offers a semantic treatment of de re attitudes that relies on a posited syntactic transformation. On this hypothesis, (85a) is syntactically ambiguous, but not between the structures in (85b) and (85c); rather, the relevant structures are as follows:

(89)   a. Mary thinks that [[the man upstairs] is a spy]

   b. [the man upstairs]$_x$ Mary thinks that [x is a spy]

Given the usual semantic values for the words in (84), these two structures result in the appropriate truth conditions without invoking any lexical ambiguity, and independently of any issues having to do with structuralism or circumstantialism. (This approach yields all the appropriate truth conditions for (86)–(87b) as well.) The question, then, is whether this syntactic transformation is independently plausible (as the earlier one was not); research in eg (May, 1985), (Aoun & Li, 1989), and (Bruening, 2001) indicates that it is.

To summarize: Cresswell and Von Stechow's semantics for de re attitudes is indeed unavailable to the circumstantialist. However, their semantics either posits a syntactic transformation for which there is no evidence, or else breaks the tight connection between semantics and syntax that forms the backbone of our compositional semantics. What's more, their semantics must postulate wild (more than n-ways for any finite n) ambiguity in **that**, and corresponding ambiguity in other complementizers as well. If there is a semantics for de re attitudes that does not have these features, it would be preferable.

Fortunately, there is: a semantics that makes use of an independently plausible syntactic transformation. Cresswell and Von Stechow's semantics should thus be abandoned. What's more, the more plausible analysis is equally compatible with structuralism and circumstantialism; thus, de re attitudes provide no evidence whatever for structuralism.

3.4.2   Direct reference

In (Soames, 1987), an argument is given to the effect that circumstantialist propositions can't be the whole story. This argument appeals to a direct-reference theory of proper names (according to which the semantic value of any name is its bearer) and a few uncontroversial assumptions.

The main thrust of the argument goes as follows: since Hesperus = Phosphorus, $[\![$**Hesperus**$]\!]$ = Hesperus, and $[\![$**Phosphorus**$]\!]$ = Phosphorus, $[\![$**Hesperus**$]\!]$ = $[\![$**Phosphorus**$]\!]$. So, since they are built up in the same way from identical semantic values, it follows that

$[\![$**'Hesperus' refers to Hesperus and 'Phosphorus' refers to Phosphorus**$]\!]$ = $[\![$**'Hesperus' refers to Hesperus**

These semantic values are sets of circumstances of some sort, and so they must be the same set of circumstances (here is where the auxiliary assumptions get put to work; since I won't question any of them, I'll leave them tacit) as

⟦'Hesperus' refers to Hesperus and 'Phosphorus' refers to Phosphorus and there is some x such that 'Hesperus' and 'Phosphorus' both refer to x⟧.

But now there is trouble for the use of these sets of circumstances in a story about propositional attitudes. After all, the ancients believed that 'Hesperus' refers to Hesperus and 'Phosphorus' refers to Phosphorus, and the ancients didn't believe that 'Hesperus' refers to Hesperus and 'Phosphorus' refers to Phosphorus and there is some x such that 'Hesperus' and 'Phosphorus' both refer to x. But if the **that**-clauses in these sentences denote the same proposition, as Soames argues the circumstantialist must maintain, then this difference in attitudes can't be captured. So no circumstantialist account will work for the semantics of propositional attitudes.

It is revealing to see where this argument breaks down if we attempt to press it against the structuralist. If the circumstantialist is committed to ⟦**Hesperus**⟧ = ⟦**Phosphorus**⟧, then so too should the structuralist be; that was supposed to follow simply from direct reference. What's more, since **'Hesperus' refers to Hesperus and 'Phosphorus' refers to Phosphorus** has the same syntactic structure as **'Hesperus' refers to Hesperus and 'Phosphorus' refers to Hesperus**, and since corresponding terminal nodes have the same denotations, the structuralist too is committed to the result that

⟦'Hesperus' refers to Hesperus and 'Phosphorus' refers to Phosphorus⟧ = ⟦'Hesperus' refers to Hesperus

But here the argument breaks down if we try to press it against the structuralist. It is clear that **'Hesperus' refers to Hesperus and 'Phosphorus' refers to Phosphorus and there is some x such that 'Hesperus' and 'Phosphorus' both refer to x** differs in syntax from **'Hesperus' refers to Hesperus and 'Phosphorus' refers to Hesperus**, and so they cannot denote the same proposition, on a structuralist view.

So the structuralist can prevent this argument from going all the way through; but this should be little comfort. After all, it seems that the structuralist cannot account for the following fact: the ancients believed that 'Hesperus' refers to Hesperus and 'Phosphorus' refers to Phosphorus, but the ancients didn't believe that 'Hesperus' refers to Hesperus and 'Phosphorus' refers to Hesperus. This is precisely the same sort of problem Soames pressed against the circumstantialist; it turns out to hit the structuralist too. So Soames's argument, as he presents it, either is equally a problem for the circumstantialist and the structuralist, or equally not a problem; it thus provides no argument for structuralism over circumstantialism.

49

I think there's a bit more to be said here, though. Recall §3.3.2, where we considered Hesperus/Phosphorus cases. We saw there that the circumstantialist has resources to appeal to in these cases that the structuralist lacks. Since the circumstantialist is willing to acknowledge impossible circumstances, she can allow that $\llbracket \textbf{Hesperus} \rrbracket \neq \llbracket \textbf{Phosphorus} \rrbracket$; the structuralist cannot, at least not without denying the necessity of identity, or else appealing to impossible circumstances. And so long as $\llbracket \textbf{Hesperus} \rrbracket \neq \llbracket \textbf{Phosphorus} \rrbracket$, Soames's argument can't get off the ground; this is the crucial first premise.[32] To the extent, then, that the circumstantialist has an easier time justifying $\llbracket \textbf{Hesperus} \rrbracket \neq \llbracket \textbf{Phosphorus} \rrbracket$ than does the structuralist, Soames's argument, far from being an argument *against* circumstantialism, actually gives us another reason to believe in it.

3.5   Conclusion

In the foregoing, I've argued that circumstantialism gives us a better theory of propositions than does structuralism. Here, I want to draw out one consequence of this.

We've seen that the circumstantialist distinguishes the propositions denoted by distinct contradictions by allowing that there are some (impossible) circumstances where those contradictions are true. Similarly, the circumstantialist distinguishes the propositions denoted by distinct tautologies by allowing that there are circumstances where those tautologies fail. These commitments have consequences for the semantics of negation. After all, if there are circumstances where a sentence and its negation are both true, and there are circumstances where a sentence and its negation both fail, then we can no longer maintain the classical thesis that a sentence's negation is true in a circumstance if and only if the sentence is false in that circumstance. In fact, neither direction of this thesis can be maintained. Thus, if circumstantialism is right, as I've argued here, a classical semantics for negation cannot work. Some better semantics for negation must be found. That's for another day.

---

[32] For a response to Soames's argument that makes just this point, see (Edelberg, 1994).

Chapter 4

Modal negations and De Morgan laws

Considerations spelled out in chapter 3 require that we give some non-classical semantics for negation. The purpose of this chapter is to explore logical and linguistic research on a variety of negations and negation-like operators, to get a sense of what our range of options might be. Chapter 5 will go on to provide a single semantic framework that allows us to examine this range more precisely.

There are many different inferential principles a unary sentential connective might support. The arguments in chapter 3 show that, where $-$ is the connective that represents natural-language negation, the inferences ECQ and LEM must be invalid:[1]

ECQ: $A \wedge -A \vdash \bot$

LEM: $\top \vdash A \vee -A$

But there are many other inferences that have been taken to characterize negation or negative operators, such as:

Nor: $\top \vdash -\bot$

D-Nor: $-\top \vdash \bot$

DNI: $A \vdash --A$

---

[1]Here, $\bot$ and $\top$ are 0-place connectives; $\bot$ holds in no circumstance, and $\top$ holds in every circumstance. Thus, the principles $\bot \vdash A$ and $A \vdash \top$ are valid, for any sentence A. (These are 'big F' and 'big T' respectively, in the parlance of relevant logicians; see eg (Anderson & Belnap, 1975).

DNE: $--A \vdash A$

DM1: $-(A \vee B) \vdash -A \wedge -B$

DM2: $-A \vee -B \vdash -(A \wedge B)$

DM3: $-(A \wedge B) \vdash -A \vee -B$

DM4: $-A \wedge -B \vdash -(A \vee B)$

SM: If $A \vdash B$, then $-B \vdash -A$

There are plenty more besides (for an exhausting, if not exhaustive, list, see (Vakarelov, 1989)), but these are the principles that will concern us in this chapter and in chapter 5.

In fact, for dealing with the sentential negations that occur in natural languages, the first four of these principles (Nor, D-Nor, DNI, and DNE) seem to be relatively unimportant. Although Nor and D-Nor simplify bookkeeping a bit (as we'll see in chapter 5), it's not clear that $\perp$ and $\top$ represent any particular natural-language sentences,[2] so inferences governing them may well be beside the point of an empirical investigation. Too, it's not clear that there are any grammatical clauses (in any language) that contain two occurrences of sentential negation without intervening material (like embedded clauses and the like) (see eg (Acquaviva, 1997), (Ladusaw, 1996), (Zanuttini, 2001)), and so whether sentential negation is governed by DNI or DNE may well also be beside the point.

This leaves us the four De Morgan principles and contraposition. In order to explore the De Morgan principles, we'll need to make some assumptions about the behavior of $\wedge$ and $\vee$. I'll assume that they obey Distributive Lattice Logic (DLL), so named because its algebra of propositions is a distributive lattice. DLL can be specified as follows (see (Dunn & Zhou, 2005)):

- $A \vdash A \vee B$

- $B \vdash A \vee B$

- If $A \vdash C$ and $B \vdash C$, then $A \vee B \vdash C$

- $A \wedge B \vdash A$

- $A \wedge B \vdash B$

---

[2]They are sometimes rendered as 'everything is true' and 'something is true', respectively. But that doesn't seem quite right. First of all, note that the quantifiers 'everything' and 'something' there shouldn't range over $\perp$ and $\top$ themselves, but rather all other sentences. Then, note that we haven't said anything yet to rule out a circumstance (impossible, to be sure) where everything *is* true, but even there, $\perp$ would not hold; likewise for circumstances where nothing is true but $\top$ still holds. A better way to understand these, if some informal understanding is required, is as a 0-ary disjunction and 0-ary conjunction, respectively. See eg (Restall, 2000b).

- If $A \vdash B$ and $A \vdash C$, then $A \vdash B \wedge C$

- If $A \vdash B$ and $B \vdash C$, then $A \vdash C$

- $A \vee (B \wedge C) \vdash (A \vee B) \wedge (A \vee C)$

As should be apparent, nothing non-classical is afoot here; with the addition of a few inferences governing negation, we'd have full classical logic. The discussion that follows assumes DLL throughout.[3]

In DLL, given SM, two of the De Morgan inferences (DM1 and DM2) follow immediately:

$$\frac{\dfrac{A \vdash A \vee B}{-(A \vee B) \vdash -A} \qquad \dfrac{B \vdash A \vee B}{-(A \vee B) \vdash -B}}{-(A \vee B) \vdash -A \wedge -B}$$

$$\frac{\dfrac{A \wedge B \vdash A}{-A \vdash -(A \wedge B)} \qquad \dfrac{A \wedge B \vdash B}{-B \vdash -(A \wedge B)}}{-A \vee -B \vdash -(A \wedge B)}$$

These implications hold the other way around, too; given either of DM1 or DM2, SM follows. Here's how:[4]

$$\frac{\dfrac{B \vdash A \vee B \qquad \dfrac{\dfrac{A \vdash B \qquad B \vdash B}{A \vee B \vdash B}}{-B \vdash -(A \vee B)}}{-B \vdash -A \wedge -B} \qquad -(A \vee B) \vdash -A \wedge -B}{-B \vdash -A}$$
$$\qquad \qquad \qquad \qquad \qquad -A \wedge -B \vdash -A}{-B \vdash -A}$$

$$\frac{\dfrac{A \wedge B \vdash A \qquad \dfrac{\dfrac{A \vdash A \qquad A \vdash B}{A \vdash A \wedge B}}{-(A \wedge B) \vdash -A}}{-A \vee -B \vdash -A} \qquad -A \vee -B \vdash -(A \wedge B)}{-B \vdash -A}$$
$$\qquad \qquad \qquad \qquad \qquad -B \vdash -A \vee -B}{-B \vdash -A}$$

The other two De Morgan inferences, though, are independent. Operators that validate SM (and so DM1 and DM2), but not DM3 or DM4, are sometimes called *subminimal negations*, and that's what I'll call them here.[5]

If we consider the usual 'positive' modalities in a normal modal logic—$\Box$ and $\Diamond$—we see that each distributes over one binary connective: $\Box$ over $\wedge$ and $\Diamond$ over $\vee$.[6] That is:

---

[3]Some of the discussion will depend on distribution (the final inference in the list above), and some won't. I won't be calling attention to when distribution is vital and when it isn't, though. For considerations of negation in logics without distribution, see (Hartonas, 1996), and for arguments against including distribution in a theory of content, see (Brady & Meinander, 200x).

[4]Note that these proofs assume the following rule about $-$, which amounts to replacement of provable equivalents within $-$:

$$\frac{A \vdash B \qquad B \vdash A}{-A \vdash -B}$$

[5]The classic is (Hazen, 1992); see also (Dunn & Hardegree, 2001, Ch. 3).

[6]See eg (Chellas, 1980) for details.

- $\Box(A \land B) \dashv\vdash \Box A \land \Box B$

- $\Diamond(A \lor B) \dashv\vdash \Diamond A \lor \Diamond B$

Negative connectives don't distribute in this way, but some of them 'antidistribute'—turning $\land$ to $\lor$ or vice versa. If we take DM2 and DM3 together, we have one antidistributive equivalence: $-(A \land B) \dashv\vdash -A \lor -B$. DM1 and DM4 together give us the other: $-(A \lor B) \dashv\vdash -A \land -B$. (Note that a subminimal negation validates neither antidistributive equivalence.) Of course, if we take DM1–DM4 all together, we have both antidistributive equivalences.

4.1 Modal negations

Semantics for positive modalities are studied at length in (Jónsson & Tarski, 1951) and (Jónsson & Tarski, 1952). Jónsson & Tarski show that binary relational semantics (of the sort nowadays familiar for modal logics—see eg (Chellas, 1980) for details) can be provided for 'additive' modal operators—essentially those that distribute over disjunction. Although they don't connect their work to modal logic, it's clear that this result is applicable to the $\Diamond$ operator found in normal modal logics.

I'll be describing work on negation that takes the usual modal-logic approach as its departure point, but I'll use a slightly idiosyncratic presentation, so let me pause for a moment to be explicit about the semantics for modal logics I'll be adapting here.

Definition 4.1.1. *A* modal frame *is a tuple $\langle W, R \rangle$, where*

- *W is a set, and*

- *$R : W \to \wp(W)$ is a function from W to subsets of W*

Given a language $\mathcal{L}$ that builds up atomic sentences into complex sentences by using the binary connectives $\land$ and $\lor$, and the unary connectives $\neg$, $\Box$, and $\Diamond$, we can then define a *modal model* as follows:

Definition 4.1.2. *A* modal model *is a tuple $\langle W, R, [\![\,]\!] \rangle$, where*

- *$\langle W, R \rangle$ is a modal frame, and*

- *$[\![\,]\!] : \mathcal{L} \to \wp(W)$ is a function from sentences of the language $\mathcal{L}$ to subsets of W such that:*

    ∘ *$[\![A \land B]\!] = [\![A]\!] \cap [\![B]\!]$*

- $\llbracket A \vee B \rrbracket = \llbracket A \rrbracket \cup \llbracket B \rrbracket$

- $\llbracket \neg A \rrbracket = W - \llbracket A \rrbracket$

- $\llbracket \Box A \rrbracket = \{w \in W : R(w) \subseteq \llbracket A \rrbracket\}$

- $\llbracket \Diamond A \rrbracket = \{w \in W : R(w) \cap \llbracket A \rrbracket \neq \emptyset\}$

Definition 4.1.3. *An argument with premise A and conclusion B is* valid in a model M *iff for no* $w \in W$ *is it the case that* $w \in \llbracket A \rrbracket$ *but* $w \notin \llbracket B \rrbracket$. *It's* valid on a frame F *iff it's valid on every model on that frame. And it's* valid on a class of frames C *iff it's valid on every frame in the class.*

It can be shown that the normal modal logic K is sound and complete for the class of modal models. What's more, by imposing various restrictions on the function R, we can find classes of frames that all the usual normal modal logics are sound and complete for.[7]

In these logics, $\neg$ inherits its logical properties directly from the underlying operation on sets: complementation. Assuming a classical metalanguage, this guarantees that $\neg$ will validate ECQ and LEM; and thus $\neg$, with these semantics, is unsuitable as a representation of natural-language negation.[8]

Fortunately, there are semantics in the literature for nonclassical negations. One common route to these negations is to adapt the semantics given above for $\Box$ or $\Diamond$. This section explores in more detail the results of such an adaptation.

4.1.1 Compatibility negations

Dunn (1993) gives a binary-accessibility semantics for a range of negations as follows: add a function $C : W \to \wp(W)$ to our frames, and use the following recursive truth-condition:[9]

- $x \in \llbracket - A \rrbracket$ iff $\forall w \in C(x)(w \notin \llbracket A \rrbracket)$

Negations yielded by this semantics will all satisfy SM, DM4, and Nor. To see this in the case of DM4, consider a point x in an arbitrary model built on an arbitrary frame such that $x \in \llbracket - A \wedge - B \rrbracket$. It follows immediately (by the truth-condition for $\wedge$) that $x \in \llbracket - A \rrbracket$ and $x \in \llbracket - B \rrbracket$. By the truth-condition for $-$, then, $\forall w \in C(x)(w \notin \llbracket A \rrbracket$ and $w \notin \llbracket B \rrbracket)$, and this guarantees that $\forall w \in C(x)(w \notin$

[7]For proofs, see eg (Chellas, 1980). Note as well that the usual presentation of these semantics uses a binary relation on W rather than the function R I use here. Nothing besides presentational simplicity hangs on this.

[8]Priest (2006a, Ch. 5) pushes against the idea of a classical metalanguage. I agree that it's not always necessary to use a classical metalanguage, but I think there are often good reasons to; here (and throughout) I will continue to use a classical metalanguage.

[9]Restall (1999) defends a philosophical interpretation of C, meant to demonstrate that this is an applied, not simply a pure, semantics, answering a challenge put by Copeland (1979). That's not my concern here, though.

$\llbracket A \rrbracket \cup \llbracket B \rrbracket$). By the truth-condition for $\vee$, we know $\forall w \in C(x)(w \notin \llbracket A \vee B \rrbracket$, and it follows that $x \in \llbracket -(A \vee B) \rrbracket$. Since x was arbitrary in the model, the argument from $-A \wedge -B$ to $-(A \vee B)$ must be valid in the model. But the model was arbitrary too, and so was the frame the model was based on. So any class at all of frames validates this argument, given the truth-condition for $-$ in play. Similar arguments work for Nor and SM.[10]

The class of negations that can be given semantics in this way (call them *compatibility negations*, after the usual interpretation of the function C) is quite broad; it includes classical negation, but also intuitionist negation, Johansson's minimal negation, the De Morgan negations of relevant logics, and the negations of Strong Kleene logic and Priest's logic LP. The weakest negation (in DLL) that has a compatibility-style semantics is characterized by DLL, SM, DM4, and Nor; it is not well-studied in the literature, but Dunn & Zhou (2005) call it *preminimal* negation; it is analogous to the $\Box$ operator of the normal modal logic K.

4.1.2 Exhaustion negations

Exhaustion negations bear the same relation to compatibility negations that $\Diamond$ does to $\Box$ in normal modal logics.[11] For an exhaustion negation $-$, we add a function $E : W \rightarrow \wp(W)$ to our frames,[12] and then give truth-conditions like so:

- $x \in \llbracket -A \rrbracket$ iff $\exists w \in E(x)(w \notin \llbracket A \rrbracket)$

Negations yielded by this semantics all satisfy SM, DM3, and D-Nor. To see this in the case of DM3, consider a point x in an arbitrary model built on an arbitrary frame such that $x \in \llbracket -(A \wedge B) \rrbracket$. It follows that $\exists w \in E(x)(w \notin \llbracket A \wedge B \rrbracket)$, and so $\exists w \in E(x)(w \notin \llbracket A \rrbracket$ or $w \notin \llbracket B \rrbracket)$. Thus, either $\exists w \in E(x)(w \notin \llbracket A \rrbracket)$ or $\exists w \in E(x)(w \notin \llbracket B \rrbracket)$, and so $x \in \llbracket -A \rrbracket$ or $x \in \llbracket -B \rrbracket$; that is, $x \in \llbracket -A \rrbracket \cup \llbracket -B \rrbracket$. So $x \in \llbracket -A \vee -B \rrbracket$. Since x and its model and its frame were all arbitrary, the point is general; it holds for any class of frames, given the exhaustion semantics for $-$.

The class of negations that can be given semantics in this way (exhaustion negations) is also broad; it too includes classical, De Morgan, Strong Kleene, and LP negations. Intuitionist and minimal negations are not exhaustion negations, but dual-intuitionist and dual-minimal negations are (see (Dunn & Zhou, 2005)). The weakest negation (in DLL) that has an exhaustion semantics is

---

[10]For details, see (Dunn, 1993).

[11]And compatibility negations bear the same relation to $\Box$ that exhaustion negations do to $\Diamond$.

[12]Note that the R function from the normal modal semantics, the C function from the compatibility semantics, and now this E function are all subject to precisely the same requirements: that they be functions from W to $\wp(W)$. The real difference between these operators, then, comes at the level of truth-conditions, not at the level of frames.

characterized by DLL, SM, DM3, and D-Nor; it is also not well-studied in the literature, but Dunn & Zhou (2005) call it *dual preminimal negation*; it is analogous to the $\Diamond$ operator of the normal modal logic K.

4.1.3 Ockham negations

The algebraic theory of *Ockham lattices* is developed in eg (Urquhart, 1979). Following the usual bridge between algebra and logic, we can call a negation *Ockham* when it validates SM, DM1–4, Nor, and D-Nor. (Note, as before, that SM guarantees DM1 and DM2.) Every Ockham negation is both a compatibility negation and an exhaustion negation, and every negation that is both compatibility and exhaustion is Ockham. Classical negation, De Morgan negation, Strong Kleene negation, and LP negation are all Ockham, but none of these is the weakest Ockham negation, since they all validate both DNI and DNE, while the weakest Ockham negation validates neither of these.

The Routley star semantics laid out in eg (Routley & Routley, 1972) and (Routley et al., 1982) was designed for the De Morgan negation of relevant logics. As it happens, though, DNI and DNE have to be put in by hand in the star semantics; the weakest Routley star negation is the weakest Ockham negation, validating no double-negation inferences. For details, see chapter 5.

4.1.4 Subminimal negations

Little work exists on the semantics of subminimal negations (negations that validate SM, but neither DM3 nor DM4). I suspect this is largely because they cannot be given a relational semantics. As we've seen above, compatibility and exhaustiveness semantics each validates one of DM3 and DM4. To avoid both of these inferences, some other path must be taken. This section reviews the approaches to the semantics of subminimal negations already in the literature; chapter 5 will develop a semantics different from all of these.

Hazen 1992

The classic study of subminimal negation is (Hazen, 1992). Hazen adds a subminimal negation to the positive fragment of intuitionist logic. Ordinary intuitionist negation can be defined in intuitionist logic by using the intuitionist conditional and $\bot$; simply let $-A$ be $A \supset \bot$, where $\supset$ is the intuitionist conditional. Following a remark in (Curry, 1977) and a suggestion of Humberstone's, Hazen imagines a generalization of this approach, based on a set F of propositions. This allows him to define a

negation as follows: $-A$ holds whenever $A \supset f$ holds *for any $f \in F$*.[13] Hazen goes on to show, by two proofs (one model-theoretic, one proof-theoretic), that the connective so defined is a subminimal negation—in fact, that it is the weakest subminimal negation.

Here, I briefly review the model-theoretic construction Hazen uses; his overall project may be orthogonal to the present study, but the model theory of subminimal negations is certainly not. He starts from the familiar Kripke semantics for positive intuitionist logic, which I briefly review:

Definition 4.1.4. *An* I-frame *is a tuple $\langle W, \leq \rangle$ such that:*

- *W is a set, and*

- *$\leq$ is a partial order on W*

Definition 4.1.5. *An* I-model *is a tuple $\langle W, \leq, [\![\ ]\!] \rangle$ such that:*

- *$\langle W, \leq \rangle$ is an I-frame, and*

- *$[\![\ ]\!]$ is such that:*

  - *$[\![A \wedge B]\!] = [\![A]\!] \cap [\![B]\!]$*

  - *$[\![A \vee B]\!] = [\![A]\!] \cup [\![B]\!]$*

  - *$[\![A \supset B]\!] = \{x \in W : \text{for all } y \text{ such that } x \leq y, y \notin [\![A]\!] \text{ or } y \in [\![B]\!]\}$*

From here, Hazen's semantics for subminimal negation is straightforward:

Definition 4.1.6. *A* Hazen frame *is a tuple $\langle W, \leq, F \rangle$ such that:*

- *$\langle W, \leq \rangle$ is an I-frame, and*

- *$F \subseteq Prop(W)$, where $Prop(W)$ is the set of subsets of W closed upwards along $\leq$*

The definition of a *Hazen model* is then as for an I-model, with an additional clause for negation:

- *$[\![-A]\!] = \{x \in W : \text{there is an } f \in F \text{ such that for all } y \text{ such that } x \leq y, y \notin [\![A]\!] \text{ or } y \in f\}$*

(Hazen, 1992) provides soundness and completeness proofs for positive intuitionist logic + SM with respect to this semantics. Note that, in Hazen's treatment, the features of subminimal negation are a direct fallout of the features of the intuitionist $\supset$. In chapter 5, we'll see how to provide semantics for subminimal negation directly, as it were, without the detour through $\supset$. (As we'll see, something like F is crucially involved.)

---

[13]That's a sloppy way of putting it; propositions aren't sentences. The idea will be made precise below.

Dunn & Hardegree 2001

In (Dunn & Hardegree, 2001), a unary operation $-$ on posets is defined as a *subminimal complementation* iff it meets the following criterion: if $a \leq b$ then $-b \leq -a$, where $\leq$ is the partial order of the poset. Thus, all the negations considered above are subminimal in Dunn & Hardegree's sense; I continue to use 'subminimal' to distinguish those negations that fail to validate either DM3 or DM4.[14] Dunn & Hardegree do not study the properties of weak subminimal complementations; they proceed immediately to study those that also include (the alegbraic analogue of) DM4.

Méndez et al. 2002

In (Méndez et al., 2002), a different sense of 'subminimal negation' is used; Méndez et al. mean by 'subminimal negation' simply any negation weaker than Johansson's minimal negation. Thus, preminimal negation (discussed above in §4.1.1) is a subminimal negation in their sense, though not in mine. Although I don't see anything actually false in this paper, one issue is certainly put in a way that strongly suggests a falsehood; it'll be worth a moment to clear this up.

Méndez et al. work within the logic B+, the positive fragment of the basic affixing relevant logic (for details see (Routley et al., 1982)). They define a negation connective $\sim$ as follows: $\sim A$ abbreviates $A \rightarrow F$ for a particular 0-place connective F; imposing various constraints on F then has the effect of producing various constraints on $\sim$.

Note that B+ contains as a theorem $((A \rightarrow C) \wedge (B \rightarrow C)) \rightarrow ((A \vee B) \rightarrow C)$. So if $\sim A$ is defined as $A \rightarrow F$, then B+ must also contain the theorem $(\sim A \wedge \sim B) \rightarrow \sim(A \vee B)$, the theorem version of DM4. This follows simply from the definition of $\sim$ and theorems of B+; it makes no appeal to any constraints on F of any sort. So all four of the negations Méndez et al. consider have this as a theorem. Despite this, they list it as an 'exemplar theorem' for one of their stronger systems (in particular, for a system that includes the theorem version of DNI), and do not list it as an exemplar theorem for their base system, although they do there list the theorem versions of DM1 and DM2, as well as SM. This could well create the false impression that their base system does not have DM4 as a theorem, but as we've seen, it does. Thus, none of the four kinds of negation considered in (Méndez et al., 2002) is a subminimal negation in the presently relevant sense.

---

[14]Forgive the slipping between algebra and logic here; speaking carefully, none of the negations here is subminimal in Dunn & Hardegree's sense. Rather, each of them induces an operation ϕ on their logic's Lindenbaum algebra, and ϕ is subminimal. I'll keep indulging in this kind of slipperiness without further comment.

4.2 Negative contexts

The study of varieties of negation is not limited to logic narrowly understood. Linguists too have use for several different types of negative operator, of varying strengths. This section will summarize some of the research on natural languages that invokes such operators.

4.2.1 NPIs and PPIs

Probably the most well-known of the natural-language phenomena sensitive to negation and negative-like contexts is the phenomenon of *polarity items*. To a first approximation, a *negative polarity item* (NPI) is a word or phrase that can only occur in negative contexts, and a *positive polarity item* (PPI) is a word or phrase that can only occur outside of negative contexts. Some examples of NPIs: **any**, **ever**, **lift a finger**; and some examples of PPIs: **adamant**, **rain cats and dogs**. Example sentences:

- ○ I didn't eat any potatoes

  ○ *I ate any potatoes

- ○ I never ever brush my teeth before I drink orange juice

  ○ *I always ever brush my teeth before I drink orange juice

- ○ If she lifts a finger to help us, I'll be shocked

  ○ *If I'm shocked, she'll lift a finger to help us

- ○ They were adamant about their decision

  ○ *They weren't adamant about their decision[15]

- ○ It was raining cats and dogs last night

  ○ *It wasn't raining cats and dogs last night

Note that polarity items are affected by more than just negation—negative quantifiers like **nobody** or **few oranges**, antecedents (but not consequents!) of conditionals, certain embedding verbs like **doubt** and **surprise**, and yes/no questions all allow at least some NPIs and prevent occurrence of at least some PPIs. These contexts seem to have something negative about them (hence the NPI/PPI

---

[15]Like many cases of PPIs in negative contexts, this can be rescued. If the negation in this sentence is heard as *metalinguistic*, in the sense of (Horn, 2001)—that is, as rejecting the appropriateness of the word **adamant** (this use, when spoken, is accompanied by a noticeable intonation contour on **adamant**)—then the PPI is fine. Note as well that metalinguistic negation does not license NPIs. For more, see eg (Horn, 2001), (Geurts, 1998).

terminology); it would be nice to have a more precise account of what such negativity amounts to, though.

4.2.2   Downward-entailing contexts

We can begin to develop such an account by noticing that most of these contexts are *downward entailing.* This forms the basis for theories of polarity item licensing in eg (Ladusaw, 1980), (van der Wouden, 1997), (Zwarts, 1998). A downward entailing context is a context in which one can infer from superset to subset. For example, **few oranges** creates a downward entailing context; the inference from **Few oranges got on the bus today** to **Few oranges got on the bus this morning** is valid. In contrast, **A few oranges** does not create a downward entailing context; the inference from **A few oranges got on the bus today** to **A few oranges got on the bus this morning** is not valid. Negative quantifiers, antecedents (but not consequents) of conditionals, embedding verbs like **doubt** and **surprise**, adverbs like **only**, &c. all create downward-entailing contexts.[16]

It might be tempting to conclude from this that NPIs are allowed, and PPIs disallowed, in all and only downward-entailing contexts. Unfortunately, this would be too quick. As the following examples show, different polarity items have different distributions:

- ○ Not everybody ate any soup
  - ○ *Not everybody ate their soup yet

- ○ Few of us gave anything to charity
  - ○ *Few of us gave a red cent to charity

- ○ I wonder whether you've ever been skiing
  - ○ *I wonder whether you're half bad at skiing

NPIs like 'a bit', 'a red cent', and 'half bad' sit much more easily with stronger negations, as in:

- Nobody ate their soup yet

- If you gave a red cent to charity, I'll eat my hat

- I doubt you're half bad at skiing

---

[16]There's some dispute about whether some of these cases are genuinely downward-entailing. In particular, **only** and **surprise** are probably more contentious than the others. For an argument against these being downward entailing, see (Atlas, 1997). For an argument that they are indeed downward-entailing, see (von Fintel, 1999).

Zwarts (1998) and van der Wouden (1997) divide polarity items into three groups, according to the sort of negative context that licenses (NPIs) or prevents (PPIs) their occurrence. *Weak* NPIs, such as **any** or **ever**, are allowed in any downward-entailing context, but *medium* and *strong* NPIs require stronger negations. Similarly, *strong* PPIs are forbidden in any downward-entailing context, but *medium* and *weak* PPIs are only forbidden by stronger negations.[17] The next step, then, is to get clearer on what makes a negative context strong or weak.

4.2.3 Anti-additive contexts

Consider the examples at the end of §4.2.2. What do **nobody**, conditional antecedents, and **doubt** have in common that they don't share with **not everybody**, **few of us**, and **wonder**? All of these contexts are downward-entailing. But compare:

(90)   a.   Nobody ate and nobody slept $\rightarrow$ Nobody ate or slept

       b.   Not everybody ate and not everybody slept $\not\rightarrow$ Not everybody ate or slept

(91)   a.   If you ate you should wash your bowl, and if you slept you should wash your bowl $\rightarrow$ If you ate or slept you should wash your bowl

       b.   Few of us ate and few of us slept $\not\rightarrow$ Few of us ate or slept

(92)   a.   I doubt you ate and I doubt you slept $\rightarrow$ I doubt you ate or slept

       b.   I wonder whether you ate and I wonder whether you slept $\not\rightarrow$ I wonder whether you ate or slept

Here's the difference, then: **nobody**, conditional antecedents, and **doubt**—but not **not everybody**, **few of us**, or **wonder**—validate DM4, appropriately syntactically modified.[18] Since any downward-entailing operator validates DM1, these all validate one full De Morgan equivalence: $-A \wedge -B \dashv\vdash -(A \vee B)$. I'll call operators validating this equivalence *anti-additive*.[19] Zwarts (1998) and van der Wouden (1997) argue that there is a class of NPIs—the medium ones—that require not just a downward-entailing context, but an anti-additive one. From the above examples, we can conclude

---

[17]Terminology here is not standardized; I'm following van der Wouden (1997) in my use of 'weak', 'medium', and 'strong' here.

[18]Note that some of these—the propositional attitudes and conditionals—can be plausibly thought of as propositional operators, while for others—the quantifiers—it would take a bit of a stretch. Not unstretchable, though; consider cylindric algebra (see (Henkin et al., 1985)), which treats quantifiers as propositional operators, or see (Kuhn, 1980).

[19]Again, terminology is not standardized. Both Zwarts (1998) and van der Wouden (1997) use this terminology in some places, but Zwarts (1998) also calls these 'minimal' negations, contrasting them with the 'subminimal' downward-entailing ones, while van der Wouden (1997) also calls these 'regular' negations, contrasting them (confusingly, considering he claims to follow (Zwarts, 1998)) with 'minimal' downward-entailing ones. I'll stick to 'anti-additive' and 'downward-entailing'.

that **yet**, **a red cent**, and **half bad** are medium NPIs in English. In addition, van der Wouden (1997) argues for a class of PPIs—again, the medium ones—that are disallowed in any anti-additive context, but allowed in contexts that are merely downward-entailing.

4.2.4 Anti-multiplicative contexts

Anti-multiplicative contexts are those that validate DM2 and DM3, and so the other De Morgan equivalence, $-(A \wedge B) \dashv\vdash -A \vee -B$. Such contexts are formed by eg **not everybody**, **not always**, **not all the cats**, &c. To see this, compare:

(93)   a. Not everybody dressed up and played tag $\rightarrow$ Not everybody dressed up or not everybody played tag

        b. Nobody dressed up and played tag $\nrightarrow$ Nobody dressed up or nobody played tag

(94)   a. Jim does not always dress up and play tag $\rightarrow$ Jim does not always dress up or Jim does not always play tag

        b. If Jim dresses up and plays tag, I'll be happy $\nrightarrow$ If Jim dresses up I'll be happy, or if Jim plays tag I'll be happy

(95)   a. Not all the cats dressed up and played tag $\rightarrow$ Not all the cats dressed up or not all the cats played tag

        b. I doubt you dressed up and played tag $\nrightarrow$ I doubt you dressed up or I doubt you played tag

These sentences show that anti-additive contexts don't have to be anti-multiplicative; since anti-additive contexts are all downward-entailing, they also show that downward-entailing contexts don't have to be anti-multiplicative.

Anti-multiplicative contexts appear to play little role in polarity-item licensing, except for the role they play in virtue of being downward-entailing. van der Wouden (1997, pp. 134–141) notes some possible reasons for this (as well as some possible exceptions). See also (Horn, 2001), who notes that anti-multiplicative operators don't get lexicalized, and offers some explanation of this.

4.2.5 Antimorphic contexts

An antimorphic context is a context that is both anti-additive and anti-multiplicative; thus, it validates both De Morgan's equivalences (all of DM1–DM4). In English, the best candidate for an antimorphic operator is good old 'not'. Zwarts (1998) and van der Wouden (1997) argue for the

existence of a class of NPIs—the strong ones—that can only occur in antimorphic contexts, and a class of PPIs—the weak ones—that cannot occur in antimorphic contexts but can occur in anti-additive contexts. The best example of a strong NPI in English (that I can find) is **a bit** (examples from (van der Wouden, 1997, p. 141)):

(96)   a. Chomsky wasn't a bit happy about these facts

     b. ?No one was a bit happy about these facts[20]

     c. *At most three linguists were a bit happy about these facts

Both Zwarts (1998) and van der Wouden (1997) report a relatively wide range of strong NPIs in Dutch.

4.2.6   Beyond polarity

This classification of negative contexts—following (Atlas, 1997), we can call it the *De Morgan taxonomy*—is not only of use in dealing with polarity phenomena. In particular, the identification of the class of downward-entailing contexts is important in the study of a number of phenomena. Indeed, the core thesis of (van der Wouden, 1997) is that key phenomena that have been associated with negation—litotes, negative concord, denial, &c.—are in fact better thought of as phenomena surrounding downward-entailing operators in general.

4.3   Drawing connections

In this section, I point out that each category within the De Morgan taxonomy of §4.2 corresponds to a category of the modal-negation taxonomy presented in §4.1.

4.3.1   Anti-additivity is compatibility

Anti-additive expressions validate DM1, DM2, and DM4. If we add a constant falsum $\bot$ to our natural language, we would find that anti-additives validate Nor as well: **Never** $\bot$ is a logical truth. But these are just the requirements for an operator to have a compatibility semantics.[21] Thus, we can provide compatibility semantics for all our anti-additives; they validate the key inferences.

---

[20]van der Wouden (1997) reports a * judgment for this sentence, but to my ear it's at worst questionable. Still, it's decidedly worse than (96a).

[21]It might at first seem that, if we add a constant verum $\top$, that anti-additives would also satisfy D-Nor; that is, that **Never** $\top$ would be logically untrue. But this is to forget the possibility of an empty domain; it might be that never $\top$, if there aren't any times.

This point is so far purely formal; here are a few rough comments about interpretation. In one common interpretation of the relational semantics for modal logic, the points of evaluation are *possible worlds* and the relation is one of *relative possibility*. Usual applications of compatibility semantics for negation don't interpret points of evaluation as possible worlds, though, since possible worlds are negation-consistent and negation-complete, and one of the key uses of compatibility semantics is in handling structures where these conditions fail. One interpretation of the points of evaluation, more usual in substructural logics than in modal logics, is as information states.[22] Unlike possible worlds, information states are often neither negation-complete nor negation-consistent. Also, unlike possible worlds, information states can enter into meaningful inclusion relations—it's typical, in frame semantics for substructural logics, to treat information states as partially ordered, such that for any information states x, y and any sentence A, if $x \in [\![A]\!]$ and $x \leq y$, then $y \in [\![A]\!]$.[23]

With the points of evaluation interpreted as information states, the relation C can be interpreted as *relative compatibility*.[24] This gives us a semantics that makes good sense for sentential negation (see (Restall, 1999)), but not so much for anti-additives like **never**. Treating the points of evaluation as information-states-at-times, though, and treating C as the relation that holds between the times in a history, though, yields plausible results, though: **never** p holds at an information-state-at-a-time x when, for all information-states-at-times y in the same history as x, p doesn't hold at y. Similar interpretations will work for other anti-additives.

4.3.2 Anti-multiplicativity is exhaustion

Anti-multiplicative expressions validate DM1, DM2, and DM3. If we add a constant verum $\top$ to our natural language, we would find that anti-multiplicatives validate D-Nor as well: **Not always** $\top$ is a logical untruth. But these are just the requirements for an operator to have a exhaustion semantics.[25] Thus, we can provide exhaustion semantics for all our anti-multiplicatives; they validate the key inferences.

---

[22]See eg (Restall, 2000b), (Dunn, 1993).

[23]The usual way to ensure this condition is to require it to hold for atomic sentences and then prove that it holds for complex sentences as well. This ordering is typically most important in handling semantics for things other than negation (but see (Restall, 2000a)). Chapter 5 provides a semantics that includes the information ordering, for maximum compatibility with other substructural semantics.

[24]*Relative consistency* would make more sense, since it is often up for debate whether A is compatible with $\neg A$, but it's never up for debate whether A is *consistent* with $\neg A$—it isn't. But I follow established usage.

[25]A similar issue as in fn.21 arises for anti-multiplicatives and Nor. The issue is resolved in the same way.

4.3.3 Antimorphicness is Ockhamness

Since antimorphic expressions are both anti-additive and anti-multiplicative, they validate all of DM1–DM4, Nor, and D-Nor; they thus meet the requirements for Ockham negation, and can be given Routley star semantics.

4.3.4 Downward-entailingness is subminimality

The weakest members of the De Morgan taxonomy, though, the merely downward-entailing expressions, cannot be given a compatibility semantics (since they don't validate DM4) or an exhaustion semantics (since they don't validate DM3). And this rules out their being given a Routley star semantics as well. As we've seen in §4.1.4, the only semantics currently available for subminimal negations in a language with $\wedge$ and $\vee$ is that of (Hazen, 1992), which involves a detour through intuitionist $\supset$. It would be nice to be able to give a semantics for subminimal negation directly.

However, similar problems have been faced—and solved—before. Compare the case of positive modalities; consider a unary modality $\triangle$ such that if $A \vdash B$ then $\triangle A \vdash \triangle B$. Note that this is the positive analogue of SM (call it SM-A)—we might call $\triangle$ a *subminimal affirmation*. There are several characteristic principles $\triangle$ might or might not satisfy:

P-Nor: $\triangle \bot \vdash \bot$

D-P-Nor: $\top \vdash \triangle \top$

D1: $\triangle(A \vee B) \vdash \triangle A \vee \triangle B$

D2: $\triangle A \vee \triangle B \vdash \triangle(A \vee B)$

D3: $\triangle(A \wedge B) \vdash \triangle A \wedge \triangle B$

D4: $\triangle A \wedge \triangle B \vdash \triangle(A \wedge B)$

In virtue of SM-A, $\triangle$ must validate D2 and D3:

$$\frac{\dfrac{A \vdash A \vee B}{\triangle A \vdash \triangle(A \vee B)} \qquad \dfrac{B \vdash A \vee B}{\triangle B \vdash \triangle(A \vee B)}}{\triangle A \vee \triangle B \vdash \triangle(A \vee B)}$$

$$\frac{\dfrac{A \wedge B \vdash A}{\triangle(A \wedge B) \vdash \triangle A} \qquad \dfrac{A \wedge B \vdash B}{\triangle(A \wedge B) \vdash \triangle B}}{\triangle(A \wedge B) \vdash \triangle A \wedge \triangle B}$$

In normal modal logics, $\lozenge$ additionally validates P-Nor and D1, while $\square$ additionally validates D-P-Nor and D4—and operators with these properties can be explored using the usual relational semantics.[26] But what about operators that stop at SM-A?

One traditional way to give semantics for such operators is through the device of *neighborhood semantics* explored in eg (Chellas, 1980) (as *minimal semantics*) or (Hansen, 2003). Similarly, neighborhood semantics can be used to explore the behavior of subminimal negations, although to the best of my knowledge this has not been done before. In particular, all the studies of neighborhood semantics I'm familiar with use neighborhoods to add connectives to a Boolean base; this base, then, includes the classical theory of negation. In addition, all the studies I'm familiar with either impose SM-A, or no monotonicity constraints at all; none impose SM. In chapter 5, however, I bring neighborhood semantics to bear on subminimal (and other) negations, providing a single systematic framework into which the full De Morgan taxonomy fits. To do this, it's important to adapt neighborhood semantics to a non-Boolean framework. As in this chapter, I stick with DLL. Thus, the semantics arrived at involves a mix of neighborhood and substructural features, and its exploration requires a mix of neighborhood and substructural techniques.

---

[26]Operators that validate D1 and D2, and so distribute over disjunction (in the sense that $\triangle(A \vee B) \dashv\vdash \triangle A \vee \triangle B$), are sometimes called *additive*; Jónsson & Tarski's (1951) key result was to show that every Boolean algebra with additive operators could be given a relational semantics.

Chapter 5

Neighborhood negations

As we've seen in the last chapter, there is good linguistic reason to categorize negations (and negative operators in general) by which De Morgan laws they support. The weakest negative operators (merely downward monotonic) support only two De Morgan laws;[1] medium-strength negative operators support a third;[2] and strong negative operators support all four.

As we've also seen, techniques familiar from modal logic are of great use in giving unifying theories of negative operators. In particular, Dunn's (1990) distributoid theory allows us to generate relational semantics for many negations. However, the requirements of distributoid theory are a bit too strict for use in modeling the weakest negations. For a relational semantics to work, an operator must either distribute or antidistribute over either conjunction or disjunction; but the merely downward monotonic operators do not. Thus, a unifying semantics cannot be had in distributoid theory.

In the (more familiar) study of positive modalities, there is a parallel result. Normal necessities distribute over conjunction, and normal possibilities over disjunction. When these distributions break down, a relational semantics is no longer appropriate. Here, there is a somewhat familiar solution: neighborhood semantics. In this chapter, I'll adapt neighborhood semantics to the less familiar case of negative modalities, showing how it can be used to give a single semantic framework appropriate to all the pertinent sorts of negative operators.

---

[1] $-(A \vee B) \vdash -A \wedge -B$ and $-A \vee -B \vdash -(A \wedge B)$

[2] Anti-additives, like *never*, support $-A \wedge -B \vdash -(A \vee B)$; anti-multiplicatives, like *not always*, support $-(A \wedge B) \vdash -A \vee -B$.

5.1 Neighborhood Semantics

5.1.1 Modal Neighborhoods

Neighborhood semantics for modal logic is a generalization of the relational semantics we looked at briefly in chapter 4. (As in that chapter, I'll stick to unary modalities, but neighborhood semantics generalize to n-ary modalities just as relational semantics do.) The structures on which a modal language $\mathcal{L}$ is interpreted are *modal neighborhood frames*:

Definition 5.1.1. *A* modal neighborhood frame *is a tuple* $\langle W, N \rangle$, *where:*

- W *is a set, and*

- $N : W \rightarrow \wp(\wp(W))$ *is a function from members of W to sets of sets of members of W*

Definition 5.1.2. *A* modal neighborhood model *is a tuple* $\langle W, N, [\![\ ]\!] \rangle$, *where:*

- $\langle W, N \rangle$ *is a modal neighborhood frame, and*

- $[\![\ ]\!] : \mathcal{L} \rightarrow \wp(W)$ *is such that:*

    ○ $[\![A \wedge B]\!] = [\![A]\!] \cap [\![B]\!]$

    ○ $[\![A \vee B]\!] = [\![A]\!] \cup [\![B]\!]$

    ○ $[\![\neg A]\!] = W - [\![A]\!]$

    ○ $[\![\Box A]\!] = \{ x \in W : [\![A]\!] \in N(x) \}$

As is shown in (Chellas, 1980), every modal frame F has an equivalent modal neighborhood frame F', in the following sense: for any modal model $\langle F, [\![\ ]\!] \rangle$, $\langle F', [\![\ ]\!] \rangle$ is a modal neighborhood model. Since these models use the same $[\![\ ]\!]$ function, they have all the same sentences holding at all the same points. But the converse does not hold—there are modal neighborhood frames that have no modal-frame equivalents. What's more, certain inferences valid on the class of all modal frames are not valid on the class of neighborhood frames; among them is the distribution of $\Box$ over $\wedge$: $\Box(A \wedge B) \dashv\vdash \Box A \wedge \Box B$.

Since we want to explore logics in which certain De Morgan laws fail, and since De Morgan laws are the negative relatives of this sort of distribution, it behooves us to explore an adaptation of neighborhood semantics to the case of negative connectives.

5.1.2 Substructural Neighborhoods

Here, I'll be exploring the addition of neighborhood structure to frames and models suitable for substructural logics. First, let's take a look at these frames and models without the additional neighborhood structure.[3] I'll assume that $\wedge$ and $\vee$ distribute over each other; that is, that the distributive laws

(97) a. $A \vee (B \wedge C) \dashv\vdash (A \vee B) \wedge (A \vee C)$

     b. $A \wedge (B \vee C) \dashv\vdash (A \wedge B) \vee (A \wedge C)$

are satisfied.[4] This allows us to use frames suitable for *distributive lattice logic*, or DLL. DLL can be presented as follows:

(98) Axioms

     a. $A \vdash A$

     b. $A \wedge B \vdash A,\ A \wedge B \vdash B$

     c. $A \vdash A \vee B,\ B \vdash A \vee B$

     d. $A \wedge (B \vee C) \vdash (A \wedge B) \vee (A \wedge C)$

     e. $A \vdash \top,\ \bot \vdash A$

(99) Rules

     a. $A \vdash B, A \vdash C / A \vdash B \wedge C$

     b. $A \vdash C, B \vdash C / A \vee B \vdash C$

     c. $A \vdash B, B \vdash C / A \vdash C$

Note that there are no negation rules in DLL. This makes DLL a suitable base from which to explore the logical behavior of various negations.

One way to provide frames for DLL is as follows:

Definition 5.1.3. *A* DLL-frame *is a tuple* $\langle W, \leq \rangle$ *such that:*

- W *is a non-empty set, and*

- $\leq$ *is a partial order on* W.

---

[3]For details on these frames and models, see (Restall, 2000b).

[4]But see (Brady & Meinander, 200x), which questions distribution for modeling content.

You can think of the set W as the frame's information states, or, if you prefer, ways things could (or couldn't) be, or whatever your favorite way of thinking of world-like indices is. Here, I'll talk in terms either of information states, or if I'm feeling extra-vacuous, simply *points*. On an understanding in terms of information, I take $\leq$ to be the relation of informational inclusion; for $x, y \in W$, $x \leq y$ iff all the information supported by x is also supported by y (but not necessarily vice versa).

Because of this relation of informational inclusion, not just any subset of W is suitable to represent a proposition; only those subsets closed upwards along $\leq$ are. If $X \subseteq W$ is not so closed, then there are $x \in X, y \notin X$ such that $x \leq y$. Since $x \leq y$, y must support all the information supported by x, but then x's being in X can't represent its supporting any information, since that information would have to be carried by y, but $y \notin X$. So such an X can't represent a proposition. Because of this, we'll have frequent use for the set of a frame's *propositions*:

Definition 5.1.4. *For a frame F, the set $Prop(F)$ of F's propositions is $\{X \subseteq W : If\ x \in X\ and\ x \leq y\ then\ y \in X\}$*

We can use these frames to model DLL as follows: take a language $\mathcal{L}$ built from a countable set At of atoms using the connectives $\wedge, \vee$ (both binary) and $\top, \bot$ (both nullary).

Definition 5.1.5. *A DLL-model is a tuple $\langle W, \leq, [\![\ ]\!] \rangle$ such that:*

- $\langle W, \leq \rangle$ *is a DLL-frame F, and*

- $[\![\ ]\!] : \mathcal{L} \to Prop(F)$ *is such that:*

  - $[\![A \wedge B]\!] = [\![A]\!] \cap [\![B]\!]$

  - $[\![A \vee B]\!] = [\![A]\!] \cup [\![B]\!]$

  - $[\![\top]\!] = W$

  - $[\![\bot]\!] = \emptyset$

Now it's easy to build up to a definition of validity:

Definition 5.1.6. *A DLL-model M is a* counterexample *to a consecution $A \vdash B$ iff, on M, $[\![A]\!] \not\subseteq [\![B]\!]$*

Definition 5.1.7. *A consecution $A \vdash B$ is* valid on a frame F *when no model on F is a counterexample to it.*

Definition 5.1.8. *A consecution $A \vdash B$ is* valid on a class of frames *when it is valid on every frame in the class.*

Theorem 5.1.9. *The logic DLL is sound and complete for the class of all DLL-frames.*

Proof. See (Restall, 2000b). □

Adding Neighborhoods

We add neighborhoods to interpret an additional connective; since we'll be modeling varieties of negation, it'll be a unary connective $-$. Adding neighborhoods to DLL-frames is straightforward:

Definition 5.1.10. *A* neighborhood frame *is a tuple $F = \langle W, \leq, N \rangle$ such that:*

- $\langle W, \leq \rangle$ *is a DLL-frame, and*

- $N : W \to \wp(Prop(F))$ *is a function such that if $x \leq y$, then $N(x) \subseteq N(y)$*

Each neighborhood frame determines a function $m_N : Prop(F) \to Prop(F)$ as follows: $m_N(X) = \{x \in W : X \in N(x)\}$. Note that its value really is always in $Prop(F)$; this is the purpose of the restriction on N in Definition 5.1.10. Now we can define neighborhood models:

Definition 5.1.11. *A* neighborhood model *is a tuple $\langle W, \leq, N, [\![\ ]\!] \rangle$ such that:*

- $\langle W, \leq, N \rangle$ *is a neighborhood frame F, and*

- $[\![\ ]\!] : \mathcal{L} \to Prop(F)$ *is as for DLL-models, with the additional proviso:*

  - $[\![-A]\!] = m_N([\![A]\!]) = \{x \in W : [\![A]\!] \in N(x)\}$

What is the logic of neighborhood frames?

Theorem 5.1.12. *The logic DLL plus the following rule is sound and complete for the class of all neighborhood frames:*

- $A \dashv\vdash B / - A \dashv\vdash -B$

Proof. Left as an exercise—it can be cobbled together straightforwardly from (Chellas, 1980) and (Restall, 2000b). □

5.2   Negative Neighborhoods

So far, there's nothing particularly negative about the connective $-$; it's as much like a possibility or a necessity as it is like a negation, and given appropriate additional constraints, it could be developed into any of these. An important question thus arises: what are the appropriate additional constraints

to develop $-$ into a negation? Since the present goal is to develop a unified framework for a variety of negations, the question expands: which negation-like features will we explore? §5.2.1 presents a list of key features in the area. §5.2.2 focuses on one—what I'll call the *subminimal condition*—and shows how to capture it neighborhoodly. Then §5.2.3 discusses and captures the others considered.

5.2.1 Classifying Negations

This is a selection of principles from the much larger list considered in (Vakarelov, 1989); it matches the list considered in (Dunn & Zhou, 2005), and in chapter 4.

The Subminimal Condition

The first condition to consider is what I'll call the *subminimal condition*; it's the only condition to follow that takes the form of a rule rather than a consecution.

Subminimal Condition (SMC): $A \vdash B / - B \vdash -A$

This implies the neighborhood constraint $A \dashv\vdash B / - A \dashv\vdash -B$, but not vice versa; so satisfying the subminimal condition is one way for a connective to satisfy the neighborhood constraint. Note that usual positive modalities satisfy the neighborhood constraint in another way, by satisfying the obverse of the subminimal condition. For example, any normal modal operator $\star$ satisfies $A \vdash B / \star A \vdash \star B$, whether $\star$ is a possibility or a necessity. (I called the positive version *subminimal assertion* in chapter 4.)

Some unary operators (eg 'is equivalent to C' for some sentence C) satisfy the neighborhood constraint but neither the subminimal condition nor its positive counterpart, and some (eg a connective $\star$ such that $\star A \dashv\vdash \star B$ for any A, B) satisfy both the subminimal condition and its positive counterpart.

NB: As in chapter 4, from the principles of DLL together with the subminimal condition, two of the four De Morgan inferences follow:

$$\frac{\dfrac{A \vdash A \vee B}{-(A \vee B) \vdash -A} \qquad \dfrac{B \vdash A \vee B}{-(A \vee B) \vdash -B}}{-(A \vee B) \vdash -A \wedge -B}$$

$$\frac{\dfrac{A \wedge B \vdash A}{-A \vdash -(A \wedge B)} \qquad \dfrac{A \wedge B \vdash B}{-B \vdash -(A \wedge B)}}{-A \vee -B \vdash -(A \wedge B)}$$

In terms of the De Morgan taxonomy of chapter 4, we can see that a connective $-$ satisfying the subminimal condition is only downward monotonic; it is neither anti-additive nor anti-multiplicative.

This makes it suitable for exploring the logical behavior of such contexts as those created by 'few', 'seldom', 'doubt', 'surprise', &c.

Just as all the other categories in the De Morgan taxonomy are downward monotonic, so too will all the negations to be considered here satisfy the subminimal condition. Not every negation in the literature does (CLuN, Nelson's constructive negation), but many do, so this isn't too severe a restriction. Plus, since there's reason (as in eg (van der Wouden, 1997)) to think that the 'negativeness' relevant to natural language comes from downward monotonicity, focusing here will help us draw connections to natural language.

Normality and Dual-Normality

These consecutions govern the interaction of $-$ with $\top$ and $\bot$:

Nor: $\top \vdash -\bot$

D-Nor: $-\top \vdash \bot$

If Nor is valid for a negation $-$ we call the negation *normal*; if D-Nor is, we call the negation *dual-normal*. We have from the definition of a DLL-frame that $\top$ holds everywhere and $\bot$ holds nowhere. Thus, Nor tells us that $-\bot$ holds everywhere, and D-Nor that $-\top$ holds nowhere. Although $\top$ and $\bot$ seem of relatively little use for exploring the semantics of natural language, we'll see presently that Nor and D-Nor are crucial principles for understanding the relations between certain familiar negations in the logical literature.

Anti-additivity and Anti-multiplicativity

As we've seen, any negation meeting SMC will satisfy two of the four De Morgan laws. Here, we consider the other two:

Anti-$\vee$: $-A \wedge -B \vdash -(A \vee B)$

Anti-$\wedge$: $-(A \wedge B) \vdash -A \vee -B$

A negation satisfying both SMC and Anti-$\vee$ is anti-additive, since SMC provides one direction of the equivalence $-(A \vee B) \dashv\vdash -A \wedge -B$ and Anti-$\vee$ provides the other. Similarly, a negation satisfying SMC and Anti-$\wedge$ is anti-multiplicative. As was pointed out in chapter 4, anti-additive negations can be given compatibility semantics, and anti-multiplicative negations can be given exhaustivity semantics.

These principles allow us to model the De Morgan taxonomy in full: merely downward monotonic connectives satisfy SMC but neither Anti-∨ nor Anti-∧; anti-additive connectives satisfy Anti-∨ in addition, while anti-multiplicatives satisfy Anti-∧; and antimorphic operators satisfy both Anti-∨ and Anti-∧, in addition to SMC.

Double negations

There are two familiar double-negation principles to consider:

DNI: $A \vdash -{-}A$

DNE: $-{-}A \vdash A$

While many negations validate both of these inferences, there are some (eg intuitionist negation) that validate DNI but not DNE, and others (eg Da Costa's C systems[5]) validate DNE but not DNI. In addition, some very negative-looking operators discussed in (Dunn & Zhou, 2005) validate neither DNI nor DNE, while many logics (classical and relevant logics jump to mind) validate both.

ex contradictione quodlibet and the law of excluded middle

Ex contradictione quodlibet (ECQ) and the law of excluded middle (LEM) each come in a 'minimal' and a 'full' flavor:[6]

ECQm: $A \wedge -A \vdash -\top$

ECQ: $A \wedge -A \vdash \bot$

LEMm: $-\bot \vdash A \vee -A$

LEM: $\top \vdash A \vee -A$

ECQm is valid in Johansson's minimal logic, while ECQ is not. Similarly, LEMm is valid in dual-minimal logic, while LEM is not. Adding ECQ to minimal logic results in intuitionist logic; adding LEM to dual-minimal logic results in dual-intuitionist logic.

The difference between the minimal and full versions of these principles is one of normality (and dual-normality). It's apparent that ECQm + D-Nor is sufficient for ECQ, and that LEMm + Nor is

---

[5]See eg (da Costa, 1974) for details on the C systems. The C systems rely heavily on an implication connective and so do not fit neatly into the present framework.

[6]There's an important distinction between ECQ and its relative *ex falso quodlibet*, or EFQ. For my purposes, EFQ is the consecution $\bot \vdash A$, which, as part of DLL, holds in every logic we'll consider. ECQ, on the other hand, is as presented in the main text.

sufficient for LEM. Less apparent are the reverse implications, which also hold (proofs that ECQm follows from ECQ, and LEMm from LEM, are apparent and so omitted here):

Deriving D-Nor from ECQ:

$$\dfrac{\dfrac{-\top \vdash \top \qquad -\top \vdash -\top}{-\top \vdash \top \wedge -\top} \qquad \top \wedge -\top \vdash \bot}{-\top \vdash \bot}$$

Deriving Nor from LEM:

$$\dfrac{\top \vdash \bot \vee -\bot \qquad \dfrac{\bot \vdash -\bot \qquad -\bot \vdash -\bot}{\bot \vee -\bot \vdash -\bot}}{\top \vdash -\bot}$$

Thus, we can see intuitionistic logic as the dual-normalized version of minimal logic, and dual-intuitionistic logic as the normalized version of dual-minimal logic.[7]

5.2.2 Subminimal Negation

This section considers the logic that results from adding SMC to DLL, and proves it sound and complete w/r/t a certain class of neighborhood frames. This section also proves that SMC corresponds to the class of frames, in the sense that the frames on which SMC preserves validity are precisely the class of frames in question.

The class of frames in question: All those neighborhood frames $F = \langle W, \leq, N \rangle$ such that for any $X, Y \in Prop(F), x \in W$, if $X \in N(x)$ and $Y \subseteq X$, then $Y \in N(x)$. That is, these are the frames such that $N(x)$ at any state x is closed under subsets.[8]

Soundness

Soundness is immediate from the result for DLL in all but one case: SMC. We must show that if $A \vdash B$ is valid on a frame, then $-B \vdash -A$ is too.

Proof. If $A \vdash B$ is valid on a frame, then no model on that frame is a counterexample to it. Thus, for any model on the frame, $[\![A]\!] \subseteq [\![B]\!]$. Now, suppose $-B \vdash -A$ were not valid on the frame. Then there must be some model on the frame such that $[\![-B]\!] \not\subseteq [\![-A]\!]$, so there must be some point x in that model such that $x \in [\![-B]\!]$ but $x \notin [\![-A]\!]$.

Since $x \in [\![-B]\!]$, we know that $[\![B]\!] \in N(x)$, and since $x \notin [\![-A]\!]$, we know that $[\![A]\!] \notin N(x)$. But we've seen that $[\![A]\!] \subseteq [\![B]\!]$ on this model, so $N(x)$ is not closed under subsets, and this frame must not be a member of the class we're interested in. $\qquad\square$

---

[7]See (Dunn & Zhou, 2005) for more on the relation between minimal and intuitionist logic and between their duals.

[8]That is, it's closed under the subset relation *restricted to the propositions on the frame*. $N(x)$ won't contain anything that's not a member of $Prop(F)$, for any x, F.

Completeness

Completeness for neighborhood frames can be shown using variations on familiar canonical-model techniques.[9] Since section 5.2.3 will be proving completeness for a variety of systems, this section will go through the technique in some detail so that it can be deployed with minimal comment there.

We build canonical models out of *prime theories* for the logic in question:

Definition 5.2.1. *A non-empty set Γ of sentences is a* theory *for a logic L iff:*

- *If $A \in \Gamma$ and $B \in \Gamma$, then $A \wedge B \in \Gamma$, and*

- *If $A \in \Gamma$ and $A \vdash B \in L$, then $B \in \Gamma$.*

A theory Γ for L is prime *iff in addition:*

- *If $A \vee B \in \Gamma$, then either $A \in \Gamma$ or $B \in \Gamma$.*

Theories for L, then, are sets of sentences closed under conjunction and L-consequence, and prime theories are those theories that include a witness for each disjunction.

Given some set W of theories and a sentence A, we use $|A|_W$ to mean the subset of W containing all and only those theories that themselves contain A (call this A's *proof set*). I'll omit the subscript W on $|A|$ when no confusion will result (which will be pretty much always, since the concern here is with a logic L's canonical frames, all of which (as we'll see shortly) involve the same set of L-theories).

Lemma 5.2.2 (Pair Extension Lemma)**.** *If L extends DLL, then for any L-theory Γ and sentence B not in Γ, there is a prime L-theory Λ such that $\Gamma \subseteq \Lambda$, and $B \notin \Lambda$.*

Proof. See (Restall, 2000b). $\square$

Corollary 5.2.3. *If $A \vdash B$ is not valid in L, then there is a prime L-theory containing A but not B.*

Proof. Consider the L-theory $\Gamma = \{C : A \vdash C \text{ in } L\}$, together with the sentence B. Since $A \nvdash B$, $B \notin \Gamma$. By the Pair Extension Lemma, there is a prime L-theory Λ such that $\Gamma \subseteq \Lambda$ but $B \notin \Lambda$. Since $A \in \Gamma$, $A \in \Lambda$. $\square$

With these tools in hand, we're ready for canonical frames:

Definition 5.2.4. *A neighborhood frame $F = \langle W_c, \leq_c, N_c \rangle$ is* canonical *for a logic L iff:*

---

[9]The techniques used here are cobbled together from techniques used in Restall (2000b) and Chellas (1980), inter alia.

- W_c is the set of all prime theories of L,

- $\leq_c$ is the subset relation on W_c, and

- for any $x \in W_c$, $A \in \mathcal{L}$, $-A \in x$ iff $|A| \in N_c(x)$ (equivalently: $|-A| = m_{N_c}(|A|)$)

Fact 5.2.5. *This last condition is unambiguous, since if $|A| = |B|$ then $|-A| = |-B|$.*

Proof. First, note that for any A, B such that $A \not\Vdash B$, $|A| \neq |B|$. To see this, suppose without loss of generality that $A \not\vdash B$. By corollary 5.2.3, there is a prime L-theory Λ including A but not B. Since W_c is the set of all prime L-theories, $|A| \neq |B|$.

So if $|A| = |B|$ then $A \Vdash B$. And since $A \Vdash B$, $-A \Vdash -B$ (by SMC), and so any prime theory that includes $-A$ must include $-B$, and vice versa. Thus, $|-A| = |-B|$. $\square$

NB: a logic L can have many distinct canonical frames. They'll all match on W_c and $\leq_c$, but there are many different ways for a neighborhood function to be canonical. This is because the condition given in Definition 5.2.4 determines how $N_c(x)$ behaves w/r/t all proof sets (sets that are $|A|$ for some sentence A), but not otherwise. For the members of $Prop(W_c)$ that aren't proof sets, any way at all of handling them will leave us with a canonical frame.

Once we have a canonical frame, we can go on to build a canonical model:

Definition 5.2.6. *A neighborhood model $M = \langle W_c, \leq_c, N_c, [\![\ ]\!]_c \rangle$ is canonical for a logic L iff:*

- $\langle W_c, \leq_c, N_c \rangle$ *is a canonical frame for L, and*

- *for each $A \in At$, $[\![A]\!] = |A|$.*

This gives us the value of the $[\![\ ]\!]$ function at all the atoms in our language; the value for compound sentences is derived, as for all neighborhood models, by the rules in definition 5.1.11. Note that there is precisely one canonical model for each canonical frame; since there are multiple canonical frames for each logic L, there are multiple canonical models too.

We now come to a crucial fact about canonical models; it's this fact that gives them their use in proving completeness:

Lemma 5.2.7 (Truth Lemma)**.** *In any neighborhood model $M = \langle W_c, \leq_c, N_c, [\![\ ]\!] \rangle$ canonical for a logic L, for any sentence A (atomic or compound), $[\![A]\!] = |A|$.*

Proof. Proof is by induction on the formation of A. The base case, for $A \in At$, is trivial; it's given as part of definition 5.2.6.

Take the case where A is of the form $B \wedge C$ for some sentence B and C (such that the lemma holds for B and C). By 5.1.11, $[\![B \wedge C]\!] = [\![B]\!] \cap [\![C]\!]$. By the inductive hypothesis, this is the same as $|B| \cap |C|$. Since theories are closed under adjunction, $|B| \cap |C| \subseteq |B \wedge C|$; and since they are closed under consequence, $|B \wedge C| \subseteq |B| \cap |C|$. So $[\![B \wedge C]\!] = |B \wedge C|$.

Take the case where A is of the form $B \vee C$. $[\![B \vee C]\!] = [\![B]\!] \cup [\![C]\!]$. By the inductive hypothesis, this is $|B| \cup |C|$. Since theories are closed under consequence, $|B| \cup |C| \subseteq |B \vee C|$, and since all these theories are prime, $|B \vee C| \subseteq |B| \cup |C|$. So $[\![B \vee C]\!] = |B \vee C|$.

This leaves the case where A is of the form $-B$ for some sentence B such that $[\![B]\!] = |B|$. First, note that $[\![-B]\!] = m_{N_c}([\![B]\!]) = m_{N_c}(|B|)$ (by the truth-rule for $-$ and the inductive hypothesis). Now, by definition 5.2.4, $x \in |-B|$ iff $|B| \in N_c(x)$. But $m_{N_c}(|B|) = \{x : |B| \in N_c(x)\}$. So $|-B| = m_{N_c}(|B|)$, and chaining these equalities together gets us what we want. $\qquad \square$

Corollary 5.2.8. *If a consecution $A \vdash B$ is invalid in a logic L, then any neighborhood model canonical for L is a counterexample to $A \vdash B$.*

Proof. Since $A \vdash B$ is invalid, there is a prime L-theory Λ containing A but not B, by corollary 5.2.3. Since the space of states in any canonical model includes all prime theories, $|A| \not\subseteq |B|$ in such a model. By lemma 5.2.7, then, $[\![A]\!] \not\subseteq [\![B]\!]$ in such a model; that is, such a model is a counterexample to $A \vdash B$. $\qquad \square$

With corollary 5.2.8 in hand, we can lay out the canonical-model strategy for our completeness proof. We want to show that when, in our logic DLL + SMC, $A \not\vdash B$, there is some frame in the class of frames meeting our condition (that $N(x)$ be closed under subsets for any state x) such that there is a counterexample to $A \vdash B$ buildable on that frame. Since corollary 5.2.8 tells us that any canonical model is a counterexample to such a consecution, and since every canonical model is built on a canonical frame, we simply need to find some frame canonical for DLL + SMC in the relevant class of frames. Since the only difference between various canonical frames is in their neighborhood functions, this is where we need to look; the space of prime theories and the subset relation on that space won't change.

Here's a proposal that won't work (but will lead us to one that does): try the *smallest* canonical neighborhood function. That is, we try to use the function $N_{min}(x) = \{|A| : -A \in x\}$. Recall that all neighborhood functions must agree on all proof sets in $Prop(W_c)$, but can do whatever they like with non proof sets; this neighborhood function simply excludes all non proof sets from its value at every prime theory. It's surely canonical, and so the canonical model built on a frame using this neighborhood function indeed provides a counterexample to every invalid consecution. Just one

79

problem: this frame isn't in the class of frames we were concerned with, since N_{min}'s values are not closed under subsets. Consider $X = \{x \in W_c : \text{for all sentences } A \in At, x \in |A|\}$. As can be quickly verified, $X \in Prop(W_c)$. But X is not a proof set; there is no sentence B such that $X = |B|$. As such, $X \notin N_{min}(x)$ for any prime theory x. But take some atom A. Since W_c includes all prime theories, $|-A| \neq \emptyset$, so there is some prime theory y such that $|A| \in N_{min}(y)$. Now, $X \subseteq |A|$, but as we've seen, $X \notin N_{min}(y)$. So $N_{min}(y)$ isn't closed under subsets, and so our smallest canonical neighborhood frame isn't a member of the relevant class.

Nonetheless, we're close. Here's a proposal that *will* work: start with N_{min}, and add to it the right non proof sets. Which ones are the right ones? Those that are subsets of things already in N_{min}. That is, we define a neighborhood function N as follows: $N(x) = \{X : X \subseteq Y \text{ for some } Y \text{ such that } Y \in N_{min}(x)\}$. (Note that $N_{min}(x) \subseteq N(x)$, at any x.) We need to verify two things about N now: first, that N canonical, and second, that the canonical frame using N is a member of the class we wanted—that is, that $N(x)$ is closed under subsets at any $x \in W_c$.

Suppose N is not canonical. Then there must be some proof set $|A|$ such that $|-A| \neq m_N(|A|)$. Since $N_{min}(x) \subseteq N(x)$ for all x, this must be because there's some prime theory y such that $|A| \in N(y)$ but $y \notin |-A|$. Since $|A| \in N(y)$, there's some set Y such that $|A| \subseteq Y$ and $Y \in N_{min}(y)$. But $N_{min}(y)$ contains only proof sets, so $Y = |B|$ for some sentence B. Now since $|A| \subseteq |B|$, $A \vdash B$ by corollary 5.2.3, and so $-B \vdash -A$. What's more, since $|B| \in N_{min}(y)$, $y \in |-B|$; and since y is a theory, $y \in |-A|$. Contradiction. So N is indeed canonical.

Now we need to show that $N(x)$ is closed under subsets at any x. Take any $X, Y \in Prop(W_c)$ such that $X \subseteq Y$ and suppose $Y \in N(x)$ for some $x \in W_c$. There must be some $Z \in N_{min}(x)$ such that $Y \subseteq Z$. So $X \subseteq Z$ too, and $X \in N(x)$. This concludes our completeness proof for DLL + SMC.

Note that most of the proof—everything through corollary 5.2.8—holds regardless of which logic we're dealing with. So this strategy can be adapted without much work to handle a number of different logics, as we'll see. For each logic, all that must be done to prove it complete w/r/t a certain class of frames is to find a canonical frame for that logic that is a member of the class in question (or simply show that some such canonical frame exists). In section 5.2.3, we'll see how this works for a number of logics that extend DLL + SMC.

Correspondence

First, though, there's another metalogical property that holds between SMC and the present condition on frames. Not only is DLL + SMC sound and complete w/r/t the class, but the condition

SMC *corresponds* to the subset-closure condition on frames, in the following sense: the frames that satisfy the condition are precisely those on which SMC preserves validity.[10]

Proof. LTR: Take a frame F on which SMC does not preserve validity. That is, there is a consecution $A \vdash B$ valid on the frame such that the consecution $-B \vdash -A$ is not valid on the frame. Since $A \vdash B$ is valid, it must be that, on any model built on this frame, $[\![A]\!] \subseteq [\![B]\!]$; and since $-B \vdash -A$ is not valid, it must be that there is some model built on this frame such that $[\![-B]\!] \not\subseteq [\![-A]\!]$. Consider this model. It contains a point x such that $x \in [\![-B]\!]$ and $x \notin [\![-A]\!]$. By the truth-caluse for $-$, we have it that $[\![B]\!] \in N(x)$ and $[\![A]\!] \notin N(x)$. Since we know $[\![A]\!] \subseteq [\![B]\!]$, this frame must violate the subset condition.

RTL: Take a frame F that violates the subset condition. This frame contains a point x such that there are some $X, Y \in Prop(F)$, $X \subseteq Y$, such that $Y \in N(x)$ and $X \notin N(x)$. Clearly, $p \wedge q \vdash p$ is valid on this frame, since it's valid on any of our frames; but $-p \vdash -(p \wedge q)$ is not valid on this frame. Consider a model built on this frame such that $[\![p]\!] = Y$ and $[\![q]\!] = X$. We know that $[\![p \wedge q]\!] = Y \cap X = X$. Since $Y \in N(x)$, $x \in [\![-p]\!]$, but since $X \notin N(x)$, $x \notin [\![-(p \wedge q)]\!]$. So this model is a counterexample to $-p \vdash -(p \wedge q)$.

$\square$

5.2.3   Variations

This section considers the logics that result by extending DLL + SMC with one rule at a time from the menu in section 5.2.1. Since we know from section 5.2.2 that SMC is only valid on frames closed under subsets, all the frames in this section will be assumed to meet this condition; in particular, when I talk about the class of frames meeting a certain condition, I mean the class of frames *closed under subsets* meeting the additional condition.

Normality and Dual-normality

A *normal* logic, as we've seen, is one where $\top \vdash -\bot$ is valid; that is, where $-\bot$ holds everywhere. Since $[\![\bot]\!] = \emptyset$ in any model, this suggests the following frame condition: for any point x, $\emptyset \in N(x)$.

And in fact, this is just the frame condition we'll use; the logic DLL + SMC + Nor is sound and complete w/r/t the class of frames that meet this condition, and the condition corresponds to the consecution Nor. Soundness is immediate. For completeness, we show that this logic has a canonical

---

[10]Correspondence is a much-investigated metalogical property; see eg (Blackburn et al., 2001), (van Benthem, 1984a). It's usually explored in terms of formula-validity (or, in this context, consecution-validity) rather than rules' preserving validity, but the extension to the present case is straightforward.

frame meeting the condition. As usual W_c and $\leq_c$ take care of themselves. We use a familiar method to reach a suitable N_c; start with the minimal canonical neighborhood function and close it under subsets.

More explicitly, we again take $N_{min}(x) = \{|A| : -A \in x\}$, and define $N_c(x)$ as $\{X : X \subseteq Y$ for some Y such that $Y \in N_{min}(x)\}$. Note that, although the definition is similar to the one we used in the completeness proof for DLL + SMC, these are not the same functions as before; DLL + SMC + Nor has a different (smaller) set of prime theories, which affects the domain of these functions as well as the values of $|A|$. Nonetheless, the same reasoning leads us to the conclusion that this new N_c is canonical for DLL + SMC + Nor, and that it's closed under subsets. It remains to show that $\emptyset \in N_c(x)$ for every $x \in W_c$. We know that $|\top| = W_c$ (since theories are non-empty, and $A \vdash \top$ for any A), and so $|-\bot| = W_c$, since $\top \vdash -\bot$ in this logic. We also know that $|\bot| = \emptyset$. So $\emptyset = |\bot| \in N_{min}(x)$ for any x, and since $N_{min}(x) \subseteq N_c(x)$, $\emptyset \in N_c(x)$.

Correspondence is likewise quick to prove. It's clear that Nor is valid in any frame meeting this condition. And given a frame not meeting the condition, we can construct a counterexample to Nor; in fact, we can't help but construct such a counterexample, since $[\![\bot]\!] = \emptyset$ in any model. Whatever point x in this frame is such that $\emptyset \notin N(x)$, $x \notin [\![-\bot]\!]$. And since $x \in [\![\top]\!]$ always, this frame is a counterexample to Nor.

A *dual-normal* logic, recall, is one where $-\top \vdash \bot$ is valid; that is, where $-\top$ holds nowhere. Since $[\![\top]\!] = W$ in any model, this suggests the following frame condition: for any point x, $W \notin N(x)$. Proofs that DLL + SMC + D-Nor is sound and complete w/r/t frames meeting this condition, and that D-Nor corresponds to the condition, are parallel to the proofs in the case of Nor.

Anti-additivity and Anti-multiplicativity

To get a syntactic handle on De Morgan laws, we need to think about the relations between negation, conjunction, and disjunction. Appropriately enough, then, to get a semantic handle on them, we need to consider the relations between the neighborhood function N, set intersection, and set union. Some relations are immediate, given that $N(x)$ is closed under subsets (where X, Y are propositions on the frame in question):

- If $X \cup Y \in N(x)$, then $X \in N(x)$ and $Y \in N(x)$.

- If $X \in N(x)$ or $Y \in N(x)$, then $X \cap Y \in N(x)$.

These properties directly reflect the two De Morgan laws that follow from SMC. What about the other two De Morgan laws? They can be captured in the following frame conditions:

fA∨: If $X \in N(x)$ and $Y \in N(x)$, then $X \cup Y \in N(x)$; that is, $N(x)$ is closed under finite unions.

fA∧: If $X \cap Y \in N(x)$, then $X \in N(x)$ or $Y \in N(x)$; that is, $Prop(W) - N(x)$ is closed under finite intersections.

Note that these are simply the converses of the conditions that preceded them, just as the De Morgan laws we're calling A∨ and A∧ are converses of the De Morgan laws that follow directly from SMC. Soundness is immediate; let's take a look at the completeness proofs for these.

First, A∨. We again start from $N_{min}(x) = \{|A| : - A \in x\}$, and define $N_c(x) = \{X : X \subseteq Y$ for some Y such that $Y \in N_{min}(x)\}$.[11] As before, we find that N_c is indeed canonical, and is closed under subsets.

We can also show that it meets the frame condition fA∨. Consider a prime theory x such that $X \in N_c(x)$ and $Y \in N_c(x)$. We know there must be proof sets $|A|$ and $|B|$ such that $X \subseteq |A|$, $Y \subseteq |B|$, and $|A|, |B| \in N_{min}(x)$. So $x \in | - A|$, and $x \in | - B|$; thus $x \in | - A \wedge - B|$ (since theories are closed under adjunction). By the consecution A∨, then, $x \in | - (A \vee B)|$, so $|A \vee B| \in N_{min}(x)$. Since we're only dealing with prime theories, $|A \vee B| = |A| \cup |B|$, so $|A| \cup |B| \in N_{min}(x)$. Finally, since $X \subseteq |A|$ and $Y \subseteq |B|$, $X \cup Y \subseteq |A| \cup |B|$, and so $X \cup Y \in N_c(x)$.

But this canonical neighborhood function, despite its usefulness so far, won't work for A∧. Instead, we come at it from the other direction; instead of starting from the *minimal* canonical neighborhood function and adding some non proof sets, we start from the *maximal* canonical neighborhood function and remove some non proof sets.

Here's how it goes: let $N_{max}(x) = \{X \in Prop(W_c) : X \neq |A|$ for any A such that $x \notin | - A|\}$. For all the proof sets $|A|$ in $Prop(W)$, this matches the minimal function; $|A| \in N_{min}(x)$ iff $|A| \in N_{max}(x)$. But whereas all non proof sets are excluded from $N_{min}(x)$, all non proof sets are *included* in $N_{max}(x)$. This would give us trouble with closure under subsets; we might have a non proof set included despite its being a superset of an excluded proof set. So we need to tweak a bit— let $N_c(x) = \{X : X \in N_{max}(x)$ and there is no $Y \in Prop(W_c)$ such that $Y \notin N_{max}(x)$ and $Y \subseteq X\}$.[12] Now, there are some things to show: that N_c is canonical, that it's closed under subsets, and finally, that it meets the frame condition fA∧.

First: N_c is canonical.

---

[11] Again, remember that this is a different function, since we've again changed which logic is in question; now we're considering canonical frames for DLL + SMC + A∨, which will be built on a different set of prime theories.

[12] The same N_c could be reached in multiple steps as follows: define $N_{max}(x)$ as above. Let $M(x) = Prop(W_c) - N_{max}(x)$. Close $M(x)$ under *supersets* to get $L(x)$; that is, let $L(x) = \{X : X \supseteq Y$ for some Y such that $Y \in M(x)\}$. Finally, define $N_c(x)$ as $Prop(W_c) - L(x)$.

Proof. Assume it's not. Then there's some proof set $|A|$ such that $|-A| \neq m_{N_c}(|A|)$. Since $N_c(x) \subseteq N_{max}(x)$ for all x, this must be because there's some prime theory y such that $|A| \notin N_c(y)$ but $-A \in y$. Since $-A \in y$, $|A| \in N_{max}(y)$, so there must be some $X \in Prop(W_c)$ such that $X \notin N_{max}(y)$ and $X \subseteq |A|$. But the only members of $Prop(W_c)$ not in $N_{max}(y)$ are proof sets, so it must be that $X = |B|$ for some B. Now, since $|B| \subseteq |A|$, $B \vdash A$, and so $-A \vdash -B$. What's more, since $|B| \notin N_{max}(y)$, $y \notin |-B|$, and so $y \notin |-A|$. Contradiction. $\square$

Second: N_c is closed under subsets.

Proof. Suppose otherwise. Then there's some $X, Y \in Prop(W_c)$ and some $x \in W_c$ such that $X \subseteq Y$, $Y \in N_c(x)$, but $X \notin N_c(x)$. Since $Y \in N_c(x)$, we know both that $Y \in N_{max}(x)$, and that there is no $Z \in Prop(W_c)$ such that $Z \notin N_{max}(x)$ and $Z \subseteq Y$. In particular, X can't be such a Z; and since $X \subseteq Y$, it must be that $X \in N_{max}(x)$. So since $X \notin N_c(x)$, there is a $W \in Prop(W_c)$ such that $W \notin N_{max}(x)$ and $W \subseteq X$; but then $W \subseteq Y$, and $Y \notin N_c(x)$. Contradiction. $\square$

Third: $Prop(W_c) - N_c(x)$ is closed under finite intersections.

Proof. Suppose otherwise. Then there are $X, Y \in Prop(W_c)$, $x \in W_c$ such that $X \notin N_c(x)$, $Y \notin N_c(x)$, but $X \cap Y \in N_c(x)$. Since the only members of $Prop(W_c)$ not in $N_{max}(x)$ are the proof sets $|A|$ such that $-A \notin x$, this means that there are some sentences A, B such that $|A| \subseteq X$, $|B| \subseteq Y$, and $-A, -B \notin x$. Since neither $-A$ nor $-B$ is in x, and since x is prime, $-A \vee -B \notin x$. Since $-(A \wedge B) \vdash -A \vee -B$, and since x is a theory, $-(A \wedge B) \notin x$; that is, $|A \wedge B| \notin N_{max}(x)$. We know $|A \wedge B| = |A| \cap |B|$, and so $|A| \cap |B| \notin N_{max}(x)$. Since $|A| \subseteq X$ and $|B| \subseteq Y$, $|A| \cap |B| \subseteq X \cap Y$, so $X \cap Y \notin N_c(x)$. Contradiction. $\square$

The frame conditions fA$\vee$ and fA$\wedge$ also correspond to the consecutions A$\vee$ and A$\wedge$.

Fact 5.2.9. *The consecution A$\vee$ is valid on a frame iff the frame meets the condition fA$\vee$.*

Proof. LTR: Take an arbitrary frame F that doesn't meet the condition fA$\vee$. It includes some point x such that $N(x)$ isn't closed under finite unions, and so $N(x)$ can't be closed under binary unions.[13] So there are some $X, Y \in Prop(F)$ such that $X, Y \in N(x)$ but $X \cup Y \notin N(x)$. Now we show that $-p \wedge -q \vdash -(p \vee q)$ is not valid on this frame. Consider a model built on this frame such that

---

[13]If it were, it'd be closed under all finite unions. It's worth noting that this is the only appeal in these proofs to *finite* unions, rather than arbitrary. A modified frame condition requiring closure under arbitrary unions would still allow for soundness, completeness, and the RTL direction of correspondence, but the proof of full correspondence would break down here.

$[\![p]\!] = X$ and $[\![q]\!] = Y$. On such a model, $x \in [\![-p \wedge -q]\!]$, but $x \notin [\![-(p \vee q)]\!]$, so the model is a counterexample.

RTL: Take an arbitrary frame F on which A$\vee$ isn't valid. There is some model M on this frame and some sentences A and B such that on M, $[\![-A \wedge -B]\!] \nsubseteq [\![-(A \vee B)]\!]$; thus, there is a point x such that $x \in [\![-A \wedge -B]\!]$ and $x \notin [\![-(A \vee B)]\!]$. We know that $x \in [\![-A]\!]$ and $x \in [\![-B]\!]$, so $[\![A]\!], [\![B]\!] \in N(x)$. But it must be that $[\![A \vee B]\!] \notin N(x)$, so $[\![A]\!] \cup [\![B]\!] \notin N(x)$, and $N(x)$ is not closed under finite intersections. $\qquad \square$

Fact 5.2.10. *The consecution A$\wedge$ is valid on a frame iff the frame meets the condition fA$\wedge$.*

Proof. LTR: Take an arbitrary frame F that doesn't meet the condition fA$\wedge$. It includes some point x such that $Prop(F) - N(x)$ isn't closed under finite intersections, and so $Prop(F) - N(x)$ can't be closed under binary intersections.[14] So there are some $X, Y \in Prop(F)$ such that $X \notin N(x)$, $Y \notin N(x)$, but $X \cap Y \in N(x)$. Now we show that $-(p \wedge q) \vdash -p \vee -q$ is not valid on this frame. Consider a model built on F such that $[\![p]\!] = X$ and $[\![q]\!] = Y$. On such a model, $x \in [\![-(p \wedge q)]\!]$, but $x \notin [\![-p \vee -q]\!]$, so the model is a counterexample.

RTL: Take an arbitrary frame F on which A$\wedge$ isn't valid. There is some model M on this frame and some sentences A and B such that on M for some x, $x \in [\![-(A \wedge B)]\!]$ but $x \notin [\![-A \vee -B]\!]$. From this, we can see that $[\![A \wedge B]\!] = [\![A]\!] \cap [\![B]\!] \in N(x)$, that $[\![A]\!] \notin N(x)$, and that $[\![B]\!] \notin N(x)$. Since both $[\![A]\!]$ and $[\![B]\!]$ are in $Prop(F) - N(x)$, but $[\![A]\!] \cap [\![B]\!] \in N(x)$, F does not meet the condition fA$\wedge$. $\qquad \square$

Double negations

Consider our double-negation consecutions:

DNI: $A \vdash --A$

DNE: $--A \vdash A$

We can capture these as follows. On a frame F, where $X \in Prop(F)$, and $x \in W$,

fDNI: If $x \in X$, then $\{y : X \in N(y)\} \in N(x)$.

fDNE: If $\{y : X \in N(y)\} \in N(x)$, then $x \in X$.

Completeness is still proving tricky here. But correspondence holds:

---

[14]As before, this is the only appeal to fA$\wedge$'s use of *finite* rather than arbitrary intersections.

Fact 5.2.11. *The consecution DNI is valid on a frame iff the frame meets the condition fDNI.*

Proof. LTR: Take an arbitrary frame F that doesn't meet fDNI. There is some $X \in Prop(F)$ such that for some $x \in X, \{y : X \in N(y)\} \notin N(x)$. Recall that $m_N(X) = \{y : X \in N(y)\}$, so $m_N(X) \notin N(X)$. We show that $p \vdash --p$ is not valid on this frame. Consider a model such that $[\![p]\!] = X$. Then $[\![-p]\!] = m_N(X)$, so $[\![-p]\!] \notin N(x)$. Thus, $x \in [\![p]\!]$, but $x \notin [\![--p]\!]$; the model is a counterexample.

RTL: Take an arbitrary frame F on which DNI isn't valid. There is some model M on F and some sentence A such that on M for some x, $x \in [\![A]\!]$ but $x \notin [\![--A]\!]$. Thus, $x \notin m_N[\![-A]\!]$, so $[\![-A]\!] \notin N(x)$. Since $[\![-A]\!] = \{y : [\![A]\!] \in N(y)\}$, $\{y : [\![A]\!] \in N(y)\} \notin N(x)$. Thus, fDNI is violated. $\qquad\square$

Fact 5.2.12. *The consecution DNE is valid on a frame iff the frame meets the condition fDNE.*

Proof. LTR: Take an arbitrary frame F that doesn't meet fDNE. There is some $X \in Prop(F)$ such that for some $x \notin X$, $\{y : X \in N(y)\} \in N(x)$; that is, $m_N(X) \in N(x)$. We show that $--p \vdash p$ is not valid on F. Construct a model on F such that $[\![p]\!] = X$. Since $[\![-p]\!] = m_N([\![p]\!])$, $[\![-p]\!] \in N(x)$, so $x \in [\![--p]\!]$. But since $x \notin X$, $x \notin [\![p]\!]$. So this model is a counterexample.

RTL: Take an arbitrary frame F on which DNE isn't valid. There is some model M on F and some sentence A such that on M for some x, $x \in [\![--A]\!]$ but $x \notin [\![A]\!]$. Thus, $[\![-A]\!] \in N(x)$. Since $[\![-A]\!] = \{y : [\![A]\!] \in N(y)\}$, $\{y : [\![A]\!] \in N(y)\} \in N(x)$. But since $x \notin [\![A]\!]$, fDNE is violated. $\qquad\square$

ECQ and LEM

Recall the consecutions ECQ and LEM, and their minimal relatives ECQm and LEMm:

ECQ: $A \wedge -A \vdash \bot$

ECQm: $A \wedge -A \vdash -\top$

LEM: $\top \vdash A \vee -A$

LEMm: $-\bot \vdash A \vee -A$

Let's approach ECQ first. Since $\bot$ holds nowhere on any model, ECQ tells us that A and $-A$ can't hold at the same place anywhere on any model. This is easy enough to enforce via a frame condition: simply require of each frame F that for any $X \in Prop(F)$, $X \cap m_N(X) = \emptyset$. The minimal version is a bit more complex: for any $X \in Prop(F)$, for any $x \in X \cap m_N(X)$, $W \in N(x)$. Remember, if we impose the frame condition for Dual-Normality, no point x can be such that

$W \in N(x)$; this makes these conditions equivalent. In the absence of Dual-Normality, however, they come apart.[15]

Things work similarly in the case of LEM. Since $\top$ holds everywhere on every model, LEM tells us that at least one of A and $-A$ must hold at each point in any model. This can be enforced as follows: on a frame F, for any $X \in Prop(F)$, $X \cup m_N(X) = W$. As with ECQ, the minimal variation is a bit more complex: for any $X \in Prop(F)$, for any $x \in W - (X \cup m_N(X))$, $\emptyset \notin N(x)$. If we impose the frame condition for Normality, every x must be such that $\emptyset \in N(x)$, so given Normality these conditions are equivalent.

To sum up these frame conditions:

fECQ: for $X \in Prop(F)$, $X \cap m_N(X) = \emptyset$

fECQm: for $X \in Prop(F)$, for any $x \in X \cap m_N(X)$, $W \in N(x)$

fLEM: for $X \in Prop(F)$, $X \cup m_N(X) = W$

fLEMm: for $X \in Prop(F)$, for any $x \in W - (X \cup m_N(X))$, $\emptyset \notin N(x)$

I'll show completeness for the ECQs first, then the LEMs.

Ad ECQm. Consider the canonical neighborhood function gotten by closing the smallest canonical neighborhood function under subsets: $N(x) = \{X : X \subseteq |A| \text{ for some } A \text{ such that } -A \in x\}$. Assume the condition fECQm is violated. Then there must be some $X \in Prop(F)$ such that, for some $x \in X \cap m_N(X)$, $W \notin N(x)$. Since $x \in m_N(X)$, we know $X \in N(x)$; but then $X \subseteq |A|$ for some A such that $-A \in x$. Since $x \in X$, $x \in |A|$, so $A \in x$. Since both A and $-A$ are in x, $A \wedge -A \in x$, and so $-\top \in x$. But since $|\top| = W$, this means that $W \in N(x)$. Contradiction.

Ad ECQ. As above until $A \wedge -A \in x$. Then ECQ requires that $\perp \in x$, but we know $\perp \notin x$. Contradiction.

Ad LEMm. Consider the canonical neighborhood function gotten by pruning the largest canonical neighborhood function so it's closed under subsets: $N(x) = \{X : X \not\supseteq |A| \text{ for any } A \text{ such that } -A \notin x\}$. Assume the condition fLEMm is violated. Then there is some $X \in Prop(F)$, $x \in W - (X \cup m_N(X))$ such that $\emptyset \in N(x)$. Since $|\perp| = \emptyset$, $|\perp| \in N(x)$, so $|\perp| \not\supseteq |A|$ for any A such that $-A \notin x$; thus, $-\perp \in x$. By LEMm, $B \vee -B \in x$ for every sentence B; that is, $x \in |B|$ or else $-B \in x$. Now, since $x \notin m_N(X)$, $X \notin N(x)$. But then $X \supseteq |C|$ for some C such that $-C \notin x$. Since $-C \notin x$, but we know that $C \vee -C \in x$, it must be that $x \in |C|$. Since $X \supseteq |C|$, $x \in X$. But then $x \notin W - (X \cup m_N(X))$. Contradiction.

---

[15]This has a clear syntactic analogue as well; see §5.2.1 for details.

Ad LEM. Again, use $N(x) = \{X : X \not\supseteq |A|$ for any A such that $-A \notin x\}$, and assume fLEM is violated. Then there is some $X \in Prop(F)$ such that there is some $x \notin X \cup m_N(X)$. Since $x \notin m_N(X)$, $X \notin N(x)$. But then $X \supseteq |C|$ for some C such that $-C \notin x$. The argument proceeds as above.

Correspondence holds as well:

Fact 5.2.13. *The consecution ECQm is valid on a frame iff the frame meets the condition fECQm.*

Proof. LTR: Take an arbitrary frame F that doesn't meet fECQm. There's some $X \in Prop(F)$ such that for some x, $x \in X \cap m_N(X)$ and $W \notin N(x)$. We show that $p \wedge -p \vdash -\top$ is not valid on F. Build a model on F such that $\llbracket p \rrbracket = X$. Since $\llbracket \top \rrbracket = W$, $x \notin \llbracket -\top \rrbracket$, but nonetheless, since $x \in \llbracket p \rrbracket \cap m_N(\llbracket p \rrbracket)$, $x \in \llbracket p \wedge -p \rrbracket$. So this model is a counterexample.

RTL: Take an arbitrary frame F on which ECQm isn't valid. There's some model M on F such that for some sentence A on M, for some x, $x \in \llbracket A \wedge -A \rrbracket$ but $x \notin \llbracket -\top \rrbracket$. Since on any model $\llbracket \top \rrbracket = W$, it must be that $W \notin N(x)$. But $x \in \llbracket A \rrbracket \cap \llbracket -A \rrbracket$, which means $x \in \llbracket A \rrbracket \cap m_N(\llbracket A \rrbracket)$. So fECQm is not met. $\square$

Fact 5.2.14. *The consecution ECQ is valid on a frame iff the frame meets the condition fECQ.*

Proof. Trivial modifications of above. $\square$

Fact 5.2.15. *The consecution LEMm is valid on a frame iff the frame meets the condition fLEMm.*

Proof. LTR: Take an arbitrary frame F that doesn't meet fLEMm. There's some $X \in Prop(F)$ such that for some x, $x \in W - (X \cup m_N(X))$ but $\emptyset \in N(x)$. We show that $-\bot \vdash p \vee -p$ is not valid on F. Build a model on F such that $\llbracket p \rrbracket = X$. Since $\llbracket \bot \rrbracket = \emptyset$, $x \in \llbracket -\bot \rrbracket$, but nonetheless, since $x \notin \llbracket p \rrbracket \cup m_N(\llbracket p \rrbracket)$, $x \notin \llbracket p \vee -p \rrbracket$. So this model is a counterexample.

RTL: Take an arbitrary frame F on which LEMm isn't valid. There's some model M on F such that for some sentence A on M, for some x, $x \in \llbracket -\bot \rrbracket$ but $x \notin \llbracket A \vee -A \rrbracket$. Since on any model $\llbracket \bot \rrbracket = \emptyset$, it must be that $\emptyset \in N(x)$. But $x \in \llbracket A \rrbracket \cup \llbracket -A \rrbracket$, which means $x \in \llbracket A \rrbracket \cup m_N(\llbracket A \rrbracket)$. So fLEMm is not met. $\square$

Fact 5.2.16. *The consecution LEM is valid on a frame iff the frame meets the condition LEM.*

Proof. Trivial modifications of above. $\square$

5.2.4 Combinations/Consilience

Anti-additivity and anti-multiplicativity together

Consider the logic DLL + SMC + A∨ + A∧. This logic includes a negation that satisfies all four De Morgan laws, although it is not a De Morgan negation.[16]

As one might suspect, we can simply combine the conditions given in section 5.2.3, repeated here:

A∨: If $X \in N(x)$ and $Y \in N(x)$, then $X \cup Y \in N(x)$; that is, $N(x)$ is closed under finite unions.

A∧: If $X \cap Y \in N(x)$, then $X \in N(x)$ or $Y \in N(x)$; that is, $Prop(W) - N(x)$ is closed under finite intersections.

Again, our task in proving completeness is to find a canonical neighborhood function for DLL + SMC + A∨ + A∧ that meets both of these conditions. For these purposes it'll be useful to note, what's so far been entirely implicit, the algebraic structure of the space of propositions.

Take the set of prime theories of this logic W_c with its canonical ordering $\leq_c$, and define as usual the space of propositions $Prop(W_c)$ over the set. Now consider the structure $\mathcal{A} = \langle Prop(W_c), \subseteq, \cup, \cap \rangle$. $\mathcal{A}$ is a (complete distributive) lattice, with partial order $\subseteq$, join $\cup$, and meet $\cap$.

Now consider the values of a neighborhood function here. What do we know about $N(x)$ for arbitrary x? First, $N(x) \subseteq Prop(W_c)$. We know more. We require $N(x)$ to be closed under subsets; closed downward along $\mathcal{A}$'s partial order. In addition, imposing our semantic requirement for anti-additivity gives us that $N(x)$ is closed under finite unions. So $N(x)$ is a quasi-ideal in the lattice $\mathcal{A}$.[17]

We should also think about $Prop(W_c) - N(x)$, the complement of the neighborhood function's value. Since $N(x)$ is closed downwards along $\subseteq$, its complement is closed upward. And imposing our requirement for anti-multiplicativity gives us that $Prop(W_c) - N(x)$ is closed under finite intersections, so it's a quasi-filter in $\mathcal{A}$.[18]

Now, a quasi-ideal is *prime* iff whenever it contains a meet $A \cap B$, it also contains either A or B (or both), and a quasi-filter is *prime* iff whenever it contains a join $A \cup B$, it also contains either A

---

[16] De Morgan negations in addition support Nor, D-Nor, and both double negation consecutions. In Urquhart (1979), Urquhart explores the algebraic behavior of 'dual homomorphic operators'. These are closer to the negation of this section, but DHOs are required to satisfy Nor and D-Nor, and so are Ockham (see chapter 4); the present negation is not.

[17] A quasi-ideal in a lattice is a subset of the lattice that's closed downward and closed under meets. An ideal is a non-empty quasi-ideal; nothing here requires $N(x)$ to be non-empty.

[18] A quasi-filter in a lattice is a subset of the lattice that's closed upwards and closed under joins. A filter is a non-empty quasi-filter; nothing here requires $Prop(W_c) - N(x)$ to be non-empty.

or B or both. It happens that a quasi-ideal in a lattice is prime iff its complement is a quasi-filter in the lattice, and vice versa;[19] so the combination of our three requirements—subset-closure, anti-additivity, and anti-multiplicativity—is equivalent to a single requirement: $N(x)$ must be a prime quasi-ideal on $Prop(W)$ for all x. It's this fact that will be key for finding a canonical frame that meets the condition.

We combine our techniques from before. We start with $N_{min}(x) = \{|A| : -A \in x\}$ and $N_{max}(x) = \{X : X \neq |A|$ for any A such that $-A \notin x\}$. We close $N_{min}(x)$ under subsets to get $I(x)$, and close $N_{max}(x)$'s complement under supersets to get $F(x)$. That is, let $I(x) = \{X : X \subseteq Y$ for some $Y \in N_{min}(x)\}$, and $F(x) = \{X : X \supseteq Z$ for some $Z \notin N_{max}(x)\}$. $(X, Y, Z \in Prop(W_c).)$ So far, this is very like what we did for each of these conditions individually.[20] We're halfway home; $I(x)$ is a quasi-ideal on $\mathcal{A}$, and $F(x)$ is a quasi-filter.

Fact 5.2.17. $I(x)$ and $F(x)$ don't overlap.

Proof. Assume $I(x) \cap F(x) \neq \emptyset$, for some prime theory x. Then there is a $Z \notin N_{max}(x)$ and a $Y \in N_{min}(x)$ such that $Z \subseteq Y$. Since $Z \notin N_{max}(x)$, $Z = |A|$ for some A such that $-A \notin x$, and since $Y \in N_{min}(x)$, $Y = |B|$ for some B such that $-B \in x$. Since $|A|_{W_c} \subseteq |B|_{W_c}$, $A \vdash B$, and so $-B \vdash -A$. But then, since $-B \in x$ and $-A \notin x$, x is not a theory. Contradiction. $\square$

Nonetheless, $I(x)$ and $F(x)$ still may not be each other's complement, so there's another step needed. We use Zorn's lemma to guarantee the existence of our target $N_c(x)$. Consider the set $Q(x)$ of quasi-ideals on $\mathcal{A}$ that are disjoint from $F(x)$. Every chain in this set has an upper bound in $Q(x)$ (the chain's union), so $Q(x)$ itself has a maximal element $M(x)$; it's quick to show that $M(x)$ is not only a quasi-ideal disjoint from $F(x)$, but that it's also prime.[21] $M(x)$ serves as the value of our canonical neighborhood function at x; since it's a prime quasi-ideal on $\mathcal{A}$, the canonical frame based on it meets our condition; and since $N_{min}(x) \subseteq M(x) \subseteq N_{max}(x)$, $M(x)$ is indeed a canonical neighborhood function.

Deriving compatibility/exhaustiveness semantics

Under what conditions do these neighborhood semantics reduce to the more familiar (and less general) compatibility or exhaustiveness semantics explored in eg (Dunn & Zhou, 2005)?

---

[19]Proof left as exercise.

[20]Note that $I(x)$ corresponds to the $N_c(x)$ we used in the completeness proof for DLL + SMC + A$\vee$, while $F(x)$ corresponds to the *complement* of the $N_c(x)$ used for DLL + SMC + A$\wedge$.

[21]Proof of this is similar to the proof of Lemma 5.2.2.

Recall:[22]

Definition 5.2.18. *A* compatibility frame *is a tuple $F = \langle W, \leq, C \rangle$ such that:*

- *$\langle W, \leq \rangle$ is a DLL-frame, and*

- *C is a binary relation on W such that if xCy, $x' \leq x$, and $y' \leq y$, then $x'Cy'$*

Definition 5.2.19. *A* compatibility model *is a tuple $M = \langle W, \leq, C, [\![\ ]\!] \rangle$ such that:*

- *$\langle W, \leq, C \rangle$ is a compatibility frame, and*

- *$[\![\ ]\!]$ is the usual sort of denotation function, with the following clause for $-$:*

  ○ *$[\![-A]\!] = \{x : \{y : xCy\} \cap [\![A]\!] = \emptyset\}$*

Definition 5.2.20. *An* exhaustiveness frame *is a tuple $F = \langle W, \leq, E \rangle$ such that:*

- *$\langle W, \leq \rangle$ is a DLL-frame, and*

- *E is a binary relation on W such that if xEy, $x \leq x'$, and $y \leq y'$, then $x'Ey'$*

Definition 5.2.21. *An* exhaustiveness model *is a tuple $M = \langle W, \leq, E, [\![\ ]\!] \rangle$ such that:*

- *$\langle W, \leq, E \rangle$ is an exhaustiveness frame, and*

- *$[\![\ ]\!]$ is the usual sort of denotation function, with the following clause for $-$:*

  ○ *$[\![-A]\!] = \{x : \{y : xEy\} \not\subseteq [\![A]\!]\}$*

For this section, we'll need a notion of *equivalence* of models on a language $\mathcal{L}$:

Definition 5.2.22. *Two models $M = \langle W, \leq, R, [\![\ ]\!] \rangle$ and $M' = \langle W', \leq', R', [\![\ ]\!]' \rangle$ are* equivalent *on $\mathcal{L}$ iff:*

- *$W = W'$;*

- *$\leq\ =\ \leq'$; and*

- *for every $A \in \mathcal{L}$, $[\![A]\!] = [\![A]\!]'$*

---

[22]This presentation is the same as that in (Dunn & Zhou, 2005), and so differs slightly from that in chapter 4; the differences are purely notational.

Note that no explicit use is made of R or R' in this definition, nor is any restriction made to neighborhood, compatibility, or exhaustiveness models. R can be a neighborhood function, or a compatibility or exhaustiveness relation, and R' can be any of these as well; they do not have to match.

We also need a notion of frame-equivalence:

Definition 5.2.23. *Two frames $F = \langle W, \leq, R \rangle$ and $F' = \langle W', \leq', R' \rangle$ are* equivalent *on $\mathcal{L}$ iff:*

- *$W = W'$*

- *$\leq = \leq'$*

- *for any models $M = \langle F, [\![\ ]\!] \rangle$ and $M' = \langle F', [\![\ ]\!]' \rangle$, if $[\![\ ]\!]$ and $[\![\ ]\!]'$ agree on all atoms, then M is equivalent to M'.*

Now, as we saw in chapter 4, the weakest negation that can be captured with the compatibility semantics is that of DLL + SMC + Nor + A∨, and the weakest that can be captured with the exhaustiveness semantics is that of DLL + SMC + D-Nor + A∧. This might lead us to expect every frame meeting fSMC, fNor, and fA∨ to be equivalent to some compatibility frame, and *mutatis mutandis* for exhaustiveness.

This is on the right track, but it's not quite the case. The trick comes with fA∨ and fA∧. Let's look at fA∨ first; the points are precisely parallel. To prove our correspondence results, it was important that fA∨ required $N(x)$ at each x to be closed under *finite* unions only, not arbitrary unions. That is, there are neighborhood frames on which $N(x)$ at each x is closed under finite unions, but on which not all the $N(x)$ are closed under arbitrary unions; on these frames A∨ is still valid.

To define a compatibility frame equivalent to a given neighborhood frame, though, it's vital that $N(x)$ at each x be closed under *arbitrary* unions. I'll point it out when it comes up in the proof.

Fact 5.2.24. *Let F be any neighborhood frame $F = \langle W, \leq, N \rangle$ meeting fSMC, fNor, and such that $N(x)$ is closed under arbitrary unions for all $x \in W$. Define a binary relation C on W as follows: xCy iff for all $X \in N(x)$, $y \notin X$. $F' = \langle W, \leq, C \rangle$ is a compatibility frame equivalent to F.*

Proof. First we show that F' is a compatibility frame. It suffices to show that C is a compatibility relation. Suppose xCy, $x' \leq x$, and $y' \leq y$. Since $x' \leq x$, we know that $N(x') \subseteq N(x)$, so $x'Cy$. Now suppose $\neg x'Cy'$. Then there is an $X \in N(x')$ such that $y' \in X$. But since $X \in Prop(F)$, X is closed upwards under $\leq$, so $y \in X$. Thus, $\neg x'Cy$. Contradiction.

Second we show that F' is equivalent to F. Proof is by induction on length of formula; the only nontrivial case is the clause for $-$. It must be shown that $m_N(X) = \{x : \{y : xCy\} \cap X = \emptyset\}$; that is, that for $X \in Prop(F)$, $X \in N(x)$ iff $\{y : xCy\} \cap X = \emptyset$.

LTR: Suppose otherwise. Then $X \in N(x)$ and there is a $z \in X$ such that xCz. But then for all $Y \in N(x)$, $z \notin Y$. Contradiction.

RTL: Suppose $\{y : xCy\} \cap X = \emptyset$. Then either $X = \emptyset$ or not. If $X = \emptyset$, then $X \in N(x)$ by fNor. On the other hand, if $X \neq \emptyset$, then for each $z_i \in X$ there is some $Z_i \in N(x)$ such that $z_i \in Z_i$. Clearly $X \subseteq \bigcup \{Z_i\}$. Since $N(x)$ is closed under arbitrary unions, $\bigcup \{Z_i\} \in N(x)$, and since it meets fSMC, this means $X \in N(x)$. $\qquad \square$

A similar fact holds for exhaustiveness.

Fact 5.2.25. *Let F be any neighborhood frame $F = \langle W, \leq, N \rangle$ meeting fSMC, fD-Nor, and such that $Prop(F) - N(x)$ is closed under arbitrary intersections for all $x \in W$. Define a binary relation E on W as follows: xEy iff for all $X \in Prop(F) - N(x)$, $y \in X$. $F' = \langle W, \leq, E \rangle$ is an exhaustiveness frame equivalent to F.*

Proof. First we show that F' is an exhaustiveness frame. It suffices to show that E is an exhaustiveness relation. Suppose xEy, $x \leq x'$, and $y \leq y'$. Since $x \leq x'$, $N(x) \subseteq N(x')$. So $x'Ey$. Now suppose $\neg x'Ey'$. Then for some $X \in Prop(F) - N(x')$, $y' \notin X$. Since X must be closed upwards under $\leq$, $y \notin X$. So $\neg x'Ey$. Contradiction.

Second we show that F' is equivalent to F. Proof is by induction on formula length; again, the only nontrivial case is the clause for $-$. It suffices to show that for $X \in Prop(F)$, $X \in N(x)$ iff $\{y : xEy\} \nsubseteq X$.

LTR: Suppose otherwise. Then $X \in N(x)$, and if $X = W$, contradiction (with fD-Nor), so $X \neq W$. Now, for each $z_i \notin X$, there is a $Z_i \in Prop(F) - N(x)$ such that $z_i \notin Z_i$. Since everything not in X is also not in $\bigcap \{Z_i\}$, $\bigcap \{Z_i\} \subseteq X$. Since $Prop(F) - N(x)$ is closed under arbitrary intersections, $\bigcap \{Z_i\} \in Prop(F) - N(x)$. And since $N(x)$ meets fSMC, it must be that $X \notin N(x)$. Contradiction.

RTL: Suppose $\{y : xEy\} \nsubseteq X$. Then there is a z such that xEz but $z \notin X$. So for all $Y \in Prop(F) - N(x)$, $z \in Y$. Since $z \notin X$, and $X \in Prop(F)$, it must be that $X \in N(x)$. $\qquad \square$

5.3 Deriving the Routley star

This section has two parts. In §5.3.1, I develop the theory of SM-algebras (the algebras naturally generated by neighborhood frames meeting fSMC), and show how to build frames from the proper prime filters of these algebras.[23] These frames give rise to more algebras, into which the original SM-algebras can be embedded. In §5.3.2, I restrict my attention to a variety of SM-algebras, the Ockham lattices discussed in (Urquhart, 1979), and frames that meet the conditions fSMC, fNor, fD-Nor, f$A\vee$, and f$A\wedge$, which I call Ockham frames. I discuss the relation between these frames and these lattices, and use the techniques developed here to shed light on Routley star semantics for negation, showing why Routley star-defined negations have the logical properties they in fact have.

5.3.1 SM-algebras

Definition 5.3.1. *An* SM-algebra $\mathfrak{A} = \langle A, \wedge, \vee, \bot, \top, - \rangle$ *is a complete bounded distributive lattice with meet $\wedge$, join $\vee$, least element $\bot$, greatest element $\top$, and operation $-$ such that whenever $a \leq b$, $-b \leq -a$.*

Definition 5.3.2. *For any neighborhood frame $F = \langle W, \leq, N \rangle$ meeting fSMC, its N-algebra F^+ is the natural algebra of propositions on F; that is, $F^+ = \langle Prop(F), \cap, \cup, \emptyset, W, m_N \rangle$. (Recall $m_N(a) = \{x \in W : a \in N(x)\}$.)*

Fact 5.3.3. *For any neighborhood frame $F = \langle W, \leq, N \rangle$ meeting fSMC, its N-algebra F^+ is an SM-algebra.*

Proof. First, F^+ is a complete bounded distributive lattice. Second, for $a, b \in Prop(F)$, suppose $a \subseteq b$ and $x \in m_N(b)$; that is, $b \in N(x)$. Since F meets fSMC, $a \in N(x)$, thus $x \in m_N(a)$, and so $m_N(b) \subseteq m_N(a)$. $\square$

Now here's the plan. From any SM-algebra $\mathfrak{A}$, we define the class $\mathfrak{A}_+$ of its PPF-frames; we then go on to show that for any $F \in \mathfrak{A}_+$, $\mathfrak{A}$ can be embedded in F^+; this is the neighborhood analogue of the key result in (Jónsson & Tarski, 1951). To start this up, let's get a lemma and a few more definitions on the table.

Lemma 5.3.4 (Proper Prime Filter Lemma)**.** *Let $\mathfrak{A}$ be an SM-algebra, $a \in A$, and F a proper filter of $\mathfrak{A}$ such that $a \notin F$. Then there is a proper prime filter G of $\mathfrak{A}$ such that $F \subseteq G$ and $a \notin G$.*

---

[23]The technique is derived from the modal technique of ultrafilter extensions; see (Blackburn et al., 2001) for details.

Proof. Consider the set X of proper filters of $\mathfrak{A}$ that don't contain a. (X is nonempty, since at least $F \in X$.) X is closed under unions of chains, so by Zorn's Lemma X has a maximal element G. Since $G \in X$, G is a proper filter of $\mathfrak{A}$ and $a \notin G$, so it only remains to show that G is prime.

Suppose G is not prime. Then there are $b, c \in A$ such that $b \vee c \in G$, but $b \notin G$ and $c \notin G$. Since G is maximal in X, there must be $d \in G$ such that $b \wedge d \leq a$, and some $e \in G$ such that $c \wedge e \leq a$. So $d \wedge e \in G$, and since $b \vee c \in G$, this gets us $(d \wedge e) \wedge (b \vee c) \in G$, and so $(d \wedge e \wedge b) \vee (d \wedge e \wedge c) \in G$. But since $(d \wedge e \wedge b) \vee (d \wedge e \wedge c) \leq a$, it must be that $a \in G$, and so $G \notin X$. Contradiction. So G is prime. $\qquad\square$

Fact 5.3.5. *Given some $D \subseteq A$, there is a smallest filter $|D\rangle$ containing D: $|D\rangle = \{a \in A :$ there are $d_0, \ldots, d_n \in D$ such that $d_0 \wedge \ldots \wedge d_n \leq a\}$. If D has the* finite meet property *(if there is no finite subset $\{d_0, \ldots, d_n\}$ of D such that $d_0 \wedge \ldots d_n = \bot$), then $|D\rangle$ is proper.*

Definition 5.3.6. *Given an algebra $\mathfrak{A}$ let $PPF\mathfrak{A}$ be the set of $\mathfrak{A}$'s proper prime filters. For $a \in A$, let $r(a) = \{v \in PPF\mathfrak{A} : a \in v\}$.*

Now we're ready to define the class A_+ of $\mathfrak{A}$'s PPF-frames:

Definition 5.3.7. *Let $\mathfrak{A} = \langle A, \wedge, \vee, \bot, \top, - \rangle$ be an SM-algebra. A frame $F = \langle PPF\mathfrak{A}, \subseteq, N \rangle$ is a PPF-frame of $\mathfrak{A}$ (that is, $F \in \mathfrak{A}_+$) iff:*

- *$PPF\mathfrak{A}$ is the set of $\mathfrak{A}$'s proper prime filters;*

- *$\subseteq$ is the ordinary subset relation; and*

- *for $a \in A$, $u \in PPF\mathfrak{A}$, $r(a) \in N(u)$ iff $-a \in u$.*

NB: although $PPF\mathfrak{A}$ and $\subseteq$ are fully determined given an SM-algebra $\mathfrak{A}$, N is not. N's behavior is pinned down for $r(a)$ for all $a \in A$, but not otherwise. That is, for all members of $Prop(PPF\mathfrak{A})$ that aren't $r(a)$ for some a, it doesn't matter how N treats them; $F \in \mathfrak{A}_+$ all the same. (This should remind you of canonical neighborhood models.) Note that it's not in general even true that all these F meet fSMC.

Theorem 5.3.8. *Let $\mathfrak{A} = \langle A, \wedge, \vee, \bot, \top, - \rangle$ be an SM-algebra, and consider any $F = \langle PPF\mathfrak{A}, \subseteq, N \rangle \in \mathfrak{A}_+$. Then the function $r : A \to Prop(PPF\mathfrak{A})$ such that $r(a) = \{u \in PPF\mathfrak{A} : a \in u\}$ is an embedding of $\mathfrak{A}$ into F^+.*

Proof. There's a series of facts to be shown:

- Ad $r(a \wedge b) = r(a) \cap r(b)$:

$$r(a \wedge b) = \{u \in PPF\mathfrak{A} : a \wedge b \in u\}$$
$$= \{u \in PPF\mathfrak{A} : a \in u \text{ and } b \in u\}$$
$$= \{u \in PPF\mathfrak{A} : a \in u\} \cap \{u \in PPF\mathfrak{A} : b \in u\}$$
$$= r(a) \cap r(b)$$

- Ad $r(a \vee b) = r(a) \cup r(b)$:

$$r(a \vee b) = \{u \in PPF\mathfrak{A} : a \vee b \in u\}$$
$$= \{u \in PPF\mathfrak{A} : a \in u \text{ or } b \in u\}$$
$$= \{u \in PPF\mathfrak{A} : a \in u\} \cup \{u \in PPF\mathfrak{A} : b \in u\}$$
$$= r(a) \cup r(b)$$

- Ad $r(\bot) = \emptyset$:

$$r(\bot) = \{u \in PPF\mathfrak{A} : \bot \in u\}$$
$$= \emptyset$$

- Ad $r(\top) = PPF\mathfrak{A}$:

$$r(\top) = \{u \in PPF\mathfrak{A} : \top \in u\}$$
$$= PPF\mathfrak{A}$$

- Ad $r(-a) = m_N(r(a))$:

$$r(-a) = \{u \in PPF\mathfrak{A} : -a \in u\}$$
$$= \{u \in PPF\mathfrak{A} : r(a) \in N(u)\}$$
$$= m_N(r(a))$$

- Ad $a \leq b$ iff $r(a) \subseteq r(b)$: Relations between $\leq$, $\wedge$, and $\vee$, as well as $\subseteq$, $\cap$, and $\cup$, cover this one. It follows quickly that r is injective.

So r is an embedding. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

5.3.2 Ockham lattices

Routley star semantics for negation has proven tremendously fruitful in the development of relevant logics (see eg (Routley et al., 1982)). Without any special restrictions, a Routley star semantics results in a negation obeying SMC, Nor, D-Nor, A$\vee$, and A$\wedge$. This might lead one to think that any model on any Ockham frame can be given a Routley star semantics, but this is not so. Instead, given any model on any Ockham frame, one can use the frame to generate an algebra, use that algebra to generate a new frame, and define a model on that new frame which imitates the initial model, in the sense that for every state in the initial model, there is a state in the new model that verifies precisely the same sentences. This new model can then be given a Routley star semantics.

The trick amounts to this: when a frame is Ockham, its neighborhood function takes on a special sort of value: proper prime ideals on the frame's N-algebra. The complement of a proper prime ideal in an algebra is a proper prime filter, and proper prime filters are state-like things; they are closed upward in the lattice ordering (think logically closed), closed under meet (think adjunction), and prime (think prime). Any point x, then, determines a proper prime ideal $N(x)$, and so a proper prime filter, and so a point. The only problem is that the point in question might not be in the frame yet. So it stands to reason that if there were only enough states to do the job, any model on an Ockham frame could be given a Routley star semantics; the trouble only arises when there is no state in the frame corresponding to a particular proper prime filter. But by working with a frame's N-algebra, and that algebra's PPF-frames, I show here that for every Ockham frame, there is another Ockham frame whose states are the proper prime filters of the first frame's N-algebra. What's more, for every model on the first frame, there is a model on the second that imitates it,

and that imitating model can be given (equivalently) a neighborhood semantics or a Routley star semantics. Let's begin.

Definition 5.3.9. *An* Ockham lattice $\mathfrak{A} = \langle A, \wedge, \vee, \bot, \top, - \rangle$ *is a distributive lattice such that:*

- $-\bot = \top$;

- $-\top = \bot$;

- $-(a \vee b) = -a \wedge -b$;

- $-(a \wedge b) = -a \vee -b$; *and*

Fact 5.3.10. *In an Ockham lattice, if $a \leq b$, then $-b \leq -a$.*

Proof. If $a \leq b$, then $a \wedge b = a$ (by lattice properties), so $-(a \wedge b) = -a$. But $-(a \wedge b) = -a \vee -b$, so $-a = -a \vee -b$, and thus $-b \leq -a$ (by lattice properties). $\square$

Definition 5.3.11. *An* Ockham frame $F = \langle W, \leq, N \rangle$ *is a neighborhood frame that meets the conditions fSMC, fNor, fD-Nor, fA$\vee$, and fA$\wedge$.*

We proceed to show that for any Ockham lattice $\mathfrak{A}$, there is an Ockham frame $F \in \mathfrak{A}_+$. A lemma we'll use on the way:

Lemma 5.3.12 (Prime Ideal Lemma)**.** *Let $\mathfrak{A} = \langle A, \wedge, \vee, \bot, \top \rangle$ be a bounded distributive lattice, F be a filter on $\mathfrak{A}$, and I be an ideal on $\mathfrak{A}$, such that F and I are disjoint. Then there is a prime ideal J on $\mathfrak{A}$ such that $I \subseteq J$ and J is disjoint from F.*

Proof. Consider $X = \{Y : Y$ is an ideal of $\mathfrak{A}$ disjoint from $F\}$. X is closed under unions of chains, so X has a maximal element J. Clearly J is an ideal of $\mathfrak{A}$, and is disjoint from F. We proceed to show that J is prime.

Suppose otherwise. Then there are $b, c \in A$ such that $b \wedge c \in J$, $b \notin J$, and $c \notin J$. Since J is maximal in X, there must be some $d \in J$ such that $d \vee b \geq g$ for some $g \in F$, and there must be some $e \in J$ such that $e \vee c \geq h$ for some $h \in F$. So $d \vee e \in J$, and thus $(d \vee e) \vee (b \wedge c) \in J$. By distribution, $(d \vee e \vee b) \wedge (d \vee e \vee c) \in J$; but $(d \vee e \vee b) \wedge (d \vee e \vee c) \geq g \wedge h$, so $g \wedge h \in J$. But since F is a filter, and $g, h \in F$, $g \wedge h \in F$. So $J \notin X$. Contradiction. So J is prime. $\square$

Theorem 5.3.13. *For any Ockham lattice $\mathfrak{A} = \langle A, \wedge, \vee, \bot, \top, - \rangle$, there is an Ockham frame $^\star\mathfrak{A}_+ \in \mathfrak{A}_+$.*

Proof. First, we define $^\star\mathfrak{A}_+ = \langle PPF\mathfrak{A}, \subseteq, N \rangle$. $PPF\mathfrak{A}$ and $\subseteq$ are as expected. It'll take a few steps to define N. Recall that for $a \in A, r(a) = \{v \in PPF\mathfrak{A} : a \in v\}$. For any $u \in PPF\mathfrak{A}$, let

- $Min(u) = \{r(a) : -a \in u\}$

- $Max'(u) = \{r(a) : -a \notin u\}$

- $I(u) = \{X \in Prop(PPF\mathfrak{A}) : X \subseteq Y$ for some $Y \in Min(u)\}$

- $F(u) = \{X \in Prop(PPF\mathfrak{A}) : X \supseteq Y$ for some $Y \in Max'(u)\}$

We show that $I(u)$ is disjoint from $F(u)$. If not, there is some $X \in I(u) \cap F(u)$, so some $Y, Z \in Prop(PPF\mathfrak{A})$ such that $Y \subseteq X \subseteq Z$, $Z \in Min(u)$, $Y \in Max'(u)$. So $Z = r(a)$ for some a such that $-a \in u$, and $Y = r(b)$ for some b such that $-b \notin u$. Since $Y \subseteq Z$, $r(b) = \{v \in PPF\mathfrak{A} : b \in v\} \subseteq r(a) = \{v \in PPF\mathfrak{A} : a \in v\}$. It must then be that $b \leq a$ (by Lemma 5.3.4), so $-a \leq -b$ (since $\mathfrak{A}$ is an Ockham lattice). Thus, u isn't a filter. Contradiction. So $I(u)$ is disjoint from $F(u)$.

$I(u)$ is an ideal on $Prop(PPF\mathfrak{A})$, and $F(u)$ a filter.[24] So by Lemma 5.3.12, there is some prime ideal $N(u)$ on $Prop(PPF\mathfrak{A})$ such that $I(u) \subseteq N(u)$ and $N(u)$ is disjoint from $F(u)$. Since $N(u)$ is a prime ideal for every u, $^\star\mathfrak{A}_+$ is an Ockham frame.

It only remains to show that $^\star\mathfrak{A}_+ \in \mathfrak{A}_+$; for this, it suffices that $r(a) \in N(u)$ iff $-a \in u$. RTL: if $-a \in u$, then $r(a) \in Min(u)$. Since $Min(u) \subseteq I(u) \subseteq N(u)$, $r(a) \in N(u)$. LTR: if $-a \notin u$, then $r(a) \in Max'(u)$. Since $Max'(u) \subseteq F(u)$, $r(a) \in F(u)$. Since $F(u)$ is disjoint from $N(u)$, it follows that $r(a) \notin N(u)$. So $^\star\mathfrak{A}_+ \in \mathfrak{A}_+$. $\square$

Lemma 5.3.14. *For any Ockham frame* $F = \langle W, \leq, N \rangle$, *its N-algebra* $\mathfrak{A} = \langle Prop(W), \cap, \cup, \emptyset, W, m_N \rangle$ *is an Ockham lattice.*

Proof. We know from earlier results that $\mathfrak{A}$ is distributive. Since F meets condition fNor, $m_N(\emptyset) = W$. Since F meets condition fD-Nor, $m_N(W) = \emptyset$. Since F meets condition fA∨, $m_N(A \vee B) \supseteq m_N(A) \wedge m_N(B)$, and since it meets fSMC, the reverse inclusion holds as well. Since F meets condition fA∧, $m_N(A \wedge B) \subseteq m_N(A) \vee m_N(B)$, and since it meets fSMC, the reverse inclusion holds as well. So $\mathfrak{A}$ is an Ockham lattice. $\square$

Theorem 5.3.15. *Let* $F = \langle W, \leq, N_F \rangle$ *be any Ockham frame,* $\mathfrak{A} = \langle A, \cap, \cup, \emptyset, W, - \rangle$ *its N-algebra (an Ockham lattice), and* $^\star\mathfrak{A}_+ = \langle PPF\mathfrak{A}, \subseteq, N \rangle$ *the Ockham frame defined from* $\mathfrak{A}$ *as in Theorem 5.3.13. For any* $w \in W$, *let* $s(w) = \{a \in A : w \in a\}$. *Then for any model built on* F, *there is a model built on* $^\star\mathfrak{A}_+$ *such that for all* $w \in W$, w *and* $s(w)$ *satisfy precisely the same sentences.*

---

[24]Proof left as exercise.

Proof. • First, some properties of s:

    ○ s is a function from W into $PPF\mathfrak{A}$; this takes showing that for any $w \in W$, $s(w)$ is a prime proper filter of $\mathfrak{A}$. Since $w \in W$, $W \in s(w)$, and since $w \notin \emptyset$, $\emptyset \notin s(w)$. If $X \in s(w)$ and $Y \in s(w)$, then $x \in X$ and $x \in Y$. So $x \in X \cap Y$, and $X \cap Y \in s(w)$; thus $s(w)$ is closed under $\cap$. If $X \in s(w)$ and $X \subseteq Y$, then since $w \in X$, $w \in Y$, so $Y \in s(w)$; thus $s(w)$ is closed upwards along $\subseteq$. So $s(w)$ is a proper filter on $\mathfrak{A}$. Finally, $s(w)$ is prime: if $X \cup Y \in s(w)$, then $w \in X \cup Y$, so $w \in X$ or $w \in Y$, so either $X \in s(w)$ or $Y \in s(w)$.

    ○ s is injective, and in fact, it preserves order: for $w, x \in W$, $w \leq x$ iff $s(w) \subseteq s(x)$.

    LTR: Suppose $w \leq x$ and $a \in s(w)$. Since $a \in A$ and $A = Prop(W)$, a must be closed upwards along $\leq$; and since $w \in a$, it must be that $x \in a$. So $a \in s(x)$.

    RTL: Suppose $s(w) \subseteq s(x)$ but $w \not\leq x$. So for all $a \in A$, if $w \in a$, then $x \in a$. Consider $b = \{y \in W : w \leq y\}$. Since b is closed upwards along $\leq$, $b \in A$. But $w \in b$, and $x \notin b$. Contradiction. So $w \leq x$.

• Second, we need to show that, for all $a \in A$, $m_N(r(a)) = r(-a)$. That is, we must show that for all $v \in PPF\mathfrak{A}$, $r(a) \in N(v)$ iff $-a \in v$.[25]

LTR: Suppose $-a \notin v$. Then $r(a) \in F(v)$, and since $F(v)$ is disjoint from $N(v)$, $r(a) \notin N(v)$.

RTL: Suppose $-a \in v$. Then $r(a) \in I(v)$, and since $I(v) \subseteq N(v)$, $r(a) \in N(v)$.

• Third, we come to models. Take an arbitrary assignment $[\![\ ]\!]_F$ of members of A to propositional variables; this is an initial valuation on F. Define $[\![p]\!]_{PPF} = r([\![p]\!]_F)$, for all atomic p. We show by induction that, for all sentences $A \in \mathcal{L}$, $[\![A]\!]_{PPF} = r([\![A]\!]_F)$. The base case—propositional variables—is given.

    ○ For $\wedge$: $[\![A \wedge B]\!]_{PPF} = [\![A]\!]_{PPF} \cap [\![B]\!]_{PPF}$ by the usual modeling conditions. This in turn is equal to $r([\![A]\!]_F) \cap r([\![B]\!]_F)$ by the inductive hypothesis, which is equal to $r([\![A]\!]_F \cap [\![B]\!]_F)$ by Theorem 5.3.8. This, of course, is $r([\![A \wedge B]\!]_F)$.

    ○ For $\vee$: Precisely parallel to $\wedge$.

    ○ For $-$: $[\![- A]\!]_{PPF} = m_N([\![A]\!]_{PPF})$ by the usual modeling conditions. This is equal to $m_N(r([\![A]\!]_F))$ by the inductive hypothesis, which in turn is equal to $r(-[\![A]\!]_F)$ by the second main part of this proof. That's equal to $r([\![- A]\!]_F)$ by modeling conditions.

---

[25]It will help to recall from the proof of Theorem 5.3.13: for any $u \in PPF\mathfrak{A}$, $Min(u) = \{r(b) : -b \in u\}$, $Max'(u) = \{r(b) : -b \notin u\}$, $I(u) = \{X \in Prop(PPF\mathfrak{A}) : X \subseteq Y$ for some $Y \in Min(u)\}$, and $F(u) = \{X \in Prop(PPF\mathfrak{A}) : X \supseteq Y$ for some $Y \in Max'(u)\}$. Thus, $I(u) = \{X \in Prop(PPF\mathfrak{A}) : X \subseteq r(b)$ for some b such that $-b \in u\}$, and $F(u) = \{X \in Prop(PPF\mathfrak{A}) : X \supseteq r(b)$ for some b such that $-b \notin u\}$.

- Finally, we show that $w \in a$ iff $s(w) \in r(a)$, and so that $w \in \llbracket A \rrbracket_F$ iff $s(w) \in \llbracket A \rrbracket_{PPF}$. Recall that $s(w) = \{b \in A : w \in b\}$, and $r(a) = \{v \in PPF\mathfrak{A} : a \in v\}$.

  LTR: Suppose $w \in a$. Then $a \in s(w)$, and so $s(w) \in r(a)$.

  RTL: Suppose $w \notin a$. Then $a \notin s(w)$, and so $s(w) \notin r(a)$.

Since $w \in \llbracket A \rrbracket_F$ iff $s(w) \in \llbracket A \rrbracket_{PPF}$, and $\llbracket \ \rrbracket_F$ was arbitrary, the theorem is proved. $\qquad\square$

We can give this model a semantics using the Routley star. In Routley star semantics for negation, we define a unary operation * on states, and give the compositional clause for negation as follows: $\llbracket - A \rrbracket = \{w : w^\star \notin \llbracket A \rrbracket\}$.

Lemma 5.3.16. *If $\mathfrak{A} = \langle A, \wedge, \vee, \bot, \top, - \rangle$ is an Ockham lattice, then for $v \in PPF\mathfrak{A}$, let $v^\star = \{b \in A : -b \notin v\}$. Then $v^\star \in PPF\mathfrak{A}$.*

Proof. We show that $v^\star$ meets the conditions to be a prime proper filter on $\mathfrak{A}$:

- Since $\mathfrak{A}$ is Ockham, $- \top \notin v$, so $\top \in v^\star$.

- Since $\mathfrak{A}$ is Ockham, $- \bot \in v$, so $\bot \notin v^\star$.

- If $- a \notin v$ and $- b \notin v$, then $- a \vee - b \notin v$. Since $\mathfrak{A}$ is Ockham, $-(a \wedge b) \notin v$. So if $a, b \in v^\star$, $a \wedge b \in v^\star$.

- If $- a \notin v$ and $a \leq b$, then $- b \notin v$, since $- b \leq - a$. So if $a \in v^\star$ and $a \leq b$, then $b \in v^\star$.

- If $-(a \vee b) \notin v$, then $- a \notin v$ or $- b \notin v$ (since $\mathfrak{A}$ is Ockham). So if $a \vee b \in v^\star$, then $a \in v^\star$ or $b \in v^\star$.

$\qquad\square$

Corollary 5.3.17. *For any Ockham frame $F = \langle W, \leq, N_F \rangle$, consider its N-algebra $\mathfrak{A} = \langle A, \cap, \cup, \emptyset, W, - \rangle$, and the Ockham frame $^\star\mathfrak{A}_+ = \langle PPF\mathfrak{A}, \subseteq, N \rangle$. Take $V^\star$ to be defined on $PPF\mathfrak{A}$ as in Lemma 5.3.16. For any model on F, there is a model on $^\star\mathfrak{A}_+$ that ignores the neighborhood function N and instead uses the Routley star semantics for negation, such that for all $w \in W$, w and $s(w)$ satisfy precisely the same sentences.*

Proof. Proof is just as in Theorem 5.3.15, with the following modifications:

- Define $m_\star(X) : Prop(PPF\mathfrak{A}) \to Prop(PPF\mathfrak{A})$ as $m_\star(X) = \{v \in PPF\mathfrak{A} : v^\star \notin X\}$. We show that, for all $a \in A$, $m_\star(r(a)) = r(-a)$. That is, we show that for all $v \in PPF\mathfrak{A}$, $a \notin v^\star$ iff $- a \in v$. This is immediate from the definition of *.

101

- In the induction at the end of the proof, the negation clause must now read: $\llbracket A \rrbracket_{PPF} = m_\star(\llbracket A \rrbracket_{PPF})$ by the Routley star modeling conditions. This is equal to $m_\star(r(\llbracket A \rrbracket_F))$ by the inductive hypothesis, which is in turn equal to $r(-\llbracket A \rrbracket_F)$ by the last bullet point. That's equal to $r(\llbracket A \rrbracket_F)$ by modeling conditions.

The rest of the proof goes through unchanged.

$\square$

To sum up: not every Ockham frame F has a Routley star semantics. But every Ockham frame F determines a unique Ockham frame $^\star F^+{}_+$, and every model on F can be imitated by a Routley star model on $^\star F^+{}_+$, in the sense that every state in F has a corresponding state in $^\star F^+{}_+$ that verifies precisely the same sentences.

Along the way, we've developed tools for building algebras out of frames and frames out of algebras. These tools correspond closely to tools (such as ultrafilter extensions) developed for modal logics in eg (Blackburn et al., 2001), but they are peculiar to neighborhood semantics. Just as these tools have proven fruitful for modal explorations of all stripes, it is my hope that they will prove fruitful for further exploration of neighborhood models.

5.4   Summing up: a new kite of negations

Dunn (1993) gives a 'kite of negations': a Hasse diagram of several logics of negation and the inclusions among them, all capturable in the compatibility semantics of §5.2.4. Dunn & Zhou (2005) fill in the dual side of this kite, producing a diagram that gives the inclusions among negations capturable in either the compatibility or exhaustiveness semantics of §5.2.4. In figure 5.1, I give yet another kite (which doesn't look so much like a kite anymore), displaying inclusion relations among the negations in Dunn & Zhou's extended kite, plus the additional negations occurring in this chapter, those that cannot be captured by either the compatibility or the exhaustiveness semantics. The soundness and completeness results above allow for easy proofs of non-containment.

The compatibility negations here are: CL (classical), K3 (strong Kleene), LP (Priest's (1979) Logic of Paradox), Int (intuitionist), Min (Johansson's minimal), Quas (Dunn & Zhou's quasiminimal), DM (De Morgan), Pre (Dunn & Zhou's preminimal), and Ock (Ockham—the negation of Ockham lattices). The exhaustiveness negations are CL, K3, LP, D Int (dual intuitionist), D Min (dual minimal), D Quas (dual quasiminimal), DM, D Pre (dual preminimal), and Ock. The remaining negations are named after their characteristic inferences—in all cases except AA+AM, these
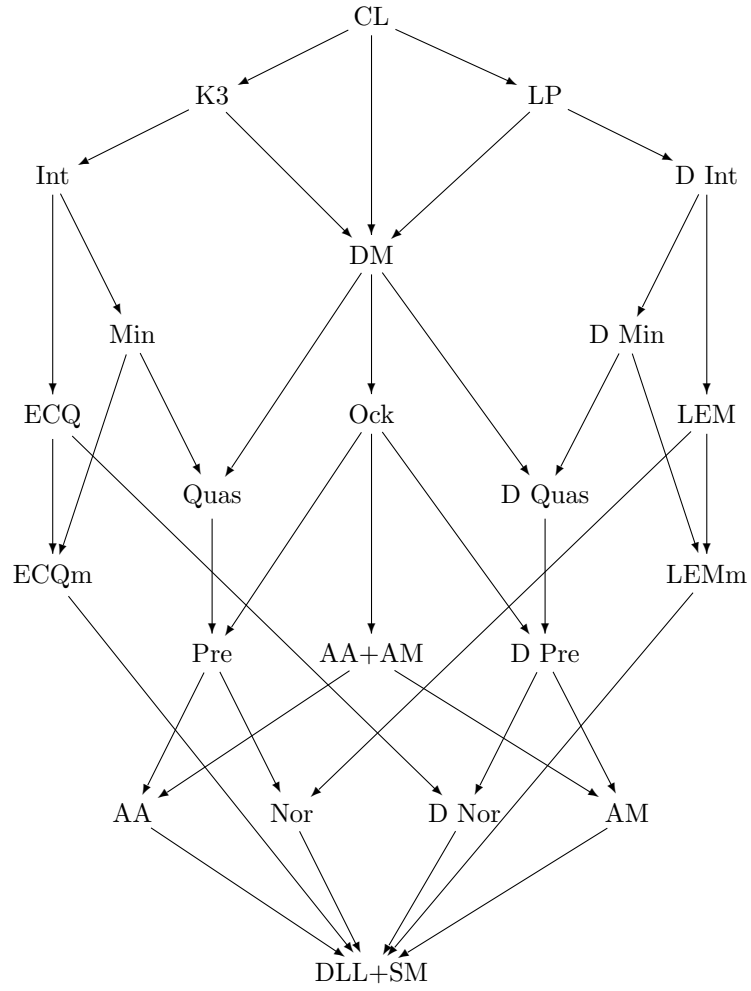
Figure 5.1: A new kite of negations

logics are obtained by adding a single consecution to DLL+SM. All the shown containments are proper; the neighborhood semantics in this chapter provides easy proofs of non-containment.

Chapter 6

Sorting out the sorites

Supervaluational theories of vagueness have achieved considerable popularity in the past decades, as seen in eg (Fine, 1975), (Lewis, 1970). This popularity is only natural; supervaluations let us retain much of the power and simplicity of classical logic, while avoiding the commitment to strict bivalence that strikes many as implausible.

Like many nonclassical logics, the supervaluationist system SP has a natural dual, the subvaluationist system SB, explored in eg (Hyde, 1997), (Varzi, 2000).[1] As is usual for such dual systems, the classical features of SP (typically viewed as benefits) appear in SB in 'mirror-image' form, and the nonclassical features of SP (typically viewed as costs) also appear in SB in 'mirror-image' form. Given this circumstance, it can be difficult to decide which of two dual systems is better suited for an approach to vagueness.[2]

The present paper starts from a consideration of these two approaches—the supervaluational and the subvaluational—and argues that neither of them is well-positioned to give a sensible logic for vague language. §6.1 presents the systems SP and SB and argues against their usefulness. Even if we suppose that the general picture of vague language they embody is accurate, we ought not arrive at systems like SP and SB. Instead, such a picture should lead us to truth-functional systems like strong Kleene logic (K_3) or its dual LP. §6.2 presents these systems, and argues that supervaluationist and subvaluationist understandings of language are better captured there; in particular, that a dialetheic approach to vagueness based on the logic LP is a more sensible approach. §6.3 goes on to consider

---

[1] Although there are many different ways of presenting a supervaluational system, I'll ignore these distinctions here; my remarks should be general enough to apply to them all, or at least all that adopt the so-called 'global' account of consequence. (For discussion, see (Varzi, 2007).) Similarly for subvaluational systems.

[2] The situation is similar for approaches to the Liar paradox; for discussion, see eg (Beall & Ripley, 2004), (Parsons, 1984).

the phenomenon of higher-order vagueness within an LP-based approach, and §6.4 closes with a consideration of the sorites argument itself.

6.1 S'valuations

Subvaluationists and supervaluationists offer identical pictures about how vague language works; they differ solely in their theory of truth. Because their overall theories are so similar, this section will often ignore the distinction between the two; when that's happening, I'll refer to them all as s'valuationists. §6.1.1 presents the shared portion of the s'valuationist view, while §6.1.2 goes on to lay out the difference between subvaluational and supervaluational theories of truth, and offers some criticism of both the subvaluationist and supervaluationist approaches.

6.1.1 The shared picture

It's difficult to suppose that there really is a single last noonish second, or a single oldest child, &c.[3] Nonetheless, a classical picture of negation seems to commit us to just that. After all, for each second, the law of excluded middle, $A \vee \neg A$, tells us it's either noonish or not noonish, and the law of non-contradiction, $\neg(A \wedge \neg A)$, tells us it's not both. Now, let's start at noon (since noon is clearly noonish) and move forward second by second. For a time, the seconds are all noonish, but the classical picture seems to commit us to there being a second—just one second!—that tips the scale over to non-noonishness.

Many have found it implausible to think that our language is pinned down that precisely, and some of those who have found this implausible (eg Fine (1975)) have found refuge in an s'valuational picture. The key idea is this: we keep that sharp borderline (classicality, as noted above, seems to require it), but we allow that there are many different places it might be. The s'valuationists then take vagueness to be something like ambiguity: there are many precise extensions that a vague predicate might have, and (in some sense) it has all of them.[4] It's important that this 'might' not be interpreted epistemically; the idea is not that one of these extensions is the one, and we just don't know which one it is. Rather, the idea is that each potential extension is part of the meaning of the vague predicate. Call these potential extensions 'admissible precisifications'.

The phenomenon of vagueness involves a three-part structure *somehow*; on this just about all

---

[3]Although not un-supposable; see eg (Sorensen, 2001), (Williamson, 1994) for able defenses of such a position.

[4]NB: S'valuationists differ in the extent to which they take vagueness to be like ambiguity, but they all take it to be like ambiguity in at least this minimal sense.

theorists are agreed. Examples: for a vague predicate F, epistemicists (eg Williamson (1994)) consider 1) things that are known to be F, 2) things that are known not to be F, and 3) things not known either to be F or not to be F; while standard fuzzy theorists (eg Machina (1972), Smith (2008)) consider 1) things that are absolutely F, or F to degree 1, 2) things that are absolutely not F, or F to degree 0, and 3) things that are neither absolutely F nor absolutely not F. S'valuationists also acknowledge this three-part structure: they talk of 1) things that are F on every admissible precisification, 2) things that are not-F on every admissible precisification, and 3) things that are F on some admissible precisifications and not-F on others.

Consider 'noonish'. It has many different precisifications, but there are some precisifications that are admissible and others that aren't. (100a)–(101b) give four sample precisifications; (100a) and (100b) are admissible precisifications for 'noonish', but (101a) and (101b) aren't:[5]

(100)   a. $\{x : x$ is between 11:40 and 12:20$\}$

        b. $\{x : x$ is between 11:45 and 12:30$\}$

(101)   a. $\{x : x$ is between 4:00 and 4:30$\}$

        b. $\{x : x$ is between 11:40 and 11:44, or x is between 12:06 and 12:10, or x is between 12:17 and 12:22$\}$

There are at least a couple ways, then, for a precisification to go wrong, to be unadmissible. Like (101a), it might simply be too far from where the vague range is; or like (101b), it might fail to respect what are called *penumbral connections*. 12:22 might not be in every admissible precisification of 'noonish', but if it's in a certain admissible precisification, then 12:13 ought to be in that precisification too. After all, 12:13 is more noonish than 12:22 is. Since this is a connection within the penumbra of a single vague predicate, we can follow Fine (1975) in calling it an *internal* penumbral connection.

Admissible precisifications also must respect *external* penumbral connections. The key idea here is that the extensions of vague predicates sometimes depend on each other. Consider the borderline between green and blue. It's sometimes claimed that something's being green rules out its also being blue. If this is so, then no admissible precisification will count a thing as both green and blue, even if some admissible precisifications count it as green and others count it as blue. In order to handle this

---

[5]At least in most normal contexts. Vague predicates seem particularly context-sensitive, although they are not the only predicates that have been claimed to be (see eg (Recanati, 2004), (Wilson & Sperber, 2002)). For the purposes of this paper, I'll assume a single fixed (non-wacky) context; these are theories about what happens *within* that context. Some philosophers (eg Raffman (1994)) have held that taking proper account of context is itself sufficient to dissolve the problems around vagueness, but that sort of view faces its own difficulties (see eg (Stanley, 2003)).

phenomenon in general, it's crucial that we count precisifications not as precisifying one predicate at a time, but instead as precisifying multiple predicates simultaneously. That way, they can give us the requisite sensitivity to penumbral connections.

S'valuational models

Now that we've got the core of the idea down, let's see how it can be formally modeled.[6] An SV model M is a tuple $\langle D, I, P \rangle$ such that D, the domain, is a set of objects; I is a function from terms in the language to members of D; and P is a set of precisifications: functions from predicates to subsets of D.[7]

We then extend each precisification $p \in P$ to a valuation of the full language. For an atomic sentence Fa, a precisification $p \in P$ assigns Fa the value 1 if $I(a) \in p(F)$, 0 otherwise. This valuation of the atomics is extended to a valuation of the full language (in $\land$, $\lor$, $\neg$, $\forall$, $\exists$) in the familiar classical way. In other words, each precisification is a full classical valuation of the language, using truth values from the set $V_0 = \{1, 0\}$.

Then the model M assigns a value to a sentence simply by collecting into a set the values assigned to the sentence by M's precisifications: $M(A) = \{v \in V_0 : \exists p \in P(p(A) = v)\}$. Thus, models assign values to sentences from the set $V_1 = \{\{1\}, \{0\}, \{1, 0\}\}$. Note that so far, these values are uninterpreted; they work merely as a record of a sentence's values across precisifications. $M(A) = \{1\}$ iff A gets value 1 on every precisification in M; $M(A) = \{0\}$ iff A gets value 0 on every precisification in M; and $M(A) = \{1, 0\}$ iff A gets value 1 on some precisifications in M and 0 on others.

6.1.2 Differences in interpretation and consequence

The s'valuationists agree on the interpretation of two of these values: $\{1\}$ and $\{0\}$. If $M(A) = \{1\}$, then A is true on M. If $M(A) = \{0\}$, then A is false on M. But the subvaluationist and the supervaluationist differ over the interpretation of the third value: $\{1,0\}$.

The supervaluationist claims that a sentence must be true on *every* precisification in order to be true *simpliciter*, and false on *every* precisification in order to be false *simpliciter*. Since the value $\{1,0\}$

---

[6]There are many ways to build s'valuational models. In particular, one might not want to have to *fully* precisify the language in order to assign truth-values to just a few sentences. Nonetheless, the approach to be presented here will display the logical behavior of s'valuational approaches, and it's pretty simple to boot. So we can get the picture from this simple approach.

[7]And from propositional variables directly to classical truth-values, if one wants bare propositional variables in the language. Vague propositional variables can be accommodated in this way as well as precise ones.

records a sentence's taking value 1 on some precisifications and 0 on others, when $M(A) = \{1, 0\}$, A is neither true nor false on M for the supervaluationist.

The subvaluationist, on the other hand, claims that a sentence has only to be true on *some* precisification to be true *simpliciter*, and false on *some* precisification in order to be false *simpliciter*. So, when $M(A) = \{1, 0\}$, the subvaluationist says that A is both true and false on M.

Both define consequence in the usual way (via truth-preservation):

(102) $\Gamma \vDash \Delta$ iff, for every model M, either δ is true on M for some $\delta \in \Delta$, or γ fails to be true on M for some $\gamma \in \Gamma$.[8]

Since the subvaluationist and the supervaluationist differ over which sentences are true on a given model (at least if the model assigns $\{1,0\}$ anywhere), this one definition results in two different consequence relations; call them $\vDash_{SB}$ (for the subvaluationist) and $\vDash_{SP}$ (for the supervaluationist).

$\vDash_{SB}$ and $\vDash_{SP}$

One striking feature of these consequence relations (much-advertised in the case of supervaluations, at least) is their considerable classicality. For example (where $\vDash_{CL}$ is the classical consequence relation):

(103) $\vDash_{CL} A$ iff $\vDash_{SB} A$ iff $\vDash_{SP} A$

(104) $A \vDash_{CL}$ iff $A \vDash_{SB}$ iff $A \vDash_{SP}$

(103) tells us that these three logics have all the same logical truths, and (104) tells us that they have all the same explosive sentences.[9] What's more, for any classically valid argument, there is a corresponding SB-valid or SP-valid argument:

(105) $A_1, \ldots, A_i \vDash_{CL} B_1, \ldots, B_j$ iff $A_1 \wedge \ldots \wedge A_i \vDash_{SB} B_1, \ldots, B_j$

(106) $A_1, \ldots, A_i \vDash_{CL} B_1, \ldots, B_j$ iff $A_1, \ldots, A_i \vDash_{SP} B_1 \vee \ldots \vee B_j$

Let's reason through these a bit.[10] Suppose $A_1, \ldots, A_i \vDash_{CL} B_1, \ldots, B_j$. Then, since every precisification is classical, every precisification p (in every model) that verifies all the As will also

---

[8]Note that this is a multiple-conclusion consequence relation. One can recover a single-conclusion consequence relation from this if one is so inclined, but for present purposes the symmetrical treatment will be more revealing. See eg (Restall, 2005b) for details, or (Hyde, 1997) for application to s'valuations. See also (Keefe, 2000) for arguments against using multiple-conclusion consequence, and (Hyde, 2006) for a response.

[9]*Explosive* sentences are sentences from which one can derive any conclusions at all, just as logical truths are sentences that can be derived from any premises at all. It's a bit sticky calling them 'logical falsehoods', as may be tempting, since some sentences (in SB at least) can be false without failing to be true. And I want to shy away from 'contradiction' here too, since I'm going to mean by that a sentence of the form $A \wedge \neg A$, and such a sentence will be explosive here but not in the eventual target system.

[10]This paragraph proves only the LTR directions, but both directions indeed hold; see (Hyde, 1997) for details.

verify one of the Bs. Consider the same argument subvaluationally; one might have all the premises true in some model (because each is true on some precisification or other), without having all the premises true in the *same* precisification; thus, there's no guarantee that any of the Bs will be true on any precisification at all. On the other hand, if one simply conjoins all the premises into one big premise, then if it's true in a model at all it guarantees the truth of all the As on (at least) a single precisification, and so one of the Bs must be true on that precisification, hence true in the model.

Similar reasoning applies in the supervaluational case. If all the As are true in a model, then they're all true on every precisification; nonetheless it might be that none of the Bs is true on every precisification; all the classical validity guarantees is that each precisification has some B or other true on it. But when we disjoin all the Bs into one big conclusion, that disjunction must be true on every precisification, so the argument is SP-valid.

Note that (105) and (106) guarantee that $\vDash_{\text{SB}}$ matches $\vDash_{\text{CL}}$ on single-premise arguments, and that $\vDash_{\text{SP}}$ matches $\vDash_{\text{CL}}$ on single-conclusion arguments. It is apparent that there is a close relationship between classical logic, subvaluational logic, and supervaluational logic. What's more, for every difference between SB and CL, there is a dual difference between SP and CL, and vice versa. This duality continues as we turn to the logical behavior of the connectives:

(107)　　a.　$A, B \nvDash_{\text{SB}} A \wedge B$

　　　　　b.　$A, \neg A \nvDash_{\text{SB}} A \wedge \neg A$

(108)　　a.　$A \vee B \nvDash_{\text{SP}} A, B$

　　　　　b.　$A \vee \neg A \nvDash_{\text{SP}} A, \neg A$

(107a) (and its instance (107b)) is dual to (108a) (and its instance (108b)). It is often remarked about supervaluations that it's odd to have a disjunction be true when neither of its disjuncts is, but this oddity can't be expressed via $\vDash_{\text{SP}}$ in a single-conclusion format.[11] Here, in a multiple-conclusion format, it becomes apparent that this oddity is an oddity in the supervaluational consequence relation, not just in its semantics. And of course, there is a parallel oddity involving conjunction for the subvaluationist.

Disjunction and conjunction can be seen as underwriting existential and universal quantification, respectively, so it is no surprise that the oddities continue w/r/t quantification. A sample:

(109)　　a.　$Fa, Fb, \forall x (x = a \vee x = b) \nvDash_{\text{SB}} \forall x (Fx)$

---

[11]This, essentially, is Tappenden (1993)'s 'objection from upper-case letters'. With multiple conclusions, there's no need for upper-case letters; the point can be made in any typeface you like.

(110) a. $\exists x(Fx), \forall x(x = a \lor x = b) \nvdash_{\text{SP}} Fa, Fb$

The cause is as before: in the SB case there is no guarantee that the premises are true on the same precisification, so they cannot interact to generate the conclusion; while in the SP case there is no guarantee that the same one of the conclusions is true on every precisification, so it may be that neither is true *simpliciter*. In the supervaluational case, consequences for quantification have often been noted;[12] but of course they have their duals for the subvaluationist.

What to say about borderline cases?

This formal picture gives rise to certain commitments about borderline cases. (I assume here that every theorist is committed to all and only those sentences they take to be true.) Assume that 12:23 is a borderline case of 'noonish'. The subvaluationist and the supervaluationist agree in their acceptance of (111a)–(111b), and in their rejection of (112a)–(112b):

(111) a. 12:23 is either noonish or not noonish.

       b. It's not the case that 12:23 is both noonish and not noonish.

(112) a. It's not the case that 12:23 is either noonish or not noonish.

       b. 12:23 is both noonish and not noonish.

On the other hand, they disagree about such sentences as (113a)–(113b):

(113) a. 12:23 is noonish.

       b. 12:23 is not noonish.

The subvaluationist accepts both of these sentences, despite her rejection of their conjunction, (112b). On the other hand, the supervaluationist rejects them both, despite her acceptance of their disjunction, (111a). So the odd behavior of conjunction and disjunction observed above isn't simply a theoretical possibility; these connectives misbehave every time there's a borderline case of any vague predicate.

    The biggest trouble for either the subvaluationist or the supervaluationist is presumably justifying their deviant consequence relations, especially their behavior around conjunction and universal quantification (for the subvaluationist) or disjunction and existential quantification (for the supervaluationist). At least prima facie, one would think that a conjunction is true iff both its conjuncts

---

[12]For example, consider the sentence 'There is a last noonish second'. It is true for the supervaluationist, but there is no second x such that 'x is the last noonish second' is true for the supervaluationist.

are, or that a disjunction is true iff one disjunct is, but the s'valuationists must claim that these appearances are deceiving.

The trouble is generated by the lack of truth-functional conjunction and disjunction in these frameworks. Consider the subvaluational case. If A is true, and B is true, we'd like to be able to say that $A \wedge B$ is true. In some cases we can, but in other cases we can't. The value of $A \wedge B$ depends upon more than just the value of A and the value of B; it also matters how those values are related to each other precisification to precisification. It's this extra dependence that allows s'valuational approaches to capture 'penumbral connections', as argued for in Fine (1975). Unfortunately, it gets in the way of sensible conjunctions and disjunctions.

6.2 LP and K$_3$

This trouble can be fixed as follows: we keep the s'valuationist picture for atomic sentences, but then use familiar truth-functional machinery to assign values to complex sentences. This will help us retain more familiar connectives, and allow us to compute the values of conjunctions and disjunctions without worrying about which particular conjuncts or disjuncts we use.

6.2.1 The shared picture

The informal picture, then, is as follows: to evaluate atomic sentences, we consider all the ways in which the vague predicates within them can be precisified. For compound sentences, we simply combine the values of atomic sentences in some sensible way. But what sensible way? Remember, we're going to end up with three possible values for our atomic sentences—{1}, {0}, and {1,0}— so we need sensible three-valued operations to interpret our connectives. Here are some minimal desiderata for conjunction, disjunction, and negation:

(114) Conjunction:

      a. $A \wedge B$ is true iff both A and B are true.

      b. $A \wedge B$ is false iff either A is false or B is false.

(115) Disjunction:

      a. $A \vee B$ is true iff either A is true or B is true.

   b. $A \vee B$ is false iff both A and B are false.

(116) Negation:

a. $\neg A$ is true iff A is false.

b. $\neg A$ is false iff A is true.

These desiderata alone rule out the s'valuationist options: as is pointed out in (Varzi, 2000), SB violates the RTL directions of (114a) and (115b), while SP violates the LTR directions of (114b) and (115a).

As before, we'll have two options for interpreting {1,0}: we can take it to be both true and false, like the subvaluationist, or we can take it to be neither true nor false, like the supervaluationist. Since the above desiderata are phrased in terms of truth and falsity, it might seem that we need to settle this question before we find appropriate operations to interpret our connectives. It turns out, however, that the same set of operations on values will satisfy the above desiderata whichever way we interpret {1,0}.

LP/K_3

These are the operations from either strong Kleene logic (which I'll call K_3; see eg (Kripke, 1975)) or Priest's Logic of Paradox (which I'll call LP; see eg (Priest, 1979)). Consider the following lattice of truth values:

$$\{1\}$$

$$\{1,0\}$$

$$\{0\}$$
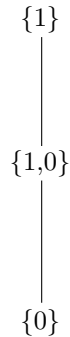
Take $\wedge$ to be greatest lower bound, $\vee$ to be least upper bound, and $\neg$ to reverse order (it takes {1} to {0}, {0} to {1}, and {1,0} to itself). Note that these operations satisfy (114a)–(116b); this is so whether {1,0} is interpreted as both true and false or as neither true nor false. For example, consider (115a). Suppose we interpret {1,0} LP-style, as both true and false. Then a disjunction is true (has value {1} or {1,0}) iff one of its disjuncts is: RTL holds because disjunction is an *upper* bound, and LTR holds because disjunction is *least* upper bound. On the other hand, suppose we interpret {1,0} K_3-style, as neither true nor false. Then a disjunction is true (has value {1}) iff one of its disjuncts is: again, RTL because disjunction is an *upper* bound and LTR because it's *least* upper bound. Similar reasoning establishes all of (114a)–(116b). So the LP/K_3 connectives meet our desiderata.

Differences in interpretation and consequence

There are still, then, two approaches being considered. One, the K_3 approach, interprets sentences that take the value {1,0} on a model as neither true nor false on that model. The other, the LP approach, interprets these sentences as both true and false on that model. This section will explore the consequences of such a difference.

As before, consequence for both approaches is defined as in (102) (repeated here as (117)):

(117) $\Gamma \vDash \Delta$ iff, for every model M, either δ is true on M for some $\delta \in \Delta$, or γ fails to be true on M for some $\gamma \in \Gamma$.

And as before, differences in interpretation of the value {1,0} result in differences about 'true', and so different consequence relations (written here as $\vDash_{K_3}$ and $\vDash_{LP}$).

First, we should ensure that the connectives behave appropriately, as indeed they do, in both K_3 and LP:

(118) a. $A, B \vDash_{LP} A \wedge B$

b. $A \vee B \vDash_{LP} A, B$

(119) a. $A, B \vDash_{K_3} A \wedge B$

b. $A \vee B \vDash_{K_3} A, B$

As you'd expect given this, so do universal and existential quantification:

(120) a. $Fa, Fb, \forall x(x = a \vee x = b) \vDash_{LP} \forall x(Fx)$

b. $\exists x(Fx), \forall x(x = a \vee x = b) \vDash_{LP} Fa, Fb$

(121) a. $Fa, Fb, \forall x(x = a \vee x = b) \vDash_{K_3} \forall x(Fx)$

b. $\exists x(Fx), \forall x(x = a \vee x = b) \vDash_{K_3} Fa, Fb$

Both consequence relations have other affinities with classical consequence, although neither is fully classical:

(122) a. $\vDash_{CL} A$ iff $\vDash_{LP} A$

b. $A \vDash_{CL}$ iff $A \vDash_{K_3}$

(123) a. $A, \neg A \nvDash_{LP} B$

b. $\neg A, A \vee B \nvDash_{LP} B$

(124) a. $A \nvDash_{K_3} B, \neg B$

b. $A \nvDash_{K_3} A \wedge B, \neg B$

(122a) tells us that LP and classical logic have all the same logical truths, while (122b) tells us that K_3 and classical logic have all the same explosive sentences. (123) shows us some of the nonclassical features of $\vDash_{LP}$; note that the failure of Explosion in (123a) does not come about in the same way as in SB (by failing adjunction), since adjunction is valid in LP, as recorded in (118a). (123b) points out the much-remarked failure of Disjunctive Syllogism in LP. Dual to these nonclassicalities are the non-classicalities of K_3 given in (124).

6.2.2 Vagueness and Ambiguity

As we've seen, one clear reason to think that LP and K_3 are better logics of vagueness than SB and SP is the sheer sensibleness of their conjunction and disjunction, which SB and SP lacked. LP and K_3 thus allow us to give an s'valuation-flavored picture of vague *predication* that doesn't interfere with a more standard picture of *connectives*. But there's another reason why at least some s'valuationists should prefer the truth-functional approach recommended here, having to do with ambiguity.

As we've seen, the s'valuational picture alleges at least some similarities between vagueness and ambiguity: at a bare minimum, they both involve a one-many relation between a word and its potential extensions. Some s'valuationists (eg Keefe (2000)) stop there, but others (eg Fine (1975) in places, Lewis (1982)) go farther, claiming that vagueness is actually a species of ambiguity. For these authors, there is an additional question worth facing: what's ambiguity like?

Non-uniform disambiguation

Here's one key feature of ambiguity: when an ambiguous word occurs twice in the same sentence, it can be disambiguated in different ways across its occurrences. For example, consider the word 'plant', which is ambiguous between (at least) *vegetation* and *factory*. Now, consider the sentence (125):

(125)   Jimmy ate a plant, but he didn't eat a plant.

It's clear that (125) has a noncontradictory reading; in fact, it has two, assuming for the moment that 'plant' is only two-ways ambiguous. 'Plant' can take on a different disambiguation at each of its occurrences, even when those occurrences are in the same sentence. If this were not the case, if multiple occurrences of an ambiguous word had to be disambiguated uniformly within a sentence, then the standard method of resolving an apparent contradiction—by finding an ambiguity—couldn't work. But of course this method does work.

Now, suppose we wanted to build formal models for an ambiguous language. They had better take this fact into account. But SB and SP cannot—they precisify whole sentences at once, uniformly. Hence, SB and SP could not work as logics for ambiguous language.[13]

LP and K_3, on the other hand, do not have this bad result. They deal with each occurrence of an ambiguous predicate (each atomic sentence) separately, and combine them truth-functionally. Thus, they avoid the bad consequences faced by s'valuational pictures. In fact, it is LP that seems to be a superior logic of ambiguous language. Here's why: typically, for an ambiguous sentence to be true, it's not necessary that *every* disambiguation of it be true; it suffices that *some* disambiguation is.[14]

Since it's clear that LP and K_3 (and LP in particular) are better logics of ambiguity than SB and SP, those s'valuationists who take vagueness to be a species of ambiguity have additional reason to adopt LP and K_3.

Asynchronous precisification

For an s'valuationist who does not take vagueness to be a species of ambiguity, the above argument applies little direct pressure to use LP or K_3, but it raises an interesting issue dividing the truth-functional approaches from the s'valuational approaches: when multiple vague predicates occur in the same sentence, how do the various precisifications of one interact with the various precisifications of the other?

Take the simplest case, where a single vague predicate occurs twice in one sentence. What are the available precisifications of the whole sentence? Suppose a model with n precisifications. On the s'valuational pictures, there will be n precisifications for the whole sentence; while on a truth-functional picture there will be n^2; every possible combination of precisifications of the predicates is available. This can be seen in the LP/K_3 connectives; for example, where $\wedge_0$ is classical conjunction,

(126) $M(A \wedge B) = \{a \wedge_0 b : a \in M(A), b \in M(B)\}$

$M(A)$ and $M(B)$, recall, are sets of classical values. $M(A \wedge B)$ is then obtained by pulling a pair of classical values, one from $M(A)$ and one from $M(B)$, conjoining these values, and repeating for every possible combination, then collecting all the results into a set. In other words, on this picture, every precisification 'sees' every other precisification in a compound sentence formed with $\wedge$; multiple predicates are not precisified in lockstep. The same holds, *mutatis mutandis*, for $\vee$.

---

[13]*Pace* Fine (1975), who, in a footnote, proposes SP as a logic for ambiguous language. As noted above, this would make it impossible to explain how one resolves a contradiction by finding an ambiguity—a very bad result.

[14]Lewis (1982) argues for LP, in particular, as a logic of ambiguity, and mentions vagueness as one sort of ambiguity.

6.2.3 What to say about borderline cases?

So much for the logical machinery. What do these approaches say about borderline cases of a vague predicate? Suppose again that 12:23 is a borderline case of 'noonish'. Consider the following list of claims:

(127)   a. 12:23 is noonish.

       b. 12:23 is not noonish.

       c. 12:23 is both noonish and not noonish.

       d. 12:23 is neither noonish nor not noonish.

       e. It's not the case that 12:23 is both noonish and not noonish.

       f. 12:23 is either noonish or not noonish.

All of these are familiar things to claim about borderline cases, although a common aversion to contradictions among philosophers means that some of them, like (127c), are more likely to be heard outside the classroom than in it. All these claims receive the value {1,0} in a model that takes 12:23 to be a borderline case of noonish. The LP partisan, then, will hold all of these to be true, while the K_3 partisan will hold none of them to be true. Which interpretation of {1,0} is more plausible, then? If we are to avoid attributing massive error to ordinary speakers (and ourselves, a great deal of the time), the LP story is far superior. Accordingly, for the remainder of the paper I'll focus in on LP, setting K_3 aside.

6.3 Higher-order vagueness

Of course, there is some well-known trouble alleged to be waiting in the wings for any three-valued approach to vagueness: the phenomenon of higher-order vagueness. This section evaluates the LP approach's response to the phenomenon, focusing first on the case of 2nd-order vagueness, and then generalizing the response to take in higher orders as well.

6.3.1 2nd-order vagueness

So far, we've seen a plausible semantics for vague predicates that depends crucially on the notion of an 'admissible precisification'. A vague atomic sentence is true iff it's true on some admissible precisification, false iff false on some admissible precisification. But which precisifications are admissible? Consider 'noonish'. Is a precisification that draws the line at 12:01 admissible, or is it

too early? It has seemed to many theorists (as it seems to me) that 'admissible' in this use is itself vague. Thus, we've run into something of the form of a revenge paradox: theoretical machinery invoked to solve a puzzle works to solve the puzzle, but then the puzzle reappears at the level of the new theoretical machinery.[15]

It would be poor form to offer an account of the vagueness of 'admissible' that differs from the account offered of the vagueness of 'noonish'. After all, vagueness is vagueness, and similar problems demand similar solutions.[16] So let's see how the account offered above applies to this particular case of vagueness.

What do we do when a precisification is borderline admissible—that is, both admissible and not admissible? We consider various precisifications of 'admissible'. This will kick our models up a level, as it were. Models (call them level-2 models, in distinction from the level-1 models of §6.2) now determine not sets of precisifications, but sets of *sets of* precisifications. That is, a level-2 model is a tuple $\langle D, I, P_2 \rangle$, where D is again a domain of objects, I is again a function from terms in the language to members of D, and P_2 is a set whose members are sets of precisifications.

Every individual precisification p works as before; it still assigns each atomic sentence A a value $p(A)$ from the set $V_0 = \{1, 0\}$. Every set of precisifications assigns each atomic sentence a value as well: a set P_1 of precisifications assigns to an atomic A the value $P_1(A) = \bigcup_{p \in P_1}\{p(A)\}$. These values come from the set $V_1 = \wp(V_0) - \emptyset = \{\{1\}, \{0\}, \{1,0\}\}$. That is, sets of precisifications work just like level-1 models, as far as atomics are concerned; they simply collect into a set the values assigned by the individual precisifications. A level-2 model $M = \langle D, I, P_2 \rangle$ assigns to every atomic A a value $M(A) = \bigcup_{P_1 \in P_2}\{P_1(A)\}$. It simply collects into a set the values assigned to A by the sets of precisifications in P_2, so it assigns values from the 7-membered set $V_2 = \wp(V_1) - \emptyset = \{\{\{1\}\}, \{\{0\}\}, \{\{1,0\}\}, \{\{1\}, \{0\}\}, \{\{1\}, \{1,0\}\}, \{\{0\}, \{1,0\}\}, \{\{1\}, \{0\}, \{1,0\}\}\}$.

In applications to vagueness, presumably only five of these values will be needed. Not much hangs on this fact in itself, but it will better show the machinery of the theory if we take a moment to see why it's likely to be so. Let's look at how level-2 models are to be interpreted. Take a model $M = \langle D, I, P_2 \rangle$. Each member P_1 of P_2 is an admissible precisification of 'admissible precisification'. Some precisifications, those that are in any admissible precisification of 'admissible precisification', will be in every such P_1. Others, those that are in no admissible precisification of 'admissible precisification', will be in no such P_1. And still others, those precisifications on the

---

[15]See eg (Beall, 2008) for a discussion of revenge.

[16]This is sometimes called the 'principle of uniform solution'. For discussion, see eg (Colyvan, 2008), (Priest, 2002).

borderline of 'admissible precisification', will be in some of the P_1s but not others.

Now let's turn to 'noonish'. 12:00 is in every admissible precisification of 'noonish', no matter how one precisifies 'admissible precisification'; 12:23 is in some admissible precisifications but not others, no matter how one precisifies 'admissible precisification';[17] and 20:00 is in no admissible precisifications of 'noonish', no matter how one precisifies 'admissible precisification'. So far so good—and so far, it all could have been captured with a level-1 model.

But there is more structure to map. Some moment between 12:00 and 12:23—let's say 12:10 for concreteness—is in every admissible precisification of 'noonish' on some admissible precisifications of 'admissible precisification', and in some admissible precisifications of 'noonish' but not others on some admissible precisifications of 'admissible precisification'. And some moment between 12:23 and 20:00—let's say 12:34—is in no admissible precisification of 'noonish' on some admissible precisifications of 'admissible precisification', and in some admissible precisifications of 'noonish' but not others on some admissible precisifications of 'admissible precisification'.

Here's a (very toy) model mapping the above structure:

(128)   a. D = the set of times from 12:00 to 20:00

          b. I = the usual map from time-names to times

          c. P_2 = { {{12:00–12:38}, {12:00–12:15}}, {{12:00–12:25}, {12:00–12:08}} }[18]

Call this model M. Now let's apply it to some atomic sentences: $M(N12:00) = \{\{1\}\}$, $M(N12:10)$ $= \{\{1\}, \{1,0\}\}$, $M(N12:23) = \{\{1,0\}\}$, $M(N12:34) = \{\{1,0\}, \{0\}\}$, and $M(N20:00) = \{\{0\}\}$. One suspects that these five values are all one needs of V_2 for (at least most) vague predicates. In order for a sentence to take the value $\{\{1\}, \{0\}, \{1,0\}\}$ on a model, the model would have to be set up so that, depending on the precisification of 'admissible precisification', the sentence could be in all admissible precisifications *or* some but not others *or* none at all. It seems unlikely that many predicates have admissible precisifications that work like this. For a sentence to take the value $\{\{1\}, \{0\}\}$ on a model, something even weirder would have to happen: the model would have to make it so that, depending on the precisification of 'admissible precisification', the sentence could be either true in all admissible precisifications or false in all of them, but there could be no admissible precisification of 'admissible precisification' that would allow the sentence to be true on some admissible precisifications but not others. This too seems unlikely. So I suspect that only five

---

[17]Again, assume a context where this is true.

[18]For simplicity, we look at only one predicate: N for 'noonish'. This set is then a set of sets of precisifications for 'noonish'. Let $\{x–y\}$ be the set of times between x and y inclusive.

of the seven members of V_2 are likely to be useful for vague predicates, although (as mentioned above) not much hangs on this.[19]

There is of course the question of interpretation: which of these values counts as true? Again, we should give the same answer here as in the case of first-order vagueness, to avoid ad-hoccery: a sentence is true iff it's true on some admissible precisification; and so it's true iff it's true on some admissible precisification, for some admissible precisification of 'admissible precisification'. That is, any sentence whose value on a model has a 1 in it anywhere—any sentence whose value isn't $\{\{0\}\}$—is true on that model.

Connectives

So that's how our atomics get their values. Of course, we need some way to assign values to compound sentences as well, and the familiar LP operations (call them $\wedge_1$, $\vee_1$, and $\neg_1$) won't work—they're defined only over V_1, but our atomics take values from V_2. Fortunately, a simple tweak will work, getting us sensible level-2 operations $\wedge_2$, $\vee_2$, and $\neg_2$ defined over V_2.

Recall one of our earlier observations about the LP connectives: in a conjunction, every precisification of one conjunct sees every precisification of the other conjunct (*mutatis mutandis* for disjunction). We can use this to define our level-2 connectives.

Consider the conjunction of two V_2 values u and v. Remember, values from V_2 are sets of values from V_1, and we already have well-behaved connectives over V_1. To get one potential V_1 value for the conjunction, we can pull a V_1 value from u and one from v, and conjoin them. If we do that in every possible way, and collect all the results into a set, we get a V_2 value appropriate to be that value of the conjunction. More formally: $u \wedge_2 v = \{u' \wedge_1 v' : u' \in u, v' \in v\}$. The same idea will work for disjunction—$u \vee_2 v = \{u' \vee_1 v' : u' \in u, v' \in v\}$—and negation—$\neg_2 u = \{\neg_1 u' : u' \in u\}$. So let's simply adopt these as our level-2 connectives.

Consequence

Level-2 models now assign values to every sentence; first the atomics, via the sets of sets of precisifications, and then to all sentences, via the level-2 connectives. What's more, we have a set $D_2 \subseteq V_2$ of *designated* values—values that count as true. (Some of them also count as false, of course.) This means that we're in a position to define level-2 consequence. We do it in the expected way: $\Gamma \vDash_2 \Delta$ iff, for every level-2 model M, either $M(\delta) \in D_2$ for some $\delta \in \Delta$, or $M(\gamma) \notin D_2$ for some $\gamma \in \Gamma$.

---

[19] Actually I don't see that anything does.

So we have a full logic erected 'up a level' from LP, as it were. At first blush, this might seem like not much of a response to the challenge of 2nd-order vagueness. After all, it seems that we simply abandoned the initial theory and adopted another. That would hardly be a persuasive defense. But in fact that's not quite what's happened; as it turns out, $\models_2 = \models_{LP}$.[20] We haven't actually abandoned the initial theory—we've just offered an alternative semantics for it, one that fits the structure of second-order vagueness quite naturally. What's more, we haven't had to invoke any special machinery to do it. Simply re-applying the first-order theory to itself yields this result.

6.3.2 Generalizing the construction

Of course, the above construction only works for 2nd-order vagueness, and there is much more to higher-order vagueness than that. In particular, just as it was vague which precisifications are admissible precisifications of 'noonish', it's vague which precisifications are admissible precisifications of 'admissible precisification'. From the above construction, of course, one can predict the reply: we'll look at admissible precisifications of 'admissible precisification of 'admissible precisification'', which is of course itself vague, and so on and so on. Let's lay out a general picture here.

Let an n-set of precisifications be defined as follows: a 0-set of precisifications is just a precisification, and a $(k+1)$-set of precisifications is a set of k-sets of precisifications. Let sets V_n of values be defined as follows: $V_0 = \{1, 0\}$, and $V_{k+1} = \wp(V_k) - \emptyset$. A level-$n$ model M_n is then a tuple $\langle D, I, P_n \rangle$ such that D is a domain of objects, I is a function from terms to members of D, and P_n is an n-set of precisifications. Consider an atomic sentence A. We build up its value $M_n(A)$ as follows: in concert with I, every precisification p assigns A a value $p(A)$ from V_0, and every $(k+1)$-set P_{k+1} of precisifications assigns A a value $P_{k+1}(A) = \bigcup_{P_k \in P_{k+1}} \{P_k(A)\}$ from V_{k+1}. $M_n(A)$ is then just $P_n(A)$. For the level-1 and level-2 cases, this is just the same as the above setup, but of course it extends much farther.

Which values count as true? By parallel reasoning to our earlier cases, any value that contains a 1 at any depth. More precisely, we can define a hierarchy D_n of sets of designated values as follows: $D_0 = \{1\}$, and $D_{k+1} = \{v \in V_{k+1} : \exists u \in v (u \in D_k)\}$.

For the connectives: we define a hierarchy $\wedge_n, \vee_n, \neg_n$ of operations as follows: $\wedge_0, \vee_0$, and $\neg_0$ are simply classical conjunction, disjunction, and negation. For values $u_{k+1}, v_{k+1} \in V_{k+1}$, $u_{k+1} \wedge_{k+1} v_{k+1} = \{u_k \wedge_k v_k : u_k \in u_{k+1}, v_k \in v_{k+1}\}$. That is, the a conjunction of sets of values is the set of conjunctions of values from those sets. Similarly for disjunction: $u_{k+1} \vee_{k+1} v_{k+1} =$

[20]For proof, see (Priest, 1984).

$\{u_k \vee _k v_k : u_k \in u_{k+1}, v_k \in v_{k+1}\}$, and for negation: $\neg_{k+1} u_{k+1} = \{\neg_k u_k : u_k \in u_{k+1}\}$. Again, this gives us just what we had before in the case where $n = 1$ or 2, but extends much farther.

We are now in a position to define a hierarchy of consequence relations $\vDash _n$ as follows: $\Gamma \vDash _n \Delta$ iff, for every n-level model M_n, either $M_n(\delta) \in D_n$ for some $\delta \in \Delta$, or $M_n(\gamma) \notin D_n$ for some $\gamma \in \Gamma$. Of course, this simply extends our earlier definition to the new level-n framework.

6.3.3   Good news

Just as in the case of second-order vagueness, this construction allows us to fully map the structure of nth-order vagueness for any n. By collecting up (with a bit of jiggering), we can come to an ω-valued model that fully maps the structure of all higher-order vagueness. What's more, just as in the 2nd-order case, we haven't affected our consequence relation at all; for every $n \geq 1$ (including ω), $\vDash _n = \vDash _1$. [21] This shows us that, although we can fully map this structure, there is in fact no need to for logical purposes; the logic we define remains unchanged. We may as well stick with the simple three-valued version. (Or any other version we like. I like the three-valued version for its simplicity, but if there's some reason to prefer another version, then by all means.)

6.4   Conclusion: the sorites

We should close by examining the familiar paradox that arises from vague language: the sorites.

Here's a sample, built on 'noonish' (still written N):

$$
\begin{array}{ll}
1. & N12{:}00 \\
2. & \forall x[(Nx) \rightarrow (N(x + 0{:}00{:}01))] \\
\hline
3. & N20{:}00
\end{array}
$$

Now, before we do any logic at all, we know a few things about this argument. We know that the first premise is true, and that the conclusion is not.[22] That leaves us just a few options: we can deny the second premise, or we can deny that $\rightarrow$ supports modus ponens.[23]

Well, what is $\rightarrow$, anyway?[24] There are many possibilities. Each possibility creates a distinct sorites argument. And, while we may think that the sorites argument must be solved in a uniform

---

[21] Proof and details can be found in (Priest, 1984). Note as well that the result can be iterated past ω into the transfinite; I don't think that'll be necessary here, since every new level is created to address the vagueness of some finite predicate.

[22] Some have denied this (eg Unger (1979)), but to take such a position seriously is to remove much of the sense of 'noonish'. And to take it seriously for every vague predicate would make it very hard indeed to talk truly at all.

[23] Or we could deny transitivity of entailment, or universal instantiation, but the LP-based solution offered here keeps to the more conservative side of the street.

[24] This question is raised forcefully in (Beall & Colyvan, 2001).

way no matter which vague predicate it's built on, we certainly should not think that it must be solved in a uniform way no matter which binary connective $\to$ it's built on.

Some examples. 1) Suppose $A \to B$ is just $A \wedge B$. Then surely modus ponens is valid for $\to$; after all, B would follow from $A \to B$ alone on this supposition. But the second premise in the sorites is obviously not true, given this interpretation: it simply claims that every moment, along with the moment one second later, is noonish. That's not at all plausible. So on this interpretation, the sorites is to be answered by rejecting the second premise. 2) Suppose on the other hand that $A \to B$ is $\neg(A \wedge \neg A)$. Then the second premise is clearly true,[25] but it just as clearly does not support modus ponens. From A and $\neg(A \wedge \neg A)$, we cannot conclude B.

Of course, $A \wedge B$ and $\neg(A \wedge \neg A)$ are silly conditionals. But they make a point: whatever our commitment to uniform solution, it does not hold when we vary the key connective in the sorites argument. We are free to reject the second premise for some readings of $\to$ and deny modus ponens for others, and this does not make our solution non-uniform in any way worth avoiding. We might both reject the premise and deny modus ponens for some readings of $\to$.[26] The one thing we cannot do is accept both the premise and the validity of modus ponens, on any single reading of $\to$.

LP as presented here includes at best a very weak conditional. Its material conditional, defined as $A \supset B := \neg(A \wedge \neg B)$, does not support modus ponens.[27] Given the theory of vague predicates advanced here, the second premise of the sorites is true if we read $\to$ as $\supset$. So the present account doesn't run into any trouble on that version of the sorites. What's more, as mentioned in (Hyde, 2001), the Stoics sometimes used this form of the argument (the form using 'for any moment, it's not the case both that that moment is noonish and that one second later isn't noonish'), precisely to avoid debates about the proper analysis of conditionals. If we do the same, no trouble ensues.

On the other hand, the most compelling versions of the sorites use the 'if. . . then' of natural language. $\supset$ isn't a very promising candidate for an analysis of a natural-language conditional, in LP or out of it, because of the well-known paradoxes of material implication (see eg (Routley et al., 1982) for details). What is the right analysis of natural-language conditionals is a vexed issue (to say the least!) and not one I'll tackle here, so this is not yet a response to the sorites built on 'if. . . then'. For now, we can see that the LP-based approach answers the material-conditional version of the sorites handily.

---

[25] at least given most theorists' commitments about the law of non-contradiction, including the commitments of the LP-based approach

[26] For example, if we read $A \to B$ simply as A.

[27] This is because modus ponens on $\supset$ is equivalent to disjunctive syllogism, which anyone who takes contradictions seriously ought to reject. See (Priest, 1979) for discussion.

What's more, it embodies the picture of vague language underlying subvaluationist and supervaluationist motivations in a more natural way than SB and SP themselves do. It also verifies much of our ordinary talk about borderline cases, contradictory and otherwise, and provides a satisfying non-ad-hoc account of higher-order vagueness. In short, LP should be considered a serious contender in the field of nonclassical approaches to the phenomenon of vagueness.

Chapter 7

Contradictions at the borders

7.1 The issue

The purpose of this essay is to shed some light on a certain type of sentence, which I call a *borderline contradiction*. A borderline contradiction is a sentence of the form $Fa \wedge \neg Fa$, for some vague predicate F and some borderline case a of F, or a sentence equivalent to such a sentence. For example, if Jackie is a borderline case of 'rich', then 'Jackie is rich and Jackie isn't rich' is a borderline contradiction. Many theories of vague language have entailments about borderline contradictions; correctly describing the behavior of borderline contradictions is one of the many tasks facing anyone offering a theory of vague language.

Here, I first briefly review claims made by various theorists about these borderline contradictions, attempting to draw out some predictions about the behavior of ordinary speakers. Second, I present an experiment intended to gather relevant data about the behavior of ordinary speakers. Finally, I discuss the experimental results in light of several different theories of vagueness, to see what explanations are available. My conclusions are necessarily tentative; I do not attempt to use the present experiment to demonstrate that any single theory is incontrovertibly true. Rather, I try to sketch the auxiliary hypotheses that would need to be conjoined to several extant theories of vague language to predict the present result, and offer some considerations regarding the plausibility of these various hypotheses. In the end, I conclude that two of the theories I consider are better-positioned to account for the observed data than are the others. But the field of logically-informed research on people's actual responses to vague predicates is young; surely as more data come in we will learn a great deal more about which (if any) of these theories best accounts for the behavior of

ordinary speakers.

7.1.1 Contradictions and borderline cases

In (Ripley, 200xb), I defend a theory of vague language based on the paraconsistent logic LP.[1] This theory predicts borderline contradictions to be true. That is, when a is a borderline case of a vague predicate F, I claim that 'a is F and a is not F' is true. Similarly, I claim that 'a is neither F nor not F' is true as well, since this follows from the former by a single De Morgan law plus an application of a double-negation rule, both of which are valid in LP. This is a dialetheist theory, since it takes some contradictions to be true.

Other theorists, of various stripes, have not been so sanguine about the truth of borderline contradictions. A few quick examples: Fine (1975) dismisses the idea in a single sentence—"Surely $P \wedge \neg P$ is false even though P is indefinite".[2] Williamson's (1994) much-discussed argument against denials of bivalence works by arguing the denier to a contradiction; assuming the denial of bivalence was initially made about a borderline case, this contradiction will itself be a borderline contradiction. If Williamson thinks this is a dialectically strong argument, as he gives every indication of, borderline contradictions had better not be true. Keefe (2000) offers: "many philosophers would soon discount the paraconsistent option (almost) regardless of how well it treats vagueness, on the grounds of ... the absurdity of p and $\neg p$ both being true for many instances of p". And Shapiro (2006) claims, "That is, even if one can competently assert Bh and one can competently assert its negation, one cannot competently contradict oneself (dialetheism notwithstanding)."[3] None of these rejections of borderline contradictions offers much in the way of argument; it's simply taken to be obvious that borderline contradictions are never true, presumably since no contradictions are ever true.[4]

Not all theorists—not even all non-dialetheist theorists—have been so quick with borderline contradictions, though. For example, fuzzy theorists allow for borderline contradictions to be par-

---

[1]See (Beall & van Fraassen, 2003) for details of this logic.

[2]Notation changed slightly; note that Fine is here treating borderline cases as "indefinite".

[3]Since, for Shapiro, the relevant cases in which one might competently assert Bh and competently assert its negation are all cases where h is a borderline case of B, this is a rejection of borderline contradictions.

[4]Williamson might claim to have an argument for his rejection of borderline contradictions: his defense of classical logic on grounds of its simplicity. Note, though, that that defense is dialectically out of line in the midst of the argument Williamson gives against denials of bivalence; why bother arguing the bivalence-denier to a contradiction, and then appeal to the truth of classical logic to reject the contradiction, when you could simply appeal to the truth of classical logic directly to counter a denial of bivalence? Presumably, Williamson thinks the rejection of borderline contradictions is dialectically more secure than his defense of the full apparatus of classical logic. Thanks to Patrick Greenough for raising this point to me.

tially (up to half) true.[5] Let's see how. The usual way of doing things assigns each sentence A a real-number truth value $\nu(A)$ from 0 to 1, inclusive. Then, the values of compound sentences are determined truth-functionally from the values of their components, as follows:

- $\nu(\neg A) = 1 - \nu(A)$

- $\nu(A \wedge B) = \min(\nu(A), \nu(B))$

- $\nu(A \vee B) = \max(\nu(A), \nu(B))$

It follows from these definitions that a contradiction (conjunction of a sentence with its own negation) can take a value as high as .5. It takes this maximum value when its conjuncts themselves each take value .5—right in the middle of a vague predicate's borderline. So a fuzzy theorist predicts borderline contradictions to be at least partially true. (This prediction is often held up as a liability of fuzzy theories; see for example (Williamson, 1994).)

7.1.2 Predictions about ordinary speakers

Smith (2008, pp. 252–253) lists ten sorts of sentence for which we don't as yet have clear empirical data about speakers' intuitions; he resists making many predictions about speakers' intuitions pending the data. At least three of his categories are borderline contradictions, in my sense, and he's right: there isn't much data on speakers' responses to them.

Some experimenters have taken brief looks at ordinary speakers' intuitions surrounding vague predicates (for example Bonini et al. (1999)), but these have primarily looked at atomic sentences, whereas the crucial action for theories of borderline contradictions is clearly in compound sentences; empirical work here is still in its infancy.

Few logically-minded theorists of vagueness, then, have bothered being very explicit about what their theories predict about ordinary speakers. In what follows, I'll talk of "contextualist" or "fuzzy" theories about ordinary speakers, but it may well be that a particular contextualist or fuzzy theorist of vagueness would not want their theory to be taken as a theory about ordinary speakers. My main goal here is to present an experiment and weigh various possible explanations for the result; as such, my comments below don't bear on any theory that is explicitly not about ordinary speakers.

---

[5]At least the usual sort of fuzzy theorists do. See for example (Smith, 2008).
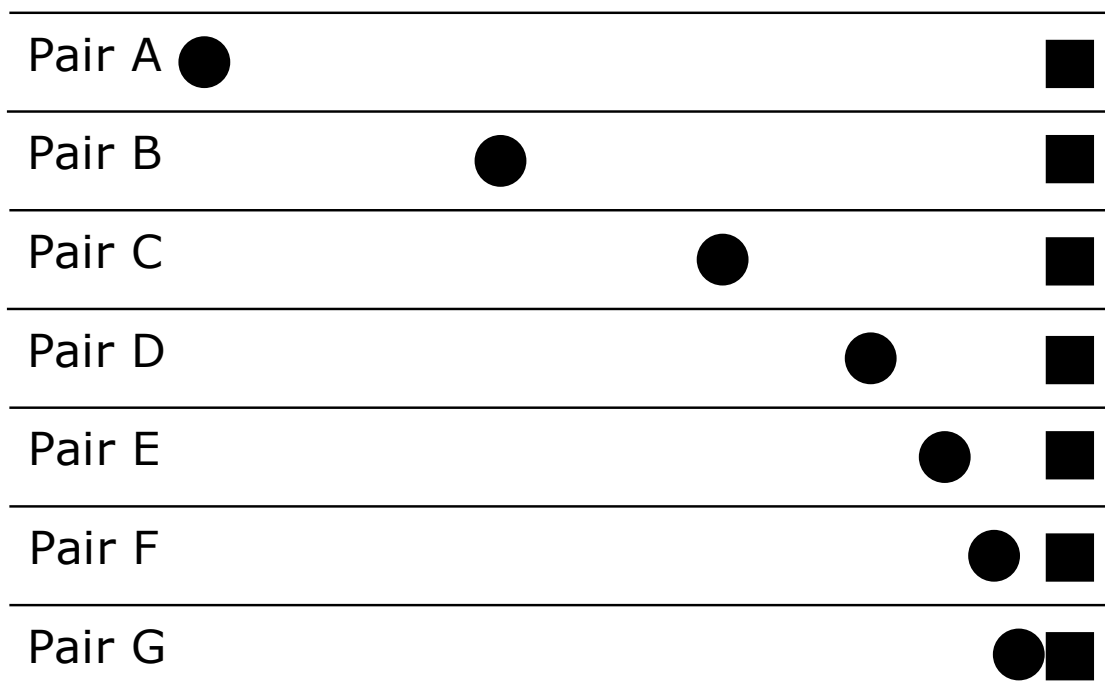
Figure 7.1: Experimental Stimulus

7.2 The experiment

To explore intuitions about contradictions in borderline cases of vague predicates, I conducted an experiment. Participants were 149 undergraduate students at the University of North Carolina. They saw a slide (projected onto a screen) with seven circle/square pairs on it, labeled 'Pair A' to 'Pair G'. In Pair A, at the very top of the slide, the circle was as far from the square as it could be, while in Pair G, at the very bottom of the slide, the circle was touching the square. In between, the remaining five pairs moved the circle bit-by-bit closer to the square. (See Figure 7.1 on page 128.)

It's difficult to tell exactly what's a borderline case of 'near the square'; as many authors have pointed out, the extension of vague predicates like 'near' is quite context-dependent, and it can be difficult to tell where the borderline is. For example, if we're discussing distances between cities, this provides a context in which the circle is near the square in *every* pair; the distance in the farthest pair is never more than the size of the screen being used, which is surely smaller than the distance between even the closest cities. Nevertheless, I take it that the context provided by this experiment is one in which: in Pair A, the circle is a clear countercase of 'near the square' (that is, it is clearly not near the square—after all, it's as far away from the square as can be projected on the screen), and in Pair G, the circle is a clear case of 'near the square'. Somewhere in between are the borderline

cases.

Participants were randomly assigned to one of four conditions. In each condition, participants were asked to indicate their amount of agreement with a particular sentence as applied to each of the seven circle/square pairs. The four conditions involved four different sentences; each participant, then, saw only one sentence and rated it seven times, once for each pair. Ratings were on a scale from 1 to 7, with 1 labeled 'Disagree' and 7 labeled 'Agree'. The four sentences were:

Conjunction, Non-elided: The circle is near the square and it isn't near the square.

Conjunction, Elided: The circle both is and isn't near the square.

Disjunction, Non-elided: The circle neither is near the square nor isn't near the square.

Disjunction, Elided: The circle neither is nor isn't near the square.

I'll discuss the difference between the elided and non-elided cases later. For now, note that each of these sentences has the form of a contradiction. The conjunctions wear their contradictoriness on their faces, while the disjunctions are a bit disguised; but one application of a De Morgan law reveals them to be contradictions as well.

7.2.1   Response Types

Over 90% of participants gave responses that fall into one of four groups. I'll call these groups *flat*, *hump*, *slope up*, and *slope down*. Here are the defining characteristics of these groups (see Figure 7.2 on page 130 for frequencies):

Flat: A *flat* response gives the same number for every question. (24 participants)

Hump: A *hump* response is not a flat response, and it has a peak somewhere between the first and last question; before the peak, responses never go down from question to question (although they may go up or remain the same), and after the peak, responses never go up from question to question (although they may go down or stay the same). (76 participants)

Slope up: A slope up response is not a flat response, and it never goes down from question to question (although it may go up or stay the same). (22 participants)

Slope down: A slope down response is not a flat response, and it never goes up from question to question (although it may go down or stay the same). (18 participants)

Other: There were a few responses that didn't fit any of these patterns. (9 participants)
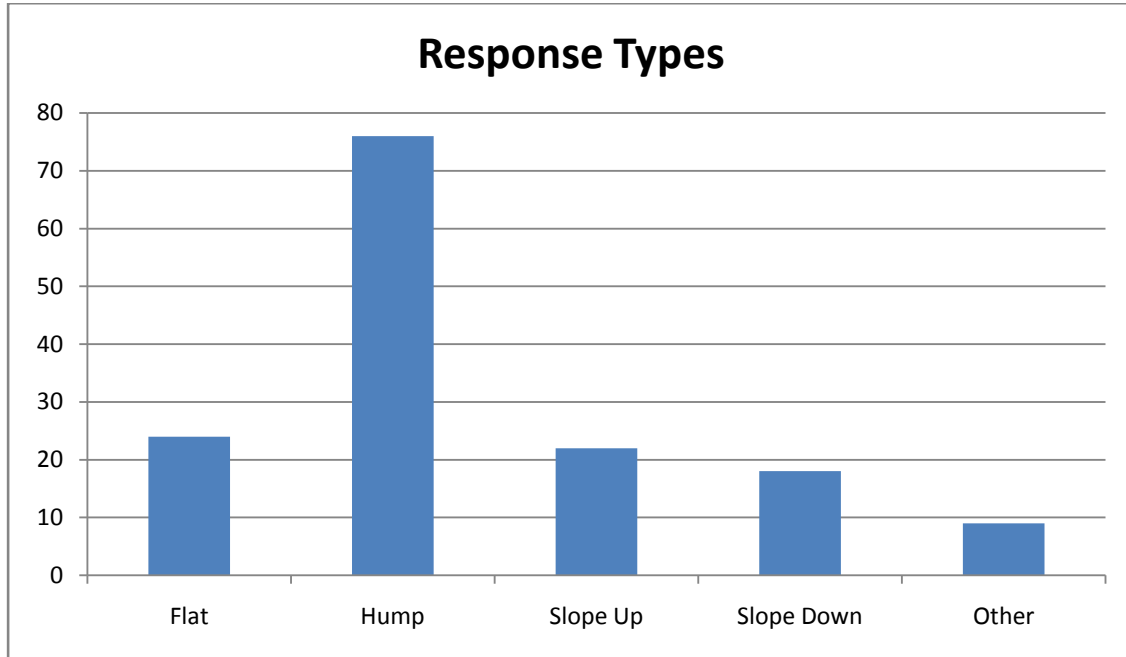
Figure 7.2: Response type frequencies

Flat responses, in particular flat 1s (14 participants), look like the sort of response that would be predicted by all those theorists who hold that no contradiction is ever true, even a bit, even in borderline cases. But the majority of responses (76/149 participants) were hump responses.

7.2.2   Agreement to contradictions

Notably, the pattern formed by the mean responses to each pair was a hump pattern (see Figure 7.3 on page 131). The highest mean occurred in response to Pair C; there the mean response was 4.1, slightly above the midpoint of the 1 to 7 scale. In other words, participants exhibit higher levels of agreement to these apparent contradictions when they are about borderline cases; they do not reject what appear to be borderline contradictions. In fact, they seem to make it to at least ambivalence (and see §7.3.5 for evidence that participants make it well past ambivalence, to full agreement in many cases).

The discussion that follows in §7.3 will consider various explanations for participants' agreement, partial or full, with these sentences. I'll focus discussion on the relatively large number of hump responses; a fuller discussion would consider potential explanations for the flat and slope groups as well.[6]

---

[6]Question type (conjunction vs. disjunction) had a significant effect on response type ($\chi^2(4, N = 149) = 11.27, p < .05$). However, this effect disappeared when the two slope response types were lumped together ($\chi^2(3, N = 149) =$
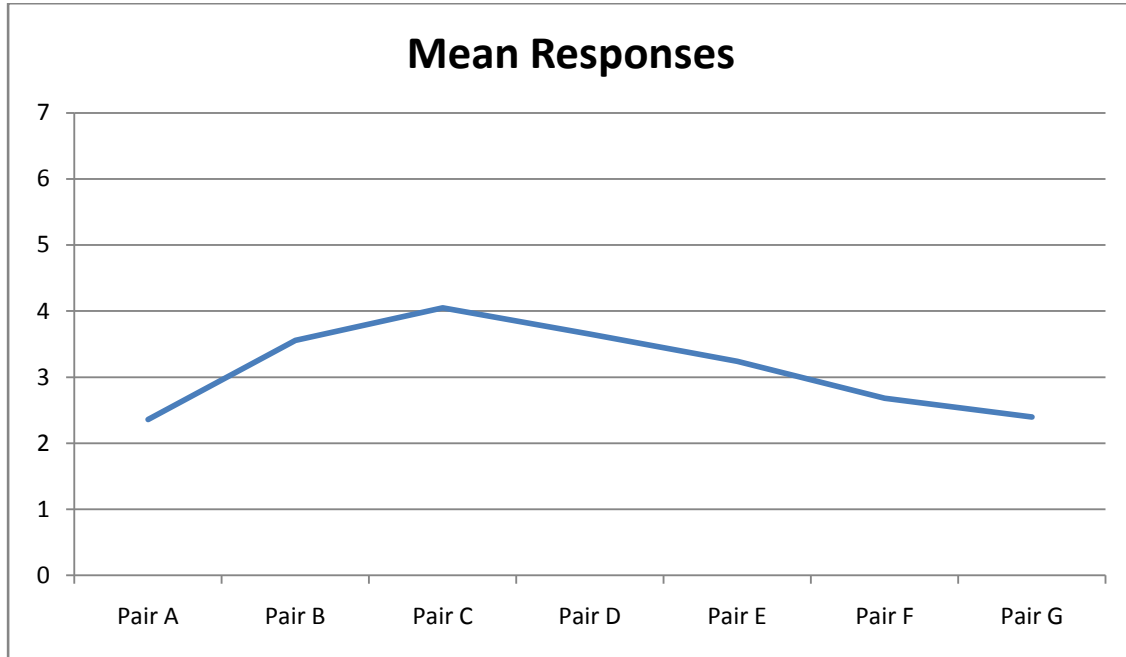
Figure 7.3: Mean responses

7.3 Interpretations

7.3.1 The options

It seems at first blush that we have substantial numbers of participants agreeing, at least somewhat, with borderline contradictions of various sorts. When it comes to interpreting this behavior, we can divide interpretations into three broad sorts: first, it might be that what participants are agreeing to isn't really a contradiction; second, it might be that they are really agreeing to a contradiction, but that they're somehow mistaken; and third, it might be that they are really agreeing to a contradiction, and are not mistaken.

7.3.2 Contextual factors

This explanation falls into the "not really a contradiction" category.

Here's one way to explain the relatively high levels of assent to sentences like 'The circle is near the square and it isn't near the square' and 'The circle neither is near the square nor isn't near the square': participants take the phrase 'near the square' to have subtly different extensions in

---

2.76, $p = .43$). Slope up responses occurred more in response to conjunctions, and slope down responses in response to disjunctions. This makes it seem as though the slope responders tended to ignore the second conjunct in each case, treating 'both near and not' as 'near' and 'neither near nor not' as 'not near'. More study would be needed to definitively interpret the slope responses.

each of its two occurrences within the sentence. If this is so, their assent to these sentences can be explained without supposing that any participants agree to a contradiction. (For my purposes here, a "contextualist" is not someone who offers any particular theory of vagueness, but rather anyone who thinks that the hump responses in the present experiment are to explained by appealing to contextual shift in the extension of 'near the square'.)

Such a contextualist theory can come in one of two flavors: it might hold that 'near the square' has these different extensions because it has different contents in each of its uses, or it might hold that 'near the square' has the same content in each of its uses, but that nonetheless it has different extensions in different contexts. Following (MacFarlane, 2009), I'll call the first flavor 'indexical contextualism' and the second 'nonindexical contextualism'. I discuss each in turn.[7]

Indexical contextualism

Indexical contextualism about vague terms is defended in (Soames, 2002). On this theory, different uses of vague terms can express different properties. This shiftiness is understood as the very same shiftiness exhibited by such indexical expressions as 'here', 'now', 'you', 'tomorrow', &c. For example, let's focus on 'you'. 'You', let's suppose, picks out a certain person: the person being addressed when it's uttered. Now, imagine someone uttering the following sentence: 'Mona sees you, and Louie sees you'. It should be clear that the two occurrences of 'you' in such an utterance might pick out different people; just imagine the context changing in the right way (that is, so that the first half of the sentence is addressed to someone different than the second half).

On an indexical contextualist theory, something just like this might be happening in the sentence 'The circle is near the square and not near the square'; the first occurrence of 'near the square' can pick out one property, and the second some other property. Of course, for this to be the case there would have to be some relevant shift in context between the two occurrences, and the indexical contextualist would have to provide some story about what the relevant context is and why it shifts.[8] Even with such a story in hand, though, the indexical contextualist runs into some difficulties with the experimental data.

The difficulty arises with the elided sentences: 'The circle both is and isn't near the square' and

---

[7]Besides the difference between indexical and nonindexical contextualism, there is another difference in the area: the difference between theories that posit sensitivity to context-of-use (sometimes called "contextualist") and theories that posit sensitivity to context-of-assessment (sometimes called "relativist"). I'll ignore that distinction; for my purposes here, I'm happy to lump the relativists in with the contextualists.

[8]Of course, this requirement is not unique to the indexical contextualist; every contextualist needs such a story. I won't be concerned with the details of such stories here—see for example (Raffman, 1996), (Shapiro, 2006), or (Soames, 1998). (NB: Raffman and Shapiro are not indexical contextualists.)

'The circle neither is nor isn't near the square'. Each of these sentences contains only one occurrence of 'near the square'. It's clear, though, that indexicals, in these circumstances, can have only a single interpretation. Compare our earlier 'Mona sees you, and Louie sees you' to 'Mona sees you, and Louie does too'. Even with the same shift in context (that is, with the second half of the sentence addressed to someone different than the first half), the second sentence *must* report that Mona and Louie see the very same person. Since there's only one occurrence of 'you', it can only pick out one person.[9]

Thus, the indexical contextualist should predict that, although participants might agree to the non-elided sentences, they should not agree to the elided sentences, since the mechanism invoked to explain participants' agreement in the non-elided cases can't operate in the elided cases. Participants simply should not agree with elided sentences. At the very least, they should agree less than they do with the non-elided sentences. This prediction is not borne out. If we consider each participants' maximum level of agreement, there is no significant difference between responses to elided and non-elided sentences.[10] Nor is there a difference in response types (flat, hump, &c.) between elided and non-elided cases.[11] If participants' agreement to these apparent contradictions, then, is to be explained by appealing to context, that context can't be operating in the way that context operates on indexicals.[12]

Nonindexical contextualism

Is there another way, then, for context to come into play? The nonindexical contextualist thinks so. I think nonindexical contextualism, suitably filled in, provides one of the more plausible explanations for the results of the present study. The task of this section will be to present some constraints that the nonindexical contextualist must satisfy to explain the observed results.

---

[9]Similar phenomena arise around (at least) demonstratives, definite descriptions, and proper names. In each of the following pairs, the first member allows a shift where the second does not:

- ○ Mary's buying that, unless Murray buys that
  ○ Mary's buying that, unless Murray does
- ○ Put your bag on the table, and your books on the table
  ○ Put your bag on the table, and your books too
- ○ Esmerelda went to the store, and Esmerelda bought some fish
  ○ Esmerelda went to the store and bought some fish

[10]As measured by a one-way ANOVA, $F(1, 148) = .24, p = .62$.

[11]$\chi^2(4, N = 149) = 1.98, p = .74$.

[12]This is similar to the argument in (Stanley, 2003), except that Stanley fails to distinguish between indexical and nonindexical contextualism. See (Raffman, 2005) for details.

To see how nonindexical contextualism works, let's consider an indexical case in more detail. Consider an utterance, by me, of the sentence 'I like to dance'. The occurrence of 'I' in that utterance refers to me, so the whole utterance has the content *Dave likes to dance*.[13] That content is (very) true, but it might have been false; it is true with regard to the world we find ourselves in, and false with regard to other possible worlds. So, in determining the extension (truth-value) of the utterance from its content, we need to take something more into account: we must consider at least which possible world we're in. The nonindexical contextualist finds a role for context in just this way—in the step from content to extension. They can offer various theories, still, about *which* contextual factors come into play; the key for the nonindexical contextualist is *when* those factors do their work.[14]

So what would a nonindexical contextualist offer as a take on the present study? Let's start with 'The circle is near the square and it isn't near the square'. The indexical contextualist held that this sentence ascribes one property ('near the square' in context 1) and the negation of *some other* property ('near the square' in context 2) to the circle; its content was thus baldly noncontradictory. But the nonindexical contextualist doesn't go this route; she'll say that the sentence ascribes one property (nearness-to-the-square) and the negation of *that very property* to the circle. In order to avoid contradiction, then, she must say that the one property has two different extensions with regard to two different contexts. Importantly, those contexts must both be at play in the interpretation of the single sentence.[15] If context is ephemeral, dependent on, say, a transient mental state of the judger (as in (Raffman, 1996)), then this should be possible. On the other hand, if context is coarser-grained, dependent on only things like world, approximate time, location, speaker, and the like, then we can see that context could not have changed mid-sentence, and so a contextualist explanation couldn't get off the ground.

By examining the elided conditions in the present study, we can see further constraints on a workable nonindexical contextualist theory. We've already seen that, for this explanation to work, the relevant features of the context in play must be relatively fine-grained. The elided conditions provide us evidence about *which* context it is that comes into play. Consider 'The circle both is and

---

[13]I ignore any possible context-sensitivity, of any sort, in 'likes to dance'.

[14]This way of framing the issue owes much to (Kaplan, 1989) and (MacFarlane, 2009).

[15]At least for indexical context-sensitivity (and why should nonindexical sensitivity differ?), it seems incontrovertible that multiple contexts can be involved in the interpretation of a single sentence. See note 9, or consider 'I am here now', which can be false if said very slowly while moving very quickly. Some authors, though, have missed this: for example Richard (1993), who writes, 'Switching contexts in the middle of interpreting a sentence is clearly contrary to the spirit, not to speak of the letter, of Kaplan's approach to indexicals.' (I'm skeptical of his reading of Kaplan.) Other authors have played it down: see (Kaplan, 1989), which makes 'I am here now' come out as a logical truth in its logic of indexicals, or (MacFarlane, 2009), which talks of context affecting whole propositions at once.

isn't near the square'. For a nonindexical contextualist explanation to work, the context relevant to determining the extension of 'near the square' cannot be the context in which 'near the square' is read by the participant. After all, there is only one such context, but the contextualist appeals crucially to a change in context between two extension-determinations. I see two options for the nonindexical contextualist: 1) it may be that participants process this sentence into some form that contains two occurrences of 'near the square' or something (conceptual material, presumably) corresponding to 'near the square'—then each separate occurrence can be affected by the context in which it occurs—or 2) it may be that participants evaluate the conjuncts one at a time, retaining only the truth-value of each conjunct after its evaluation—then each evaluation can be affected by the context in which it occurs.

I think that nonindexical contextualism, suitably filled-in in the ways described above, offers one satisfying explanation of the present results. Below, I'll consider other possible explanations.

7.3.3 Noncompositional theories

Another variety of not-really-a-contradiction explanation claims that the sentences in question are not compositionally interpreted; that 'The circle is near the square and it isn't near the square' directly expresses something like what's expressed by 'The circle is a borderline case of "near the square"'. Perhaps it's an idiom, or something like an idiom. Then participants' relatively high level of agreement could be explained without supposing that they agree to a contradiction.

The problem with such an account is that it's difficult to see why apparent contradictions would express borderline-case-ness. How would such an idiom get off the ground? Presumably because some other explanation canvassed here (in particular, one of the explanations in §§7.3.2, 7.3.5, or 7.3.6) was at one time correct; then language learners, for whatever reason, might have mistaken their elders' compositional utterances for direct claims of borderline-case-ness. This fills in the story, but it does so compositionally. Without some explanation very unlike this (lightning strike?), I don't see that a noncompositional theory can avoid essentially appealing to some compositional theory, and it seems that it will then take on the pros and cons of whatever compositional theory it chooses.

There will be a few extra cons, however. A non-compositional theory must explain why there is no significant difference in the frequency of observed hump responses between the four experimental conditions, and why there is no significant difference between the maximum responses given by participants in these conditions.[16] Do we have four closely-related idioms? If so, why? In addi-

---

[16]See notes 6, 10, and 11; and note that there was also no significant difference between maximum responses to conjunctive and disjunctive sentences ($F(1, 148) = .53, p = .47$), nor any interaction effect on maximum responses

tion, this strategy invokes an additional step: learners coming to acquire noncompositional uses of these once-compositionally-used expressions. Without further evidence, a noncompositional theory introduces needless complication; better to stick with a compositional story.

7.3.4 Error theories

So much for explanations that work on the hypothesis that what participants are agreeing to isn't a contradiction. Among theories that concede that participants are agreeing to a contradiction, error theories of various sorts are available. An error theorist holds that, while participants are in fact agreeing to real contradictions, they are wrong to do so—these contradictions are simply false.

An error theory might work something like those presented in (Eklund, 2005) and (Sorensen, 2001), according to which all competent speakers have dispositions to accept certain falsehoods involving vague predicates, or it might work in a much more informal way, supposing participants to simply be mistaken, not in virtue of being competent speakers, but just in virtue of being confused, or not paying attention, or being misled by the experiment, or some such.

The latter sort of view—the one that simply takes all the participants who agreed with some contradiction to be mistaken for some reason other than their linguistic competence—has an immediate gap: on its own, it offers us no explanation for why participants would make *these* errors and not others. It could of course be supplemented with some theory about the conditions under which people are likely to make certain errors, and then that supplemental theory could be dealt with on its own merits. I leave it to any proponents of such a supposition to provide these supplemental theories.[17]

Consider now the former sort of error theory. Eklund's view can directly explain why participants would make errors in these cases; it's part of his theory that competent speakers have a disposition to make errors in the use of vague predicates. But the errors he takes speakers to be disposed to make are not hump-style responses. Rather, he supposes that competent speakers are disposed to believe *tolerance principles* around their vague predicates. He takes his tolerance principle from (Wright, 1975); for a vague predicate F, the tolerance principle reads:

- Whereas large enough differences in F's parameter of application sometimes matter to the

---

between conjunction/disjunction and elided/non-elided (as measured by a two-way ANOVA, $F(1, 148) = 1.37, p = .25$).

[17]NB: It can't simply be that participants err randomly under certain conditions; there are very many possible response patterns that simply didn't occur, or that occurred very rarely, while the hump pattern occurred in more than half of the responses.

justice with which it is applied, some small enough difference never thus matters.[18]

But belief in a principle like this would not lead participants to give hump-style responses; rather, if it applied at all, it would lead participants to give flat responses, responses not affected by the small differences in the cases they were shown. So while Eklund predicts that participants will make a certain sort of error, he does not predict the hump-style responses given by many participants.

Sorensen (2001) faces a similar problem: although he claims that competent speakers will believe contradictions involving vague predicates, he does not predict the present results. The "contradictions" Sorensen predicts speakers to believe are sentences of the form 'If a is F, then a's successor is F too', where a and its successor are consecutive members of a sorites sequence for F. Since Sorensen is an epistemicist, he thinks there is some sharp cutoff between the Fs and the non-Fs; when a and its successor straddle this sharp cutoff, he believes this conditional to be analytically false. Nonetheless, he thinks, we believe it.

This is essentially the same as Eklund's view, except for the decision to call these tolerance conditionals "contradictions". This sort of view, if it can be made to make any predictions at all about the present study, predicts *flat* responses, not *hump* responses. So again, this style of view cannot explain the present results.

I suppose someone might hold a view like this: being a competent speaker requires us to believe contradictions like 'a both is and isn't F' when a is a borderline case of F, but nevertheless such contradictions are always false. *That* view of course would predict the hump responses obtained in the present study. But why would competent speakers believe *those* falsehoods and not others? Any view of this sort would need to answer that question. Sorensen and Eklund go to great lengths to motivate their claims that speakers believe certain falsehoods; an error theorist of this type would need some story to fill a corresponding role. I know of no error theorist who holds this kind of theory, and so I know of no error theorist who's attempted to provide such a story.

7.3.5 Fuzzy theories

A fuzzy theory can both 1) allow that participants interpreted the sentences in question as contradictions, and 2) allow that participants might not be mistaken in partial assent to such sentences. This second feature is a virtue for a few reasons. First, as we've seen in §7.3.4, no existing error theory predicts speakers to be mistaken in this way; and second, it seems a bit odd to suppose that

---

[18]F's parameter of application is the dimension along which something can vary to make it more or less F; so 'tall''s parameter of application is height, 'bald''s is amount and arrangement of hair, &c.

speakers are mistaken about what's near what, when they can see the relevant objects clearly, are deceived in no way about the distance between them, and are not under any time pressure to come to a judgment. A fuzzy theory can allow for non-mistaken (partial) assent to contradictions because on a fuzzy theory contradictions can be partially true, as we saw in §7.1.1.

At first blush, then, it appear that the fuzzy theorist has the resources to account for the responses observed. This appearance is strengthened if we look at the mean responses for each question (see Figure 7.3 on page 131): the clear cases on each end result in mean responses just above 2—very low in agreement—and the mean responses rise gradually as one approaches pair C, where the mean response is just barely above 4, the midpoint in the agreement scale. These data are very much in line with what a fuzzy theorist would most likely predict.

Appearances, though, can be deceiving. Although the mean responses to each question create a pattern congenial to the fuzzy theorist, they do so for a strikingly non-fuzzy reason. This can be brought out by considering the difference between the *maximum* of the *mean* responses (4.1) and the *mean* of the *maximum* responses (5.3). The majority of responses were hump responses, but not all humps reach their peak in response to pair C, presumably due to slight disagreements between participants on which pairs were the clearest borderline cases. Suppose now that we abstract away from *where* in the series each response peaks, and simply look at *how high* the peak is when it's reached.

If the fuzzy theorist's formalism maps directly on to participants' responses, we would expect participants' responses to these contradictions to peak somewhere around 4, the midpoint. After all, none of these sentences can ever be more than .5 true, on a fuzzy theory. But this is not what happens. In fact, more participants peak at 7—full agreement—than peak at any other response.[19] The mean of the maximum responses is 5.3—significantly above 4.[20]

The fuzzy theorist, faced with these data, should conclude that the fuzzy formalism does not map directly onto participants' responses, then. Here's a hypothesis she might offer: perhaps responses as high as 7—full agreement—can still indicate the speech act of .5-assertion. If this is so, then the fuzzy theorist can simply claim that participants who gave very high responses to these sentences were still only .5-asserting them.

I don't see that this hypothesis is untenable, but it would take some filling in. Presumably a

---

[19]Actually, more than *twice* as many peak at 7 than peak at any other response, and the second-place response is 6. Over half of participants peak at either 6 or 7. See Figure 7.4 on page 139.

[20]In fact, this is so for each of the four conditions: for conjunction, non-elided, $t(43) = 3.96, p < .001$; for conjunction, elided, $t(39) = 4.719, p < .001$; for disjunction, non-elided, $t(28) = 5.67, p < .001$; for disjunction, elided, $t(35) = 2.81, p < .01$.
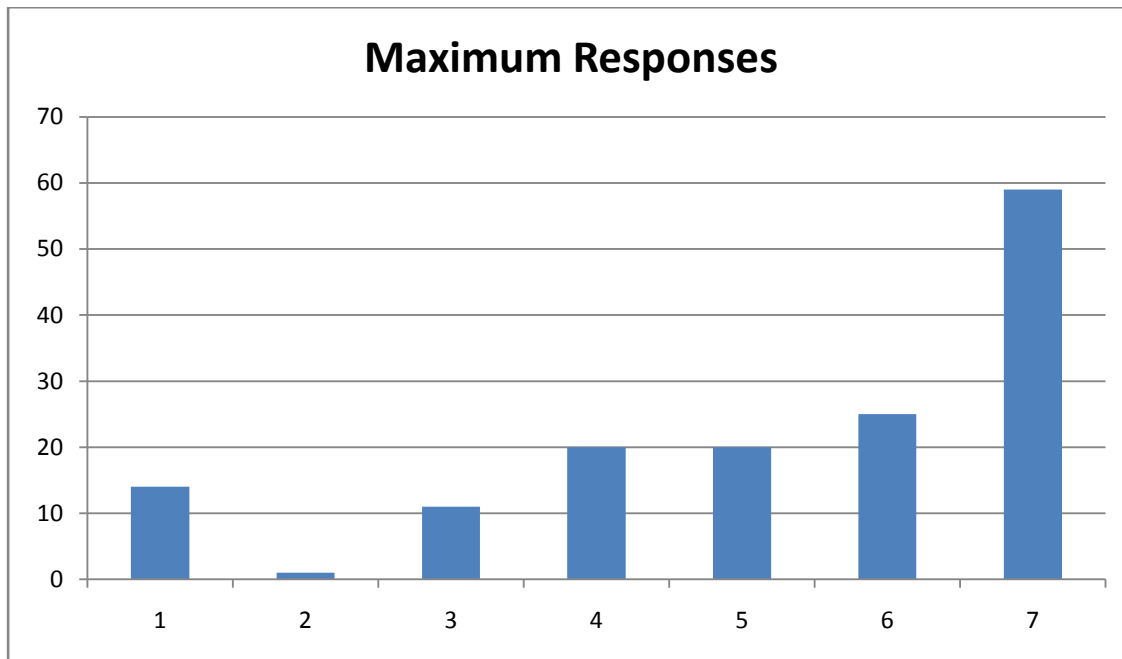
Figure 7.4: Maximum responses

response of 7 can also indicate 1-assertion (full assertion), so this hypothesis leads the fuzzy theorist to suppose that a 7-point scale from 'Disagree' to 'Agree' is not sensitive to the different degrees of assertion participants might wish to make. But if not this sort of scale, then what *would* be sensitive to those degrees? It seems that the fuzzy theorist appealing to this hypothesis would need to address that question. With an answer to that question in hand, a study like the present one could be conducted, to see whether participants really do indicate .5-assertion to these sentences.

Alternatively, the fuzzy theorist could offer an error theory of some sort. She might allow that although the highest level of assertion *appropriate* to these sentences is .5, most participants in fact evinced a higher level of assertion, and simply claim that these participants are mistaken. As we've seen, such responses are unilluminating unless conjoined with some explanation of why participants would make *these* mistakes in *these* circumstances; but there is no reason why a fuzzy theorist couldn't propose such an explanation.

7.3.6   Dialetheisms

A dialetheic theory like that presented in (Ripley, 200xb) shares some of the features of a fuzzy explanation for the present data: it can allow that, in line with appearances, participants are responding to genuine contradictions; and it can allow that these participants are not mistaken.

What's more, since a dialetheic theory predicts that the contradictions that occurred in this study are (fully) true, it can predict levels of assent higher than the midpoint values predicted by fuzzy theorists.

This is because, according to this variety of dialetheic theory, the borderline contradictions in the present study are true.[21] The circle really is both near the square and not near the square, when it's a borderline case of 'near the square'. And similarly, it's neither near the square nor not near the square, in the same circumstances. Since participants in the present study were well-positioned to see this, and since they are competent with 'near the square', conjunction, disjunction, and negation, they agreed with the borderline contradictions because they recognized them as true.

A dialetheic explanation, then, faces a quite different puzzle from the other theories we've seen. The question a dialetheist must answer is not 'Why so much assent?' but 'Why so little?'. As we've seen, the mean of the maximum responses was 5.3. Even allowing for ceiling effects, this is unlikely to represent full agreement. But if participants were well-situated to recognize the truth, and the truth is contradictory, why would they not simply fully agree to borderline contradictions? A dialetheist owes some answer here.

Since I defend a dialetheic theory of vagueness elsewhere, I'll offer a sketch of one possible answer. It's been alleged among cross-cultural psychologists that people from East Asian cultures are more open to contradictions than are people from Western cultures. These allegations, though, have often used a very wide sense of 'contradiction', much wider than that used here. For example, Peng & Nisbett (1999) count all of the following as "tolerating contradictions":

- Preferring compromise to adversarial dispute resolution

- Preferring proverbs like 'too humble is half proud' to proverbs like 'one against all is certain to fall'

- Reconciling 'most long-lived people eat fish or chicken' with 'it's more healthy to be a strict vegetarian'

Clearly, their sense of 'contradiction' is not the sense in play here; so while they may have found a very real cultural difference, their data do not show anything about cultural acceptance of contradictions, in our sense.

---

[21]It thus differs from the dialetheic theory proposed in (Hyde, 1997), which holds borderline contradictions to always be false. A Hyde-style dialetheist would presumably resort to an error theory of some variety to explain the present results.

In an attempt to connect this cross-cultural research more directly to the philosopher's idea of contradiction, Huss & Yingli (2007) ran a cross-cultural study that asked participants in Canada and China for their responses to more paradigm contradictions: the liar paradox, a reductio argument, and most importantly for my purposes here, a borderline contradiction. In particular, they presented their participants with a vignette describing a borderline case of 'raining', and asked about the sentence 'It's raining and it's not raining'.

Despite the narrower focus, the results they found were broadly in line with Peng & Nisbett's research: Huss & Yingli's Chinese participants were much more willing to agree with the contradictions they saw than were their Canadian counterparts. This suggests that Westerners hold a cultural norm against agreeing with contradictions.[22]

Suppose this to be true. Then, despite their linguistic competence pushing them to accept the borderline contradictions, subjects in the present experiment (as well as Canadian subjects in Huss & Yingli's study) may well have had their assent reduced by cultural norms. The effect would be much the same if we were to ask participants for their grammatical (rather than semantic) intuitions about sentences like 'Which table did you leave the book on?'; although ending a sentence with a preposition is perfectly grammatical in English, the cultural norm against it may well drive participants to reduce their judgments of grammaticality.[23]

Since work in cross-cultural psychology leads us to suppose that Westerners have a cultural aversion to contradictions in general, we should expect the levels of assent given by university students in North Carolina to be somewhat lower than what would be generated purely by their linguistic competence; once we take this into account, the dialetheist has a straightforward explanation for the middling levels of assent. So it seems that the dialetheist has a plausible explanation for the observed results as well.

7.4   Conclusions

When it comes to (apparent) borderline contradictions, then, it seems that the nonindexical contextualist and the dialetheist offer the two most plausible explanations of the observed results. Before I close, I want to draw some attention to the similarities between these views that allow them to succeed where other views do not. I also want to draw attention to just how hard it will be to design

---

[22]Note that if contextualism of the sort described in §7.3.2 is right, Huss & Yingli's sentence was presumably not really interpreted as a contradiction either, at least by those who agreed with it. A contextualist should then probably say that Canadians are more likely to give such a sentence a contradictory reading than Chinese.

[23]See (Labov, 1996) for examples of this sort of response.

an experiment that could distinguish between these theories.

Note that the nonindexical contextualist, to plausibly explain the results of this study, needed to invoke a relatively fine-grained notion of context. In particular, it seems that context must be able to change for a participant who sees nothing different and doesn't move. Context must thus be at least difficult to observe. Now, the nonindexical contextualist I've envisioned sticks to classical logic *at the level of extensions*. But since it's very difficult to tell when we've changed context, this means that the logic of properties we'll use to generate experimental predictions will blur across contexts. And when you blur classical logic in this way, the result is the paraconsistent logic LP. (See (Lewis, 1982) and (Brown, 1999) for details and discussion.)

On the other hand, the dialetheist view I defend in (Ripley, 200xb) holds LP to be the correct logic of vagueness even in a single context.[24] Thus, it could be quite tricky to find an experimental wedge between the two views. The key to such a wedge would come from some operationalization of the notoriously slippery term 'context'. The contextualist and the dialetheist make different predictions about what will happen in a single context. I leave this issue for future work.[25]

---

[24]NB: the dialetheist is under no obligation to use a fine-grained context, although she might find reason to.

[25]Acknowledgements...

References

Paolo Acquaviva. *The Logical Form of Negation: A Study of Operator-Variable Structures in Syntax*. Garland Publishing, New York (1997).

Anthony E. Ades & Mark J. Steedman. 'On the order of words'. *Linguistics and Philosophy* 4:517–558 (1982).

Jonas Åkerman & Patrick Greenough. 'Vagueness and non-indexical contextualism'. In Sarah Sawyer (ed.), *New Waves in the Philosophy of Language*. Palgrave Macmillan (2010).

Sam Alxatib & Jeffry Pelletier. 'The psychology of vagueness: Borderline cases and contradictions' (200x). Under submission.

Alan Ross Anderson & Nuel D. Belnap. *Entailment: The Logic of Relevance and Necessity*, volume 1. Princeton University Press, Princeton, New Jersey (1975).

Alan Ross Anderson, Nuel D. Belnap, & J. Michael Dunn. *Entailment: The Logic of Relevance and Necessity*, volume 2. Princeton University Press, Princeton, New Jersey (1992).

Joseph Aoun & Yen-hui Audrey Li. 'Scope and constituency'. *Linguistic Inquiry* 20:141–172 (1989).

Bradley Armour-Garb & JC Beall. 'Can deflationists be dialetheists?' *Journal of Philosophical Logic* pp. 593–608 (2001).

Bradley Armour-Garb & JC Beall. 'Should deflationists be dialetheists?' *Noûs* pp. 303–324 (2003).

Jay David Atlas. 'Negative adverbials, prototypical negation, and the De Morgan taxonomy'. *Journal of Semantics* 14:349–367 (1997).

Marc Authier & Lisa Reed. *Structure and Interpretation in Natural Language*. LINCOM Europa, München (1999).

Mark Baker, Kyle Johnson, & Ian Roberts. 'Passive arguments raised'. *Linguistic Inquiry* 20(2):219–251 (1989).

Ellen Barton & Ljiljana Progovac. 'Nonsententials in minimalism'. In *Ellipsis and Nonsentential Speech*, pp. 71–93. Kluwer Academic Publishers, Dordrecht, The Netherlands (2004).

Jon Barwise. 'Conditionals and conditional information'. In *The Situation in Logic*. CSLI Publications, Stanford, California (1989a).

Jon Barwise. *The Situation in Logic*. CSLI Publications, Stanford, California (1989b).

Jon Barwise. 'Situations and small worlds'. In *The Situation in Logic*. CSLI Publications, Stanford, California (1989c).

Jon Barwise. 'Information and impossibilities'. *Notre Dame Journal of Formal Logic* 38(4):488–515 (1997).

Jon Barwise & Robin Cooper. 'Generalized quantifiers and natural language'. *Linguistics and Philosophy* 4:159–219 (1981).

Jon Barwise & John Etchemendy. *The Liar: an Essay on Truth and Circularity.* Oxford University Press, Oxford (1987).

Jon Barwise & John Perry. 'Shifting situations and shaken attitudes'. *Linguistics and Philosophy* 8:105–161 (1985).

Jon Barwise & John Perry. *Situations and Attitudes.* CSLI Publications, Stanford, California (1999).

JC Beall. 'Transparent disquotationalism'. In JC Beall & Brad Armour-Garb (eds.), *Deflationism and Paradox.* Oxford, Oxford (2005).

JC Beall. 'Prolegomenon to future revenge'. In *Revenge of the Liar: New Essays on the Paradox*, pp. 1–30. Oxford University Press, Oxford (2008).

JC Beall. *Spandrels of Truth.* Oxford University Press, Oxford (2009).

JC Beall, Ross T. Brady, A. P. Hazen, Graham Priest, & Greg Restall. 'Relevant restricted quantification'. *Journal of Philosophical Logic* 35(6):587–598 (2006).

JC Beall & Mark Colyvan. 'Heaps of gluts and Hyde-ing the sorites'. *Mind* 110:401–408 (2001).

JC Beall & Bas C. van Fraassen. *Possibilities and Paradox: An Introduction to Modal and Many-valued Logic.* Oxford University Press, Oxford (2003).

JC Beall & Greg Restall. *Logical Pluralism.* Oxford University Press, Oxford (2006).

JC Beall & David Ripley. 'Analetheism and dialetheism'. *Analysis* 64(1):30–35 (2004).

Johan van Benthem. 'Correspondence theory'. In Dov M. Gabbay & Franz Guenthner (eds.), *Handbook of Philosophical Logic*, volume 2, pp. 167–247. Reidel (1984a).

Johan van Benthem. 'Review of 'On when a semantics is not a semantics' by B. J. Copeland'. *Journal of Symbolic Logic* 49(3):994–995 (1984b).

Maria Bittner. 'Cross-linguistic semantics'. *Linguistics and Philosophy* 17:53–108 (1994).

Patrick Blackburn, Maarten de Rijke, & Yde Venema. *Modal Logic.* Cambridge University Press, Cambridge (2001).

Nicolao Bonini, Daniel Osherson, Riccardo Viale, & Timothy Williamson. 'On the psychology of vague predicates'. *Mind and Language* 14:377–393 (1999).

Ross Brady. 'Rules in relevant logic — II: Formula representation'. *Studia Logica* 52:565–585 (1993).

Ross Brady. 'Rules in relevant logic — I: Semantic classification'. *Journal of Philosophical Logic* 23:111–137 (1994).

Ross Brady. *Universal Logic.* CSLI Publications, Stanford, California (2006).

Ross Brady & Andrea Meinander. 'Distribution in the logic of meaning containment and in quantum mechanics' (200x). Forthcoming.

Aldo Bressan. *A General Interpreted Modal Calculus.* Yale University Press, New Haven (1972).

Bryson Brown. 'Yes, Virginia, there really are paraconsistent logics'. *Journal of Philosophical Logic* 28:489–500 (1999).

Benjamin Bruening. 'QR obeys Superiority: Frozen scope and ACD'. *Linguistic Inquiry* pp. 233–273 (2001).

John A. Burgess & I. L. Humberstone. 'Natural deduction rules for a logic of vagueness'. *Erkenntnis* 27(2):197–229 (1987).

Rudolf Carnap. 'Meaning and synonymy in natural languages'. *Philosophical Studies* 6(3):33–47 (1955).

Rudolf Carnap. *Meaning and Necessity*. University of Chicago Press, Chicago (1956).

Bob Carpenter. *Type-Logical Semantics*. MIT Press, Cambridge, Massachusetts (1997).

David Chalmers. 'Epistemic two-dimensional semantics'. *Philosophical Studies* 118:153–226 (2004).

David Chalmers. 'Propositions and attitude ascriptions: A Fregean account'. *Noûs* (200x). Forthcoming.

Brian F. Chellas. *Modal Logic: an Introduction*. Cambridge University Press, Cambridge (1980).

Noam Chomsky. *Rules and Representations*. Columbia University Press, New York (1980).

Noam Chomsky. *Lectures on Government and Binding*. Foris, Dordrecht (1981).

Noam Chomsky. *The Minimalist Program*. MIT Press, Cambridge, Massachusetts (1995).

Noam Chomsky. 'Beyond explanatory adequacy'. In Adriana Belletti (ed.), *Structures and Beyond*, pp. 104–131. Oxford University Press, Oxford (2004).

Alonzo Church. 'A formulation of the simple theory of types'. *Journal of Symbolic Logic* 5:56–68 (1940).

Alonzo Church. *The Calculi of Lambda-Conversion*. Princeton University Press, Princeton, New Jersey (1941).

Richard A. Coates. 'Properhood'. *Language* 82(2):356–382 (2006).

Mark Colyvan. 'Vagueness and truth'. In Heather Duke (ed.), *From Truth to Reality: New Essays in Logic and Metaphysics*. Routledge, New York (2008).

B.J. Copeland. 'On when a semantics is not a semantics'. *Journal of Philosophical Logic* pp. 399–413 (1979).

Newton da Costa. 'On the theory of inconsistent formal systems'. *Notre Dame Journal of Formal Logic* 15:497–510 (1974).

Maxwell J. Cresswell. *Structured Meanings*. MIT Press, Cambridge, Massachusetts (1985).

Maxwell J. Cresswell & Arnim Von Stechow. 'De re belief generalized'. *Linguistics and Philosophy* 5:503–535 (1982).

Mark Crimmins & John Perry. 'The prince and the phone booth: Reporting puzzling beliefs'. *Journal of Philosophy* 86(12):685–711 (1989).

Haskell B. Curry. *Foundations of Mathematical Logic*. Dover, Mineola, New York (1977).

Donald Davidson. 'The logical form of action sentences'. In Nicholas Rescher (ed.), *The Logic of Decision and Action*. University of Pittsburgh Press, Pittsburgh, Pennsylvania (1967a).

Donald Davidson. 'Truth and meaning'. *Synthèse* 17:304–323 (1967b).

Donald Davidson. 'On saying that'. *Synthèse* 19:130–146 (1968).

Donald Davidson. 'Semantics for natural languages'. In *Inquiries into Truth and Interpretation*. Oxford University Press, Oxford (2001).

John Divers. *Possible Worlds*. Routledge, London (2002).

David Dowty. 'Thematic proto-roles and argument selection'. *Language* 67(3):547–619 (1991).

J. Michael Dunn. 'Gaggle theory: An abstraction of galois connections and residuation with applications to negations and various logical operators'. In *Logics in AI, Proceedings of European Workshop JELIA 1990*, pp. 31–51. LNCS, Berlin (1990).

J. Michael Dunn. 'Star and perp: Two treatments of negation'. *Philosophical Perspectives* 7:331–357 (1993).

J. Michael Dunn & Gary M. Hardegree. *Algebraic Methods in Philosophical Logic*. Oxford University Press, Oxford (2001).

J. Michael Dunn & Chunlai Zhou. 'Negation in gaggle theory'. *Studia Logica* 80:235–264 (2005).

Walter Edelberg. 'Propositions, circumstances, objects'. *Journal of Philosophical Logic* 23(1):1–34 (1994).

Matti Eklund. 'What vagueness consists in'. *Philosophical Studies* 125(1):27–60 (2005).

Paul Elbourne. 'Why propositions might be sets of truth-supporting circumstances'. *Journal of Philosophical Logic* (200x). Forthcoming.

Hartry Field. 'Tarski's theory of truth'. *Journal of Philosophy* 69:347–375 (1973).

Hartry Field. *Saving Truth from Paradox*. Oxford University Press, Oxford (2008).

Kit Fine. 'Vagueness, truth, and logic'. *Synthèse* 30:265–300 (1975).

Kai von Fintel. 'NPI licensing, Strawson entailment, and context dependency'. *Journal of Semantics* 16:97–148 (1999).

Kai von Fintel & Irene Heim. 'Intensional semantics' (2007). Unpublished. Available at http://tinyurl.com/intensional.

Graeme Forbes. *Attitude Problems: An Essay on Linguistic Intensionality*. Oxford University Press, Oxford (2006).

Danny Fox. 'On logical form'. In Randall Hendrick (ed.), *Minimalist Syntax*. Blackwell, Oxford (2003).

Gottlob Frege. 'On sense and nominatum'. In A. P. Martinich (ed.), *The Philosophy of Language*. Oxford University Press, Oxford, 4 edition (2001).

L.T.F. Gamut. *Logic, Language, and Meaning, Volume 1: Introduction to Logic*. University of Chicago Press, Chicago (1991a).

L.T.F. Gamut. *Logic, Language, and Meaning, Volume 2: Intensional Logic and Logical Grammar*. University of Chicago Press, Chicago (1991b).

Bart Geurts. 'The mechanisms of denial'. *Language* 74:274–307 (1998).

Leonard Goddard & Richard Routley. *The Logic of Significance and Context, vol. 1*. Scottish Academic Press, Edinburgh (1973).

Elena Guerzoni & Yael Sharvit. 'A question of strength: on NPIs in interrogative clauses'. *Linguistics and Philosophy* (forthcoming).

Liliane Haegeman. *Introduction to Government and Binding Theory*. Blackwell, Oxford, 2nd edition (1994).

Liliane Haegeman. *The Syntax of Negation.* Cambridge University Press, Cambridge, England (1995).

Liliane Haegeman. 'Negative preposing, negative inversion, and the split CP'. In Laurence R. Horn & Yasuhiko Kato (eds.), *Negation and Polarity: Syntactic and Semantic Perspectives.* Oxford University Press, Oxford (2000).

Liliane Haegeman & Raffaella Zanuttini. 'Negative heads and the Neg criterion'. *Linguistic Review* 8:233–52 (1991).

James A. Hampton. 'Typicality, graded membership, and vagueness'. *Cognitive Science* 31:355–383 (2007).

Helle Hvid Hansen. *Monotonic Modal Logics.* Master's thesis, University of Amsterdam, Amsterdam (2003).

Bengt Hansson & Peter Gärdenfors. 'A guide to intensional semantics'. In *Modality, Morality and Other Problems of Sense and Nonsense: Essays dedicated to Sören Halldén*, pp. 151–167. CWK Gleerup Bokförlag, Lund (1973).

Chrysafis Hartonas. 'Order-duality, negation, and lattice representation'. In Heinrich Wansing (ed.), *Negation: A Notion in Focus*, pp. 27–36. Walter de Gruyter, Berlin (1996).

Allen P. Hazen. 'Subminimal negation' (1992). Philosophy Department Preprint, University of Melbourne.

Irene Heim & Angelika Kratzer. *Semantics in Generative Grammar.* Blackwell, Oxford (1998).

Leon Henkin. 'Completeness in the theory of types'. *Journal of Symbolic Logic* 15:81–91 (1950).

Leon Henkin, J. D. Monk, & Alfred Tarski. *Cylindric Algebras: Part I.* Elsevier, North Holland (1985).

Terence Horgan. 'Robust vagueness and the forced-march sorites'. *Philosophical Perspectives* 8:159–188 (1994).

Laurence R. Horn. 'Metalinguistic negation and pragmatic ambiguity'. *Language* 61:121–174 (1985).

Laurence R. Horn. *A Natural History of Negation.* CSLI Publications, Stanford, California (2001).

Paul Horwich. *Meaning.* Oxford University Press, Oxford (1998).

G. E. Hughes & M. J. Cresswell. *A New Introduction to Modal Logic.* Routledge, New York (1996).

Brian Huss & Yi Yingli. 'Ethnologic: Empirical results' (2007). Presented at the 2007 meeting of the Canadian Philosophical Association (Saskatoon, SK).

Dominic Hyde. 'From heaps and gaps to heaps of gluts'. *Mind* 106:641–660 (1997).

Dominic Hyde. 'A reply to Beall & Colyvan'. *Mind* 110:409–411 (2001).

Dominic Hyde. 'The prospects of a paraconsistent approach to vagueness'. In *Contemporary Research on Vagueness.* Oxford University Press, Oxford (2006).

Dominic Hyde. *Vagueness, Logic, and Ontology.* Ashgate, Hampshire, England (2008).

Dominic Hyde & Mark Colyvan. 'Paraconsistent vagueness: Why not?' *Australasian Journal of Logic* 6:107–121 (2008).

Osvaldo A. Jaeggli. 'Passive'. *Linguistic Inquiry* 17(4):587–622 (1986).

Bjarni Jónsson & Alfred Tarski. 'Boolean algebras with operators, part I'. *American Journal of Mathematics* 73(4):891–939 (1951).

Bjarni Jónsson & Alfred Tarski. 'Boolean algebras with operators, part II'. *American Journal of Mathematics* 74(1):127–162 (1952).

David Kaplan. 'Demonstratives'. In Joseph Almog, John Perry, & Howard K. Wettstein (eds.), *Themes From Kaplan*, pp. 481–564. Oxford University Press, Oxford (1989).

Yasuhiko Kato. 'Interpretive asymmetries of negation'. In Laurence R. Horn & Yasuhiko Kato (eds.), *Negation and Polarity: Syntactic and Semantic Perspectives*. Oxford University Press, Oxford (2000).

Richard Kayne. *Connectedness and Binary Branching*. Foris, Dordrecht, The Netherlands (1984).

Rosanna Keefe. *Theories of Vagueness*. Cambridge University Press, Cambridge (2000).

Jeffrey C. King. *The Nature and Structure of Content*. Oxford University Press, Oxford (2007).

Ewan Klein & Ivan Sag. 'Type-driven translation'. *Linguistics and Philosophy* 8:163–201 (1985).

Edward S. Klima. 'Negation in English'. In Jerry A. Fodor & Jerrold J. Katz (eds.), *The Structure of Language*. Prentice-Hall, Englewood Cliffs, New Jersey (1964).

Hilda Koopman & Dominique Sportiche. 'The position of subjects'. *Lingua* 85:211–258 (1991).

Angelika Kratzer. 'What 'must' and 'can' must and can mean'. *Linguistics and Philosophy* 1:337–355 (1977).

Angelika Kratzer. 'Building statives'. In *Berkeley Linguistic Society 26* (2000).

Saul Kripke. 'Outline of a theory of truth'. *Journal of Philosophy* 72(19):690–716 (1975).

Saul Kripke. 'A puzzle about belief'. In Avishai Margalit (ed.), *Meaning and Use*. D. Reidel, Dordrecht, The Netherlands (1979).

Saul Kripke. *Naming and Necessity*. Harvard University Press, Cambridge, Massachusetts (1980).

Steven T. Kuhn. 'Quantifiers as modal operators'. *Studia Logica* 39(2–3):145–158 (1980).

William Labov. 'Negative attraction and negative concord'. *Language* 48(4):773–818 (1972).

William Labov. 'When intuitions fail'. In L. McNair, K. Singer, & M. Aucon (eds.), *Papers from the Parasession on Theory and Data in Linguistics, Chicago Linguistic Society 32*, pp. 77–106 (1996).

William Allen Ladusaw. *Polarity Sensitivity as Inherent Scope Relations*. Ph.D. thesis, University of Texas at Austin, Austin, TX (1980). Distributed by Indiana University Linguistics Club.

William Allen Ladusaw. 'Negative concord and 'mode of judgement''. In Heinrich Wansing (ed.), *Negation: A Notion in Focus*. Walter de Gruyter, Berlin (1996).

Joachim Lambek. 'The mathematics of sentence structure'. *American Mathematical Monthly* 65:154–169 (1958).

Richard K. Larson. 'On the double object construction'. *Linguistic Inquiry* 19(3):335–391 (1988).

Howard Lasnik. *Analyses of Negation in English*. Ph.D. thesis, MIT (1972).

Howard Lasnik, Juan Urigereka, & Cedric Boeckx. *A Course in Minimalist Syntax: Foundations and Prospects.* Blackwell, Oxford (2005).

Ernest Lepore. 'What model-theoretic semantics cannot do'. *Synthèse* 54:167–187 (1983).

Clarence Irving Lewis & Cooper H. Langford. *Symbolic Logic.* Dover, Mineola, New York (1959).

David Lewis. 'General semantics'. *Synthèse* 22:18–67 (1970).

David Lewis. *Couterfactuals.* Blackwell, Oxford (1973).

David Lewis. "Tensions'. In Milton K. Munitz & Peter K. Unger (eds.), *Semantics and Philosophy.* NYU Press, New York (1974).

David Lewis. 'Score-keeping in a language game'. *Journal of Philosophical Logic* 8:339–359 (1979).

David Lewis. 'Logic for equivocators'. *Noûs* 16(3):431–41 (1982).

David Lewis. *On the Plurality of Worlds.* Blackwell, Oxford (1986).

Martha C. Linebarger. 'Negative polarity and grammatical representation'. *Linguistics and Philosophy* 10:325–387 (1987).

Greg Littman & Keith Simmons. 'A critique of dialetheism'. In Graham Priest, JC Beall, & Bradley Armour-Garb (eds.), *The Law of Non-Contradiction: New Philosophical Essays.* Oxford University Press, Oxford (2004).

William G. Lycan. *Modality and Meaning.* Kluwer Academic Publishing (1994).

William G. Lycan. *Real Conditionals.* Oxford University Press, Oxford (2001).

John MacFarlane. 'Nonindexical contextualism'. *Synthèse* 166:231–250 (2009).

K. F. Machina. 'Vague predicates'. *American Philosophical Quarterly* 9:225–233 (1972).

Edwin D. Mares. 'Who's afraid of impossible worlds?' *Notre Dame Journal of Formal Logic* 38(4):516–526 (1997).

Edwin D. Mares. *Relevant Logic: A Philosophical Interpretation.* Cambridge University Press, Cambridge (2004).

Edwin D. Mares & André Fuhrmann. 'A relevant theory of conditionals'. *Journal of Philosophical Logic* 24:645–665 (1995).

Tim Maudlin. *Truth and Paradox.* Oxford University Press, Oxford (2004).

Robert May. *Logical Form: Its Structure and Derivation.* MIT Press, Cambridge, Massachusetts (1985).

James D. McCawley. *Everything that Linguists have Always Wanted to Know about Logic (but were ashamed to ask).* University of Chicago Press, Chicago (1993).

Joseé M. Méndez, Francisco Salto, & Pedro Méndez R. 'Four kinds of subminimal negation within the context of the basic positive logic B+'. *Logique et Analyse* 177–178:119–128 (2002).

Richard Montague. 'The proper treatment of quantification in ordinary English'. In Paul Portner & Barbara H. Partee (eds.), *Formal Semantics: The Essential Readings*, pp. 17–34. Blackwell, Oxford (2002).

Stephen Neale. *Descriptions.* MIT Press, Cambridge, Massachusetts (1990).

STEPHEN NEALE. 'Pragmatism and binding'. In ZOLTÁN GENDLER SZABÓ (ed.), *Semantics versus Pragmatics*. Oxford University Press, Oxford (2005).

TERENCE PARSONS. 'Assertion, denial, and the liar paradox'. *Journal of Philosophical Logic* 13:137—152 (1984).

BARBARA H. PARTEE. 'Opacity and scope'. In MILTON K. MUNITZ & PETER K. UNGER (eds.), *Semantics and Philosophy*. NYU Press, New York (1974).

BARBARA H. PARTEE. 'Possible world semantics and linguistic theory'. *Monist* 60:303–326 (1977).

BARBARA H. PARTEE. 'Situations, worlds, and contexts'. *Linguistics and Philosophy* 8:53–58 (1985).

BARBARA H. PARTEE. 'Noun phrase interpretation and type-shifting principles'. In JEROEN GROENENDIJK, DICK DE JONGH, & MARTIN STOKHOF (eds.), *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*, pp. 115–143. Foris, Dordrecht, The Netherlands (1987).

BARBARA H. PARTEE, ALICE TER MEULEN, & ROBERT E. WALL. *Mathematical Methods in Linguistics*. Kluwer Academic Publishers, Dordrecht, The Netherlands (1993).

BARBARA H. PARTEE & MATS ROOTH. 'Generalized conjunction and type ambiguity'. In RAINER BÄUERLE, CHRISTOPH SCHWARZ, & ARNIM VON STECHOW (eds.), *Meaning, Use and the Interpretation of Language*, pp. 361–393. Walter de Gruyter & Co., Berlin (1983).

KAIPING PENG & RICHARD NISBETT. 'Culture, dialectics, and reasoning about contradiction'. *American Psychologist* 54:741–54 (1999).

JOHN PERRY. 'The problem of the essential indexical'. *Noûs* 13:3–21 (1979).

JOHN PERRY. 'From worlds to situations'. *Journal of Philosophical Logic* 15:83–107 (1986).

JOHN PERRY. 'Possible worlds and subject matter'. In STURE ALLÉN (ed.), *Possible Worlds in Humanities, Arts, and Sciences: Proceedings of Nobel Symposium 65*. Walter de Gruyer, Berlin/New York (1989).

J. Y. POLLOCK. 'Verb movement, UG, and the structure of IP'. *Linguistic Inquiry* 20:365–424 (1989).

GRAHAM PRIEST. 'Logic of paradox'. *Journal of Philosophical Logic* 8:219–241 (1979).

GRAHAM PRIEST. 'Semantic closure'. *Studia Logica* 43:117–129 (1982).

GRAHAM PRIEST. 'Hyper-contradictions'. *Logique et Analyse* 107:237–243 (1984).

GRAHAM PRIEST. 'Minimally inconsistent LP'. *Studia Logica* pp. 321–331 (1991).

GRAHAM PRIEST. 'The import of inclosure: Some comments on Grattan–Guinness'. *Mind* 107(428):835–840 (1998).

GRAHAM PRIEST. 'Objects of thought'. *Australasian Journal of Philosophy* 78(4):494–502 (2000).

GRAHAM PRIEST. *An Introduction to Non-Classical Logic*. Cambridge University Press, Cambridge (2001).

GRAHAM PRIEST. *Beyond the Limits of Thought*. Oxford University Press, Oxford (2002).

GRAHAM PRIEST. 'What's so bad about contradictions?' In GRAHAM PRIEST, JC BEALL, & BRADLEY ARMOUR-GARB (eds.), *The Law of Non-Contradiction: New Philosophical Essays*. Oxford University Press, Oxford (2004).

GRAHAM PRIEST. *Towards Non-Being: The Logic and Metaphysics of Intentionality.* Oxford University Press, Oxford (2005).

GRAHAM PRIEST. *Doubt Truth to be a Liar.* Oxford University Press, Oxford (2006a).

GRAHAM PRIEST. *In Contradiction.* Oxford University Press, Oxford (2006b).

GRAHAM PRIEST. *An Introduction to Non-Classical Logic: From If to Is.* Cambridge University Press, Cambridge, 2 edition (2008).

GRAHAM PRIEST. 'Vague inclosures' (ms.).

GRAHAM PRIEST & RICHARD SYLVAN. 'Simplified semantics for basic relevant logics'. *Journal of Philosophical Logic* 21:217–232 (1992).

LJILJANA PROGOVAC. 'Negative polarity: A semantico-syntactic approach'. *Lingua* 86:271–299 (1992).

LJILJANA PROGOVAC. 'Negation and comp'. *Rivista di Linguistica* 5(2):329–347 (1993a).

LJILJANA PROGOVAC. 'Negative polarity: Entailment and binding'. *Linguistics and Philosophy* 16:149–180 (1993b).

WILLARD VAN ORMAN QUINE. *Word and Object.* MIT Press, Cambridge, Massachusetts (1960).

WILLARD VAN ORMAN QUINE. 'Truth by convention'. In *The Ways of Paradox and Other Essays.* Random House, New York (1966).

DIANA RAFFMAN. 'Vagueness without paradox'. *The Philosophical Review* 103(1):41–74 (1994).

DIANA RAFFMAN. 'Vagueness and context-relativity'. *Philosophical Studies* 81:175–192 (1996).

DIANA RAFFMAN. 'How to understand contextualism about vagueness'. *Analysis* 65(3):244–248 (2005).

VEIKKO RANTALA. 'Quantified modal logic: Non-normal worlds and propositional attitudes'. *Studia Logica* 41(1):41–65 (1982).

FRANÇOIS RECANATI. *Oratio Obliqua, Oratio Recta.* MIT Press, Cambridge, Massachusetts (2000).

FRANÇOIS RECANATI. *Literal Meaning.* Cambridge University Press, Cambridge (2004).

FRANÇOIS RECANATI. 'It is raining (somewhere)'. *Linguistics and Philosophy* 30:123–146 (2007).

M.K. RENNIE. *Some Uses of Type Theory in the Analysis of Language.* Research School of Social Sciences, Australian National University (1974).

GREG RESTALL. 'Simplified semantics for relevant logics (and some of their rivals)'. *Journal of Philosophical Logic* 22:481–511 (1993).

GREG RESTALL. *On Logics Without Contraction.* Ph.D. thesis, The University of Queensland (1994).

GREG RESTALL. 'Information flow and relevant logics'. In JERRY SELIGMAN & DAG WESTERSTÅHL (eds.), *Logic, Language, and Computation: the 1994 Moraga Proceedings*, pp. 463–477. CSLI Publications, Stanford, California (1995).

GREG RESTALL. 'Ways things can't be'. *Notre Dame Journal of Formal Logic* 38(4):583–596 (1997).

GREG RESTALL. 'Negation in relevant logics: How I stopped worrying and learned to love the Routley star'. In DOV GABBAY & HEINRICH WANSING (eds.), *What is Negation?*, pp. 53–76. Kluwer Academic Publishers, Dordrecht, The Netherlands (1999).

Greg Restall. 'Defining double negation elimination'. *Logic Journal of the IGPL* 8(6):853–860 (2000a).

Greg Restall. *An Introduction to Substructural Logics*. Routledge, London (2000b).

Greg Restall. 'Logics, situations, and channels'. *Journal of Cognitive Science* 6:125–150 (2005a).

Greg Restall. 'Multiple conclusions'. In Petr Hajek, Luis Valdes-Villanueva, & Dag Westerståhl (eds.), *Logic, Methodology, and Philosophy of Science: Proceedings of the Twelfth International Congress*, pp. 189–205. Kings' College Publications (2005b).

Mark Richard. 'Attitudes in context'. *Linguistics and Philosophy* 16:123–148 (1993).

David Ripley. 'Contradictions at the borders' (200xa). Under submission.

David Ripley. 'Sorting out the sorites'. In Francesco Berto, Edwin Mares, & Koji Tanaka (eds.), *Paraconsistent Logic (tentative title)* (200xb). Forthcoming.

Luigi Rizzi. *Relativized Minimality*. MIT Press, Cambridge, Massachusetts (1990).

Luigi Rizzi (ed.). *The Structure of CP and IP*. Oxford University Press, Oxford (2004).

Richard Routley. *Exploring Meinong's Jungle and Beyond*. Research School of Social Sciences, Australian National University (1980).

Richard Routley. 'The American plan completed: Alternative classical-style semantics, without stars, for relevant and paraconsistent logics'. *Studia Logica* 43:131–158 (1984).

Richard Routley & Robert K. Meyer. 'Towards a general semantic theory of implication and conditionals I'. *Reports on Mathematical Logic* 4:67–89 (1975).

Richard Routley & Robert K. Meyer. 'Every sentential logic has a two-valued world semantics'. *Logique et Analyse* 19:174–194 (1976).

Richard Routley & Robert K. Meyer. 'Towards a general semantic theory of implication and conditionals II'. *Reports on Mathematical Logic* 9:47–62 (1977).

Richard Routley, Robert K. Meyer, Val Plumwood, & Ross T. Brady. *Relevant Logics and their Rivals 1*. Ridgeview, Atascadero, California (1982).

Richard Routley & Valerie Routley. 'The semantics of first degree entailment'. *Noûs* 6(4):335–359 (1972).

Richard Routley & Valerie Routley. 'The role of inconsistent and incomplete theories in the semantics of belief'. *Communication and Cognition* 8:185–235 (1975).

Richard Routley & Valerie Routley. 'Negation and contradiction'. *Revista Colombiana de Matemáticas* 19:201–231 (1985).

Stewart Shapiro. *Vagueness in Context*. Oxford University Press, Oxford (2006).

Nicholas J. J. Smith. *Vagueness and Degrees of Truth*. Oxford University Press, Oxford (2008).

Scott Soames. 'Direct reference, propositional attitudes, and semantic content'. *Philosophical Topics* 15(1):47–87 (1987).

Scott Soames. *Understanding Truth*. Oxford University Press, Oxford (1998).

Scott Soames. 'Replies'. *Philosophy and Phenomenological Research* 65(2):429–452 (2002).

Scott Soames. 'Why propositions cannot be sets of truth-supporting circumstances'. *Journal of Philosophical Logic* 37:267–276 (2008).

ROY SORENSEN. *Vagueness and Contradiction.* Oxford University Press, Oxford (2001).

ROBERT J. STAINTON. 'In defense of non-sentential assertion'. In ZOLTÁN GENDLER SZABÓ (ed.), *Semantics versus Pragmatics*, pp. 383–457. Oxford University Press, Oxford (2005).

ROBERT STALNAKER. 'Pragmatic presuppositions'. In MILTON K. MUNITZ & PETER UNGER (eds.), *Semantics and Philosophy.* New York University Press, New York (1974).

ROBERT STALNAKER. 'Assertion'. *Syntax and Semantics* 9 (1978).

JASON STANLEY. 'Context, interest-relativity, and the sorites'. *Analysis* 63(4):269–280 (2003).

JASON STANLEY & ZOLTÁN GENDLER SZABÓ. 'On quantifier domain restriction'. In *Language in Context*, pp. 69–110. Oxford University Press, Oxford (2007).

PETER F. STRAWSON. 'On referring'. *Mind* 59:320–344 (1950).

M. SUNER. 'NPIs, island effects, and resumptive *no*' (1993). MS Cornell University.

JAMIE TAPPENDEN. 'The liar and sorites paradoxes: Toward a unified treatment'. *The Journal of Philosophy* 90(11):551–577 (1993).

ALFRED TARSKI. 'The semantic conception of truth'. *Philosophy and Phenomenological Research* 4:341–376 (1944).

ALFRED TARSKI. 'The concept of truth in formalized languages'. In *Logic, Semantics, Metamathematics.* Hackett Publishing Company, Indianapolis, Indiana (1983).

PETER UNGER. 'There are no ordinary things'. *Synthèse* pp. 117–154 (1979).

ALASDAIR URQUHART. 'Semantics for relevant logics'. *Journal of Symbolic Logic* 37(1):159–169 (1972).

ALASDAIR URQUHART. 'Distributive lattices with a dual homomorphic operation'. *Studia Logica* 38(2):201–209 (1979).

DIMITER VAKARELOV. 'Consistency, completeness, and negation'. In GRAHAM PRIEST, RICHARD ROUTLEY, & JEAN NORMAN (eds.), *Paraconsistent Logic: Essays on the Inconsistent*, pp. 328–363. Philosophia Verlag, Munich (1989).

ACHILLE VARZI. 'Supervaluationism and paraconsistency'. In DIDERIK BATENS, CHRIS MORTENSEN, GRAHAM PRIEST, & JEAN-PAUL VAN BENDEGEM (eds.), *Frontiers in Paraconsistent Logic*, pp. 279–297. Research Studies Press, Baldock (2000).

ACHILLE VARZI. 'On logical relativity'. *Philosophical Issues* 12:197–219 (2002).

ACHILLE VARZI. 'Supervaluationism and its logics'. *Mind* 116:633–676 (2007).

PHILIP WELCH. 'Ultimate truth *vis-à-vis* stable truth'. *Review of Symbolic Logic* 1(1):126–142 (2008).

TIMOTHY WILLIAMSON. *Vagueness.* Routledge, London (1994).

DEIRDRE WILSON & DAN SPERBER. 'Truthfulness and relevance'. *Mind* 111:583–632 (2002).

YOAD WINTER. *Flexibility Principles in Boolean Semantics.* MIT Press, Cambridge, Massachusetts (2002).

TON VAN DER WOUDEN. *Negative Contexts: Collocation, Polarity, and Multiple Negation.* Routledge, London (1997).

CRISPIN WRIGHT. 'On the coherence of vague predicates'. *Synthèse* 30:325–365 (1975).

Takashi Yagisawa. 'Beyond possible worlds'. *Philosophical Studies* 53(2):175–204 (1988).

Raffaela Zanuttini. *Negation and Clausal Structure*. Oxford University Press, Oxford (1997).

Raffaella Zanuttini. 'Sentential negation'. In Mark Baltin & Chris Collins (eds.), *The Handbook of Contemporary Syntactic Theory*, pp. 511–535. Blackwell, Oxford (2001).

Frans Zwarts. 'Three types of polarity'. In Fritz Hamm & Erhard Hinrichs (eds.), *Plurality and Quantification*, pp. 177–238. Kluwer Academic Publishers, Dordrecht, The Netherlands (1998).