Replacing Truth

Kevin Scharp

The Ohio State University

Version 3.5 29 Sept 2010 214,814 words

Contents

Part II: The Realignment

List of Information Boxes

Acknowledgements

I started thinking about truth and paradoxes in 1994 at about the time I realized I wanted to go to graduate school in philosophy. Needless to say, I have talked about the ideas presented here with numerous people between then and now. This book is also based on my dissertation, completed at the University of Pittsburgh in 2005. Bob Brandom was the director and he, more than anyone else, has had an impact on the shape of it. His ideas moved me before I even arrived at Pittsburgh, and his encouragement, constructive criticism, and advice will stay with me long into the future. The other members of my committee, John McDowell, Anil Gupta, and Hartry Field are also owed a huge dept of gratitude. I have learned so much from each of them.

Other people along the way who have had a positive effect on the ideas here and whom I thank include (in roughly chronological order), William Falcão Kerr, Bradford Cokelet, Rena Samole, Joel Anderson, Pauline Kleingeld, Todd Reisinger, Geoff Lampsa, Ben Henwood, Heather Matula, Fred Heller, John Koethe, Bob Schwartz, Seth Ard, Peter Guildenhuys, Sebastian Rand, Jason Leddington, Josh Andresen, Daniel Nolan, Ron Loeffler, Arthur Fine, Cristina Lafont, Michael Williams, Jürgen Habermas, Graham Hubbs, Jamsheed Siyar, Hamutal Dotan, Alp Aker, Sam Floyd, Lionel Shapiro, John Morrison, Joe Camp, Cian Dorr, Nuel Belnap, Mark Wilson, Susanna Schellenberg, Steve Leeds, Bryan Weaver, Volker Halbach, Steve Yablo, Michael Miller, Owen King, Eric Carter, Kevin Connor, Sal Florio, Stewart Shapiro, Neil Tennant, Robert Kraut, Ben Caplan, David Sanson, Craige Roberts, Declan Smithies, Abe Roth, Richard Samuels, Gabriel Uzquiano, Judith Tonhauser, Jc Beall, Stephen Schiffer, Matti Eklund, Michael Kremer, Mark Lance, John MacFarlane, Dan Scharp, Dana Scott, Brandon Fitelson, Kenny Easwaren, Crispin Wright, Marcus Rossberg, Philip Ebert, Graham Priest, Kit Fine, John Hawthorne, Roy Cook, Ori Belkind, Doug Patterson, Huw Price, Patrick Greenough, Harvey Friedman, Joe Shipman, Solomon Feferman,

Albert Visser, Jeff King, Keith Simmons, Dorit Bar-On, David Ripley, Berit Brogaard, Michael Lynch, David Chalmers, Greg Restall, Henry Jackman, Pat Suppes, Hannes Leitgeb, Gabriel Lakeman, Zachary Hamm, and anyone I have overlooked.

I taught a graduate seminar on this material at The Ohio State University in Spring 2009; the participants provided essential feedback on many of the ideas, arguments, and theories in this book.

Some of this work has been given as talks at University of Pittsburgh, University of Georgia, State University of New York at New Paltz, Wayne State University, The University of Munich, University of Melbourne, University of North Carolina Chappell Hill, University of California at Berkeley, The Ohio State University, University of Tilburg, University of Virginia, University of St. Andrews, the Australasian Association of Philosophy Meeting in Melbourne, and the Pacific American Philosophical Association Meeting in San Francisco.

Parts of "Aletheic Vengeance" made it into Chapters Eight and Twelve; thank you to Oxford University Press for permission to reprint them. Parts of "Replacing Truth" made it into Chapters Eleven and Twelve; thank you to *Inquiry* for permission to include it.

Finally, I want to express my appreciation to the members of my family for their love and for their help along the way. Thank you to Alison Duncan Kerr for her presence, her passion, her insight, and her unwavering support; my life would be unrecognizable without her.

Conventions

Single quotes are used to form the names of linguistic expressions; e.g., 'Kevin' is the name of Kevin.

Individual constants are in the same font as regular text; e.g., p is a sentence.

Corner quotes are used in conjunction with names for complex sentences; e.g., '[~p]' is a name of the negation of p.

Schematic variables are in italics; e.g., the law of excluded middle is: *p* or ~*p*.

Sentential variables are Ariel font; e.g., John believes that p.

Angle brackets are used to form names in conjunction with sentential variables; e.g., ⟨p⟩ is true iff p.

Italics are used to specify meanings of linguistic expressions; e.g., 'beer' means *beer*.

Supposing *truth* is a woman – what then? Are there not grounds for the suspicion that all philosophers, insofar as they were dogmatists, have been very inexpert about women? That the gruesome seriousness, the clumsy obtrusiveness with which they have usually approached truth so far have been awkward and very improper methods for winning a woman's heart?

—Nietzsche

Dude: Look, man, I've got certain information, all right? Certain things have come to light. And, you know, has it ever occurred to you, that, instead of, uh, you know, running around, uh, uh, blaming me, you know, given the nature of all this new shit, you know, I-I-I-I … this could be a-a-a-a lot more, uh, uh, uh, uh, uh, uh, complex, I mean, it's not just, it might not be just such a simple … uh, you know?

Lebowski: What in God's holy name are you blathering about?

Dude: I'll tell you what I'm blathering about … I've got information man! New shit has come to light! And shit, man … she *kidnapped herself*.

—The Big Lebowski

Introduction

Although the concept of truth has a venerable history, it also has a dark side that has been known for millennia: it gives rise to nasty paradoxes, the most famous of which is the liar paradox.[1] Despite the fact that these paradoxes do not surface much in everyday conversations, they pose a serious threat to us. We have discovered that they inhibit our ability to explain our own rational behavior, and the problem is so acute that researchers who study language, reasoning, and thought avoid truth like the plague; for the paradoxes of truth wreak havoc on our attempts to understand these aspects of ourselves. The simple fact that we possess this concept has become an impediment to our attempts to understand ourselves. In this work, I argue that these paradoxes are symptoms of an intrinsic defect in the concept of truth; for this reason we should replace truth for certain purposes.

Truth has had its detractors over the years. From Protagoras to Richard Rorty, thinkers have tried to downplay its importance or eliminate it altogether.[2] My reasons are wholly distinct from theirs. The case against truth contained in these pages has nothing to do with relativism or postmodernism.[3] Rather, it is because of truth's utility, value, and importance that it needs to be replaced. If it were just an antiquated ideal that enlightened agents should discard, then there would be no point in replacing it. It is an unfortunate fact that its utility, value, and importance come at a high price.

---

[11] For example, if we use the name 'Sentence (1)' for the sentence 'Sentence (1) is false', then that very sentence says of itself that it is false. We can reason intuitively that if it is true, then what it says is true, namely, that it is false. So it is true that it is false, or, more directly, it is false. On the other hand, if it is false, then what it says is false, namely that it is false. So it is false that it is false, or, more directly, it is true. Thus, we derive that it is false from the assumption that it is true, and we derive that it is true from the assumption that it is false. It takes just a couple of steps from here to the claim that Sentence (1) is both true and false. In what follows we will go through all the steps in this reasoning and the other paradoxes associated with truth in detail.

[2] See Blackburn (2005) for a discussion of this tradition.

[3] Although I end up suggesting that truth has certain features that bear a superficial similarity to those claimed by relativists, my view has none of the radically subjectivist consequences.

The problems caused by truth are severe, but they are *not* unprecedented. Once we understand that the source of the paradoxes is a conceptual defect, we can do what we have always done in these situations—replace the offender with one or more concepts, at least for certain purposes, that are free of defects. I offer a team of concepts that, together, can do truth's job without the cost of paradoxes. In addition, they can be used in an explanation of our defective concept of truth itself, thus freeing us from our predicament.

I want it to be clear, right from the start, that I do not advocate *eliminating* truth from our conceptual repertoire. I am not trying to persuade people to stop using the word 'true'. For most purposes, the risk posed by our concept of truth is negligible; so it is reasonable to use truth, despite its defect, in most situations. Only those engaged in trying to explain our thought or language will so much as notice the change. Although this revolution will be relatively quiet, it should have a significant impact on the way we think about ourselves at the most fundamental level.

0.1 Methodology

0.1.1 Some Remarks on Style

I find it unfortunate that so much analytic philosophy is written so as to be inaccessible to non-philosophers. Instead of writing only for fellow philosophers or philosophy graduate students, I have tried to write this book so that a dedicated and intelligent non-specialist could read it and get the general idea. To be sure, the book engages with many disputes that will be familiar only to professional philosophers, but they are introduced in a way that should allow others to follow along. As such, I hope that it might serve well to introduce someone to the contemporary philosophical debates about truth. Moreover, I have attempted to tie the relatively insulated debates on truth to broader issues in philosophy in order to give those who are unfamiliar with recent developments in this area a sense of just how fertile it is. I want to emphasize, however, that this book is not

primarily a survey or an introductory text—it argues for a particular theory of truth. This theory has not, as far as I know, been articulated before. However, it is necessary to understand where we are right now in our understanding of truth before we can see the right path forward.

There have been, to put it mildly, *many* suggestions for how to deal with the problems that constitute the focus of this work. I review and engage with much of it in what follows, but the chief motivation for the approach I advocate is not literature-driven; although the approach presented here is better than the existing alternatives, that fact is not the sole or even primary justification for it. Instead, I lay out what I take to be the most important phenomena associated with truth and the paradoxes to which it gives rise. It is by appreciating these phenomena that one can see what is causing the problems and how best to deal with them.

0.1.2 The Two Camps in the Analytic Tradition

Several factors make a comprehensive and accessible presentation of work on truth difficult. The first is that the literature on truth in the analytic tradition of western philosophy is split into two camps, and there is very little interaction between them. One camp tries to understand the nature of truth; i.e., what it is we are saying of something when we call it true. The other tries to figure out what to do about the liar and related paradoxes. Often what counts as common sense in one camp is regarded as highly contentious in the other.[4] Although there have been some efforts in the past decade to bring the two together, they represent a small fraction of the massive amount of work in each camp.[5]

During the twentieth century, logic, as an independent area of inquiry, blossomed. Its techniques are used throughout philosophy, but especially in response to the liar paradox. It slowly

[4] For example, the use of language-specific truth predicates (e.g., 'true-in-English'); these expressions are a topic of Chapters One, Two, Four, and Six.
[5] For example, see Beall and Armour-Garb (2005).

dawned on us that it is unbelievably difficult to say anything at all about the liar paradox without contradicting oneself. I do not know of any other area of philosophy where simply saying something *consistent* is such an accomplishment. So it seems that the precision offered by logic was a natural match for investigating the liar. This trend can easily be traced back to Alfred Tarski's work on truth in the early 1930s.[6] The received view for several decades was based on his work until doubts began to grow in the late 1960s. A lecture of Saul Kripke's that was published in 1975 shook the confidence that many had in the received view, and unleashed a flood of highly technical, mathematically sophisticated approaches to the liar that continues to this day.[7] Almost all this work is done using artificial languages and advanced logical techniques, including proof theory and model theory. Although it happens occasionally, there is not much emphasis on natural languages or language users. So this tradition is very young (in philosophical terms), highly technical, and for the most part divorced from standard issues and topics that arise in philosophy of language and metaphysics.

On the other hand, the literature focusing on the nature of truth in the twentieth century is continuous with that which preceded it. The emphasis is almost always on finding a good analysis of truth—saying exactly what truth consists in or finding a philosophically illuminating definition of 'true'. These people often frame their projects by inquiring into what we mean when we call something true. This tradition has also produced a staggering amount of work in the last sixty years, but it too is highly insulated. The received view is that whatever the right solution to the liar paradox turns out to be, it will not have a significant impact on how to understand the nature of truth.

---

[6] Tarski (1933).
[7] Kripke (1975).

Besides the complication of two massive and insulated traditions of work on truth, a further problem is that most of the literature on the liar paradox is highly technical, which makes it a daunting task to sift through it all. Mastery of this work requires knowledge of set theory and other branches of mathematical logic that most philosophers just do not have. However, to understand the challenge of providing a cohesive unified theory of truth (i.e., one that includes both a view on the nature of truth and an approach to the paradoxes), it is essential to have at least a rudimentary grasp of the pros and cons of the various approaches to the paradoxes. In what follows, I avoid technical aspects and try to present this material in an accessible way. When technical details would be helpful to specialists, they are confined to appendices.

Finally, the case I present against truth is complex and subtle. It draws from insights that belong to both camps (and seemingly unrelated areas of philosophy and linguistics), and it has implications for both camps as well. As a result, the project requires much more stage setting than usual. It requires not just a summary of each of the two traditions, but their unification into a single study of truth. My hope is that once it is clear just how deeply the insights of each tradition affect the other, it will no longer be acceptable for those working in one to ignore the other.

0.1.3 Four Key Issues

Before presenting my positive view, I emphasize four key issues that, together, have significant, unappreciated consequences for a unified theory of truth.

The key issues are:

(i) the expressive role that truth predicates play in our linguistic practice and cognitive lives,

(ii) that some versions of the paradoxes associated with truth depend on empirical facts,

(iii) the tendency for purported solutions to these paradoxes to generate new paradoxes, often called *revenge* paradoxes, and

(iv) internalizability, which is a relation that obtains between a semantic theory and a language when the semantic theory is expressible in and descriptive of that language.

The first comes from debates about the nature of truth, the second and third have their source in discussions of the liar paradox, and the fourth is new. By studying the relations among these four key issues, we shall discover that truth is an inconsistent concept, and that the other theories of truth available are, on reflection, inadequate. Thus, this work does not belong to either of the aforementioned camps; rather, it seeks to understand how the insights of one interact with the insights from the other. Only by considering them together can we arrive at the correct view of truth.

These key issues are not random. They have been chosen because they are relatively uncontroversial, yet many who write on truth pay them only lip service. Their significance has yet to be appreciated; individually, they have important consequences for various theories of truth, and together they have very significant consequences that no one seems to have noticed.

One can think of this part of the project as a massive import/export business, which occurs when a philosopher uses an insight from one side of a topical boundary to draw conclusions for issues on the other side.[8] One cannot do philosophy all at once—it helps to break it up into topics. The twentieth century has seen an unprecedented specialization in philosophy. Of course, the boundaries between topics are conventional and subject to a great deal of change, but anytime someone draws a boundary, there will be those who make their living moving things across it. Topical boundaries are no different.

I am not a fan of philosophical writings that present the author's positive theory only after giving a lengthy and often tedious survey of the literature together with a battery of objections to alternative views. In the case of truth, however, I am unaware of any survey that covers theories of

[8] I believe I heard the term 'import/export business' used in this way from Ben Caplan.

the nature of truth, approaches to the liar and related paradoxes, and combinations of the two; hence, this part of the book is essential to understanding my project. Moreover, my attempted reorientation of the debate by focusing on the four key issues is a helpful precursor to grasping my own positive view. I would have liked to just write up the positive view in a way that made clear its significance, but, given the current state of the literature, it was not to happen.

Once the importance of the key issues is appreciated, one sees just how difficult it is to present a unified account of truth—i.e., one specifying the nature of truth and an approach to the paradoxes. The view presented in this book is that the difficulty stems from the concept of truth itself. In the terminology I prefer, truth is an inconsistent concept, which means, roughly, that the principles governing it are inconsistent (given certain uncontroversial facts about the world). Not only do I think truth is an inconsistent concept, I argue that inconsistent concepts on which we rely for certain purposes ought to be replaced, at least if we desire that some concept or concepts continue to perform the function in question.

0.1.4 Concepts and Philosophy

My view is that philosophy is, for the most part, the study of inconsistent concepts (although I do not argue for that here). Once enough progress has been made to arrive at a set of relatively consistent concepts for some subject matter, it gets outsourced as a science. For the past 500 years, since the scientific revolution, philosophy has been giving birth to sciences in this way. Of course, these are huge generalizations, but it should give the reader at least some inkling of how I see this academic endeavor.[9]

[9] This view of the philosophical enterprise seems to dovetail with those voiced in Schiffer (2003) and Pettit (2004); see also Johnston (1993).

These views on philosophy and inconsistent concepts fit well with a dynamic philosophical method, which I have heard called *conceptual engineering*.[10] A precursor to this kind of philosophical project goes by the name of *explication*, and was popularized by Rudolph Carnap; it involves taking a more or less fuzzy intuitive concept and providing a more precise replacement for it.[11] Like Carnap, instead of sitting back and analyzing our concepts, as many analytic philosophers still do, I prefer to engage in the hard work of improving them. However, there are many other kinds of conceptual engineering besides explication. One can identify a conceptual confusion, where someone assumes that one coherent concept can do a certain job, but it actually requires two or more (e.g., the concept of mass as it occurs in Newtonian mechanics, which I discuss at length in Part III, is confused). Conversely, there can be cases where we assume that two distinct concepts can do two distinct jobs, but it turns out that these jobs are interrelated in a certain way, which requires a single coherent concept; there is no name for this phenomenon, but we might call it conceptual *confission* (e.g., the concepts of space and time as they occur in Newtonian mechanics get replaced by a single concept, spacetime). And there are others as well.

The very idea that our concepts might need improving is hard for some to accept. Nevertheless, I think a good case can be made that we encounter inconsistent concepts pretty frequently and we alter our conceptual scheme in response to them. The result is a conceptual revolution and a new conceptual scheme that can then be used and developed and pushed until some other part shows itself to be in need of improvement.[12] The philosophical method of conceptual engineering and the idea that philosophy is the study of inconsistent concepts go hand in hand. This work is an illustration of how to do philosophy in this way. It requires: (i) an exhaustive understanding of the

---

[10] Robert Brandom (2001) and Simon Blackburn (2001) both use this term.
[11] Carnap (1950).
[12] This talk of conceptual schemes should be taken with requisite care given the work on the scheme/content dualism by Davidson (1974); see also Child (1994), McDowell (1999) and Davidson (1999).

current state of play (e.g., what the philosophical problems are taken to be, what theories have been offered, what arguments have been given for them, and the costs and benefits of each theory), (ii) knowing the issues on which to focus (e.g., which issues are controversial and which ones are accepted, which issues and combinations of issues theories tend to get right and which ones tend to be stumbling blocks, and what legitimate issues together tell us about the right kind of theory), and (iii) a sense of which new concepts will allow us to do what we want to do without running into the old problems. These three elements of conceptual engineering correlate to the three parts of the book: (i) in Part I, "The Market", (ii) in Part II, "The Realignment", and (iii) in Part III, "The Proposal".

0.1.5 Six Philosophical Methods

More specific than this broad characterization of philosophical practice is the particular philosophical methodology to which I subscribe in constructing the positive theory in Part III. Defending a particular philosophical method is well beyond the scope of this book, but a brief look ought to help orient the reader. Analytic philosophy goes through stages of being obsessed with its own methodology; it is currently in the midst of one of these stages. It seems that there are at least the following six major kinds of philosophical methods being practiced at present.[13]

1. Conceptual Analysis is specifying illuminating apriori (i.e., knowable independently of experience) or analytic (i.e., true by virtue of their content alone) connections between some concept and other concepts. Often it is seen as doing even more: specifying the conceptual constituents of some complex concept. For example, a conceptual analysis of the concept of bachelor might be that bachelors are unmarried human males. In this case, it seems as though the concept analyzed is more

[13] This classification is very rough—the list is not exhaustive and the descriptions are in no way definitive; see the cited works for more detailed treatments.

complex than the concepts used in the analysis. However, one need not adhere to this basic/complex view of analysis. Conceptual analysis came in for some blistering attacks in the mid 20th century,[14] but despite that, it seems to be still dominant in analytic philosophy, which is so-named because of this methodology.[15]

2. *Reductive Explanation* is explaining some phenomenon by appeal to a different and usually better-understood kind of phenomena. In reductive explanation, the explanandum (i.e., the item to be explained) is wholly subsumed under the explanans (i.e., the items in terms of which it is explained). For example, reductive naturalism is a particularly popular kind of reductive explanation according to which every genuine phenomenon can be reduced to the phenomena studied by the hard sciences (and often to fundamental physics). Reductive naturalists hold that all genuine phenomena are, at root, physical phenomena. This includes consciousness, moral properties, mental states, and so on. There are versions of reductive explanation that are not naturalistic; for example, reductive phenomenalism reduces all genuine phenomena to experience.[16] The reduction of all genuine phenomena to the explanans class can be accomplished by *translation* (e.g., a reductive naturalist might say that all claims about legitimate phenomena can be translated into the vocabulary of particle physics), but it need not; a laxer reductive explanation appeals to *apriori entailment* instead of translation (e.g., a reductive naturalist of this stripe might claim that all true claims about legitimate phenomena are entailed apriori by true claims about the nature and behavior of

[14] See Quine (1951, 1960), Putnam (1962, 1971, 1975), and Kripke (1972) for examples.
[15] The most influential contemporary defense of conceptual analysis is certainly Jackson (1997); see also Jackson (2001a, 2001b), Balog (2001), Stich and Weinberg (2001), Stalnaker (2001), and Williamson (2001, 2008). See also Lewis (1994).
[16] See Carnap (1928), Quine (1951), and Sellars (1963) for more on phenomenalism.

fundamental particles).[17] The former is closely connected to conceptual analysis, while the latter is less demanding.[18]

3. Quietism is a method that avoids proposing and defending philosophical theories, and instead sees philosophical problems as the result of confusions that are often caused by misunderstanding language. The quietist attempts to rephrase or reformulate common-sense ideas (or perhaps just remind us of things we already knew) in a way that exposes the mistake and allows those taken in by the problem to see it as a pseudo-problem. There are probably two strands of quietism. The first consists of the ordinary language philosophers, like Gilbert Ryle, P. F. Strawson, and J. L. Austin; Charles Travis is a contemporary philosopher pursuing something like this project.[19] The other strand is heavily influenced by the later work of Ludwig Wittgenstein; John McDowell is probably the best-known contemporary practitioner.[20]

4. Experimental Philosophy is new on the scene; although it has historical precursors,[21] in its current form it is just a decade or so old. Experimental philosophy eschews the kind of armchair reflection and intuitions that other philosophical methods, especially conceptual analysis, take to be essential to doing philosophy. Instead, experimental philosophy advocates conducting surveys of non-philosophers' intuitions on issues of current interest in philosophy (e.g., influential thought experiments). From these results an experimental philosopher constructs a psychological theory about the source of those intuitions, and that theory is then used to support or attack various

---

[17] Even weaker reductions are familiar as well; e.g., the technical relation of supervenience is sometimes used—I discuss it in Chapters Five and Seven.

[18] See the papers in Hohwy and Kallestrup (2008) on reductive explanation; see Chalmers (2010) for much more on the "apriori entailment" version of reductive explanation. See also Block and Stalnaker (1999), Chalmers and Jackson (2001), and Gertler (2002) on conceptual analysis and reductive explanation. See the papers in de Caro and Macarthur (2004) for criticism of reductive naturalism.

[19] See Ryle (1931, 1949), Strawson (1950, 1959), Austin (1959), and Travis (2008).

[20] See Wittgenstein (1953), Malcolm (1959), and McDowell (1994, 2009). See also Zangwell (1992), Wright (1992: ch. 6, 1998), Pettit (2004), Rorty (2007), and Kuusela (2008) for discussion of quietism. Note that the term 'quietism' has come to have a negative connotation in the hands of philosophers like Blackburn (see his 1998); for this reason, one might prefer the term 'therapeutic'. However, I intend no such implication.

[21] Naess (1938) for example, which bears particular relevance to the topic of this book.

philosophical views that depend on those intuitions. Attitudes toward this new methodology run the gamut from adoration to disdain.[22]

5. *Analytic Pragmatism*, like experimental philosophy, is a reaction against conceptual analysis and reductive explanation, but it seeks a synthesis of the latter two methods with the insights of Wittgenstein, Wilfrid Sellars, and the classical pragmatists. Instead of emphasizing the relations between sets of concepts on which conceptual analysis or reductive explanation focuses, analytic pragmatism looks to relations between how words are used and the concepts those words express. The goal of an analytic pragmatist project is to specify relations between the concepts used to describe how some words are used and the concepts those words express. Although there are plenty of precursors, analytic pragmatism as a philosophical methodology is new, and its primary expositor and defender is Robert Brandom.[23]

6. *Methodological Naturalism*, as a philosophical method, is dramatically different from reductive naturalism, which is a kind of reductive explanation.[24] Methodological naturalists emphasize the similarity or continuity between science and philosophy; they suggest that philosophical problems should be approached by using the methods of the sciences and that philosophical theories should, like scientific theories, not only offer explanations, but be empirically testable. Beyond that, there is very little agreement on how to pursue methodological naturalism. Disputes about scientific methodology and the difficulty of performing experiments on philosophical topics (e.g., there is no

---

[22] See the papers in Knobe and Nicholas (2008) for more on experimental philosophy.
[23] See Brandom (2008) for the presentation of analytic pragmatism. I take the projects in Kripke (1982), Davidson (2001), Stanley (2005), Kukla and Lance (2008), and Capellen and Hawthorne (2009) to be instances of analytic pragmatism.
[24] The qualifier 'as a philosophical method' is meant to distinguish it from the view in philosophy of science, which sometimes goes under the same name, that one of the criteria for science is that it rejects supernatural explanations.

laboratory where one can study the properties of propositions) cause problems for methodological naturalist projects.[25]

0.1.6 My Philosophical Method

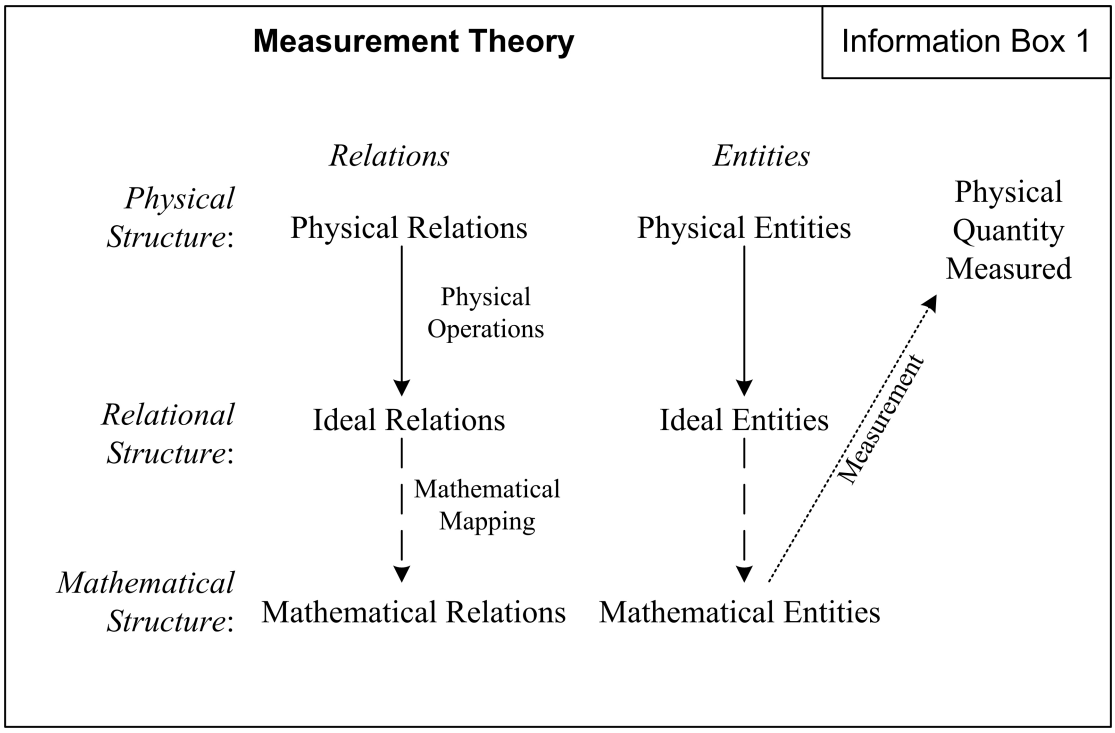
The positive view of truth I offer in Part III is not an instance of conceptual analysis, reductive explanation, quietism, or experimental philosophy. One can think of it as a work of analytic pragmatism in the sense that I look to our use of 'true' and general features of linguistic practice, focusing especially on what linguists say about communication as a condition of adequacy on a theory of truth. However, that still leaves us short of a theory of truth in any ordinary sense. One might say that once the use has been described, that is all there is to do, but I reject that conclusion entirely.

It is most accurate to say that the methodology in this book is a specific type of methodological naturalism—in particular, it is *measurement-theoretic methodological naturalism* (MTMN). I take it that this is Donald Davidson's methodology, despite the fact that he never explicitly defends it or even articulates it as such.[26] Measurement theory is the study of how formal and mathematical structures apply to the physical world; I like to think of it as somewhat analogous to set theory, but for science—it serves as an all-purpose background theory for science in the way that set theory serves as an all-purpose background theory for mathematics.[27] What I call a *measurement system* is composed of three structures and the links between them: a physical structure, which includes the phenomenon to be explained, a relational structure, which is an idealized theory of the phenomena in question,

---

[25] See Papineau (2007) for an overview of methodological naturalism. See Wilson (2006) and Maddy (2008) for examples. See also Price (2010) for what he calls subject naturalism, which I take to be very similar. See the papers in Braddon-Mitchell and Nola (2009) for a discussion of the relation between conceptual analysis and naturalism.
[26] For some of Davidson's remarks about measurement theory, see Davidson (1970: 220-221; 1997a: 130-132; 1997b: 75; 1999a: 253). For background on measurement theory, see Narens (2002, 2007), Suppes, et. al (1971, 1989, 1990).
[27] The claim about set theory as a foundation for mathematics is controversial; nothing of substance turns on this analogy.

and a mathematical structure. The links between the physical structure and the relational structure on the one hand and the links between the relational structure and the mathematical structure on the other allow one to apply the mathematical structure to the physical phenomena (see Information Box 1).[28] According to MTMN, a philosophical theory of X should be cast as a measurement system for X. That is exactly what I do for truth.[29]



This kind of methodological naturalism fits well with conceptual engineering and a view of philosophy as the study of inconsistent concepts. Trying to arrive at a measurement system for truth shows us that truth is an inconsistent concept, and, moreover, it points the way toward its replacements. The theory of the replacements is cast as a measurement theory for those concepts. Then, those concepts are employed by the measurement theory for truth. So measurement theoretic

---

[28] Throughout the book, the reader will find information boxes inserted in the text. These frequently contain diagrams, lists, or discussions of material that either illustrate some point made in the text or go beyond the text in some way or another.

[29] I present the details in Chapters One, Four, Thirteen, Fourteen, and Fifteen that are relevant to understanding the view of truth I propose, but a full-scale development and defense of MTMN will have to wait until another occasion.

methodological naturalism and conceptual engineering go hand-in-hand. I do not think this should come as a surprise, any more than the fact that physics and mechanical engineering are a natural fit.

0.2 Scope and Organization

Parts I and II present the background information and conditions on acceptable theories needed to understand and appreciate the positive account of truth given in Part III. The positive proposal I offer is multi-faceted and does not fit neatly into existing ways of classifying contemporary thinking about truth.

Part I is an attempt to survey the literature on truth in two traditions of analytic philosophy—work on the nature of truth and work on the liar paradox—and explore the combinations. *Chapter One* is a summary of the current views on the nature of truth. Next, I give an overview of the major approaches to the liar paradox; the innovation in this overview is that they are split into *philosophical approaches*, which discuss some feature of truth predicates of natural languages relevant to resolving the paradoxes, and *logical approaches*, which focus on artificial languages and propose principles that truth predicates obey and logics for languages containing truth predicates. The former are covered in *Chapter Two*, and the latter in *Chapter Three*. *Chapter Four* introduces unified theories of truth, which address the nature of truth (the topic of Chapter One), philosophical approaches (covered in Chapter Two), and logical approaches (from Chapter Three). *Chapter Five* explores connections between truth and other concepts.[30]

[30] I should mention that two excellent books, one on each tradition, have appeared recently: Wolfgang Künne's *Conceptions of Truth* (2003) surveys the debates about the nature of truth and Hartry Field's *Saving Truth from Paradox* (2008a) covers the major approaches to the paradoxes. These two books go into much more detail on most of theories I survey in Part I, and any reader interested in an in-depth discussion should consult them. Other summaries of work on the nature of truth include Kirkham (1995), Walker (1996), and Candlish and Damnjanovic (2007). Visser (2002), Beall (2006), and Cantini (2009) are overviews of work on the liar paradox.

Part II attempts a reorientation of the philosophical debates about truth, by explaining four key ideas (truth's expressive role, empirical paradoxes, revenge paradoxes, and internalizability) and their consequences. Accordingly, it consists of four main chapters and a summary; each main chapter is divided into a discussion of the key idea in question and pointing out its consequences.

Chapter Six describes the first key issue: how we use truth predicates. It is helpful to have an adjective that means *pertaining to truth*. Since there is no accepted term with this meaning, I use 'aletheic', which derives from the Ancient Greek word 'aletheia'. Thus, *aletheic practice* is our activity of uttering and interpreting sentences containing truth predicates, and our mental states involving the concept of truth. Truth predicates play a crucial *expressive* role in natural languages. They allow us to formulate certain generalizations that we otherwise would not be able to express, and they allow us to endorse propositions we cannot directly assert. Likewise, possessing the concept of truth allows us to formulate generalizations and think thoughts that would otherwise be unavailable. In addition, this chapter contains a discussion of how we use language, which is a topic known as pragmatics. Throughout the book, emphasis is given to what real people do with truth predicates, and the goal is to account for our aletheic practice. So, understanding how we use language in general is crucial.

Chapter Seven presents the second key issue—*empirical paradoxes*. Some aletheic paradoxes are empirical in the sense that the sentences in question only count as paradoxical because of some seemingly unrelated empirical facts. This strongly suggests that aletheic paradoxicality is not a syntactic phenomenon—it does not occur because we are making some mistake about which strings of letters count as sentences. It also strongly suggests that paradoxicality is not a semantic phenomenon—it does not occur because we are making some mistake about the meanings of our sentences or words. Finally, it suggests that paradoxicality is not a pragmatic phenomenon—it does not occur because we are making some mistake about the forces or implicatures of our utterances.

Instead, participants in a conversation are not usually in a position to recognize whether the sentences uttered that contain truth predicates are paradoxical.

The third key issue, the *revenge paradox* phenomenon, is the topic of *Chapter Eight*. In an effort to understand why one of the most beloved and revered members of our conceptual repertoire could cause us so much trouble, philosophers have for centuries proposed solutions to the liar paradox. However, it seems that our concept of truth takes offense at our efforts to understand it because it appears to retaliate against those who propose such solutions. It takes its revenge on us by creating new paradoxes from our own attempts to find resolution. That is, most proposed solutions to the liar paradox give rise to new, more insidious paradoxes—often called *revenge paradoxes*. For our attempts at understanding, truth rewards us with inconsistent theories, untenable logics, and a deep feeling of bewilderment. It is as if our concept of truth lashes out at us because it wants to remain a mystery. After a few run-ins with truth, many philosophers have the good sense to keep their distance. Far from being the serene, profound concept most people take it to be, those of us who think much about the liar paradox know truth to be a vengeful bully—a conceptual misanthrope.

Because the standard response to revenge paradoxes is to restrict one's approach to the liar so that it does not apply to languages capable of formulating revenge paradoxes, the issue of whether one ought to be able to provide a theory of truth that applies to the very language in which it is formulated is a topic frequently discussed in the literature on the liar paradox. This is our fourth key issue from above. However, there is no precise framework against which to have these discussions. *Chapter Nine* provides one, and the core concept of this framework is *internalizability*. A semantic theory for truth is internalizable for a language if and only if there is an extension of that language to which the theory applies and in which the theory is expressible. I argue that any semantic theory of truth should be interalizable for every natural language.

Chapter Ten summarizes the four key points and offers several conditions any acceptable unified theory of truth ought to meet.

Part III is where I set out my proposal for understanding truth. On my view, truth is an inconsistent concept. Saying what that means is the job of *Chapter Eleven*, which explains the basics of inconsistent concepts. *Chapter Twelve* includes several arguments for the claim that truth is an inconsistent concept.

Since truth is inconsistent, we need to replace our inconsistent concept of truth with two concepts, which I call *ascending truth* and *descending truth*. Not only do the replacement concepts perform the explanatory work we ask of truth, they avoid the paradoxes caused by truth as well. Moreover, one can use them as the basis for a theory of our inconsistent concept of truth itself. Thus, there are really two essential parts to the view I recommend. First, there is the prescriptive theory, which explains the replacement concepts. It says how we should expand our conceptual repertoire and provides reasons to think that this expansion will not just result in the same old problems we find with truth. Second, there is the descriptive theory, which explains our defective concept of truth—what principles it obeys and why it gives rise to paradoxes. One of the fundamental commitments of the entire project is that the descriptive theory should *not* appeal to the concept of truth; instead, it uses the replacement concepts. Accordingly, *Chapter Thirteen* contains the prescriptive theory and explains how the replacement concepts work, and *Chapter Fourteen* uses those replacement concepts to explain how truth works.

The central claim of the *prescriptive* theory is that, for certain purposes, we ought to use two new concepts, ascending truth and descending truth, instead of truth. *Ascending truth* is like truth in that the inference from a declarative sentence p to 'p is ascending true' is valid. It differs from truth in that the inference 'p is ascending true' to p is not always valid (although it is valid for the vast majority of sentences). *Descending truth* is like truth in that the inference from 'p is descending true'

to p is valid for any declarative sentence. However, it differs from truth in that the inference from p to 'p is descending true' is not always valid (although, again, it is valid for the vast majority of sentences). Together, ascending truth and descending truth can do the work we require of truth without giving rise to paradoxes of any kind. Moreover, the theory of ascending truth and descending truth is compatible with classical logic, and it imposes no expressive restrictions on the languages that contain ascending truth predicates and descending truth predicates.

The *descriptive* theory implies that a truth predicate of a natural language is assessment-sensitive, which means that it has the same content in every context of utterance, but its extension (i.e., the set of things that are true) depends on a context of assessment. Contexts of assessment model situations in which a person interprets someone's utterance. From different contexts of assessment, the truth predicate has different extensions. The descriptive theory employs the concepts of ascending truth and descending truth, and they determine how the extension of the truth predicate varies across contexts of assessment. This assessment-sensitive theory of truth solves the liar paradox, it is compatible with classical logic and all the expressive resources we have in natural language, and it does not give rise to any new paradoxes.

Finally, *Chapter Fifteen* covers many of the issues that arise when one tries to replace a concept like truth that is so central to our way of thinking about ourselves, the world, and the relationship between them. In particular, it gives an overview of the relations among the concepts of ascending truth and descending truth and many of the concepts that are closely tied to truth, including validity, meaning, assertion, knowledge, predication, and reference. A brief *Conclusion* follows Part III.

Part I

The Market

> I would prefer to introduce myself as doing conceptual engineering. For just as the engineer studies the structure of material things, so the philosopher studies the structure of thought. Understanding the structure involves seeing how parts function and how they interconnect. It means knowing what would happen for better or worse if changes were made. This is what we aim at when we investigate the structures that shape our view of the world. Our concepts or ideas form the mental housing in which we live. We may end up proud of the structures we have built. Or we may believe that they need dismantling and starting afresh. But first, we have to know what they are.
>
> —Simon Blackburn, *Think*, pp. 1-2

Chapter 1

The Nature of Truth

The first order of business is to get some sense of where we are in the study of truth. What have other people said about truth? What are the major accomplishments? What are seen as the open problems? It is only against this background that the reorientation and replacement I attempt makes sense. However, I will not be spending much time on each view in these opening chapters. The goal is a quick and dirty overview for the novice. In later parts of the book, when I discuss their merits and problems, I am much more careful about the subtleties of formulation. Those readers who already know much of the work on truth might find this breezy introduction a bit disconcerting; if you are one of those people, I invite you to continue with an open mind.

Throughout this book, we will focus on the English adjective, 'true', its synonyms, and the expressions of other languages that have roughly the same meaning. However, like most words, 'true' is ambiguous or polysemous and we care about only one of its meanings: cases like 'what Herschel said is true' or 'Albert's theory is true'. 'True' can also be used to mean something like *genuine* (e.g., in 'Milhouse is a true friend') or something like *straight* (e.g., in 'the arrow's flight was true'). It can also be used as a verb meaning *to level* or *to straighten*, as in 'the mechanic trued the bicycle wheel'). And it has other meanings as well. Here we focus on 'true' as it is applied to items with propositional content, like sentences, beliefs, and utterances. I call these words with this understood meaning *truth predicates*.

1.1 Truth bearers, Truth Aptness, and Truth Value

The word 'true' of English seems to be a one-place predicate, and most views on the nature of truth take this claim for granted. Consequently, when one uses 'true', one predicates truth of some entity or entities, and these are often called *truth bearers*. If ordinary speech is a guide, then there are many different kinds of truth bearers, including sentences, beliefs, propositions, utterances, theories, stories, songs, and probably many other things as well. It is far from clear how to make sense of this practice.

Philosophers almost always pick one kind of truth bearer and explain what it is for that kind of thing to be true; these are called *primary* truth bearers. After selecting a primary truth bearer, one of two strategies is employed. An *interpretive* strategy says that anytime someone seems to attribute truth to something that is not a primary truth bearer, then that person's claim should be interpreted as attributing truth to a primary truth bearer. According to this option, primary truth bearers are the only things that are, strictly speaking, true. For example, if a theory takes propositions to be primary truth bearers, then it might say that when someone says that some sentence is true, that person really means that the proposition expressed by that sentence is true. On the other hand, an *explanatory* strategy says that the truth of things that are not primary truth bearers should be explained in terms of the truth of primary truth bearers. For example, if propositions are taken to be primary truth bearers, then one might say that a sentence is true if and only if it expresses a true proposition. A theory like this takes the person's claim that some sentence is true at face value—it does not interpret the person as if she were talking about propositions.[1]

Even after we decide on primary truth bearers, we still have work to do. For example, assume we pick sentences. Which sentences can be true or false? All of them? That does not seem right. Questions and commands do not seem like they are true or false. Rather, it seems like only declarative sentences can be true. Even among declarative sentences, it is not obvious that all of

[1] See Rojszczak (2005) for a historical discussion of truth bearers.

them could be true or false. For example, one might think that 'Nelson is cool' is neither true nor false because it does not even attempt to represent the world. Instead, one might think that this sentence is used to express an attitude of approval toward Nelson. Indeed, some philosophers think that many declarative sentences, ethical sentences for example, are fundamentally for the purpose of expressing the speaker's attitudes.[2] Some of these philosophers claim that unless a speaker is purporting to represent the world by uttering a sentence, the sentence is neither true nor false.

It is customary to think of this as an issue concerning which things are *truth apt*, not which things are truth bearers. Although it might be tempting to think of truth aptness issues and truth bearer issues as one and the same, it seems to me that there is an important distinction here. The issue of truth aptness arises once one has already made choices about truth bearers. A truth bearer is truth-apt if it is capable of having the property of truth.[3] How is this any different from just being a truth bearer? The difference is that one's choice of truth bearers is a choice between types of objects without regard to their syntactic, semantic, or pragmatic features. The options for truth bearers are not distinguished in these terms (e.g., fact-stating sentence tokens). If one chooses sentence tokens as primary truth bearers, then there is still the issue of deciding which sentence tokens are capable of possessing truth. We do not think that sentence tokens used to produce questions or commands are true or false. Some philosophers argue that sentence tokens whose semantic presuppositions fail are neither true nor false.[4] These are truth-aptness issues. Of course, one could combine truth aptness issues and truth bearer issues into one topic, but I think that this would do a disservice to those engaged in debates about them. Expressivists who discuss which things are capable of truth are not

---

[2] This view is called *ethical expressivism*; see Ayer (1939), Blackburn (1984, 1998), Gibbard (1990), and Schroeder (2008).
[3] I have heard some people use the term 'truth aptitude' instead of 'truth aptness'. As I understand the terms, 'aptness' and 'aptitude' have very similar meanings, but the latter tends to have the connotation of ability—something that an animate entity can do—whereas the former seems to apply more readily to inanimate objects. I prefer 'aptness' since it seems odd to me to say that truth apt truth bearers have the ability to be true (i.e., they can accomplish truth if they try hard enough).
[4] See Strawson (1950).

(usually) worried about choosing between propositions and sentence tokens; they are concerned with truth-aptness.

Another way to bring out the difference between truth bearers vs. truth aptness issues is to say that the former are often explanatory (i.e., how should we go about explaining the truth and falsity of one type of entity in terms of the truth and falsity of another?), whereas the latter tend to be demarcational (i.e., which entities are in the class of those that can be true or false?). Finally, the property of being a truth bearer seems like it is fundamentally different from the property of being truth apt. For example, a single truth bearer can be truth apt in one context and not truth apt in another. If a token of 'the room is cool' is used to describe the temperature of the room in question then it is truth apt (on most accounts), but if it is used to express one's positive evaluation of the room, then many people think that it is not truth apt. However, truth-bearerhood is not something an object can have in one context and not in another. Clearly, these are different issues and deserve to be kept distinct.

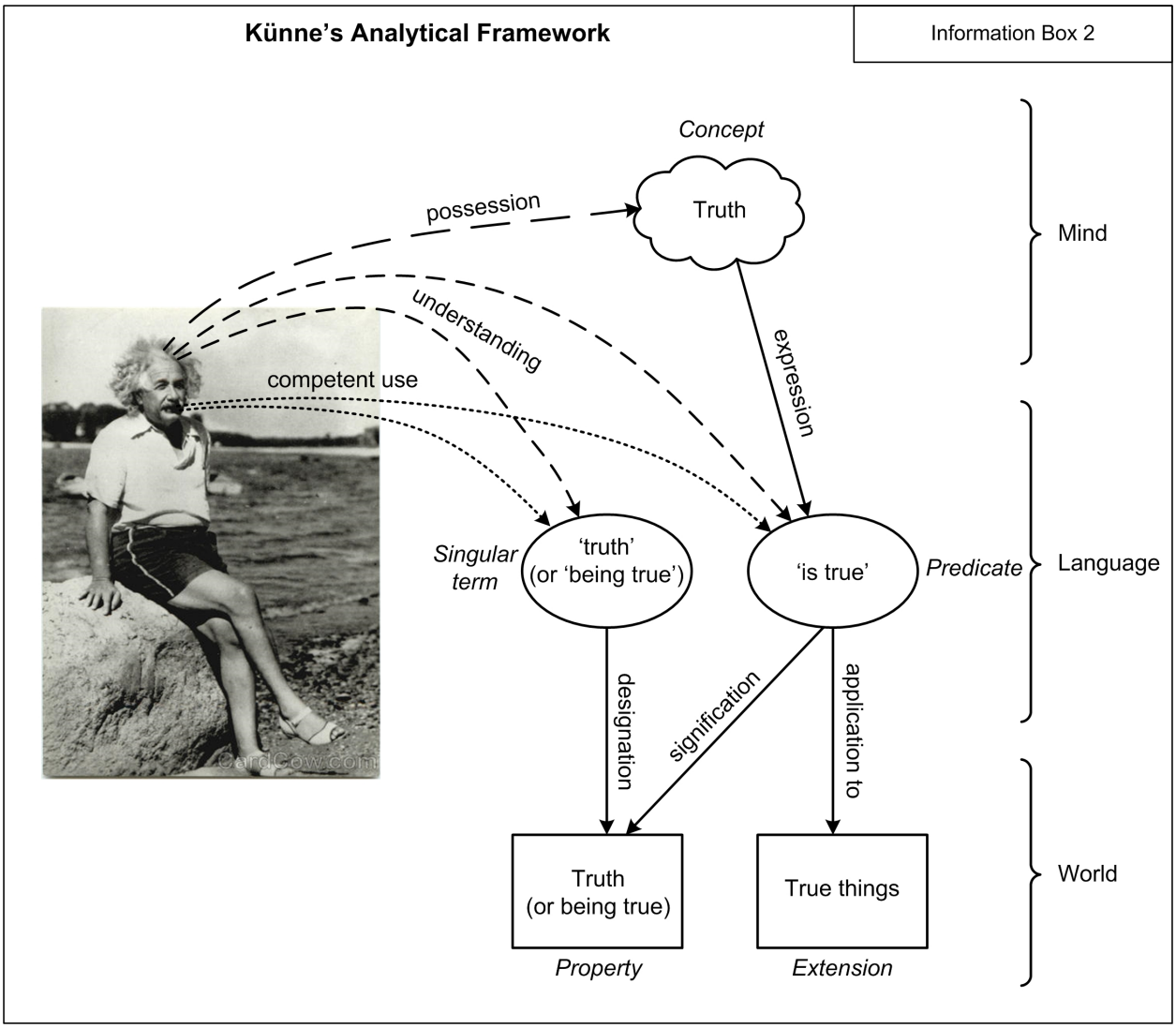
One also finds the term 'truth value' in discussions of truth quite often. Usually, truth and falsity are thought of as the only two truth values. However, there are special kinds of logic that use more than two truth values; I discuss some of them in Chapter Three.

In sum, to be a *truth bearer* is to be the kind of thing that is capable of truth or falsity; to be *truth apt* is to be a truth bearer that is capable of truth or falsity; to have a *truth value* is to be either true or false. Many think that there are truth bearers that are not truth apt (e.g., 'Nelson is cool') and that there are even truth apt truth bearers that do not have a truth value (e.g., paradoxical sentences, which are the topic of Chapter Two).

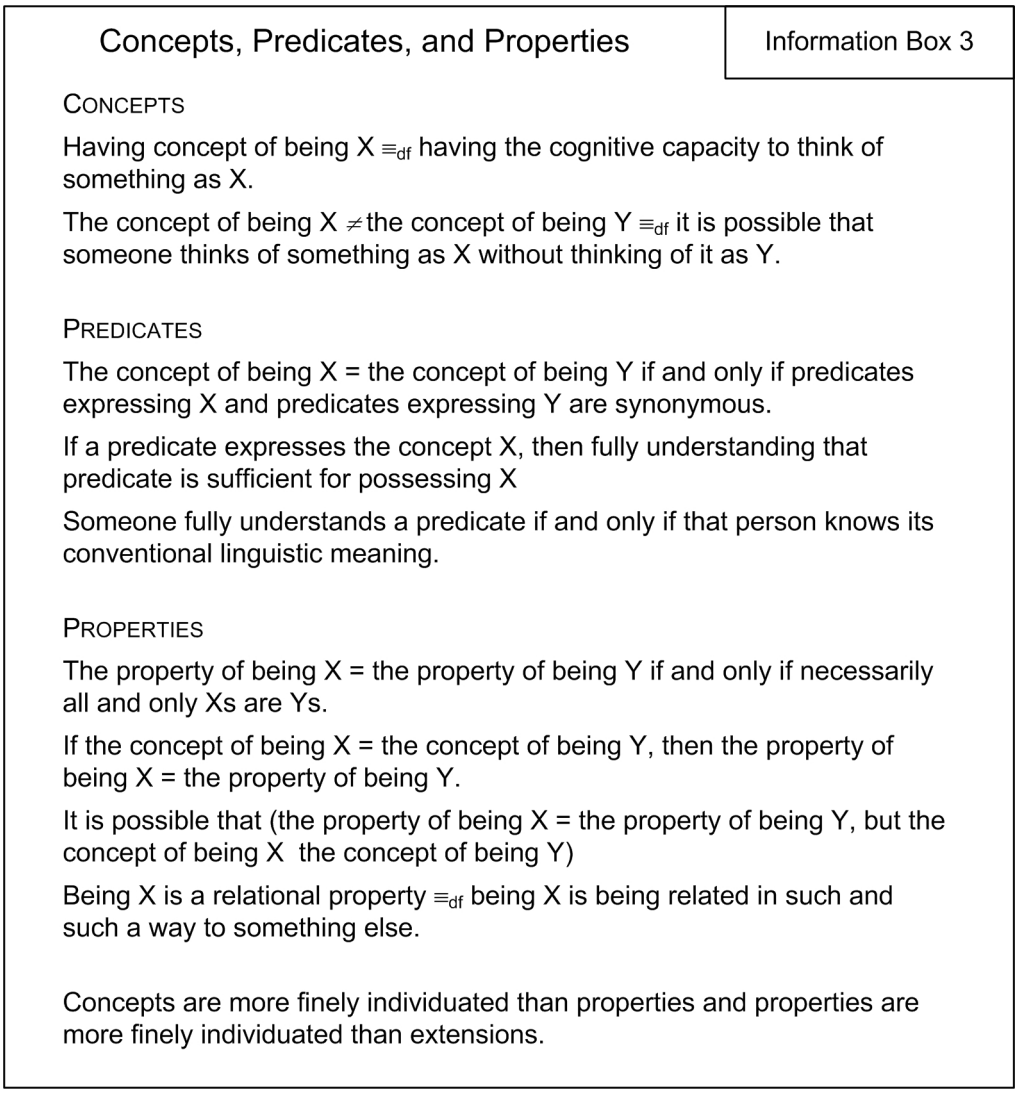
1.2 Künne's Analytical Framework

In the rest of this chapter, I present an overview of the major theories of truth—at least those in the tradition that focuses on the nature of truth instead of the liar paradox. Most of those who present theories in this tradition say that they are giving an *analysis* of truth, but that project has an effect on what one thinks about the meaning of 'true', the property of being true, and which things are true.[5] These issues can be confusing, and philosophers rarely take the time to sort them out. One notable exception is Wolfgang Künne, whose book, *Conceptions of Truth* (OUP, 2003) is an excellent introduction to work on the nature of truth (although it ignores all the work on the liar paradox). In it, Künne presents a helpful framework for understanding, comparing, and evaluating the claims made by philosophers about truth. I have put it in the form of a diagram (Information Box 2) and summarized the relevant principles (Information Box 3).

[5] At the end of this chapter I consider other ways of understanding a theory of truth.



The diagram illustrates the relations between a concept (i.e., truth), the associated predicate (i.e., 'is true') and singular term (i.e., 'truth'), the associated property (i.e., being true), and the associated extension (i.e., the set of true things). Consider Einstein above, who looks rather dashing in his beach-wear. He *possesses* the *concept* of truth, he *understands* the *predicate* 'is true' and the *singular term* 'truth', and he is a *competent user* of both these terms. Moreover, the concept of truth is *expressed by* the predicate 'is true', which *applies to* the things in its *extension* and *signifies* the *property* of being true; in addition, the singular term, 'truth', *designates* the property of being true. Throughout the book, I use the preceding italicized words as technical terms.



Accompanying the diagram is a list of principles that can be used to explain the terms above, make relevant distinctions, and apply them. For example, Einstein possesses the concept of truth if and only if he has the cognitive capacity to think of something as true. Einstein understands the predicate 'is true' if and only if he knows its meaning. If Einstein understands 'is true', then he possesses the concept of truth. A word of warning: these are extraordinarily complex topics, and this analytical framework is only intended to introduce them to the reader and provide a rough account. When substantive issues regarding these topics become arise later in the book, I take care to sort out the relevant subtleties.

1.3 Correspondence Theories

By far the most intuitive and historically popular view is that truth is correspondence with reality. When one calls some thing true, one means that reality is the way that thing says it is. So one's belief that snow is white is true if and only if the belief is about some portion of reality (i.e., snow's whiteness), and that portion of reality is the way the belief describes it (i.e., snow really is white). All correspondence theories take truth to be a *relational property* in the sense that being true is constituted by a relation between the true thing and some other thing(s). Note that this claim is different from saying that the predicate 'is true' is polyadic, which would be the case if 'is true' had an extra "input slot"; for example, some philosophers think 'is true' should be read as 'is true at w', where 'w' is a possible world (I consider these views in Section 11). We can add blanks to clarify: '__ is true at __' is a polyadic predicate, whereas '__ is true' is a monadic predicate. Correspondence theorists take 'is true' to be a monadic predicate that signifies a single property, and that property is constituted by a relation between the truth bearer and something else.

There are many versions of the correspondence theory depending on how one thinks about the correspondence relation and the things to which truths correspond. One popular view is that a true truth bearer, as a unified whole, corresponds to a fact or state of affairs in the world. Facts are notoriously difficult to explain, especially without appealing to the notion of truth, but they are usually taken to be things denoted by bare 'that' clauses (e.g., that snow is white) and described by phrases like 'snow's being white'. Correspondence theorists who go in for facts are split between those who think true truth bearers and facts can be matched up one to one and those who think a single truth bearer might correspond to many facts or many truth bearers might correspond to the

same fact. Either way, these theorists take correspondence to be a simple relation between a truth bearer and some constituent of reality.[6]

Some correspondence theorists take the items to which true truth bearers correspond to be objects. These views see truth bearers as complex items with parts; the correspondence relation is a complex relation between the parts of a truth bearer and objects in the world. For example, 'Columbus is in Ohio' contains a name, 'Columbus' and a predicate 'is in Ohio'. According to the views in question, the whole sentence is true because 'Columbus' refers to Columbus and Columbus has the property designated by 'is in Ohio' (i.e., being in Ohio). On these views, true truth bearers correspond to the world because the truth bearer's parts and its structure are systematically linked to objects in the world and their structure.[7,8]

One radical view is that the correspondence relation is just identity, which results in what have been called identity theories of truth. Russell and Moore seemed to hold something like this view in the early twentieth century, but John McDowell is responsible for bringing it back into the contemporary discussion.[9] Identity theories are somewhat limited in their choice of primary truth bearers—it seems like they have to go with propositions and say that true propositions just are facts.

An important motivation for correspondence theories is that they capture the intuitive meaning of the word 'true'. Most reputable dictionaries have something like 'agreement with reality' under the entry for 'true'. Thus, correspondence theories do well in justifying our intuitions about the

---

[6] See Moore (1899), Russell (1910), Wittgenstein (1923), Austin (1950), Armstrong (1973, 1997, 2004), Millikan (1986), David (1994), Fumerton (2002), Newman (2002), Kitcher (2002), Vision (2004), Englebretson (2006), and Marino (2006, 2008, 2010). See Kirkham (1995: 119-141) and Künne (2003: 112-126) for discussion.

[7] See Davidson (1969), Field (1972, 1973, 1974, 1986), and Devitt (1984); see Künne (2003: 94-112) for discussion. Note that Davidson retracted his view on correspondence; see Davidson (2005).

[8] Another version of the correspondence theory takes truth bearers as unified wholes to correspond to events instead of facts. This view avoids having to explain the nature of facts without appealing to truth, but it does require a way of explaining what events are; see Künne (2003: 145-148) for discussion.

[9] McDowell (1994, 2009); see also Cartwright (1987), Dodd (1995, 1996, 1999b, 2000, 2008) and Hornsby (1997, 2005). There is a dispute about whether Bertrand Russell and G. E. Moore held this view. See Künne (2003: 6-12) and Candlish and Damnjanovic (2007: 231-233) for discussion.

meaning of the truth predicate. In addition, it seems to me that another motivation for the view comes from the idea that a primary function (perhaps *the* primary function) of language and thought is to represent the world. A sentence or thought that correctly represents the world is true; one that does not is false. Thus, truth is to be explained in terms of some relation between linguistic or mental items and the world. This is a powerful line of reasoning and should not be underestimated.[10]

Correspondence theorists often explain the correspondence relation in terms of truthmaking. A *truthmaker* is that which makes a truth bearer true. A huge literature has grown up around this idea, and not all of it is concerned with correspondence theories of truth. Many philosophers treat truthmaking as a basic metaphysical phenomenon that deserves to be investigated in its own right. Correspondence theorists who go in for facts often take the truthmaking relation to be explanatory—the fact that snow is white is what explains why 'snow is white' is true. Others take it to be metaphysical—the fact that snow is white is what necessitates or forces 'snow is white' to be true. One can endorse a truthmaker principle (i.e., that all truths have truthmakers) for certain discourses but not for others; thus, it seems as if fact-based correspondence theories are committed to some truthmaker principle, while endorsing a truthmaker principle need not saddle one with a correspondence theory of truth.[11]

There are plenty of objections to correspondence theories. One is just that they fail to adequately specify (in a non-circular manner) the correspondence relation or the nature of facts. Another is the so-called slingshot argument, which purports to show that on some very plausible assumptions, all true truth bearers correspond to a single fact.[12] Others include the worry that not

---

[10] See Millikan (1986, 1990), Blackburn (1984, 2005), Wright (1992), and Lynch (2009: 21-36).
[11] For more on the truthmaker debates see Armstrong (2004), Beebe and Dodd (2005), Merricks (2007), Lowe and Rami (2009), and Schaffer (2010).
[12] See Davidson (1969) and Neale (2001) for discussion.

all true truth bearers have truthmakers, and that the correspondence theory of truth is not really about truth at all.[13]

1.4 Coherence Theories

Another relational view is the coherence theory, which states that a truth bearer is true if and only if it coheres with other truth bearers. Correspondence theories explain truth in terms of a relation to things in the world, but coherence theories explain it in terms of a relation to other truth bearers. Correspondence theories have traditionally been associated with realism (i.e., the view that reality is independent of our minds), whereas coherence theories have often gone hand in hand with idealism, the view that reality (at least in part) is constituted by mental properties or processes. Although somewhat popular in the nineteenth century, this view of truth has disappeared almost completely along with idealism.

Beliefs are usually the primary truth bearers for coherence theorist. Coherence always involves consistency, but the key to this kind of theory is explaining what more is required for a system of truth bearers to be coherent. Coherence can be a property of the set of truth bearers, it can be a relation between them, or it can be a relation between a truth bearer and a set of truth bearers. In addition, some coherence theorists suggest that logical relations like entailment or epistemic relations like justification contribute to the coherence relation. Some claim that explanatory relations also play a role.[14,15,16]

---

[13] See Künne (2003: ch. 3) for discussion and further references. See Taylor (2006) for a thorough discussion and criticism.

[14] For examples see Jocahim (1906), Bradley (1914), Blanchard (1939), Dauer (1978), Young (1995, 2001), da Costa, Bueno, and French (2005), and Dorsey (2006).

[15] Michael Lynch has recently introduced the notion of supercoherence as part of his pluralist theory of truth; see Lynch (2009: 168-180). I discuss pluralist theories below.

[16] Sometimes Donald Davidson's view that belief is intrinsically veridical gets classified as a coherence theory of truth (mostly because he titled the paper in which he announced this view "A Coherence Theory of Truth and Knowledge"),

The criticisms leveled against coherence theories are that it leads to idealism, that it confuses truth with the criterion for truth (i.e., how we determine whether something is true), and that it is impossible to specify the class of truth bearers to which one must cohere in a non-circular way.[17] Given the dearth of support for coherence theories, they do not figure prominently in the rest of the book.

1.5 Pragmatic Theories

There are precursors to the pragmatic theory of truth in earlier thinkers, but it was first presented by the American Pragmatists in the late nineteenth century and early twentieth centuries. According to this view, a truth bearer is true if and only if it is prudent to have the belief associated with that truth bearer. Prudence should be thought of as utility-based—it is prudent to have a belief just in case acting on that belief tends to satisfy the agent's desires.[18] William James spent a considerable part of his career defending this view, to little avail.

Although that is usually how the classical pragmatist theory is understood, two of the classical pragmatists, Charles Sanders Peirce and John Dewey, have views on truth that do not fit very well with this characterization. Peirce said that a truth bearer is true if and only if it is fated to be agreed upon at the end of inquiry.[19] That strikes many people as an epistemic view of truth, so perhaps Peirce belongs in the next section. Dewey alternately agreed with Peirce and claimed that truth is

---

but that is a mistake (as Davidson himself admitted). The veridicality of belief is a controversial view about the way beliefs acquire their content, not a view on the nature of truth. See Davidson (1982b, 1990, 2001).

[17] For discussion see Rescher (1973), Walker (1989), Kirkham (1995: 101-111), Wright (2003: ch. 9), McGinn (2002), Künne (2003: 381-393), and Thaggard (2007). See also Olsson (2005) for a discussion of coherence theories of truth and knowledge.

[18] James (1907, 1909) contains the clearest presentation of this theory.

[19] Peirce (1877, 1878).

just warranted assertibility, which can be read in a couple of ways, but also sounds like an epistemic view.[20]

More recently, Richard Rorty has offered what he calls a pragmatic theory of truth whose tenets are:

(i) 'true' has no explanatory uses.

(ii) we understand all there is to know about the relation of beliefs to the world when we understand their causal relations with the world; our knowledge of how to apply terms like 'about' and 'true of' falls out of a naturalistic account of linguistic behavior.

(iii) there are no relations of 'being made true' that hold between beliefs and the world.

(iv) there is no point to debates about realism and anti-realism, for such debates presuppose the empty and misleading idea of beliefs 'being made true'.[21]

These claims do not seem to have much to do with what James said about truth, and they have a much greater affinity to deflationists' theories, which I discuss below. It seems to me that no one tries to explain truth in terms of utility these days, and what are sometimes called pragmatic or pragmatist theories of truth follow Peirce and should be classified as epistemic theories.

Objections to the pragmatic theory include that it confuses truth with the criterion for truth, it makes truth dependent on our desires, and that it leads to idealism.[22]

1.6 Epistemic Theories

Epistemic theories of truth are newer than the first three. These views hold that a truth bearer is true if and only if an ideal rational agent in ideal epistemic circumstances would justify it. Obviously, this view takes primary truth bearers to be the kinds of things that can be justified, so it works best with beliefs. Epistemic theories are sometimes lumped in with coherence theories because the

[20] See Kirkham (1995: 79-101) and Brandom (1994: 285-299) for discussion.
[21] Rorty (1986); Rorty articulates these theses as part of an argument that Davidson is part of the pragmatist tradition, but it seems to me that Davidson would reject (i)—see Davidson (1990, 1996).
[22] See Kirkham (1995: ch. 3) for discussion.

coherence relation is sometimes explained in terms of epistemic notions like justification, but this is unfortunate.[23] Coherence theories are relational theories—they treat truth as a relational property—whereas epistemic theories are not. Moreover, coherence theorists are virtually extinct, but epistemic theories continue to be a focus of research.[24]

One of the earliest to propose an epistemic theory is Charles Sanders Peirce, who claimed that truth is that which is fated to be agreed upon by all who engage in the relevant inquiry.[25] Jürgen Habermas also accepted an epistemic theory of truth for many years, but has since given it up.[26] Jay Rosenberg argued for a version of Peirce's theory of truth in the 1970s, and Cheryl Misak has recently given an extended defense of Peirce's version of the epistemic theory.[27] Robert Almeder's recent book defends a version of the epistemic theory and considers its usefulness in dealing with the problem of skepticism (i.e., whether our beliefs about the external world are ever justified).[28]

A motivation for epistemic theories often comes from realism debates, which have occupied a huge swath of the philosophical landscape. The contrast between realism and idealism is often cashed out in terms of correspondence theories vs. coherence theories of truth, but the realism / anti-realism distinction is often identified with the correspondence theory vs. epistemic theory divide. The realism/idealism debate is about the ultimate nature of reality—whether it is independent of or dependent on minds. Within professional philosophy, this debate is mostly dead.[29] On the other hand, the realism/anti-realism debate is alive and well—it is about whether language and thought can represent reality in a way that goes beyond our capacity to verify. Perhaps the most familiar defender of this project is Michael Dummett, who explains truth in terms of

---

[23] See Walker (1989) for an example.
[24] See Dummett (1959, 1978, 1999, 2002) and Tennant (1997).
[25] Peirce (1894).
[26] Habermas (1973, 2003).
[27] Rosenberg (1974), Misak (2004).
[28] Almeder (2010).
[29] However, John McDowell (1994, 2009), Robert Brandom (1994, 2009), and Jonathon Schaffer (2010a) have argued that analytic philosophy still has plenty to learn from this debate.

warranted assertibility so as to get a global theory of meaning on which the law of excluded middle (i.e., p or ~p) is not valid.[30] Others who have pursued this programme include Hilary Putnam and Neil Tennant.[31]

Another group developing epistemic theories are pluralists, who only endorse an epistemic notion of truth for certain discourses. For example, Crispin Wright suggests that the concept of superassertibility would work well as a concept of truth for ethical discourse, discourse about humor, and others. Wright writes:

> [A]nother property constructible out of assertibility which is both absolute and, so it is plausible to think, may not be lost, is the property of being justified by some (in principle accessible) state of information and then *remaining* justified no matter how that state of information might be enlarged upon or improved. ... A statement is superassertible, then, if and only if it is, or can be, warranted and some warrant for it would survive arbitrarily close scrutiny of its pedigree and arbitrarily extensive increments to or other forms of improvement of our information.[32]

Notice that the act of assertion, which is a kind of utterance, has nothing to do with the definition of 'superassertibility'; instead, it is defined in terms of warrant.[33] Superassertibility is designed as a truth predicate for claims about which it would be odd or incoherent to say that they might be true even though no one could ever discern their truth. For example, it would be strange to say that some joke was funny even though no one ever found it funny and no one ever came to know that it was funny. The reason is that it seems that whether something counts as funny depends on how we react to it. Wright's claim is that the notion of superassertibility works as a concept of truth for

---

[30] Dummett (1991).
[31] Putnam (1981) and Tennant (1997).
[32] Wright (1992: 42-3). See also Wright (2003), Kenyon (1999), and Tomasi (2006) for discussion.
[33] One finds this kind of dual use of 'assertibility' throughout contemporary analytic philosophy—it seems to me that it comes from an ambiguity in 'warranted assertion'. On the one hand, a speaker's assertion is warranted only if that speaker is justified in performing that action (asserting) in the conversation in question. On the other hand, a speaker's assertion is warranted only if the speaker has warrant for the proposition asserted (i.e., a good reason to believe it). The difference between the two readings is the difference between reasons for action (assert*ing*) and reasons for belief (what is assert*ed*). Wright's notion of superassertibility should get the second reading. Wilfrid Sellars makes a similar point with respect to experience in Sellars (1954).

claims like these. I am unaware of anyone who uses superassertibility as the basis for an epistemic theory of truth in general.

Epistemic theories come in for plenty of criticism. Objections include that they have to appeal to a non-epistemic notion of truth to explain ideal circumstances.[34] Another perennial objection is that they fall prey to Fitch's paradox, which is that, assuming some uncontroversial principles about knowledge, one can prove that if all truths are knowable (as the epistemic theorist holds), then all truths are known. That is a huge problem for any epistemic view of truth, and it is the topic of considerable research both by epistemic theorists and by many others as well.[35]

1.7 Deflationist Theories

Deflationism has come to be something like the received view in philosophy, which is a major shift over the past half-century. There are many versions of deflationism and it is hard to give an overall characterization that does justice to them all. One thing they have in common is a negative claim: there is no substantive analysis of truth. Thus, according to deflationism, correspondence theories, coherence theories, and the others rest on a false assumption. Deflationists also think that a principle known as the T-schema is central to a philosophical explanation of truth.

(Schema T) b is true if and only if **p**.

In this schema, 'b' is a name or description of a truth bearer, and '**p**' is a placeholder for a translation of the content of b into the language used to formulate Schema T. In our case, that language is English, so if b is the sentence 'snow is white', then we get the following instance of Schema T: 'snow is white' is true if and only if snow is white. If b is Langdon Alger's belief that Squishees are

[34] See Plantinga (1982), Wright (2000), and Nolt (2008).
[35] See Kirkham (1995: chs. 7 and 8), Alston (1998: ch. 7), Wright (2000), Kunne (2003: ch. 7), Fox (2008) for discussion of epistemic theories; see the papers in Salerno (2009) for discussion of Fitch's paradox.

delicious, then we get the following instance: Langdon Alger's belief that Squishees are delicious is true if and only if Squishees are delicious.

In a recent collection on deflationism, the editors, Jc Beall and Bradley Armour-Garb, define deflationism as the view that instances of the T-schema are conceptually and explanatorily fundamental.[36] They are conceptually fundamental in that they do not follow from definitional relations holding between truth and other concepts. They are explanatorily fundamental because there is no unifying account of why they hold and they explain the expressive role of truth predicates. The expressive role of truth predicates is a focus of Chapter Six, but it involves the fact that we use truth predicates to indirectly assert or deny propositions that we cannot directly assert or deny, and the fact that we use truth predicates to formulate certain generalizations that would otherwise be inexpressible.

One important point that deflationists seem to agree on is that if deflationism is true, then it is illegitimate to appeal to truth in philosophical explanations. If that is correct, then that is a big cost since truth is one of the most popular explanatory tools in the philosophers' kit. It is often invoked as part of the standard explanations of knowledge, meaning, assertion, belief, validity, objectivity, and rationality, just to name a few. The extent to which a deflationist can accept theories of these concepts that appeal to truth is a very delicate issue, and I discuss it in Chapter Six.

With so many philosophers endorsing deflationism, it should not come as a surprise that there are many versions of the view. Information Box 4 summarizes important work on the four most prominent families: expressivism, disquotationalism, minimalism, and prosententialism.

---

[36] Beall and Armour-Garb (2005: 3-6).

Varieties of Deflationism

Information Box 4

Aletheic Expressivism: prima facie truth attributions are not assertions, and they do not purport to describe the world; instead they express a commitment or attitude of the speaker.

- Ayer (1936)—truth attributions express the attitude of acceptance.
- Strawson (1950)—truth predicates function only as devices of endorsement.
- Kraut (1993)—truth attributions express a commitment of the speaker, but there is still a substantive difference between truth apt truth bearers and non-truth-apt truth bearers.
- Price (2003)—truth attributions express commitment to a norm of assertion that goes beyond warranted assertibility.
- Schroeder (2010)—truth attributions express a commitment of the speaker, and one can give a commitment-based semantics for 'true'.

Disquotationalism: sentences are primary truth bearers, and the sentential T-sentences for a language (or for sentences one understands) characterize the truth predicate for that language (or one's idiolect).

- Quine (1970)—truth predicate serves to cancel quotation marks.
- Leeds (1978)—defends naturalistic instrumentalism, which holds that representational relations between language and the world play no role in explaining language; disquotationalism explains the utility of truth predicates.
- Williams (1986)—disquotationalism is epistemologically and metaphysically neutral.
- McGee (1993)—defends idiolectic disquotationalism (i.e., T-sentences for my idiolect are necessary and analytic).
- Field (1994)—defense of methodological deflationism, introduces pure disquotationalism, extended disquotationalism, and quasi-disquotationalism.
- Williams (1999)—disquotationalism is compatible with Davidson's truth-conditional theory of meaning.
- Halbach (2002)—defends modalized disquotationalism (i.e., combines pure disquotationalism with theory of necessity to get a stronger theory that avoids several of the main objections to disquotationalism).
- Azzouni (2002)—offers a theory of anaphorically unrestricted quantifiers and uses them to formulate a version of disquotationalism that permits truth attributions to sentences of other languages.

Minimalism: propositions are primary truth bearers, and the propositional T-sentences characterize the truth predicate.

- Ramsey (1927)—advocates the redundancy theory, which works only for truth operators (i.e., 'it is true that').
- Horwich (1990)—presents minimalism as a theory of truth predicates.
- Soames (1999)—defends minimalism together with Kripke's approach to the liar.
- Hill (2002)—defends substitutionalism, which uses substitutional quantifiers to get a stronger theory that avoids some of the objections to minimalism.

Prosententialism: sentences containing 'true' inherit their content anaphorically from sentences that do not contain 'true'.

- Williams (1969)—analyzed sentences containing 'true' by introducing Prior's prosentential devices along with a sentential quanifier.
- Grover, Camp, Belnap (1975)—introduced the notion of prosentences as analogous to pronouns and argued that sentences containing truth predicates are prosentences.
- Brandom (1988)—truth expressions are prosentence-forming operators.
- Lance (1996)—the prosentential theory of truth is compatible with representational (e.g., truth-conditional, or causal, or nomic) theories of meaning.

Expressivism about truth is similar to expressivist views in other areas of philosophy—proponents distinguish between words that purport to represent the world and words that serve some other purpose, which is usually said to be expressing an attitude or commitment of the speaker.[37] Expressivists about truth differ on what exactly is expressed using truth predicates, but they all agree that when a speaker calls something true, that speaker is not describing that thing; in particular, there is no property of truth that the speaker is attributing to the thing in question.

Disquotationalism is the view that the truth predicate acts like the opposite of quotation marks, so the sentence "snow is white' is true' is equivalent to the sentence 'snow is white'. The sense of equivalence is strong enough to preserve what the sentences are about—disquotationalists say that when a speaker asserts that 'snow is white' is true, that person is not attributing some property, truth, to a sentence, but rather is asserting that snow is white.[38]

Disquotationalists usually treat sentences as primary truth bearers, although Hartry Field appeals to computational roles for that purpose.[39] They also claim that the instances of the T-schema (where b is a primary truth bearer) are necessary. That is, even in possible worlds where 'snow is white' does not mean that snow is white, the T-sentence, "snow is white' is true if and only if snow is white', is true.[40] How can this be? The key to understanding it is that a disquotational theory of truth is really a theory of a language-specific truth predicate (e.g., 'true-in-English'), and each language-specific truth predicate gets its own disquotational theory. A language-specific truth predicate applies only to sentences of a particular language. So "snow is white' is true-in-English' is true, but "Schnee ist weiss' is true-in-English' is false, even though the German sentence 'Schnee ist weiss' is true. If L is a language, then a disquotational theory of truth-in-L is just the set of T-

---

[37] See Blackburn (1984), Gibbard (1990), and Schroeder (2007) for examples.
[38] See Quine (1970) and Field (1994a) for this view.
[39] Field (1994a).
[40] Field (1994a) and McGee (1993).

sentences for sentences of L (e.g., "snow is white' is true-in-L if and only if snow is white'). In other possible worlds where 'snow is white' has some other meaning, it is still true that "snow is white' is true-in-English if and only if snow is white', where 'English' refers to the language as it is spoken in the actual world. As a consequence, disquotationalists have difficulty explaining the relation between truth predicates of natural language and the language-specific truth predicates on which their theories focus. I discuss the consequences of these features of disquotationalism at length at the end of this chapter and in Chapter Six.

Minimalism is very similar to disquotationalism in that both theories consist of lists of T-sentences, but minimalism uses propositions as primary truth bearers instead of sentences.[41] As such it does not have as much trouble with language-specific truth predicates, but that benefit is purchased at the cost of having to give an account of propositions. It turns out that many popular views on propositions explain them in terms of truth, and since most philosophers agree that deflationism is incompatible with theories that cast truth in an explanatory role, minimalists have to give an account of propositions without appealing to truth. It is unclear whether this problem for the minimalist is more or less of a burden than the affiliated problems facing disquotationalists.

The fourth major family of deflationary theories is prosententialism, which says that sentences containing truth predicates behave very much like pronouns.[42] For example, in the sentence 'Barney drank a beer and he fell off the stool', 'he' is a pronoun that inherits its semantic content from the name 'Barney', which occurs earlier in the sentence. This phenomenon is called *anaphora*, and it is pervasive in natural languages. Prosententialists say that a sentence like "snow is white' is true' anaphorically inherits its semantic content from 'snow is white' in the same way. The theory gets its

---

[41] Horwich (1982, 1997, 1998, 1999, 2001, 2004, 2005, 2006, 2008, 2009)
[42] See C.J.F. Williams (1976, 1992), Grover, Camp and Belnap (1976), Brandom (1994), and Lance (1997), Scharp (2009b) and Brandom (2009). See also Armour-Garb and Woodbridge (2010), which defends a version of prosentantialism.

name from the claim that "snow is white' is true' is a pro*sentence*, just as 'he' is a pro*noun*. Prosententialists offer subtle and complicated theories for explaining how more complex sentences containing 'true' work (e.g., 'Everything Moe said last night is true').

Prosententialists typically take sentences to be primary truth bearers, but anything that can enter into anaphoric relations would do. According to prosententialists, when someone asserts "snow is white' is true', one is simply asserting that snow is white since the prosentence inherits all its content from its antecedent. Thus, prosententialists claim that although 'is true' is grammatically a predicate, it does not behave like a predicate when it comes to semantics (i.e., the way it contributes to the meanings of the sentences in which it occurs).

Given its popularity and revolutionary consequences, it is no wonder that deflationism has come in for plenty of criticism. Information Box 5 details many of the most common objections that have appeared. More than in any other theory, the disputes about deflationism are subtle, complex, and interrelated. To the novice, this literature can seem like a maze. For this reason, I include a detailed account of how the debates about these objections have unfolded and a preliminary description of the relations between some of them in an appendix to this chapter.

Common Objections to Deflationism Information Box 5

1. Modality and T-sentences: deflationism implies that the T-sentences as necessary, but it seems that they are false in other possible worlds (e.g., where we use our words differently).

2. Normativity: a deflationist about truth cannot explain why truth is a norm of assertion and inquiry.

3. Indeterminacy: deflationists cannot explain indeterminacy (presupposition failure, vagueness, etc.) since indeterminacy is typically thought of as being neither true nor false.

4. Meaning: a deflationist about truth cannot accept a truth-conditional theory of meaning since the latter employs an inflationary notion of truth.

5. Success: deflationists cannot explain why true beliefs lead to satisfying our desires or why true theories lead to accurate predictions.

6. Non-Factualism: deflationists cannot endorse any form of non-factualism because non-factualism requires inflationist theory of truth-aptness and deflationism requires deflationary theory of truth-aptness.

7. Maximal sets of T-sentences: deflationism cannot avoid paradoxical T-sentences by appealing to a maximally consistent set of T-sentences for a language.

8. Ideology: deflationism requires that the T-sentences fix the meaning of 'true'; but if that is correct, then one would have to have to master every concept expressible in a language to possess the concept of truth for that language.

9. Generality: deflationist theories of truth cannot derive important truths about truth (e.g., a conjunction is true if and only if both conjuncts are true).

10. Foreign sentences: deflationism is incapable of explaining our practice of attributing truth to sentences of other languages or sentences we do not understand.

11. Conservativeness: deflationism should be a conservative theory of truth (i.e., for sentences not containing a truth predicate, the theory does not allow one to prove anything that cannot already be proven without it), but conservative theories of truth are inadequate.

1.8 Modest Theories

Modest theories of truth have not attracted much attention, but they attempt to keep most of the benefits of deflationism without the costs. Many of them take their inspiration from the famous quote of Aristotle: "To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, and of what is not that it is not, is true."[43] Probably the most well-developed modest theory is due to Künne, which defines truth in the following way:

(Modest) $\forall x$ (x is true if and only if $\exists p$ (x=$\langle p \rangle$ and p)).

In this formulation there are two kinds of variables. 'x' is a regular individual variable that is a placeholder for a singular term. 'p' is a sentential variable (placeholder for a sentence) and '$\langle p \rangle$' is a name for whatever sentence takes the place of 'p'.[44] For example, that snow is white is true iff there is a proposition p such that that snow is white = $\langle p \rangle$ and p. There is such a p, namely, the proposition that snow is white. Thus, an instance of the right hand side of (Modest) is 'that snow is white = the proposition that snow is white and snow is white'.[45]

The benefit of a modest theory over inflationist theories is that it does not offer an analysis of truth in terms of some other concepts, so it does not run into problems specifying the nature of correspondence, coherence, or epistemically ideal conditions in the right way. The benefit of a modest theory over deflationism is that the modest theory is more powerful than minimalism or disquotationalism since it is not just a list of T-sentences. Further, it is more intuitive than expressivism or prosententialism since it implies that truth predicates behave semantically just like other genuine descriptive predicates.[46]

---

[43] Aristotle (*Metaphysics* 1011b25).
[44] Künne (2003: 333-374).
[45] Notice the presupposition that propositions are primary truth bearers.
[46] Other modesty theorists include Carnap (1942: 187), Kneale (1972), Mackie (1973: 52-55), and Alston (1996: ch. 1). It seems to me that the theory presented in Gupta and Belnap (1993) should be classified as a modest theory as well. They claim that truth is defined by the T-sentences, but they think that this makes truth a "circularly defined concept"; accordingly, they offer a theory of these concepts and apply it to truth. The result is not a version of deflationism, but it seems to have much in common with other modest theories. I discuss their view in Chapters Two, Three, and Four.

There is very little secondary literature on modest theories, but given the nice combination of features they seem to have, they should attract more attention in the future. One worry is that the interplay between the individual and sentential quantification might cause some problems depending on how they are explained.[47]

1.9 Pluralist Theories

Pluralist theories of truth are the new kids on the block, and the have yet to be developed very extensively. They, like modest theories, attempt to keep the positive aspects of deflationism without the problems. Crispin Wright deserves the credit for introducing this idea into analytic philosophy in his influential Waynflete lectures that were published in 1992 as *Truth and Objectivity*.[48] There Wright criticizes deflationist theories and, he claims that a theory of truth should take the form of a group of platitudes about truth; however, it might be the case that these platitudes are not specific enough to identify a unique property. Wright offers the following list of platitudes:

(i) To assert something is to present it as true.

(ii) Truth apt statements have truth apt negations.

(iii) To be true is to correspond to the facts.

(iv) A statement might be true without being justified, and vice versa.

(v) Truth is absolute and has does not admit of degrees.

(vi) Truth is timeless.

(vii) Truth is objective.[49]

Wright concedes that this list is not exhaustive and that there is room for dispute about exactly which principles should count as platitudes. He argues that in different discourses, different

[47] See Hofweber (2005) for details.
[48] Wright (1992).
[49] Wright (1992: ch. 1).

properties can play the role of truth because these properties satisfy the platitudes. For example, it might be that in a discussion about Jupiter, *correspondence to facts* is the property had by true truth bearers, whereas in a conversation about the moral permissibility of abortion, true truth bearers do not correspond to anything, but they are ideally justifiable. On this view, there is a single concept, truth, and the predicate 'is true' is univocal (i.e., it has a single meaning—it is not ambiguous), but the truth predicate signifies different properties in different discourses (likewise, the singular term 'truth' denotes different properties in different discourses). Wright concentrates on properties and says little about how to think about the application relation (i.e., whether truth predicates have the same extension in every discourse) on this view.[50]

Another version of pluralism is due to Michael Lynch, who claims that there is a single property signified by a truth predicate (and presumably a single extension as well) in every discourse, but it is a functional property in that it can be instantiated by different underlying properties, including correspondence to facts, ideal justifiability, and so on.[51] Lynch agrees with Wright that the properties capable of being truth properties are those that satisfy certain platitudes about truth. For Lynch these are:

(Objectivity) The belief that p is true if and only if with respect to the belief that p, things are as they are believed to be.

(Norm of Belief) It is prima facie correct to believe that p if and only if the proposition that p is true.

(End of Inquiry) Other things being equal, true beliefs are a worthy goal of inquiry.[52]

The main difference between Wright and Lynch is that Wright thinks that truth predicates signify different properties in different discourses, whereas Lynch claims that there is a single property,

[50] Note that Wright calls his view 'minimalism', but it is distinct from the minimalism that is a brand of deflationism. Wright knowingly gave his substantive theory of truth the same name as Horwich's deflationary theory because it is "what would-be deflationists like Horwich ought to advocate" (Wright 1992: 13n13). Philosophy has enough ambiguities without them being intentionally introduced, so I use 'pluralism'.
[51] Lynch (2001, 2004, 2005, 2008, 2009)
[52] Lynch (2009: ch. 1).

being true, but it is metaphysically realized by different properties in different discourses; Wright's pluralism is at the level of signification or designation (predicate-property relation, singular term-property relation, respectively), and Lynch's pluralism is at the level of realization (property-property relation).[53]

There are plenty of issues surrounding aletheic pluralism, and since this is the youngest family of theories on the scene, it is no wonder that it will take a while to sort them out. One issue is how to deal with compound truth bearers whose components are from different discourses (e.g., 'Jupiter is a planet and abortion is morally permissible') and with complex truth bearers that quantify over different subject-matters (e.g., everything Armen said yesterday is true). Mixed inferences involving claims from different discourses are also a problem.[54] Another worry is that pluralist theories inherit most of the problems of the other theories of truth (e.g., for discourses where the truth property is realized by the property of corresponding to facts, the pluralist has the same trouble as the correspondence theorist in explaining the correspondence relation).

Two other views might also be called pluralist. The first is what Matthew McGrath calls *weak deflationism*, which is that a correspondence theory of truth is correct for sentences, and minimalism is correct for propositions. So this is a kind of pluralism about truth that is based on they type of truth bearer involved.[55] Again, it inherits most of the problems of correspondence theories as well as those of minimalism.

The final pluralist view is that the predicate 'is true' and the singular term 'truth' are ambiguous and have two (or perhaps more) distinct meanings. This view has been defended by Max Kölbel, but it also surfaces from time to time in other writings.[56] Kölbel argues that one of the meanings of

---

[53] Lynch (2001, 2009)
[54] Lynch argues that his realization-pluralism does a better job with these than Wright's signification-pluralism.
[55] McGrath (2000)
[56] Kölbel (2008); see also McGee (2005b).

'true' is a deflationist meaning and the other is a correspondence meaning. One problem with this view is linguists have several diagnostics for ambiguity and 'true' seems to fail all of them (more on this in Chapter Six).

1.10 Alternatives to Analysis

Not all those who offer theories of the nature of truth are in the business of analysis; some are engaged in a more modest task. For instance, Wright makes the following observation:

> [N]otwithstanding the fact that it rationalizes many of the moves made, and doubtless reflects therefore the intentions of many of the protagonists, the conception of the traditional debate about truth as centered upon reductive analysis of the concept is not best fitted to generate the most fruitful interpretation of it. To see this, suppose for the sake of argument that the indefinabilists are right, that 'true'—like, say, 'red'—admits of no illuminating conceptual breakdown. It is striking that philosophical discussion of color has hardly been silenced by the corresponding point about the concept red or basic color concepts generally. … So, consistent with its indefinability—if it is indefinable—a similar range of issues can be expected to arise in connection with truth. … Whatever [truth bearers] we have in mind, we can ask whether one of them being true is in any way an *implicitly relational* matter—and if so, what are the terms of the relation; whether it is a *response-dependent* matter, or in any other way dependent on subjectivity or a point of view; whether there is indeed nothing generally in which the truth of such an item consists—whether an inventory of all the properties to be found in the world would include mention of *no such thing as truth*.[57]

I agree with this point. Many of the major theories of truth have been proposed and defended as analyses of the concept of truth, but we can consider how they fare at other explanatory tasks. One question is: how else should they be treated? If they are not giving an analysis of the concept, what might they be doing?

1.10.1 Künne's Classification

[57] Wright (2003: 244-5). Künne (2003) makes a similar point.

Künne's analytical framework provides a range of options. All the traditional theories of truth offer something like the following universally quantified biconditional:

For all x, x is true if and only if x is F

Künne suggests the following five options for interpreting this claim:

(1) 'is F' expresses a concept *coextensive* with the concept of truth (same extension)

(2) 'is F' expresses a concept *necessarily coextensive* with the concept of truth (same intension)

(3) 'is F' expresses a concept that *can be known apriori to be coextensive* with the concept of truth (same intension)

(4) 'is F' expresses a concept that *is self-evidently coextensive* with the concept of truth (same intension)[58]

(5) 'is F' expresses a concept that *is identical* to the concept of truth (same sense)

Satisfying condition *n* is necessary but not sufficient for satisfying condition *n*+1.[59] Option (5), the strongest reading, is the one on which a theory of truth offers an analysis of the concept of truth. The other four are less demanding and, thus, can be thought of as alternative readings of theories of truth in line with Wright's comments.

(1) is hardly demanding at all, requiring only a concept that happens to apply to all and only true truth bearers. (2), (3) and (4) are much more strict and demand a predicate with the same extension as truth in all possible worlds . (3) and (4) impose additional epistemic constraints as well. Although I am not going to consider the four alternative readings of all the theories discussed above, the reader is invited to do so.

1.10.2 Davidson's Theory of Truth

---

[58] Two predicates express self-evidently coextensive concepts if and only if nobody who fully understands them can believe that one of them applies to (does not apply to) a certain entity x without believing that the other one applies to (does not apply to) x.
[59] Künne (2003: 25-27).

Giving up on the project of analysis not only opens up alternative readings of the theories of truth above, but it also permits new kinds of theories. The one most important and influential of these is Donald Davidson's theory. Davidson argues that although we cannot give an analysis or reductive explanation of truth, we can trace the relationship between truth and other important concepts, like meaning, belief, and rationality.

Instead of using one of the universally quantified biconditionals discussed above, Davidson, inspired by Tarski's writings on truth, uses an axiomatic theory of truth. An axiomatic theory of truth is just a set of sentences that contain truth predicates and is closed under logical consequence (i.e., any sentence that is a logical consequence of some sentences in the set is also in the set). The sentences in the theory are taken to be the principles that truth predicates obey. Davidson uses an axiomatic theory that has different principles for sentences of different forms; for example, a subject-predicate sentence is true iff[60] the thing referred to by the subject term is in the extension of the predicate, a negation is true iff the sentence negated is not true, and a conjunction is true iff both conjuncts are true. From the axiomatic theory of truth (for a certain language), one can derive a T-sentence for each sentence of the language in question: b is true-in-L if and only if q (where $\langle\mathsf{q}\rangle$ is a translation of b into the metalanguage).

This axiomatic theory of truth is quite different from any of the theories of truth discussed above. It does not give a conceptual analysis of truth in terms of some other (perhaps more basic) concepts. One might be (justifiably) unimpressed if that was all a Davidsonian could say about truth. However, Davidson's view is that the axiomatic theory should be given empirical content, and this tells us much about the concept of truth. Although the details are subtle and complex, the

[60] I frequenly use 'iff' as short for 'if and only if'.

basic idea is that a theory of truth plays an important role in an overall theory of a rational agent's beliefs, desires, and language.

The first step in providing empirical content to the axiomatic theory is realizing that an axiomatic theory of truth (for some language) serves as a meaning theory for that language, where a meaning theory specifies the meanings of each of the sentences of the language.[61] If enough external constraints are placed on the application of the axiomatic theory, then the T-sentences (e.g., "snow is white' is true if and only if snow is white') that are derivable from it specify the meanings of the sentences of that language. That is, the sentence on the right hand side specifies the meaning of the sentence called true on the left hand side.

Davidson also has an account of the external constraints on the axiomatic theory that ensure the T-sentences specify meanings of the language in question; this can be thought of as a general method for applying the axiomatic theory to a particular language or language user. He uses the thought experiment of the radical interpreter to satisfy this demand.

The radical interpreter is supposed to be a person in the situation of trying to understand another person (who I call the target). The radical interpreter has to figure out what the target's sentences mean even though she does not understand the target's language and does not have recourse to a translator or dictionary or any other similar tool. The radical interpreter can use only publically available evidence to figure out what the target's sentences mean. She begins with observable evidence about the target (i.e., the rational entity that speaks the language in question), which includes certain attitudes the target takes toward the sentences of the language (holding-true). The interpreter also has access to which events in the target's environment cause the target to hold-true certain sentences of the language (the distal stimuli). From this basis, the radical interpreter has

---

[61] This claim is controversial; see Foster (1976), Davidson (1976, 1990), and Lepore and Ludwig (2005, chs. 4, 8, 9) for discussion.

to construct an axiomatic theory (which includes specifying the referents of singular terms and the extensions of predicates) for the target's language and a set of beliefs held by the target. Davidson describes a complex sequence of steps the radical interpreter performs to arrive at this goal. One crucial aspect of the construction is that the radical interpreter uses a principle of charity: in the vast majority of cases, when the target holds-true a sentence in a given circumstance, the sentence is true in that circumstance. The radical interpreter is able to go from relatively thin evidence (sentences held-true) to a relatively rich explanatory structure (beliefs and meanings) by assuming that that structure has certain features characterized by the theory of truth.

Although Davidson does not formulate it in these terms, he does draw attention to a problem that is analogous to the one faced by the radical interpreter—how to attribute beliefs and desires to a rational agent on the basis of that agent's (non-verbal) behavior. The latter is a problem addressed by Bayesian decision theory as developed by Frank Ramsey, who lays out a procedure by which a person (whom we might call the *radical rationalizer*) begins with a target's preferences and arrives at a set of degrees of belief and a set of degrees of desire for the target. The formal theory in this case is the theory of probability (for degrees of belief) and the theory of utility (for degrees of desire). One crucial aspect of the construction is that the radical rationalizer assumes that the target maximizes expected utility. The radical rationalizer is able to go from relatively thin evidence (ordinal preferences) to a relatively rich explanatory structure (degrees of belief and degrees of desire) by assuming that that structure has certain features characterized by the theory of probability and the theory of utility.[62]

Davidson argues that Bayesian decision theory and the theory of interpretation need to be combined into a single unified theory of rationality. The theory of interpretation takes the agent's desires for granted in assigning meanings and beliefs, and the Bayesian decision theory takes the

[62] See Ramsey (1926) and Davidson (1974b, 1976a, 1990).

meanings of the agent's sentences for granted when assigning degrees of belief and degrees of desire. Moreover, the three concepts (belief, desire, and meaning) are equally unavailable to someone in the position of the radical interpreter and they are all on equal footing in terms of explanation.

To construct a unified theory for a target, a theorist begins with a particular relational attitude, preferring-true, that obtains between the target and pairs of sentences. From this basis, Davidson shows how to arrive at a meaning theory for the target's language (in the form of a Tarskian axiomatic theory of truth), a theory of the target's degrees of belief (in the form of a theory of subjective probability), and a theory of the target's degrees of desire (in the form of a theory of utility). The theorist uses both principles from above (i.e., charity and expected utility maximization) in addition to another, the *requirement of total evidence for inductive reasoning*: give credence to the hypothesis supported by all available relevant evidence. The procedure for constructing the unified theory is a combination of the radical interpreter's procedure and the radical rationalizer's procedure.[63]

Davidson's unified theory is really a unification of formal semantics (used to explain meaning) and formal epistemology (used to explain beliefs and desires). His fundamental question is: what do rational entities have to be like in order for them to be able to interpret one another? His answer is that the unified theory—which uses the axioms of probability to as constraints on degrees of belief, the axioms of utility theory as constraints on degrees of desire, and the axioms of a Tarskian theory of truth as constraints on meaning—describes what a rational agent's attitudes and linguistic competence have to be like so that her mental states and utterances are interpretable by another.

For Davidson, truth cannot be analyzed or reduced to something more primitive—he is not engaged in conceptual analysis or reductive explanation. However, a theory of truth describes certain essential features of rationality; it is an integral part of a unified theory of meaning, belief,

[63] Davidson (1974b, 1980b, 1990, 1996).

desire, and action. Moreover, Davidson claims that these aspects of the unified theory, along with other principles involved (e.g., the principle of charity, the principle of continence, the principle of total evidence), are partly constitutive of the concepts of truth, belief, desire, meaning, action, and rationality. In addition, the unified theory provides a way of giving an empirical interpretation to the axiomatic theory of truth at its heart.

It seems to me that Davidson's unified theory and developing an analogy he draws between the unified theory and measurement theory illuminates many of his other philosophical views. He does not make much of this, so I will not bother with it here. However, I say much more about Davidson's view and this analogy later in the book; indeed, it is the account of the nature of truth I end up advocating in Chapter Fourteen.

1.11 'True in', 'True at', and 'True for'

So far, the entire discussion has been about a monadic truth predicate, which combines with only one singular term to make a sentence. However, there are several polyadic truth predicates bouncing around the literature, and they can be somewhat confusing. In this section I discuss 'true in L' (where 'L' is the name of a language), 'true in C' (where 'C' is the name of a context), 'true in M' (where 'M' is the name of a model, which is a kind of mathematical structure), 'true at w' (where 'w' is the name of a possible world), and 'true for P' (where 'P' is the name of some parameter, like a person, a group of people, a culture, a time period, or a conceptual scheme).

The first kind of polyadic truth predicate is a language-specific truth predicate (LS truth predicate), like 'true-in-English', as discussed above in connection with disquotationalism. These truth predicates apply only to sentences, with the assumption that sentences are individuated coarsely so that 'Bret rang', which is a sentence of German meaning that Bret wrestled, and 'Bret

rang', which is a sentence of English meaning that Bret rang, are identical.[64] LS truth predicates are the dominant focus of disquotationalists and of those who work on the liar paradox. Indeed, we will see that they play a crucial role in the standard response to revenge paradoxes (this is a topic of Chapters Eight and Nine). Anyone who offers a theory of truth that takes LS truth predicates as its primary explanandum owes us a story about how they relate to natural language truth predicates. I argue in Chapter Six that the stories offered so far are inadequate.

'True-in-a-context' is a polyadic truth predicate that plays an important role in semantic theories for context-dependent expressions, which include indexicals (e.g., 'here', 'I'), demonstratives (e.g., 'this'), pronouns (e.g., 'he', 'it'), quantifiers (e.g., 'all', 'some', 'most'), and gradable adjectives (e.g., 'tall', 'flat'). Following David Kaplan, we can distinguish two kinds of meaning for these kinds of expressions, their *character*, which is invariant, and their *content*, which differs from use to use and is determined by features of the context.[65] There are other linguistic phenomena that require a notion of truth-in-a-context; for example, any ambiguities must be resolved before a sentence has a truth-conditional content, so it might be that 'Gil is at the bank' is true if 'bank' means *financial institution*, but false if 'bank' means *river side*. Instead of saying that the same sentence is true and false, we say that the sentence is true in one context and false in another. John Macfarlane sums this up well when he writes "the condition for a sentence to be true at a context is the central semantic fact we need to know if we are to *use* the sentence and understand others' uses of it. Truth at a context is the point at which semantics makes contact with pragmatics, in the broad sense—the study of the use of language."[66]

How are 'true' and 'true-in-a-context' are related? Sentences, qua noises, or marks on paper or bits of data, are not true or false any more than walnuts or plankton are true or false. It is only as

---

[64] This example is similar to one in Sawyer (1999).
[65] Kaplan (1989).
[66] MacFarlane (forthcoming d: 77).

meaningful vehicles that sentences are true or false. Sentences that have context-dependent features are meaningful in the way required for being true or false. However, there are two ways to think about how to assign truth values to these sentences: (i) truth values are assigned to pairs consisting of a sentence and a context, or (ii) truth values are assigned to a sentence relative to each context. Either way, the context supplies the additional information required for truth values to be assigned.

One might wonder exactly what a context is, and this is a legitimate issue since the term is used in so many ways. Furthermore, for reasons associated with what is known as the principle of compositionality (i.e., the meaning of a complex sentence is determined by its structure and the meanings of its parts), one needs to distinguish between a context and an index. I put off these technical details until they are needed in Chapter Seven.

Truth in a model is a technical notion used in mathematical logic, philosophical logic, and formal semantics.[67] It is based on the notion of a model, which is a certain mathematical structure. Models are assigned to languages, but only certain artificial languages that have an exactly specified syntax can have a model. For certain purposes, we can treat fragments of natural language as if they have an exactly specified syntax, but this assumption is unrealistic if taken literally. A model of a language contains two parts—a domain of objects, and an assignment function that maps each expression of the language onto a set-theoretic object. For example, a predicate (e.g., 'is a doctor) gets a subset of the domain, a name (e.g., 'Bob') gets an object from the domain, and a term function (e.g., 'the father of') gets assigned a function from the domain to the domain. Logical terms (e.g., 'not', 'and', 'or', 'all', 'is identical to') do not get assignments; instead, they get special clauses in the definition of truth in a model.

The definition of truth in a model is usually recursive (although more complex definitions are sometimes used—we will see one in Chapter Thirteen); it takes the atomic sentences as its base.

[67] One also sees 'truth in an interpretation' instead.

Atomic sentences are those without any logical vocabulary (e.g., 'the father of Bob is a doctor'), and whether they are true in a model or false in a model is determined by the relations between the set theoretic objects assigned to its constituents (e.g., 'the father of Bob is a doctor' is true in a model M iff the output of the function M assigns to 'the father of' when the object M assigns to 'Bob' is given as input is in the set M assigns to 'is a doctor'). The recursive step of the definition of truth in a model is split into clauses for each bit of logical vocabulary in the language (e.g., a conjunction, p and q, is true in a model M iff p is true in M and q is true in M).

It is important to understand the difference between truth and truth in a model. Truth is an everyday notion that can be applied freely to all sorts of truth bearers for all kinds of reasons. Saying that a sentence is true in a model presupposes that the sentence belongs to an particular kind of artificial language, that there is a particular model of that language, and that one is treating the expressions of that language as being about the items in the domain of that model. Since the model and the artificial language are just mathematical objects, truth in a model is just a mathematical concept. These assumptions are all quite dubious when it comes to natural languages. Still, we will see that truth in a model is a very helpful tool.

The fourth polyadic truth predicate to be explained is 'true at a world'. Philosophers and logicians frequently use the idea of possible worlds to explain modal vocabulary and phenomena ('modal' usually means *pertaining to necessity and possibility*). Here 'world' means an entire universe, not just Earth. There are many debates about the nature of possible worlds; some treat them as mere technical devices in logic and formal semantics, some take them to be ways the actual world could have been, and some claim that they are concrete things that exist independently of the actual world. Whatever one's view, philosophers find it helpful to think about the truth values of truth bearers if the actual world had been different. For example, instead of reading this book right now, you could have been sleeping or eating or doing countless other things. We can consider whether various

sentences, beliefs, or propositions would have been true or false if you had in fact done one of those other things. In general, truth at a world can be thought of as truth if the world had been different. It is common to think that the truth value of a sentence depends both on what it means and how the world is. The notion of truth at a world allows us to consider what happens when one keeps the meaning of the sentence fixed and varies the way the world is. This notion allows one to consider the truth value of a single proposition, say that Wiggum is chief of police, in different circumstances of evaluation. This proposition is true in some worlds (where Wiggum is chief of police) and false in others (where Wiggum's life took some other path).

Caution is required when thinking about truth at a world since many mathematical models employ a framework of possible worlds for various purposes in logic and formal semantics. In these cases, truth at a world is given an exact mathematical definition as part of a definition of truth in a model. I discuss the details of how this works in Chapters Thirteen and Fourteen.

Finally, we have the 'true for' locution. This term is almost always associated with relativism, which holds that some class of (or perhaps all) truth bearers are not absolutely true or false, but only true or false relative to some standard or parameter. For example, moral relativists often take ethical claims (e.g., that abortion is wrong) to be true or false only relative to a person or culture.[68] To make the relativity explicit we can say that 'abortion is wrong' is true for Tim, but not true for Apu. Relativists think that these truth bearers, by themselves, are not true or false; in order to assign a truth value to them, one needs additional information about the relevant standard or parameter. They mark this feature by using the phrase 'true for X', where 'X' is a placeholder for the relevant standard or parameter.

There are many varieties of relativism, and they have been a part of the conversation in Western Philosophy for millennia. Throughout most of that time, relativism has been treated as a self-

[68] E.g., Harman (2000: chs. 1-5).

refuting or absurd position—the kind of thing one calls an opposing view in an attempt to discredit it. Recently, however, there has been a resurgence of interest in relativism, as its promoters have offered new, technically advanced formulations of the view and suggested new applications for it in explaining puzzling linguistic phenomena.[69] I discuss the details in Chapter Fourteen.

I have explained that there are several kinds of polyadic truth predicates, many reasons for using them, and many applications for them. However, throughout this work, I take it that it is a mistake to interpret everyday uses of truth predicates in natural language as covertly expressing one of the polyadic truth concepts. Instead, the monadic truth concept is the one everyday speakers use. I argue for this position in the case of language-specific truth predicates in Chapter Six, and it seems to me that the argument given there generalizes to the others, but I do not go to the trouble of laying out the details.

[69] See Lasersohn (2005, 2008, 2009), Kölbel (2002), MacFarlane (2003, 2005a, 2005b, 2007a, 2007b, 2008, 2009, forthcoming a, forthcoming d), and Hales (2006). See also Cappelen and Hawthorne (2009, forthcoming a, forthcoming c), the collection by Garcia-Carpenterio and Kölbel (2008) and the special issue of *Synthese* 166: 2 on relative truth edited by Berit Brogaard (January, 2009).

Appendix: Objections to Deflationism

Below are eleven prominent objections to deflationism listed in the order in which they appeared in the literature. Following each objection is a list of important works on that objection in chronological order. After the objections, one will find in Information Box 6 a diagram of some of the relations between the objections and a description of each relation.

1. *Modality and T-sentences*: deflationism implies that the T-sentences are necessary, but it seems that they are false in other possible worlds (e.g., where we use our words differently).
   - Lewy (1947)—perhaps the first to present this worry as an objection to Tarski.
   - Field (1986)—presents a version of this objection.
   - McGee (1993)—endorses necessity of T-sentences; it is a condition on truth serving as device of generalization.
   - Field (1994)—endorses necessity of T-sentences; it is a condition on truth serving as device of generalization.
   - David (1994)—argues that T-sentences are not necessary.
   - Soames (1999)—claims that the necessity of T-sentences supports propositions as primary truth bearers.
   - Halbach (2002)—makes modal status of T-sentences a part of his modalized disquotationalism.

2. *Normativity*: a deflationist about truth cannot explain why truth is a norm of assertion and inquiry.
   - Dummett (1959)—argues that deflationism cannot account for the point of having a truth predicate or that truth is a normative goal of assertion and inquiry.
   - Blackburn (1984)—deflationism cannot account for point of aiming at truth.
   - Wright (1992)—argues that on deflationism, a truth bearer is true if and only if it is assertible, but assertibility does not obey convention T; only a concept that obeys convention T is normative in the way truth is, but such a notion will be substantive.
   - Horwich (1994)—deflationism need not explain the normative aspect of truth; instead, this is a feature of assertion.
   - Tennant (1995)—Wright's (1992) argument does not work for intuitionist deflationist.
   - Rorty (1995)—sides with Davidson against Wright that there is no norm of truth beyond its role in a theory of meaning.
   - Searle (1995)—deflationism cannot account for point of aiming at truth.
   - Shapiro and Taschek (1997)—Wright (1992) argument does not work for an intuitionist deflationist.
   - Clark (1997)—criticism of Blackburn (1984) and Searle (1995).
   - Wright (1999)—reformulates argument from Wright (1992).
   - Dodd (1999a)—there is no norm of truth for assertion—the truth predicate is being used to generalize.
   - Miller (2001)—Wright (1999) is confused about deflationism and the property of truth.
   - Engel (2002)—defends Wright's objection.
   - Price (2003)—truth predicates express a norm that is stronger than assertibility, but deflationism cannot accommodate this norm.
   - McGrath (2003)—criticism of Price (2003); the norm is one of assertion, not truth.

3. *Indeterminacy*: deflationists cannot explain indeterminacy (presupposition failure, vagueness, etc.) since being indeterminate is typically thought of as being neither true nor false.

Dummett (1959)—argues that deflationism is committed to bivalence.
Kripke (1975)—proposes a deflationist theory of truth with truth-value gaps.
McGee (1991)—proposes a deflationist theory of truth and a theory of definite truth based on Kripke (1975).
David (1994)—argues that deflationism owes us theory of indeterminacy.
Field (1994)—offers a deflationist theory of indeterminacy.
Simmons (1999)—approaches to the liar are incompatible with deflationism because they either deny intersubstitutability or appeal to indeterminacy.
Leeds (2000)—offers a deflationist theory of indeterminacy.
Holton (2000)—argues that deflationism is compatible with indeterminacy (as long as it has a new kind of conditional).
Beall (2000)—criticism of conditional in Holton (2000); offers alternative conditional.
Beall (2002)—Dummett (1959) argument confuses exclusion negation and choice negation.
Glanzberg (2003b)—argues that deflationism is incompatible with solutions to the liar paradox.
Gupta (2005)—argues that the liar offers no additional problem for deflationism.
Field (2008)—proposes a deflationist theory of truth and a theory of determinate truth based on Kripke (1975).
Greenough (2010)—argues that deflationists cannot accept truth-value gaps.

4. *Meaning*: a deflationist about truth cannot accept a truth-conditional theory of meaning since the latter employs an inflationary notion of truth.

Dummett (1959)—claims that a Tarskian truth definition can either serve as an explication of truth if one takes meaning for granted or as a theory of meaning if one takes truth for granted, but it cannot do both.
Field (1986)—argues that deflationism is incompatible with truth-conditional theories of meaning.
Horwich (1990)—argues that deflationism requires a non-truth-conditional theory of meaning.
McGee (1993)—argues that deflationism is incompatible with a truth-conditional theory of meaning.
Field (1994)—argues that deflationism is incompatible with truth-conditional theories of meaning.
Brandom (1994)—accepts that deflationism is incompatible with a truth-conditional theory of meaning; offers an inferentialist theory of meaning.
Lance (1997)—argues that the prosentential theory of truth is compatible with truth-conditional theories of meaning.
Horwich (1998)—offers a use theory of meaning as compatible with deflationism.
Kemp (1998)—argues if each sentence expresses a proposition then meaning is truth conditions if and only if truth is *not* a property.
Williams (1999)—argues that deflationism is compatible with Davidson's truth-conditional theory of meaning.
Dummett (1999)—reformulates argument from Dummett (1959).
Bar-On, Horisk, and Lycan (2000)—argue for truth-conditional theories of meaning and questions whether deflationists cannot accept them.
Kölbel (2001)—argues that deflationism is compatible with Davidson's truth-conditional theory of meaning.

Williams (2002)—claims that deflationist should endorse inferentialist theory of meaning (but Davidson's theory of meaning is covertly inferentialist).
Hershfield and Soles (2003)—criticism of Williams (1999); Davidsonian truth conditional theory of meaning requires truth to be an explanatory concept.
Price (2003)—claims that objections to non-truth-conditional theories of meaning undermine deflationism.
Patterson (2005)—explains incompatibility between deflationism and truth-conditional theories of meaning as depending on distinguishing between object language and metalanguage.
Patterson (2006)—criticism of Bar-On, Horisk, and Lycan (2000).
Horisk (2007)—evaluates several combinations of truth-conditional semantics and deflationism.

5. *Success*: deflationists cannot explain why true beliefs lead to satisfying our desires or why true theories lead to accurate predictions.
Field (1972)—argues that usefulness of truth in explaining success suggests that it is naturalistically reducible.
Leeds (1978)—points out that usefulness alone does not suggest naturalistic reduction; instead it needs to play a causal-explanatory role in laws as well.
Putnam (1978)—claims that truth does play a causal-explanatory role in laws (e.g., success).
Williams (1986)—suggests that truth's role in laws just serves its generalizing function; anything that can be explained with truth can be explained without it.
Field (1986)—argues that truth plays a substantive role in laws that cannot be accounted for by its generalizing function.
Horwich (1990)—explains how deflationist can explain truth's role in laws.
Field (1994)—retraction of Field (1972, 1986).
Leeds (1995)—criticism of Field (1986).
Kitcher (2002)—argues that correspondence explanations of success are better than deflationist explanations.
Damjanovic (2005)—argues that deflationism is incompatible with Jackson/Pettit theory of causal explanation.
Leeds (2007)—argues that scientific realism does not need correspondence truth.
Maddy (2007)—argues that correspondence truth is not needed for naturalistic explanations.

6. *Non-Factualism*: deflationists cannot endorse any form of non-factualism because non-factualism requires inflationist theory of truth-aptness and deflationism requires deflationary theory of truth-aptness.
Boghossian (1990a)—offers these this objection, and argues that deflationism is a variety of non-factualism, so it is self-refuting.
Devitt (1990)—claims that Boghossian (1990) misinterprets deflationism.
Boghossian (1990b)—reply to Devitt (1990).
Devitt and Rey (2001)—reply to Boghossian (1990b).
Wright (1992)—argues that deflationism is incompatible with expressivism.
Kraut (1993)—presents deflationism as compatible with substantive theory of truth-aptness and criticizes Boghossian (1990).
Horwich (1994)—claims that deflationism is incompatible with expressivism.
Field (1994b)—offers deflationist theory of non-factualism based on Gibbard's expressivism.
Smith (1994)—argues that deflationism supports expressivism and criticizes Wright (1992) and Horwich (1993).

Divers and Miller (1994)—argues that deflationism is incompatible with expressivism and criticizes Smith (1994).
Jackson, et. al. (1994)—argues that deflationism is compatible with non-cognitivism.
Drier (1996)—claims that deflationism does not help expressivism solve the Frege-Geach problem and criticizes Horwich (1993).
Burgess (1997)—distinguishes deflationism about truth from deflationism about truth aptness.
Wedgewood (1997)—defends Wright (1992) from Smith (1994) and Jackson et. al. (1994).
Blackburn (1998)—argues that deflationism is compatible with expressivism and quasi-realism; criticism of Wright (1992) and Boghossian (1990).
Wright (1998)—reformulates argument in Wright (1992); criticism of Blackburn (1998).
Holton (2000)—argues deflationism is compatible with indeterminacy (as long as it has a new kind of conditional) but not with expressivism.
Richard (2008)—offers detailed investigation of truth and non-factual discourse.

7. *Maximal sets of T-sentences*: deflationism cannot avoid paradoxical T-sentences by appealing to a maximally consistent set of T-sentences for a language.
McGee (1992)—proved that there are many maximally consistent sets of T-sentences, they are incompatible with one another, they overlap only on "truth-teller" sentences, and they are not recursively axiomatizable.
Weir (1996)—suggests non-classical logic for the deflationist to block argument in McGee (1992).
Gauker (2001)—claims McGee (1992) shows that any way of avoiding paradoxical T-sentences will be incompatible with Gödel's incompleteness theorem.
Gauker (2003)—argues that deflationist needs a special notion of consequence to avoid paradoxical T-sentences.
Armour-Garb and Beall (2003)—claim that there is no way for the deflationist to exclude paradoxical T-sentences.
Leitgeb (2005)—offers theory of dependence as an alternative to maximally consistent sets of T-sentences.

8. Ideology: deflationism requires that the T-sentences fix the meaning of 'true'; but if that is correct, then one would have to have to master every concept expressible in a language to possess the concept of truth for that language.
Gupta (1993)—presents this objection to disquotationalism and minimalism.
David (1994)—endorses this objection against disquotationalism.
Hill (2002)—argues that deflationist need not accept that in order to possess a concept one must possess all concepts in terms of which it is defined.
Künne (2003)—endorses this objection.
Gupta (2005)—replies to Hill (2002).

9. *Generality*: deflationist theories of truth cannot derive important truths about truth (e.g., a conjunction is true if and only if both conjuncts are true).
Gupta (1993)—presents this objection to disquotationalism and minimalism.
Horwich (1999)—adds an infinitary rule to minimalism to deal with the problem.
Halbach (2000)—argues that by attributing particular modal status to the T-sentences, one can derive generalizations.
Field (2001)—appeals to a theory of schematic variables to derive generalizations.

Armour-Garb (2004)—claims that Horwich's solution to the generalization problem is incompatible with his solution to the liar paradox.

10. *Foreign sentences*: deflationism is incapable of explaining our practice of attributing truth to sentences of other languages or sentences we do not understand.

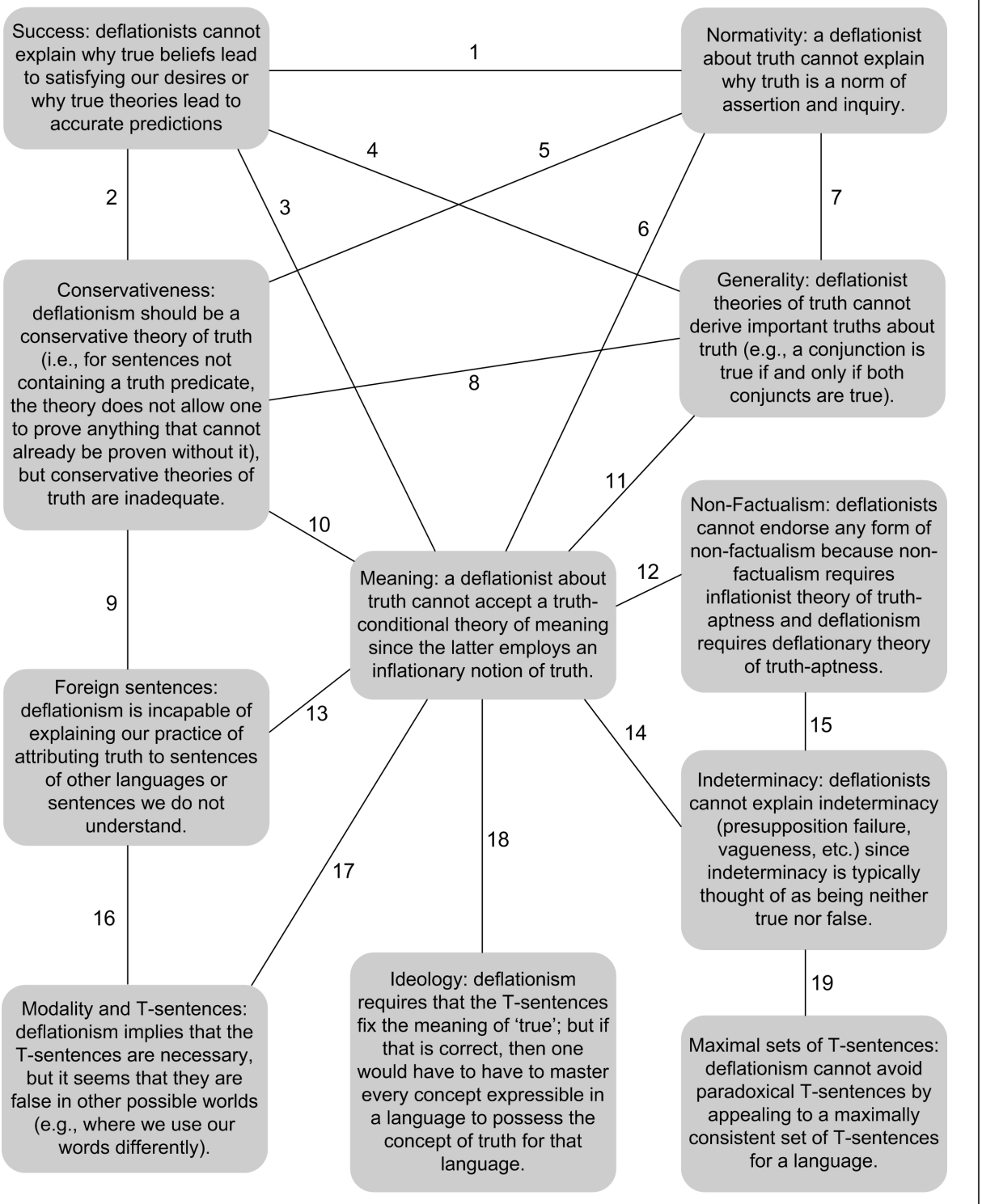
McGee (1993)—offers translation solution to the problem.
David (1994)—presents this objection against disquotationalism.
Field (1994)—introduces extended disquotationalism to address this problem.
Resnik (1997)—suggests that disquotationalism avoid the problem by invoking the polyglot (i.e., the amalgam of all possible human languages).
Field (2001)—replies to Shapiro (2003) by giving up quasi-disquotationalism and advocating a new view on truth bearers.
Williams (2002)—argues that disquotationalist will have to invoke meanings as part of extended disquotationalism.
Künne (2003)—offers several versions of this objection that focus on Field (1994).
Shapiro (2003)—argues against extended disquotationalism by giving an example of untranslatable sentences; also shows that deflationist is committed to unacceptable notion of logical consequence.
Shapiro (2005)—reviews argument in Shapiro (2003); criticism of Resnik (1997) and Field (2001).
Ebbs (2009)—offers a deflationist account of truth attributions to foreign sentences.

11. *Conservativeness*: deflationism should be a conservative theory of truth (i.e., for sentences not containing a truth predicate, the theory does not allow one to prove anything that cannot already be proven without it), but conservative theories of truth are inadequate.

Horsten (1995)—suggests that deflationist theory of truth should be conservative.
Shapiro (1998)—claims that deflationism should be conservative, but a conservative theory of truth requires a notion of logical consequence the deflationist cannot accept.
Field (1999)—argues that there is no reason for the deflationist to accept conservativeness; criticism of Shapiro (1998).
Ketland (1999)—argues that deflationism should be conservative, but a conservative theory of truth will not be able to prove generalizations about truth.
Halbach (1999)—considers relation between deflationism and infinite conjunctions; discusses reducibility of truth and conservativeness.
Ketland (2000)—modifies earlier objection for extended disquotationalism.
Halbach (2001)—distinguishes two kinds of conservativeness; deflationism is not strongly conservative, contra Shapiro (1998), and if deflationism should prove generalizations about truth it isn't weakly conservative either.
Tennant (2002)—argues that the deflationist can keep conservativeness and offer alternative explanations of what a conservative theory cannot explain (e.g., generalizations, reflection principles, consequence relations, Gödel phenomena, etc.); criticism of Shapiro (1998) and Ketland (1999).
Shapiro (2002)—re-evalutation of argument in Shapiro (1998); replies to Field (1999), Halbach (2001), and Tennant (2002).
Ketland (2005)—criticism of Tennant (2002): alternative explanations fail when it comes to reflection principles.
Tennant (2005)—replies to Ketland (2005).

Cieslinski (2009)—argues that deflationist is committed to reflection principles, which make deflationism non-conservative.
Bayes (2009)—argues against Ketland on the definability of deflationary truth.
Ketland (2010)—response to Bays (2009).
Tennant (2010)—argues that deflationist can use schematic reflection principles.
Ketland (2010)—reply to Cieslinski (2009).

Relations between Objections to Deflationism Information Box 6



1. These are both explanatory objections and they are sometimes confused with one another—a successful assertion is true, a successful inquiry reaches truth, a successful scientific theory is true, true

beliefs are likely to lead to successful fulfillment of desires. It is hard to decide which of these are more like the others. Putnam focused on one pair, Dummett and Wright focused on the other. One difference might be direction of fit.

2. A deflationist reply to the Success objection is "anything you can explain with truth you can explain without it", but that depends on a conservativeness claim. There is also a connection between explanation, reduction, and conservativeness.
3. Beliefs have propositional content and if true beliefs tend to help us satisfy our desires, then presumably it is because of their content. Even if one accepts that 'true' plays only a generalizing role in formulating "success" claims, it still might be plausible that only because beliefs have truth-conditional contents that that tend to help us satisfy our desires.
4. These are both explanatory objections. Deflationists typically say that 'true' plays merely a generalizing function in "success" claims, but for some reason they do not say this about other generalizations.
5. There is a tight connection between the Success objection and the Normativity objection (#1), and there's a connection between the Success objection and the Conservativeness objection (#3); same goes here.
6. Assertions have propositional content just as beliefs. Philosophers often take content to be normative. Can content really be normative without it being explained in terms of truth conditions?
7. These are both explanatory objections. Deflationists typically say that 'true' plays merely a generalizing function in "normative" claims, but for some reason they do not say this about other generalizations.
8. There is a very tight connection here—if deflationism can explain generalizations about truth that fall under the Generalization objection, then deflationism is not conservative. So the pressure to explain these generalizations is pressure to be non-conservative.
9. One problem with deflationism being conservative is the consequence relation this requires, and a problem with untranslatable sentences being called true is the consequence relation this requires, so both objections turn on consequence relations. Also, Conservativeness objections are complicated by the move from pure to extended disquotationalism (a move intended to deal with foreign sentences).
10. There are many connections between Gödel phenomena (which figure in Conservativeness objections) and whether meaning is truth-conditional. Gödel phenomena force a distinction between proof theory and model theory, each of which can be used to explain meaning. Only model-theoretic meaning is truth-conditional.
11. Do the generalizations in question hold by virtue of the meanings of the logical connectives, or by virtue of the meaning of the truth predicate? Deflationists might offer a very different account of the relation between logical words and truth functions. So the issue of whether meaning involves truth conditions plays a big role in what one thinks needs to be explained in the generalizations in question.
12. Non-factualists often hold that certain sentences are not truth-apt because of what they mean, so there is a connection between thinking about meaning in terms of truth conditions and the possibility of non-factualism. Moreover, even if the deflationist can provide some characterization of non-factualism, it is an open question as to whether the deflationist can explain the apparent contrast between factual and non-factual *meaning* without appealing to truth-conditions.
13. Deflationists often appeal to meaning or translation when explaining how truth applies to foreign sentences. One issue is whether the deflationist can get the right truth values of foreign sentences without appealing to a truth-conditional account of translation. There is also a whole constellation of issues surrounding indeterminacy of translation and truth-conditional views on meaning that affect the deflationists' strategy for dealing with foreign sentences.
14. Indeterminacy is often thought to affect truth-aptness because of the meaning of the indeterminate sentence in question. So this connection is similar to #12.
15. There is a very tight connection here since non-factualism and indeterminacy are often hard to distinguish. Disputes about one often spill over into the other. An obvious connection is that both non-factualism and indeterminacy are thought to lead to truth-value gaps.
16. Disputes about the modal status of T-sentences often turn on the contrast between homophonic and translational T-sentences. Also, some deflationists think that there is a difference in modal status between them.

17. The meaning of sentences containing 'true' is often thought to be a factor in the modal status of the T-sentences. So, one's views on whether meaning includes truth conditions will impact one's views on whether the T-sentences are necessary, apriori, or analytic.
18. One's views on the concepts one requires in order to understand some claim are tied to one's views on how best to explain the meaning of that claim; if meaning does not include truth conditions, the Ideology objection might not work.
19. One issue that comes up in the Indeterminacy objections is the liar paradox since a very popular way of dealing with the liar appeals to indeterminacy. The liar paradox is the major issue when it comes to deciding which T-sentences a deflationist should include in her theory.

Chapter 2

Philosophical Approaches to Paradox

Now that we have seen an overview of what philosophers have said about the nature of truth, let us turn to the other major tradition to see what has been said about the liar and other paradoxes affecting truth. The main innovation in my presentation of this material is the distinction between philosophical approaches and logical approaches to the paradoxes. Philosophical approaches, the topic of the present chapter, describe some feature of natural language truth predicates that "solves" the paradoxes and (one hopes) explains why we were taken in by them in the first place. Logical approaches are concerned entirely with modeling natural language truth predicates by way of artificial languages—they investigate consistent (or at least non-trivial) combinations of logical principles and principles governing truth; I present them in Chapter Three. Both philosophical approaches and logical approaches use the techniques of formal semantics and mathematical logic; the former focus on semantic features of truth predicates (e.g., ambiguity, context-dependence, indeterminacy), while the latter concentrate on the formal properties of languages containing truth predicates (e.g., obeying classical logic, obeying bivalence, etc.). Few, if any, presentations of this material adhere to this distinction, which I take to be essential to a proper understanding of our current predicament regarding truth and the paradoxes. After outlining these two families of approaches, I consider combinations of them in Chapter Four.

Almost every philosopher working in the tradition that focuses on the liar paradox treats sentences as primary truth bearers and they usually construct artificial languages rather than considering natural languages. Of course, when pushed, they will admit that solving the liar paradox as it arises in natural language is the goal, but natural languages are very messy, and it is much easier

to construct theories for artificial ones.[1] With this habit comes the risk of irrelevance—it can be very difficult to scale up a theory designed for an artificial language so that it works for natural languages.

This chapter begins with a description of the paradoxes that affect the concept of truth; it turns out that the liar is not the only one, or even the most insidious. Following that are two sections, one on the problems posed by the paradoxes and one on the projects one might engage in when addressing them. The rest of the chapter summarizes the philosophical approaches to the paradoxes.

2.1 Aletheic Paradoxes

A *paradox* is some reasoning that begins with intuitively acceptable assumptions and proceeds via intuitively acceptable steps, but arrives at an intuitively unacceptable conclusion. The liar paradox is certainly the best-known paradox affecting truth, but there are several others as well. A crucial component in understanding each paradox is appreciating precisely which principles are at work. The most common are the following (I refer to these throughout the book as the *primary aletheic principles*):

(T-In) If p then $\langle p \rangle$ is true.

(T-Out) If $\langle p \rangle$ is true, then p.

(Sub) If $\langle p \rangle = \langle q \rangle$, then $\langle p \rangle$ is true if and only if $\langle q \rangle$ is true.

In the formulation of these principles, the letter 'p' is being used as a sentential variable. The most common kind of variable is an individual variable, which is a place-holder for a name. With sentential variables, the letter is a place-holder for a sentence. These principles also contain angle

[1] Tarski (1933, 1944) is a notable exception.

brackets, which are used to formulate names for whatever sentence replaces the sentential variable. So an instance of (T-In) would be: 'if snow is white, then 'snow is white' is true; a sentence, 'snow is white' has replaced '⟨p⟩', and the name of that sentence, ' 'snow is white' ', has replaced '⟨⟨p⟩⟩'.

(T-In) and (T-Out) are intuitively acceptable simply because it would be very strange to assert some sentence and also assert that it is not true; likewise, it would be rather odd to assert that some sentence is true, but deny that sentence itself. (Sub) is just a principle that seems to hold of any genuine predicate—to deny it would be to say that a single sentence could be true when called by one name and false when called by another.

There are three well-known paradoxes that use both (T-In) *and* (T-Out): the liar paradox, Curry's paradox, and Yablo's paradox. The liar paradox has been known in one form or another for millennia. Epimenides is thought to have come up with one version around 600 B.C.E., but the most common formulation is probably due to Eubulides (around 300 B.C.E.). One instance concerns the following sentence:

(1) (1) is false.

Using some logic and the primary aletheic principles, we can derive '(1) is true if and only if (1) is false', from which '(1) is true and (1) is false' follows:

1.	Assume (1) is true	
2.	'(1) is false' is true	[(Sub) from 1]
3.	(1) is false	[(T-Out) from 2]
4.	If (1) is true, then (1) is false	[logic from 1-3][2]
5.	Assume (1) is false	
6.	'(1)' is false' is true	[(T-In) from 5]
7.	(1) is true	[(Sub) from 6]
8.	If (1) is false, then (1) is true	[logic from 5-7]
9.	(1) is true if and only if (1) is false	[logic from 4, 8][3]
10.	(1) is true and (1) is false	[logic from 9][4]

[2] The inference rule is *conditional proof*: if one can derive q from assumption p, then one may conclude 'if p then q'.
[3] The inference rule for the *biconditional*: '*p* iff *q*' is equivalent to 'if *p* then *q* and if *q* then *p*'.
[4] 'p and not p' follows from '*p* iff not *p*'. Assume '*p* or not *p*'. 'if *p* then *p*' is a tautology and 'if not *p* then *p*' follows from the biconditional; these two let us derive '*p* or *p*' from the assumption, which is equivalent to *p*. Likewise, 'if not *p* then

There are many sentences that can be used in place of (1) (e.g., '(1) is not true', 'the negation of (1) is true', etc.), and many ways of reasoning to the unacceptable conclusion.[5]

Curry's paradox focuses on conditional sentences. Consider the sentence:

(2) If (2) is true, then 0=1.

Using some logic and the truth rules, we can derive '0=1' as follows:

1.	Assume (2) is true.	
2.	'if (2) is true, then 0=1' is true	[(Sub) from 1]
3.	If (2) is true, then 0=1	[(T-Out) from 2]
4.	0=1	[logic from 1, 3]
5.	If (2) is true, then 0=1	[logic from 1-4]
6.	'if (2) is true, then 0=1' is true	[(T-In) from 5]
7.	(2) is true	[(Sub) from 6]
8.	0=1	[logic from 5, 7]

One could replace '0=1' with any absurd claim; the point is that using the truth principles above and logic, one can derive anything by reflecting on sentences like (2). Curry's paradox has two interesting features—it does not use negation or the notion of falsity at all, and (T-Out) is used inside a subproof, but (T-In) is not (that is different from the reasoning in the liar, where (T-Out) and (T-In) were both used in subproofs).[6] These issues will come up again below.

Yablo's paradox concerns a sequence of sentences instead of just a single sentence. Consider the sequence of sentences:

(3.0) For k>0 (3.k) is not true.

(3.1) For k>1 (3.k) is not true.

(3.2) For k>2 (3.k) is not true.

…

---

not *p*' is a tautology and 'if *p* then not *p*' follows from the biconditional; these two let us derive 'not *p* or not *p*' from the assumption, which is equivalent to 'not *p*'. Putting these two together, we get '*p* and not *p*'.

[5] The argument above is not elegant or perspicuous; I have presented the reasoning in this way for accessibility. Although all three aletheic principles are needed, the logical principles involved can be varied considerably (there are intuitionistic versions, relevant versions, etc.). These issues are treated in Chapter Three.

[6] Curry (1942).

Each one of the sentences in this sequence says that all the ones that come after it are not true.

Using some logic and the truth principles, we can derive a contradiction as follows:

1.	Assume For some x, (3.x) is true.	
2.	(3.n) is true	[logic from 1—letting 'n' be a name]
3.	'∀k>n, (3.k) is not true' is true	[(Sub) from 2]
4.	∀k>n, (3.k) is not true	[(T-Out from 3]
5.	(3.n+1) is not true	[logic from 4][7]
6.	∀k>n+1, (3.k) is not true	[logic and arithmetic from 4]
7.	'∀k>n+1, (3.k) is not true' is true	[(T-In from 6]
8.	(3.n+1) is true	[(Sub) from 7]
9.	(3.n+1) is true and (3.n+1) is not true	[logic from 5, 7]
10.	For all x, (3.x) is not true.	[logic from 1-9][8]
11.	(3.0) is not true	[logic from 10]
12.	∀k>0, (3.k) is not true	[logic from 10]
13.	'∀k>0, (3.k) is not true' is true	[(T-In) from 12]
14.	(3.0) is true	[(Sub) from 13]
15.	(3.0) is true and (3.0) is not true	[logic from 11, 14]

Notice that none of the sentences in the sequence is self-referential—they refer only to later sentences in the sequence.[9] There is some disagreement about whether Yablo's paradox *really* involves self-reference or not; I do not take this to be a significant issue.[10]

One common reaction to these paradoxes is to blame (T-In) and (T-Out); one might wonder whether there are similar principles that are weaker. Consider the following rules:

(T-Intro) $\mathsf{p} \vdash \langle \mathsf{p} \rangle$ is true.

(T-Elim) $\langle \mathsf{p} \rangle$ is true $\vdash \mathsf{p}$.

One may read the '$\vdash$' (called a *single turnstile*) as 'entails'. Although they seem to be the same, (T-In) and (T-Out) are statements formulated using conditionals, whereas (T-Intro) and (T-Elim) are

---

[7] Letting 'n+1' name the successor of n.
[8] The inference rule is *reductio*: if one can derive a contradiction from an assumption p, then one may conclude 'not p'.
[9] Yablo (1993c).
[10] See Priest (1997), Sorenson (1998), Beall (2001c), and Schlenker (2007).

inference rules. It turns out that inference rules can be weaker than their associated conditional statements. In classical logic, there is no difference between them, but in some non-classical logics the inference rules are valid, but the conditional statements are not true. We will see examples in the next chapter.

As I mentioned, both (T-Out) and (T-In) are used in the derivation of an absurd claim in the liar paradox, Curry's paradox, and Yablo's paradox. Also, at least one of those rules is used in a subproof in each of those paradoxes.[11] A subproof occurs when, in the course of some reasoning, one presents a hypothesis and argues as if that hypothesis is correct in an effort to show what follows from it; reductio ad absurdum arguments and conditional proofs are examples of reasoning that require subproofs (see the above arguments for examples). Given this fact about the paradoxes, one might wonder whether one of the principles (T-In) or (T-Out) could be weakened so that it cannot be used in subproofs.

There are similar principles that are even weaker than the inference rules (T-Intro) and (T-Elim). (T-Intro) and (T-Elim) are inference rules that can be used at any time in an argument, but there are rules that are restricted. For example, some rules can only be used in categorical reasoning, not in hypothetical reasoning. That is, one can use these rules on results that have already been proven, but one cannot use them on mere assumptions or anything that depends on a mere assumption. We can call these *categorical rules* in contrast with inference rules (which I take to apply in both categorical and hypothetical cases). Those pertaining to truth are the following:

(T-Enter) If $\vdash \mathsf{p}$, then $\vdash \langle \mathsf{p} \rangle$ is true.

(T-Exit) If $\vdash \langle \mathsf{p} \rangle$ is true, then $\vdash \mathsf{p}$.

---

[11] This claim presupposes a certain kind of proof theory (natural deduction); in others (e.g., Hilbert systems), there is no such thing as a sub-proof.

One can read these as 'if p is assertible, then 'p is true' is assertible', and 'if 'p is true' is assertible, then p is assertible'.[12] Note the difference between them and the formulation of (T-Intro) and (T-Elim). (T-Enter) and (T-Exit) may not use the rules inside subproofs or hypothetical reasoning—they can be used only on items that have been derived in the course of the reasoning. For example, if I have already proven '2+2=4' from some set of axioms, then (T-Enter) tells me that "2+2=4' is true' is also provable from them. On the other hand, if I just suppose that 2+2=5, then (T-Intro) allows me to infer "2+2=5' is true', but (T-Enter) does not—it may not be used on mere suppositions or hypotheses.[13]

Throughout the discussion, it is crucial to keep the distinction between conditional statements (e.g., (T-In) and (T-Out)), inference rules (e.g., (T-Intro) and (T-Elim)), and categorical rules (e.g., (T-Enter) and (T-Exit)) firmly in mind. Conditional statements are the strongest of these principles, inference rules are weaker (in some logics), and categorical rules are the weakest.

It turns out that one can still generate Curry's paradox using (T-Out) and (T-Enter) and a variant of the same reasoning works for (T-In) and (T-Exit); so weakening only one of the conditional statements to a categorical rule is not enough to avoid Curry's paradox.[14] We will see in the next chapter that weakening both conditional statements to categorical rules does work, but (like most approaches) it comes at a heavy cost.

Instead of weakening both (T-In) and (T-Out), one might consider allowing exceptions to just one of them. By far the most popular choice is to allow exceptions to (T-In). One reason is that an exception to (T-In) would allow asserting a sentence q without asserting that q is true. In the history of philosophy, there have been lots of views that have this consequence; for example, I mentioned

---

[12] Those with training in logic are used to reading '$\vdash$' as 'provable'. That is consistent with the reading in the text if one takes provability to be the basis for assertibility in mathematical and logical contexts.

[13] Take a moment to look back at the arguments in the liar paradox, Curry's paradox and Yablo's paradox—all of them use hypothetical reasoning.

[14] This result was proved in Montague (1963).

in the last chapter that moral irrealists sometimes say that claims about moral goodness are neither true nor false even though they are assertible. On the other hand, an exception to (T-Out) would allow denying a sentence q even while asserting that q is true. Given that a proponent of this view would probably accept (T-In), that would lead one to assert that both q and its negation are true. Having a sentence and its negation both be true seems much more counterintuitive than having a sentence that one accepts even though it is not true.

However, Montague's paradox requires only (T-Out) in conjunction with several other intuitively plausible principles about truth. Thus, it affects even those who allow exceptions to (T-In). If T is the theory with the following axioms, then T is inconsistent:

(i) If ⟨p⟩ is true, then p.

(ii) ⟨If ⟨p⟩ is true, then p⟩ is true.

(iii) Tautologies are true.

(iv) If a conditional is true, then if its antecedent is true, its consequent is true.

(v) Principles of arithmetic are true.[15]

These are all very intuitive principles about truth. Montague's argument is complex, but it uses a sentence like:

(4) 'if Q then (4) is true' is true

where Q is the conjunction of certain principles of arithmetic. Therefore, accepting (T-Out) along with (ii)-(v) still results in paradox, even if one allows exceptions to (T-In).

What if one allows exceptions to both (T-Out) and (T-In)? McGee's paradox does not require either (T-Out) or (T-In), but it does appeal to some other intuitive principles about truth. Let T be the theory with the following axioms:

---

[15] Montague proves the theorem for Robinson's Arithmetic, Q, which is a very weak theory (i.e., it is finitely axiomatizable). I am assuing classical logic throughout this exposition, but one can generate Montague's result intuitionistically; see Tennant (forthcoming).

(i) If $\vdash \mathsf{p}$, then $\vdash \langle \mathsf{p} \rangle$ is true.[16]

(ii) If $\langle \sim \mathsf{p} \rangle$ is true, then $\langle \mathsf{p} \rangle$ is not true.

(iii) If all instances of a generalization are true, then the generalization is true.

(iv) If a conditional is true, then if its antecedent is true, its consequent is true.

(v) Principles of arithmetic are true.[17]

McGee's paradox is that T is ω-inconsistent. ω-inconsistency is weaker than inconsistency, but it is still pretty bad. It is derivable from an ω-inconsistent theory that each number has a certain property and that there exists some number that does not have that property. A consequence is that if one accepts an ω-inconsistent theory, then one cannot give arithmetic vocabulary its standard meaning. McGee's reasoning is complex, but it uses a sentence like:

(5) For some n, the result of applying the truth predicate n times to (5) is not true.

McGee's paradox shows that there are paradoxes associated with truth even if one rejects both of the truth principles, (T-In) and (T-Out).

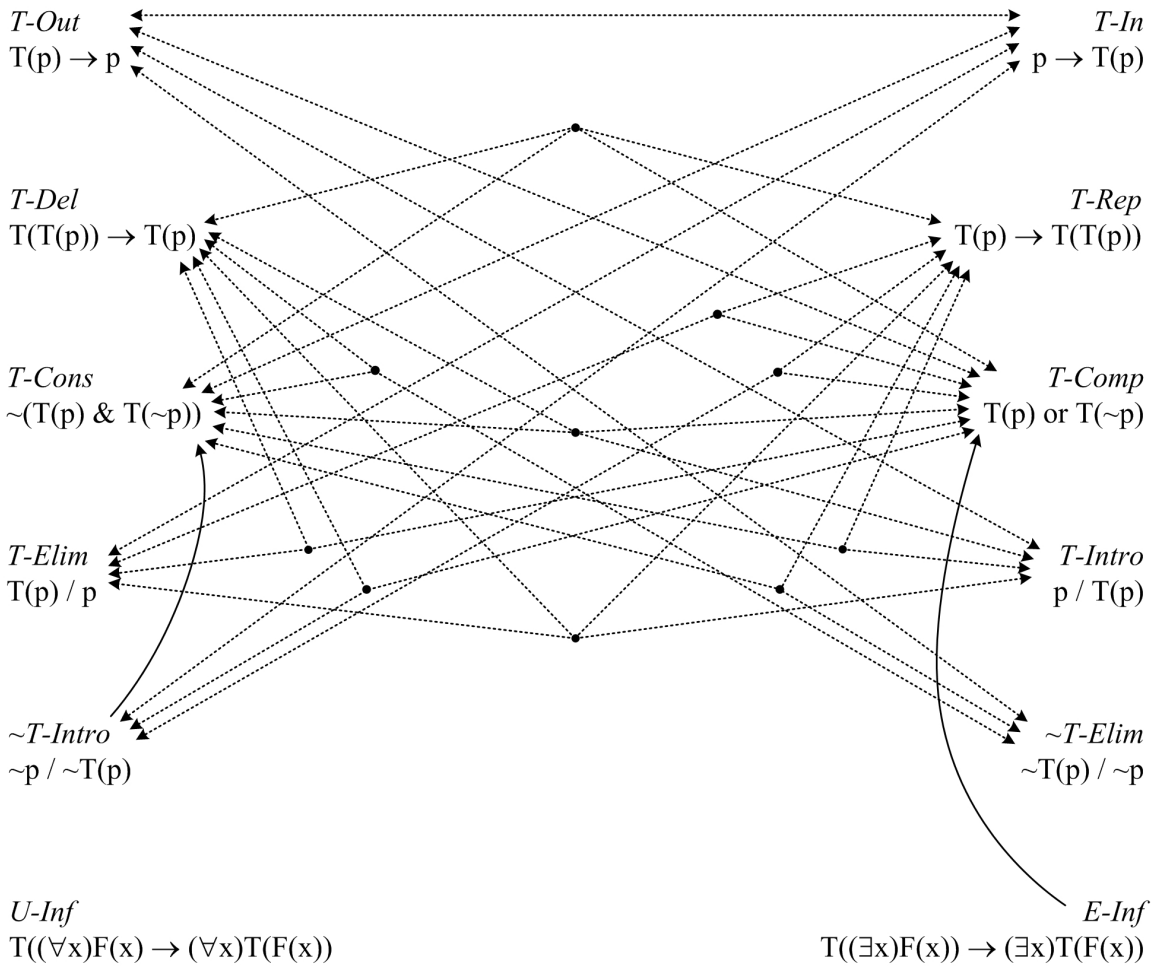
At this point, one might wonder whether there are other paradoxes that involve combinations of (T-In), (T-Out), and other principles involving truth. Harvey Friedman and Michael Sheard took on the unenviable task of determining every consistent and inconsistent combination of twelve principles including (T-In), (T-Out), (T-Intro), and (T-Elim).[18] Information Box 7 displays their results (any combination that is not labeled inconsistent is consistent). There are many more principles involving truth that Friedman and Sheard did not include in their survey; see Appendix 1 of Chapter Three for details. We will consider some of these in Part III.

---

[16] Notice that this is what I called (T-Enter) above.
[17] McGee proves the theorem for Robinson's Arithmetic plus the claim that successor is 1-1 and zero is not a successor, which is a very weak theory (i.e., it is finitely axiomatizable); see McGee (1985).
[18] Friedman and Sheard (1987). See also Friedman and Sheard (1988) and Leigh and Rathjen (2010).

Aletheic Principles (from Friedman and Sheard)

Information Box 7



KEY

Entailment

Inconsistency

ASSUMPTIONS

T(p → q) & T(p) → T(q)
T(p) if p is a tautology
T(p) if p is a truth about constants, free variables, or primitive recursive function symbols

So far, we have seen the paradoxes of the liar, Curry, Yablo, Montague, and McGee. There are also blends of these paradoxes; e.g., a paradox generated by an infinite sequences of conditionals, which is a blend of Curry's paradox and Yablo's paradox.[19]

Two other versions of the liar paradox are worth mentioning. First, consider the following two sentences:

(6a) (6b) is false.

(6b) (6a) is true.

Both these sentences are paradoxical and together are known as a *liar pair*. One can generate liar triples, etc. as well. In addition, liar pairs can consist of sentences from different languages or even different truth bearers.

Second, contingently paradoxical sentences are especially troubling. Consider an overhead projector that is currently showing only the sentence:

(7a) The sentence on the blackboard is false.

If it just so happens that the blackboard in question has written on it only the sentence:

(7b) The sentence on the overhead projector is true.

then both these sentences are paradoxical (they constitute a liar pair). Notice that if the empirical facts had been different, then they would not have been paradoxical. This phenomenon casts doubt on our ability to determine when a truth bearer is paradoxical in everyday conversational contexts; it is the topic of Chapter Seven.[20]

---

[19] See Sorenson (1998), Beall (1999), and Schlenker (2007) for examples.

[20] There are two other significant results that deserve mentioning. First is Vann McGee's insight that any sentence is equivalent to some T-sentence, and the consequence that for any consistent set of sentences, there is a maximally consistent set of T-sentences that entails them; see McGee (1992). Second is Greg Restall's aletheic paradox that uses no logical terms whatsoever—merely principles of truth and logical consequence along with a notion of truth value or proposition; see Restall (2010). McGee's result poses a problem for deflationist responses to the liar paradox that do not resort to non-classical logics (I discuss this problem in Chapter Seven). Restall's result poses a problem for any non-classical approach to the aletheic paradoxes.

There are plenty of other paradoxes that bear a similarity to the liar but pertain to concepts other than truth (e.g., predication, reference, propertyhood, set membership, well-ordering, and vagueness). I do discuss some of these other paradoxes in Chapter Fifteen, but my focus in this work is on the paradoxes specifically associated with truth. From here on I use 'aletheic paradoxes' as a general term for them.

2.2 Problems

When presented with a paradox, people's reactions differ considerably. Some treat it as a harmless trick or puzzle, while others find it compelling and see it as exposing flaws in our most fundamental concepts and beliefs. Most of those who write on paradoxes either say nothing of the problems they pose or remark in passing that they seem to follow from obvious assumptions by valid inference rules. For the purposes of assessing work on the paradoxes, I find it quite helpful to get straight on the multiple problems they pose and the variety of projects that one can undertake in addressing them.

The first problem, which I will call *the derivation problem*, concerns the fact that one can derive contradictions from seemingly impeccable assumptions via seemingly unimpeachable inference rules. It seems obvious that the primary aletheic principles are true, that contradictions are false, and that valid inference rules preserve truth; thus, it seems that there must be some fault in the derivations associated with the paradoxes. However, the problem, if there is one, has been exceedingly difficult to find. It is important to note the difference between a paradoxical item (e.g., sentence (1)), a paradox, and the contradiction that follows from the associated derivation. The paradoxical item (even if it is a sentence) is not derived; rather, one derives two contradictory claims about the status of the paradoxical item. Thus, the contradiction is that sentence (1) is true and sentence (1) is not true. The paradox is the fact that one can derive such a conclusion from certain assumptions about

truth (e.g., the primary aletheic principles). A solution to the derivation problem would be an account of what is wrong with the derivations associated with the aletheic paradoxes.

The aletheic paradoxes pose a major problem for anyone engaged in interpretation, which is the practice of determining what something or someone means. Although we rarely pay attention to them, paradoxical sentences are present in our language, and we find ourselves accepting paradoxical propositions. Those facts cause headaches for anyone trying to investigate the nature of language or thought. The problem is that the principles involved in the derivation of the aletheic paradoxes are part of any plausible theory of a language containing a truth predicate and part of any plausible theory of mental states for someone possessing the concept of truth. Therefore, attempts to characterize a language by specifying the meanings of its sentences results in the interpreter accepting contradictions. Consequently, any plausible theory of language or thought turns out to be inconsistent when paradoxical items are present. The problem is exacerbated by the fact that many theories of language and thought appeal to the concept of truth; indeed, one of the most important and influential theories implies that the meaning of a sentence or the content of a belief determines its truth conditions (i.e., the conditions under which it is true). It is disturbing that many theories of language and thought attempt to characterize a paradoxical subject matter using the same concept that generates the paradoxes. Further, this problem occurs *even for theories that do not appeal to truth at all*—as long as paradoxical truth bearers are in their domain, there is a serious problem. I call this the *interpretation problem*.

Finally, the paradoxes pose a problem for the coherence of our conceptual seheme. Many agree that truth is one of the most important and fundamental concepts we have. Moreover, the primary aletheic principles that lead to paradox are absolutely basic to the concept of truth; they are principles we rely on anytime we use the concept. Further, the logical principles involved in the paradoxes are basic to the way we reason, and we rely on them in not just everyday reasoning, but in

scientific and mathematical reasoning as well. Thus, there does not seem to be an easy way (or even a radically complex way) to solve either of the previous two problems without drastically altering our linguistic and conceptual practice. Perhaps the root of the derivation problem and the interpretation problem lies with inherent flaws in our beliefs, our practice, or perhaps our concept of truth. I call this *the conceptual problem*. A solution to it would be to either adopt the least damaging way to alter our practice so that paradoxical items are rendered benign or explain how we could rationally go on participating in problematic practices, employing incoherent concepts, or adopting inconsistent beliefs.

To sum up, there are at least three problems posed by the aletheic paradoxes:

(i) *the derivation problem*: how can one derive a contradiction from what seem to be logical and conceptual truths?

(ii) *the interpretation problem*: how can a language that contains all the ingredients for the aletheic paradoxes have intuitive semantic and pragmatic features?

(iii) *the conceptual problem*: how can our conceptual scheme be coherent given that it contains all the ingredients for the aletheic paradoxes?

2.3 Projects

There are so many facets to the aletheic paradoxes and such a wide range of writings on them, that philosophers have spent some time reflecting on the kinds of things one might be doing when writing about them. One obvious goal is to explain why they occur, which often takes the form of pointing an accusatory finger at one of the principles used to derive them; we can call this the *diagnostic project*.[21] It offers a solution to the derivation problem.

Another worthwhile project is to explain the semantic and pragmatic features of languages containing truth predicates and the semantic features of thoughts involving the concept of truth; this

[21] This term comes from Chihara (1979).

is often called the *descriptive project*, and it targets the interpretation problem.[22] The descriptive project has several sub-projects. First, one might want a theory that specifies the semantic properties of truth predicates and sentences containing them that accords, as well as possible, with the intuitions of native speakers; call this the *semantic project*. Second, one might want to explain how native speakers arrive at their intuitions about the semantic features of sentences containing truth predicates; call this the *psychological project*.[23] Also, it turns out that approaches to the aletheic paradoxes have a great deal of trouble specifying, in the languages under consideration, the semantic features of sentences of those languages; it is very tempting to say "well, we just can't say anything about that" when dealing with some aspect of the aletheic paradoxes. So, an important part of the descriptive project is showing how we can use the natural language under consideration to characterize the semantic features of the sentences belonging to that very language. Call this the *exhaustive characterization project*.[24] Finally, since natural languages seem to obey the logical principles involved in the paradoxes, and they have paradoxical sentences and truth predicates that seem to obey all the aletheic principles involved in the paradoxes, it is not at all clear how they avoid being trivial. A trivial language is one that has a trivial consequence relation (i.e., everything follows from every set of sentences) and every sentence is both true and false. Nevertheless, few things are as abhorrent as the idea that our natural language is trivial; so why is it not trivial? Answering this question is pursuing the *non-triviality project*.[25] So there are at least four parts to the descriptive project (i.e., solving the interpretive problem).

Since the paradoxes are caused by principles that almost anyone would accept (before realizing that they lead to contradiction), some philosophers advocate changing some aspect of our linguistic

---

[22] This term comes from Gupta (1982) and Yablo (1985).
[23] This term comes from Yablo (1985).
[24] This term comes from Beall (2006).
[25] This term comes from Beall (2006).

and cognitive practice in light of them. Some think we should change our logic, others say we should give up some deeply-held principle about truth, still others say we should replace our concept of truth with some other concept(s). These theorists are pursuing the *prescriptive project*, which specifies the changes we should make in our conceptual scheme.[26] The prescriptive project should ideally be paired with at least a rudimentary account of our linguistic and cognitive practice *as it is now*, before the proposed change, but that does not always happen. A prescriptive project would offer a solution to the conceptual problem posed by the paradoxes.[27]

To sum up, there are at least three main projects one might pursue in offering an approach to the aletheic paradoxes:

(i) *the diagnostic project*: specify where the reasoning goes wrong in the aletheic paradoxes and explain why we have been fooled by the culprit for so long.

(ii) *the descriptive project*: explain the content and use of truth bearers that involve the concept of truth; it has several subprojects:

  (ii-a) *the semantic project*: explain how to provide a semantics for a natural language that has all the ingredients for the aletheic paradoxes.

  (ii-b) *the psychological project*: explain how users of a natural language attribute semantic properties to it.

  (ii-c) *the exhaustive characterization project*: explain how to assign semantic properties to all the elements of a natural language by using that language.

  (ii-d) *the nontriviality project*: explain why natural languages are not trivial.

(iii) *the prescriptive project*: explain what changes we should make to our natural language and conceptual schemes in light of the aletheic paradoxes.

---

[26] Gupta (1982) calls this the *normative project* and Chihara (1979) calls it the *treatment project*.

[27] Another goal is to find the most efficient and elegant ways of blocking aletheic paradoxes in artificial languages studied by mathematicians and logicians; this is the *preventative project* (the term comes from Chihara (1979)). Since my main focus is natural language, I don't discuss this project.

I use these terms (and those for the problems posed by the paradoxes) repeatedly throughout the rest of the book, and I use the generic term 'approach' for any attempt to pursue one of these projects.

2.4 Philosophical Approaches

I have already remarked that there are several factors that make an overview of work on the liar paradox extremely challenging, but let me elaborate. One factor is the very high level of difficulty in understanding the technical details, given that almost every approach utilizes high-powered mathematical techniques that are unfamiliar to most philosophers (and effectively inaccessible given the pressures of modern academic life and the time it would take to master them).[28] For example, transfinite inductive (recursive) constructions resulting in fixed points are a staple of this literature. A second issue is that many theorists offer philosophical interpretations of their mathematical structures, which are designed to connect a natural language truth predicate to the technical theory. However, each mathematical structure can be (and often is) given different philosophical interpretations by different theorists. For example one might interpret a revision theory (a technical mathematical construction) by using it to model a circularly defined natural language truth predicate, or by using it to model a partially defined truth predicate, or by using it to model a context-dependent truth predicate. The technical mathematical constructions and their philosophical interpretations can be mixed and matched in myriad ways. Even when keeping a philosophical interpretation fixed, there is the fact that one mathematical structure can model a truth predicate in many ways. For example, people use phrases like 'Kripke's minimal fixed point' to mean quite different things. Finally, the use of artificial languages by most theorists is a complicating factor

[28] Field (2008a) claims that his book is "fairly non-technical" in the preface, but he starts the first chapter with Gödel's Diagonalization lemma, which is a very complicated result from mathematical logic. The funny thing is that, compared to most of the literature on the liar paradox, Field's book really is "fairly non-technical".

since the relation between the artificial language and natural language is rarely mentioned, much less given the emphasis it deserves. It turns out that this relation can dramatically affect the features of an approach.

This chapter surveys philosophical views of how truth predicates of natural language work in light of the aletheic paradoxes, while the next summarizes the logico-mathematical technical apparatuses that are used to model natural language truth predicates and the principles governing them. I begin with the philosophical approaches to the liar paradox since the mathematical models are easier for the uninitiated to appreciate after a bit of philosophical motivation. The philosophical approaches are divided up by the features they attribute to truth predicates of natural languages.

2.4.1 Grammaticality

One approach to the aletheic paradoxes holds that the paradoxical sentences are not syntactically well-formed. The justification for this view is that if one attempts to replace 'this sentence' in 'this sentence is false' with the quote-name of the whole sentence, then one is left with a sentence in which 'this sentence' occurs again. Further attempts to replace it lead to a regress. Only a handful of philosophers have advocated this view, and there are none in the last fifty years with which I am familiar.[29]

This approach addresses all three problems and constitutes a diagnostic project and a descriptive project. It is not associated with any logical approaches since it implies that there are no paradoxical sentences.

An obvious problem is that the paradoxical sentences are well-formed according to standard theories of syntax and we have no independent reason to think that these accounts of syntax should

[29] Jorgensen (1953, 1955) and Kattsoff (1955).

be revised. For example, there does not seem to be anything wrong with 'this sentence has five words'.[30]

2.4.2 Meaningfulness

Some theorists have said that paradoxical sentences are syntactically well-formed, but meaningless. One justification for this view is that if one wants to understand a truth or falsity attribution, usually one looks to the target of that truth attribution to see what truth or falsity is being attributed to. However, with paradoxical sentences, they attribute truth or falsity to themselves, making it difficult to interpret them.[31] Another justification for this approach comes from the prosentential version of deflationism. If sentences containing 'true' inherit their content anaphorically from antecedently contentful sentences, then paradoxical sentences never inherit anything.[32]

A radical approach in this category is that the aletheic paradoxes show that *all* expressions of *all* languages capable of formulating them are meaningless. The reasoning is that the aletheic paradoxes render inconsistent any semantic theory for these languages; thus, there is no consistent theory to underwrite the meanings of these expressions.[33]

The meaningfulness approach offers solutions to all three problems and constitutes a diagnostic project and a descriptive project. It is sometimes associated with logical approaches (e.g., the inner inductive strong Kleene theory) that it uses to define which sentences containing 'true' are meaningful.[34]

---

[30] For more information on theories of syntax, see Givón (2001) and van Valin (2001).
[31] See Mackie and Smart (1953, 1954), Ushenko (1957), Skinner (1959), Ziff (1960), Ross (1969), Goldstein and Goddard (1980), Keene (1983), Sorensen (2001, 2005), and Englebretsen (2006: 161-166).
[32] See Grover (1977), Brandom (1994: 321-322), Beall (2001), and Armour-Garb (2001).
[33] As far as I know, Douglass Patterson is the only one who accepts this view; see Patterson (2006, 2007a, 2007b, 2009). I discuss it in Chapter Twelve.
[34] Brandom (1994).

This approach (the moderate version anyway) was once very popular, but has since fallen out of favor. An obvious problem is that standard theories of meaningfulness imply that paradoxical sentences are meaningful, and we have no independent reason to think that they are in error in this regard.[35] Also, paradoxical sentences seem to have inferential roles, and they can be accepted, rejected, challenged, and defended, just like any other meaningful sentence.[36] Finally, I argue in Chapter Seven that the existence of empirically paradoxical sentences casts considerable doubt on meaningfulness approach.

2.4.3 Assertibility

Another approach that is similar to the first two focuses on the pragmatic status of utterances of paradoxical sentences. Although paradoxical sentences are well-formed and meaningful, they cannot be asserted. This view has long been defended by Laurence Goldstein and recently by John Kearns.[37]

It does not seem like this approach addresses the derivation problem; instead, it focuses on the interpretation problem. It constitutes a descriptive project of sorts, but it does not say anything about the meanings or truth values of paradoxical sentences.

Again, it does not seem that standard theories of assertion imply that paradoxical sentences are not assertible, and empirical paradoxicality poses a threat (I cover this point in Chapter Seven).[38] Moreover, the assertibility approach suffers from a kind of incompleteness. Even if paradoxical

---

[35] See de Swart (1998), Chierchia and McConnell-Ginet (2000), and the papers in Davis and Gillon (2004) for more information on semantics and meaningfulness.
[36] Toms (1956) and Pollack (1977) make this point.
[37] Prior (1958, 1961), Richards (1967), Martinich (1983), Goldstein (1985, 1986a, 1986b, 1992, 1999, 2008), and Kearns (2007).
[38] See Pagin (2007) and Weiner (2007) for overviews of theories of assertion.

sentences cannot be asserted, they are still meaningful and so attempting to specify their semantic values will land one in the aletheic paradoxes.[39]

2.4.4 Intensionality

Brian Skyrms offered a novel approach to the semantic paradoxes in a string of papers from the early 1970s to the early 1980s.[40] His strategy was to deny the substitutivity of identicals for truth contexts. That is, Skyrms proposed systems in which the substitution principle (Sub) fails; for example, there are truth bearers p and q such that p=q, but the theory implies that p is true and q is not true. This is a radical maneuver that amounts to denying that truth values are attributed to truth bearers at all. Instead, on Skyrms' account, truth bearers under descriptions would have truth values. Under different descriptions, a truth bearer might have different truth values. This maneuver allows one to specify, in the object language, the semantic statuses of paradoxical sentences without inconsistent results. The trick is that one can attribute paradoxicality to a paradoxical sentence by using only some of its names, whereas some of its names will not work.

This approach offers solutions to all three problems, and it constitutes a diagnostic project and a descriptive project.

Losing the substitutivity of identity is an extremely hard pill to swallow, however; it gives up the idea that 'true' and its synonyms are genuine predicates. Moreover, to extend this approach to languages with quantification, one has to restrict universal instantiation as well. Despite the favorable results, these costs have proven too much for most everyone.

2.4.5 Epistemicism

---

[39] Cohen (1961) makes this point.
[40] Skyrms (1970a, 1970b, 1984).

Epistemicism is a view that is most familiar from disputes about the sorites paradox, which affects vague predicates. An expression is v*ague* if and only if it has borderline cases (i.e., it neither definitely applies nor definitely fails to apply to some entities); e.g., 'bald' and 'heap' seem to be vague. The sorites paradox for, say 'heap', results from three claims: (i) there are heaps of grains of sand, (ii) there are groups of grains of sand that do not constitute heaps, (iii) adding or subtracting a single grain of sand to/from a group of grains of sand does not affect whether that group is a heap. All three seem utterly plausible, but repeated application of the process described by (i) will eventually turn any heap of sand into something that is not a heap (or vice versa). So (i)-(iii) are inconsistent (given very weak assumptions about the logic involved). The sorites paradox is probably about as old as the liar paradox (and if ancient sources are correct, they were even discovered by the same philosopher, Eubulides of Miletus).[41]

As one might guess, there are *many* approaches to the sorites paradox. Perhaps the most scorned in contemporary philosophy is epistemicism—the view that, contrary to common sense, vague predicates have sharp boundaries after all, it is just that we can never know where those boundaries are. For example, there is a number such that if a bunch of grains of sand has that many members, then it constitutes a heap, whereas if you take just one grain away, then it is no longer a heap; however, we do not know and cannot know what that number is.[42]

With respect to the aletheic paradoxes, epistemicism is the view that one or more of the truth principles fail in each paradox, but we do not know which ones. The epistemicist can keep classical logic, the claim that every truth bearer is either true or false, and the claim that no truth bearer is both true and false without having to decide which truth principles fail on which occasions.

---

[41] See the papers in Keefe and Smith (1999), Graff and Williamson (2002), and Dietz and Moruzzi (2010) for discussion of vagueness and the sorites paradox.
[42] See Williamson (1994) and Sorensen (2001) for defenses of epistemicism.

Epistemicism as an approach to the aletheic paradoxes is largely associated with Paul Horwich's work on minimalism, although Horwich himself has yet to fully endorse this approach.[43] The motivation for epistemicism comes largely from the combination of deflationism and classical logic. Deflationism emphasizes the importance and independence of (T-In) and (T-Out) in understanding our concept of truth. Indeed, many deflationist theories of truth consist entirely of the T-sentences for a given language. However, if the deflationist is to keep classical logic and avoid having an inconsistent theory because of the aletheic paradoxes, she needs to prevent paradoxical T-sentences from being part of her theory. It turns out that there are some daunting technical problems with this strategy, and there are many cases where it is not clear which of several T-sentences should be avoided.[44] Instead of trying to solve these problems, some deflationists opt for epistemicism: we know that the theory does not contain all the T-sentences, and we know that it does contain lots of the non-paradoxical T-sentences, but we do not (and cannot) know exactly which T-sentences it contains for difficult cases. A consequence is that the paradoxical sentences are either true or false, but we do not (and cannot) know which.

This approach is hard to classify, but I suppose it offers a solution to all three problems together with a justification for refusing to engage in either the diagnostic project or the descriptive project.

The main worry voiced against epistemicism is that it just seems crazy, although it is difficult to turn this attitude into a proper objection. One such attempt might be: our linguistic expressions have their semantic features by virtue of how we use them, and we do not use truth predicates in a way that would determine that paradoxical sentences are simply true or in a way that would determine that they are simply false. Only an unacceptable view on how language works would stipulate that semantic features of an expression go beyond those determined by its use (how would

---

[43] See Armour-Garb and Beall (2005) and Restall (2005) for discussion.
[44] See McGee (1992) and Weir (1996) for discussion.

they get those semantic features? Magic?). Whatever the reason, epistemicism is not taken seriously by most of those who work on the aletheic paradoxes.

2.4.6 Ambiguity

A linguistic expression is *ambiguous* if and only if it has two or more determinate, independent meanings. The standard example in English is 'bank', which can mean *effluvial embankment* or *financial institution* (of course, it has other meanings as well). It is important to distinguish ambiguity from vagueness and context-dependence. I discussed vagueness in the previous subsection; an expression is *context-dependent* if and only if its content depends on the context in which it is used; e.g., 'here' and 'tall' are context-dependent. It is common to distinguish the meaning (sometimes called *character*) of a context-dependent expression, which is constant, from its *content*, which varies. According to my usage, a linguistic expression can be ambiguous without being vague or context-dependent.[45]

The idea that truth predicates are ambiguous has been around for a long time and is fairly popular despite the fact that it is rarely defended in print. It seems to me that this approach is frequently the default way of linking natural languages with mathematical models that employ multiple truth predicates or logical operators.

Perhaps the most famous use of ambiguity as a philosophical approach to the aletheic paradoxes is described in Saul Kripke's paper on truth. Kripke does not endorse the view, but calls it the *Orthodox Approach*, and formulates it as a foil to his own. The Orthodox Approach uses a hierarchy of very specific truth predicates that are constructed using a technique from Alfred Tarski. The "bottom" of the hierarchy is a truth predicate that applies only to sentences that do not contain truth predicates at all. The next member applies only to sentences that do not contain truth

[45] See Zwicky and Saddock (1975), Cruse (1986), Atlas (1989), Gillon (2004), Wasow, Perfors, and Beaver (2005), and Kenedy (2010) on ambiguity.

predicates and sentences that contain the bottom truth predicate. Generalizing, each predicate in the hierarchy applies only to sentences that either contain no truth predicates or contain truth predicates lower in the hierarchy.[46] Of course, natural languages have no such hierarchy of truth predicates; instead, the Orthodox Approach says that truth predicates of natural languages are ambiguous and can have the meaning of any of the predicates in the hierarchy.[47] The Orthodox Approach purports to solve all three problems and constitutes a diagnostic and descriptive project. It blocks the paradoxes because all the paradoxical sentences are straightforwardly false—they attribute truth (of some level in the hierarchy) to themselves, which the hierarchy construction forbids.

Other theorists use different hierarchies in approaches to the liar; some generate hierarchies of negations,[48] or conditionals,[49] or meaningfulness predicates,[50] or determinacy operators.[51] These can be linked to natural languages by saying that 'not' or 'if…then…' or 'meaningful' or 'determinate' are ambiguous. The result is similar to the Orthodox Approach, but it posits the ambiguity in a different place.

Kripke's objections to the Orthodox Approach have been extremely influential and constitute a turning point of sorts in attitudes towards the Tarskian hierarchy. Kripke argued that the Orthodox Approach cannot accommodate certain uses of truth predicates, it cannot deal with certain empirically paradoxical liar pairs, and that it suffers from several technical problems as well. I discuss these objections in detail in Chapter Seven.

---

[46] I discuss this hierarchy in the next chapter.
[47] See Kripke (1975: 695) for discussion of this interpretation; Kripke attributes this interpretation to Parsons (1974), but see also Williamson (2000b). Early versions of ambiguity views include Toms (1956), Wormell (1958), and Herzberger (1966), but some of these seem to confuse ambiguity with context-dependence; see Rozeboom (1957), Mackie (1973), and Gaifman (1992) for discussion.
[48] Cook (2008), Schleneker (forthcoming).
[49] Myhill (1984)
[50] McDonald (2000)
[51] Field (2008a).

2.4.7 Context Dependence

The idea that truth predicates are context dependent and our failure to keep track of this feature is the cause of the aletheic paradoxes is very popular. One major challenge for defenders of this view is to specify the various contents that sentences containing 'true' can have and explain how the contents of these sentences depend on the contexts in which they are used. Similar theories are currently popular in other areas of philosophy; epistemological contextualism (i.e., the view that 'knows' is context dependent) is probably the most familiar.[52]

There are two main families of context dependence approaches: those that locate the source of context dependence in the truth predicate itself, and those that find it in some other constituent of sentences containing 'true'. Theories of the first type imply that 'true' is a context dependent expression similar to demonstratives (e.g., 'that'), indexicals (e.g., 'here'), gradable adjectives (e.g., 'tall'), and genitives (Ned's house). On a theory of the first type, 'true' has a fixed meaning (often called its *character*), but its content on an occurrence of its use depends on the context in which it is used. That is, the contribution an occurrence of 'true' makes to the truth conditions of a sentence token in which it occurs depends on the context in which that sentence token is uttered. Once any ambiguities have been resolved and the context-dependent contents of any non-semantic terms have been determined, there is the additional variability in the semantic content of the sentence token due to the presence of the truth predicate. In particular, the extension of the truth predicate changes from context to context. What would count as a paradoxical sentence token in a given context is

[52] See Cohen (1986, 1999), DeRose (1992, 1995, 2002), Lewis (1996), and the papers in Preyer and Peter (2005) for discussion. See Hawthorne (2004) and Stanley (2005) for criticism.

eliminated from the extension of 'true' in that sentence token. Paradoxical sentence tokens are treated as either false or gappy in the contexts in question.[53]

The second type of contextual theory of truth implies that, while 'true' is not a directly context dependent expression, sentence tokens in which 'true' occurs do display context dependence. One way to think of this context dependence is to treat truth as a predicate of propositions and claim that an attribution of truth to a sentence is actually an attribution of truth to the proposition expressed by that sentence. On this view, attributions of truth to sentences have a hidden quantifier. For example, let 'p' be the name of a sentence token. The truth attribution 'p is true' becomes 'the proposition expressed by p is true'; on a Russellian interpretation of definite descriptions, the latter becomes 'there is a unique proposition such that it is expressed by p and it is true'.[54] Quantifiers are known to display context dependence in their domains.[55] That is, the set of objects over which a quantifier ranges is determined, in part, by the context in which it is used. It is this quantificational context dependence that is claimed to be present in sentences that contain truth predicates. The appeal to propositions can be replaced with talk of schemes for interpreting sentences, which contain domains and are determined by the context.[56]

Theories of this type solve the liar paradox by claiming that one sentence token does not express a proposition at all, but another token of the same type does express a proposition. For example, if p is a token of 'p is not true', then p actually says that it is not the case that there exists a unique proposition such that it is expressed by p and it is true. Assume that p is in context C. In context C, the scope of the hidden quantifier in p is a certain set of propositions. In C, there is no proposition in this set for p to express. However, if one asserts 'p is not true' then this action changes the

---

[53] For examples, see Thomason (1976), Burge (1979a, 1982a, 1982b), Gaifman (1983, 1992, 2000), Hodges (1986), Barwise and Etchemendy (1987), Koons (1992, 2000b), Simmons (1993, 2000, 2003), and Cantini (1995).
[54] See Russell (1905); for commentary see Neale (1990, 2005); non-Russellian views are compatible with this approach, but I use the Russellian view throughout as an example.
[55] See Stanley and Szabo (2000), Bach (2000), and Villalta (2003) for discussion.
[56] Parsons (1974, 1983) and Glanzberg (2001, 2004).

context. In the new context C′, the quantifier ranges over a more inclusive set of propositions, one of which is the proposition that it is not the case that there exists a proposition such that it is expressed by p and it is true.

Most contextual theories of truth (of either type) incorporate some sort of hierarchy in order to accommodate changes in context. For example, Tyler Burge's theory employs a Tarskian hierarchy of truth predicates as a repository of contents for 'true'; in a given context, 'true' has the content of one of the Tarskian truth predicates. That is, whereas the Orthodox Approach described in the previous subsection combines the Tarskian hierarchy with ambiguity in the natural language truth predicate, Burge combines the Tarskian hierarchy with context dependence in the natural language truth predicate. Others who offer hierarchical contextual theories include Charles Parsons, who employs a hierarchy of interpretation schemes and Michael Glanzberg, who employs a hierarchy of contexts. A notable exception is Keith Simmons, who offers a theory of the first kind on which each contextual use of 'true' is at the same level as every other one. However, each contextual use of 'true' also has some gaps—sentences that are neither true nor false in that context; Simmons calls these *singularities*. He provides a set of rules for determining which singularities are appropriate for which contexts and he is explicit about the fact that each contextual use of 'true' should be thought of as applying to all truth apt truth bearers except the singularities.

One of the driving forces behind the context dependence approach is that it handles cases like:

(8) (8) is not true.

(9) (8) is not true.

'(8)' and '(9)' are to be thought of as names of the particular physical sentence tokens on this page of this particular physical document. Further, (8) and (9) are tokens are of the same type. However, on one interpretation, while (8) is a liar sentence and thus, paradoxical, (9) is a comment on (8) to the effect that, because it is paradoxical, it is not true. Thus, on this interpretation, although they are

two tokens of the same type, one is true and the other is paradoxical. This phenomenon is known as *the two-line puzzle*.[57]

A related phenomenon concerns the reasoning that accompanies (8). An intuitive argument supports the claim that sentence (8) is true if and only if it is not true. Thus, sentence (8) seems to be paradoxical. It is natural to think that paradoxical sentences are not true because assuming that are true leads to contradiction. Thus, one might conclude that because sentence (8) is paradoxical, it is not true (indeed, the argument for this point is above). At this point, one might reread sentence (8) and realize that it says of itself that it is not true. We have just argued that it is not true; thus, sentence (8) accurately describes its own status—it says that it is not true and, indeed, it is not true. Hence, sentence (8) must be true (for that is what we say about sentences that say that such and such is the case when such and such actually is the case). I refer to this as *the strong liar reasoning*.[58]

All the context dependence views offer solutions to the three problems posed by the paradoxes and constitute diagnostic and descriptive projects. In addition, they make sense of the two-line puzzle and they validate the strong liar reasoning.

One obvious problem with context dependence approaches is that speakers of natural languages do not, even upon reflection, think that truth predicates are context dependent. This kind of problem is known as "the semantic blindness objection", and it has been a thorn in the side of epistemological contextualism (i.e., the view that 'knows' is context dependent).[59] I take up this issue in Chapter Seven.

Another persistent problem is revenge paradoxes. I have not talked much about these, but they are quickly becoming *the* issue to discuss when it comes to aletheic paradoxes. Typically a revenge

---

[57] See Hazen (1987), Gaifman (1992), Juhl (1997), Clark (1999), Goldstein (1999, 2001), Weir (2000, 2002), and Gupta (2001).
[58] See Kearns (1970), Parsons (1974), Burge (1979a), Hazen (1987), Gaifman (1992, 2000), Gupta (2001), and Glanzberg (2001, 2003, 2004, 2005) for discussion.
[59] See Schiffer (1996), Hofweber (1999), Hawthorne (2004), Stanley (2005), DeRose (2006), and Capellan (2007).

paradox occurs for an approach when one can use some idea or concept from that approach to formulate a new kind of paradoxical sentence that the approach cannot handle. For context-dependence approaches, a revenge paradox would stem from a sentence like:

(10) (10) is not true in any context,

which quantifies over contexts. I have much more to say about revenge paradoxes in Chapters Eight and Nine.[60]

2.4.8 Indeterminacy

Perhaps the single most popular and intuitive philosophical approach to the aletheic paradoxes is to say that paradoxical sentences are neither true nor false, or that they do not have a standard truth value, or they are indeterminate in some way. This view is frequently voiced as the idea that paradoxical sentences are *truth value gaps*, or just *gaps* for short.

There are many ways of implementing this idea, and one must use extreme care to avoid confusion. The main source of confusion is over the terms 'false' and 'gap'. There are several ways of defining 'false', and they are not equivalent. The two definitions relevant for this discussion are negation of truth (i.e., p is false if and only if '~p' is true) and truth of negation (p is false if and only if p is not true). When one says that paradoxical sentences are neither true nor false, one could mean that neither they nor their negations are true (this is the "truth of negation" reading of falsity) or that it is not the case that they are true and it is not the case that they are not true (this is the "negation of truth" reading of falsity). So, once one considers these two readings of falsity, one arrives at two readings of 'truth value gap'. The former reading of truth value gap (a sentence such

[60] For criticism of Burge, see Gupta (1982) and Simmons (1993); for criticism of Gaifman, see Simmons (1993) and Yi (1999); for criticism of Barwise and Etchemendy, see Gupta (1989), Grim (1991), McGee (1991), Gaifman (1992), and Priest (1993); for criticism of Simmons, see Antonelli (1996), Hardy (1997), and Beall (2003). For discussion of revenge liars for contextual theories see Hazen (1987), Hinckfuss (1991), Juhl (1997), Clark (1999), Weir (2000, 2001, 2002), and Newhard (2009).

that it and its negation are true) is compatible with classical logic, whereas the latter (a sentence such that it is not true and is not not true) is not compatible with classical logic. Thus, the reading one gives to 'truth value gap' has a massive impact on the approach to the paradoxes one endorses. Also, given the two readings of 'false' and 'gap', there is tremendous confusion surrounding these approaches to the liar. In what follows, I use the phrase 'classical gap' to mean a sentence such that it is not true and its negation is not true, and I use the phrase 'non-classical gap' to mean a sentence p such that 'p is true or p is not true' is not true.[61]

Although the idea that paradoxical sentences are not true and not false has been around for a long time, it was not until the late 1960s that it began to show up in well-developed approaches to the liar.[62] One of the first well-developed approaches is due to Bas van Fraassen, who adapted his view of presupposition failure to apply to languages that have paradoxical sentences. His idea is that sentences presuppose bivalence (i.e., every truth bearer is either true or false) and paradoxical sentences presuppose contradictions and, thus, cannot have truth values. His technique is called *supervaluations*, and it has proven to be influential in a wide range of topics. A *valuation* is a way of assigning truth values to all the sentences of some language. However, in some cases, there is more than one way to assign truth values to certain sentences given how the world is and how we use the language in question. Van Fraassen suggests that we look at all the acceptable ways of assigning truth values to sentences and construct a *supervaluation*, which assigns truth to a sentence if and only if every valuation assigns it truth, and which assigns falsity to a sentence if and only if every valuation assigns it falsity. If there are sentences that are assigned truth by some valuations and falsity by others, then the supervaluation does not assign them anything—they are left indeterminate. The assignments of a supervaluation are sometimes called *supertruth* and *superfalsity* to

---

[61] When Hartry Field says that his view does not posit gaps, he means classical gaps; see Field (2003b: 270).
[62] See Ryle (1951), Fitch (1964), and Martin (1967, 1968) for examples.

distinguish them from the assignments of the individual valuations.[63] I discuss the logical aspects of this approach in the next chapter.

Another indeterminacy approach is found in Kripke's 1975 paper, which is without question, the single most important paper on the aletheic paradoxes since Tarski's work; the approach it presents has been immensely popular. Kripke describes a procedure by which a truth predicate is introduced into a first-order language.[64] The novel feature of Kripke's approach is that the truth predicate is only partially defined. That is, its extension and anti-extension are not jointly exhaustive, so some sentences are in the extension of the truth predicate, some are in the anti-extension of the truth predicate, and some are in neither. Paradoxical sentences end up in neither, and for this reason are called 'gaps'; it depends on how one interprets Kripke's construction whether they are classical gaps or non-classical gaps.

At the center of Kripke's discussion is the notion of *grounding*, which has proven to be an important concept in discussions of the aletheic paradoxes. Consider the following passage:

> In general, if a sentence such as ['Most (i.e., a majority) of Nixon's assertions about Watergate are true'][65] asserts that (all, some, most, etc.) of the sentences of a certain class *C* are true, its truth value can be ascertained if the truth values of the sentences in the class *C* are ascertained. If some of these sentences themselves involve the notion of truth, their truth value in turn must be ascertained by looking at other sentences, and so on. If ultimately this process terminates in sentences not mentioning the concept of truth, so that the truth value of the original statement can be ascertained, we call the original sentence grounded; otherwise ungrounded.[66]

Some examples should help clarify the notion of groundedness. Every sentence that does not contain a truth predicate (or other semantic expression) is grounded. Any sentence that attributes truth to only grounded sentences is grounded. For example, 'All the axioms of Robinson Arithmetic

---

[63] See van Fraassen (1968, 1970a, 1970b). For discussion see Skyrms (1968), Herzberger (1970), and Kearns (1970), Kripke (1975), and McGee (1991).
[64] Kripke (1975).
[65] I have changed 'false' to 'true' in this example to correct a mistake in Kripke's text.
[66] Kripke (1975: 693-694).

are true' is grounded since none of the axioms of Robinson Arithmetic contain truth predicates. Further, "snow is white' is true or 'snow is white' is not true' is grounded because its truth value depends on the truth value of 'snow is white', which does not contain a truth predicate.[67]

The procedure that Kripke outlines begins by assigning sentences that do not contain semantic predicates to either the extension of 'true' or to the anti-extension of 'true' in the usual way; since paradoxical sentences all contain 'true' or 'false', this step is uncontroversial. Then sentences that assign truth or falsity to those already in either the extension or anti-extension of 'true' get put in either the extension or anti-extension. Then the same step repeats; more and more sentences are put in the extension of 'true', and more and more sentences are put into its anti-extension. There are several complicated issues having to do with the logical scheme one uses to interpret the logical vocabulary in this process, but I discuss them in connection with logical approaches in the next chapter. Kripke proves that, eventually, the process comes to an end and no more sentences are put in either the extension or anti-extension; the end result is called a *fixed point*. I use 'eventually' loosely because Kripke defines the construction so that it can have an infinite number of steps. One needs to understand a bit of set theory to understand how the construction works, which I put off until the next chapter as well. Kripke shows how to define this procedure for several different logical schemes including supervaluations. He also defines a whole system of fixed points that differ depending on the logical scheme one uses and the sentences one starts with. This paper initiated a tidal wave of technical work on the aletheic paradoxes that has yet to subside.[68]

One problem with Kripke's work is that, if he is interpreted as offering a non-classical logic, then languages he works with are very difficult to reason with since they do not have well-behaved

---

[67] Skinner (1959) is the first place I have found the idea of grounding, and Herzberger (1970) is the first to use the term as far as I can tell.

[68] For discussion see Kindt (1978), Davis (1979), Hazen (1981), Feferman (1982, 1991), Yablo (1982, 1985), Woodruff (1984), Visser (1984), Fitting (1986), Burgess (1986), Kremer (1988), McGee (1991: chs. 4 and 5), Gupta and Belnap (1993: chs. 2 and 3), Halbach (1997), Maudlin (2004), and Field (2008a, ch. 3).

conditionals. On the other hand, one can use his constructions to formulate a theory that is compatible with classical logic, but it will imply that one of the truth principles (specifically (T-In)) has exceptions. In addition, it is impossible to say, in one of these languages, that paradoxical sentences are not true and not false—indeed, one cannot offer any characterization of their semantic status in the language. To do that, one must use an expressively stronger language. Also, the languages Kripke considers do not allow certain kinds of logical terms in them or else his construction will never reach a fixed point. Finally, approaches to the aletheic paradoxes based on Kripke's constructions are subject to revenge paradoxes. In this case, if one adds the predicate 'is a truth value gap' to the language in question, then the theories based on Kripke's constructions are inconsistent. That is, one will be able to formulate a sentence:

(11) Either (11) is false or (11) is a truth value gap.

and one can prove that these theories imply that (11) is true and either false or a gap.[69]

Three other philosophers deserve to be mentioned in this subsection for developing Kripke's indeterminacy approach to the liar. The first is Vann McGee, whose approach is explicitly a prescriptive project; that is, he wants to create a new concept of truth that can be used without falling into paradox. His replacement concept of truth is vague, and McGee endorses a familiar theory of vagueness that distinguishes between being X (for some vague term $\langle X \rangle$) and being definitely X. For truth, that means a distinction between truth and definite truth. He uses one of Kripke's constructions, but he interprets it in a novel way (it is a symmetric classical theory—to be discussed in the next chapter) that does not allow (T-In) or (T-Out), but it does allow (T-Entry) and (T-Exit). McGee argues that his theory of truth and theory of definite truth can both be expressed in the language to which they apply. That makes his one of the only approaches that does not

[69] I discuss this point at length in Chapter Eight.

require an expressively richer language to assign semantic statuses to all the sentence of the language in question.[70]

Tim Maudlin offers a theory based on Kripke's work as well, but Maudlin interprets one of the constructions so that it is compatible with classical logic; his view is that paradoxical sentences are classical gaps (he has to allow exceptions to (T-In) for this to work). He uses an interesting analogy between boundary value problems and the interpretive problem to justify his choice of theory, and he shows how to introduce a groundedness predicate into the language without getting revenge paradoxes, which is an important advance over Kripke's theory. Maudlin also shows that his theory can classify all the sentences of the languages he considers without recourse to an expressively richer language, which, again, is progress.[71]

Criticisms of Maudlin's theory include often focus on the fact that his theory implies that some sentences of his theory are ungrounded (and thus untrue). Maudlin addresses the latter worry by arguing that some ungrounded sentences are assertible even though they are not, strictly speaking, true.[72]

Hartry Field's theory belongs in the category of indeterminacy approaches as well. He too is inspired by Kripke's work, but he takes the idea that the primary aletheic principles (T-In) and (T-Out) should be true very seriously, so he offers a theory on which paradoxical sentences are non-classical gaps. That is, his theory is not compatible with classical logic. Field's big move is to introduce a well-behaved conditional into the languages Kripke considers. He then uses that conditional to define a determinacy operator, which can be used together with the truth predicate to classify all the sentences of the resulting language. His theory does not need an expressively richer language at all (just like Maudlin and McGee), it accepts both (T-In) and (T-Out) (unlike Maudlin

---

[70] McGee (1991: chs. 7-9).
[71] Maudlin (2004, 2006a, 2006b, 2007).
[72] Maudlin (2004: ch. 7 and 8).

and McGee), and his theory has no untoward consequences about its own truth (unlike Maudlin). These benefits are purchased at the price of giving up classical logic, which makes it difficult to interpret Field's theory. Perhaps he is engaged in the descriptive project and thinks our natural languages have non-classical logics; or maybe he is engaged in the prescriptive project and thinks we should change our logic to permit us to accept his theory.[73] Either way, McGee, Maudlin, and Field offer three distinct ways of extending Kripke's work on the aletheic paradoxes.

It seems to me that Neil Tennant's view on the aletheic paradoxes could also be classified as an indeterminacy approach, but it is very different from the others in this category. Tennant shows that the paradoxical reasoning in each aletheic paradox is not normalizable, which means that it cannot be put into normal form.[74] Tennant accepts the primary truth rules and all the logical principles involved in the paradoxical reasoning, but imposes a global constraint (normalizability) on valid arguments; it follows that the reasoning in the paradoxes does not meet one of the necessary conditions for being evaluated as valid. Thus, according to Tennant, there is something wrong with the reasoning in each paradox, but it cannot be traced to any particular principle used.[75] His view requires a non-classical logic as well, which I discuss in the next chapter.

2.4.9 Circularity

Anil Gupta and Hans Herzberger independently arrived at the idea that one could use (T-In) and (T-Out) as rules for generating a sequence of truth-value assignments for a language.[76] This idea eventually matured into the revision approach, which I discuss in the next section on logical approaches. Gupta, together with his later collaborator, Nuel Belnap, offered a philosophical

---

[73] See Shapiro (2010) and Field (2010b) for discussion.
[74] See Troelstra and Schwichtenberg (2000: ch. 6) for more information. Roughly, an argument is in normal form if it does not have any superfluous steps.
[75] See Tennant (1982, 1995, MS1, MS2).
[76] Gupta (1982) and Herzberger (1982).

interpretation to complement that logical approach. Their philosophical interpretation is that truth is a circularly defined concept, which I describe here.[77]

Usually a definition of some term is thought of as having two parts—the *definiens*, which is the thing being defined, and the *definiendum*, which is what it is being defined as. For example, the International Astronomical Union's definition of 'planet' is:

> Something is a planet if and only if it is a celestial body that is (a) in orbit around the Sun, (b) has sufficient mass for its self-gravity to overcome rigid body forces so that it assumes a hydrostatic equilibrium (nearly round) shape, and (c) has cleared the neighborhood around its orbit.[78]

Everything on the left hand side of the 'if and only if' is the definiens and everything on the right hand side is the definiendum. Traditionally, for a definition to be legitimate, the definiens cannot appear in the definiendum. That is, one cannot legitimately define some term by using that very term. However, Gupta and Belnap argue that we can relax this condition on definitions and allow circular definitions, where the definiens occurs in the definiendum.

If one takes all the T-sentences for some language (or a suitable generalization like Künne's modest theory) to be a definition of the truth predicate for that language, then one takes truth to be circularly defined. The reason is that sentences that contain truth predicates, like 'Goldbach's conjecture is true', will have T-sentences with truth predicates on the right hand side (e.g., 'Goldbach's conjecture is true' is true if and only if Goldbach's conjecture is true). Thus, if we take the totality of T-sentences to be a definition of truth, then truth is circularly defined.[79]

As part of their theory of circularly defined concepts, Gupta and Belnap claim that the meaning of a word that expresses a circular concept is captured by a rule of revision. Given a hypothesis

---

[77] See also Patterson (2010), who advocates this view. Field also describes truth as circular in Field (2008a: 7, 8, 11n11, 15, 19); he seems to think that truth predicates display indeterminacy because they are circular, but he does not elaborate on this claim and the idea that the concept of truth is circularly defined plays no role in his philosophical theory of indeterminacy or his logical approach to the aletheic paradoxes.

[78] IAU (2006).

[79] Gupta and Belnap (1993: chs. 4-7).

about such a word's semantic features, a rule of revision specifies a more accurate hypothesis about its semantic features. The rule of revision for a circular predicate P implies that if we assume F has a certain extension, then we can determine a different extension for P. That is, the rule of revision specifies the extension of P under different assumptions about the extension of P. Although circular concepts do not have fixed semantic features, one can use the revision rule for a given circular concept to acquire information about its semantic features by considering its behavior during repeated applications of the rule. For example, if we begin with an arbitrary extension for P and apply the revision rule for P over and over, we generate a sequence of extensions for P. Roughly, if a certain object b always ends up in the extension of P and stays there through repeated applications of the revision rule to different starting extensions, then b satisfies P. Likewise, if a certain object c always ends up outside the extension of P and stays there through repeated applications of the revision rule to different starting extensions, then c does not satisfy P. We can say that b is *stably* P and that c is *stably* not P.[80]

In addition to this way of extracting information about a circularly defined term's semantics, Gupta and Belnap provide a theory for how to reason with circularly defined terms as well. The key is to treat the 'if and only if' in a circular definition as weaker than a material 'if and only if'. One then keeps track of the stage in the revision procedure when reasoning. The result is that this approach to the aletheic paradoxes allows exceptions to (T-In) and (T-Out), but it does validate (T-Intro) and (T-Elim), which makes it a weakly classical theory in the classification scheme of the next chapter.[81]

The circularity approach offers a solution to all three problems posed by the paradoxes and constitutes a diagnostic and descriptive project. Problems with it include the exceptions to (T-In)

---

[80] Gupta (1982) uses 'stable', but Gupta and Belnap (1993) use 'categorical' instead; however, I use 'categorical' in a different way.
[81] Gupta and Belnap (1993); see also Gupta (1982, 1990, 1997b, 2000).

and (T-Out) and the fact that it requires an expressively richer language to classify all the sentences of a particular language. Moreover, it is subject to revenge paradoxes; if one adds a stability predicate to the language in question, then the theory becomes inconsistent. I discuss these points in Chapters Six and Eight.

2.4.10 Inconsistency

Perhaps the first inconsistency approach comes from Tarski, who claimed that natural languages are inconsistent because they have all the ingredients for the aletheic paradoxes.[82] It was (and still is) far from clear what Tarski meant by this, but he certainly used this claim as a justification for focusing on artificial languages instead of natural languages. This kind of claim is more of an excuse for avoiding the diagnostic project and the descriptive project instead of a genuine approach to the paradoxes. Either way, Tarski's claims about natural language have been the subject of debate ever since, and they have definitely inspired others in this category.[83]

The year 1979 saw two papers that defined the two major families of inconsistency approaches. One was Graham Priest's "The Logic of Paradox", and the other was Charles Chihara's "The Liar Paradox: A Diagnostic Investigation".[84] I explain them in order.

Priest suggests that we should treat the reasoning in the aletheic paradoxes as valid and accept their conclusions—that is, accept that some contradictions are true. He calls this view *dialetheism*. Just to be clear, one could accept that some contradictions are true without accepting any contradictions if one rejected (T-Out). Then one would accept that some sentence 'p and [~p]' is true without thereby accepting p and [~p]. Dialetheism, as Priest defines it, requires accepting

---

[82] Tarski (1933).
[83] For discussion see Stroll (1954), Herzberger (1966, 1967), Sinisi (1967), Hugly and Sayward (1980), Soames (1999), Eklund (2002a), Ray (2003), and Patterson (2006).
[84] Priest (1979) and Chihara (1979).

contradictions, along with (T-In) and (T-Out). Indeed, according to Priest, the very meaning of the word 'true' is constituted by (T-In) and (T-Out), so they are true by definition. Once we accept that and the logical principles involved in the paradoxes, we are forced to accept contradictions (as the reasoning in the paradoxes shows).

In classical logic, everything follows from a contradiction (a rule called *Ex Falso Quodlibet* or sometimes *explosion*).[85] Thus, if one accepts classical logic and one accepts a contradiction, then one accepts everything. Although Priest accepts some contradictions, he does not accept everything; so, he suggests a non-classical logic that does not validate explosion. His logic is called LP (for "Logic of Paradox"), and I discuss it in the next chapter. It is in the family of paraconsistent logics.[86]

Dialetheism has had a couple of converts, most notably Jc Beall, whose recent book defends a version that is compatible with disquotationalism.[87] No matter what one thinks about the view, Priest really deserves credit here for an unbelievable act of courage. When he began defending dialetheism thirty years ago, it was ridiculed mercilessly, but today, due to his efforts, it is a view that everyone who works on the paradoxes is forced to take seriously. It counts as a diagnostic and descriptive project. A common objection is that it is irrational to accept a contradiction so dialetheism is irrational. Another is that dialetheism is subject to revenge paradoxes just like many of the other approaches.[88]

The other major family of inconsistency approaches rejects dialetheism, but accepts that either natural languages or the concept of truth are inconsistent. Chihara deserves the credit for putting this approach on the map. All the other approaches to the paradoxes suggest that we should give up one of the principles involved in the paradoxical reasoning. However, Chihara argues, these

---

[85] Assume '*p* and not *p*'. Separate to get *p* and 'not *p*'. From *p*, '*p* or *q*' follows. Using disjunctive syllogism, *q* follows from '*p* or *q*' and 'not *p*'. This is known as the "Lewis Proof"; see Lewis and Langford (1959).
[86] See Priest (2006a, 2006b) for the definitive statement and defense of the view.
[87] Beall (2009); see also Armour-Garb and Beall (2001, 2002, 2003b), Woods (2003), and Brady (2006).
[88] See the papers in Priest, Beall, and Armour-Garb (2005) for discussion.

principles are constitutive of the concepts involved, so giving them up would be to give up the concepts in question. According to Chihara, the concept of truth is itself responsible for the aletheic paradoxes—truth is an inconsistent concept because its constitutive principles, (T-In) and (T-Out), are inconsistent (given certain facts about languages that contain truth predicates). Chihara offers only a diagnostic project, and he says very little about how to understand inconsistent concepts or what to do once one discovers that one's concepts are inconsistent. That is, he is not engaged in the descriptive or prescriptive projects at all, and he says nothing about how to deal with the any of the problem posed by the paradoxes. Still, several theorists (discussed below) have adopted the idea that we need an approach to the paradoxes that takes truth to be an inconsistent concept but rejects dialetheism .[89]

Note that if one is to accept Chihara's view that some concepts have inconsistent constitutive principles without falling into dialetheism, then one must accept that it is possible that a concept's constitutive principles are false. For many people unfamiliar with disputes about defective concepts, this can sound very counterintuitive. It seems to me that the counterintuitiveness comes from confusion about what constitutive principles are supposed to do. Pick, for an example, the claim that cats are animals, and assume that this is a constitutive principle for the concept of a cat.[90] One way to think of constitutive principles is as conditions the world puts on the intelligibility of a concept or a word; on this view, if there is a concept of a cat, then cats are animals. So constitutivity is factive on this view—constitutive principles have to be true. I take it that this notion of constitutivity is the target of W. V. Quine's attack on analyticity that has been so influential in analytic philosophy.[91] However, there is another job for constitutive principles, which is as a guide to concept possession and linguistic competence. On this view, if someone possesses the concept

---

[89] Chihara (1973, 1979, 1984).
[90] See Putnam (1962) for a discussion of this example.
[91] Quine (1951, 1960: ch. 2).

of a cat (or alternatively, if someone's word 'cat' means *cat*), then that person believes that cats are animals. On this construal, the claim that cats are animals can be constitutive of the concept of a cat and false if, say, cats turn out to be cleverly disguised robots.

Paul Boghossian has spent the last decade or so arguing for a distinction like this and championing the latter conception of constitutive principles.[92] Even so, there are good reasons for thinking that the relationship between concept possessors and constitutive principles is not belief or acceptance, but this is a topic best left until later (Chapter Eleven). Whether one thinks there are constitutive principles or not, an inconsistency theorist who follows in Chihara's footsteps is committed to some notion of constitutive principles on which they can turn out to be false.

Matti Eklund's work on the aletheic paradoxes is probably the best-developed and best-known follow-up to Chihara's view .[93] Eklund provides a theory of truth on which truth is an inconsistent concept.[94] He argues that by virtue of our semantic competence, we accept the premises and the inference rules that lead to the liar paradox. Eklund phrases his analysis in terms of inconsistent languages, but I prefer to concentrate on concepts because of the flexibility this allows. One can think of an inconsistent language as one that expresses an inconsistent concept. For Eklund, a concept is inconsistent if and only if the set of constitutive principles for it is inconsistent. He resists the temptation to think of constitutive principles as true or unrevisable. To clarify this claim, he introduces the notions of competence dispositions and culprits. One's *competence dispositions* are belief-forming dispositions that one has by virtue of one's semantic competence. A *culprit* is the false premise or invalid inference used in the derivation of the contradiction in a paradox. One can

---

[92] Boghossian (1996, 1997, 2000, 2001, 2003a, 2003b).
[93] Other inconsistency theorists include Mates (1951), Harman (1986), Yablo (1993a, 1993b), Ludwig (2001), Hill (2002: 118-120), Patterson (2006), and Ludwig and Badici (2007). Note that Patterson already showed up under the "Meaningfulness" banner. The reason is that he thinks that natural languages are inconsistent and that all expressions of inconsistent languages are meaningless. Schiffer (2003) suggests that most philosophical problems are caused by inconsistent concepts, which is a sentiment I endorse in the Introduction.
[94] Eklund (2002a).

say that one's competence dispositions lead one to accept the culprit of the paradox because the set of cognitive meaning-constitutive sentences associated with the concept in question is inconsistent. If a concept displays this phenomenon, then the paradox associated with it is said to *exert pull*. One of Eklund's central theses is that the liar paradox exerts pull.

Eklund's suggestion for a semantics for inconsistent concepts is to define an acceptable assignment of semantic values to expressions of an inconsistent language, L, as one that makes true a weighted majority of the constitutive principles for L. Eklund says very little on how to determine whether an assignment is acceptable and on the weighting function that should be used. That sounds similar to supervaluations, but the big difference (according to Eklund) between his theory and supervaluation semantics is that, for supervaluation semantics, the semantic values of the expressions in question are determined by considering a collection of assignments and constructing one on the basis of their shared properties. In contrast, Eklund's theory determines semantic values by considering a collection of assignments and picking one (or more) of them on the basis of which ones satisfy the constitutive principles associated with the expressions of the language. On the former, an expression has a certain semantic feature if all the various acceptable assignments imply that it does; on the latter, an expression has a certain semantic feature if the claim that it does is the one most compatible with the constitutive principles.[95]

It will not come as a surprise that the approach I present and defend in this book is an inconsistency approach that follows Chihara and Eklund instead of Priest. The big difference between my view and those of the other inconsistency theorists is that I advocate replacing our inconsistent concept of truth with a team of replacement concepts that will do the work we require of truth without giving rise to the paradoxes. That is the goal of Part III.

---

[95] See Eklund (2002a, 2002b, 2007, 2008a, 2008b).

Chapter 3

Logical Approaches to Paradox

The previous chapter surveyed the philosophical approaches to the aletheic paradoxes. Philosophical approaches pursue one of the projects for addressing the paradoxes as they occur in natural language, which usually means that they identify some feature of natural language truth predicates that has been overlooked or disregarded by those puzzled by the paradoxes. This chapter covers logical approaches, which pursue one of the projects for addressing the paradoxes by specifying the principles truth predicates obey and a logic for languages that contain truth predicates; they are almost always formulated with respect to artificial languages. It is much easier to give technical treatments of artificial languages and they allow one to prove various things about one's theory of truth (e.g., that it is consistent relative to a background mathematical theory). It is my view that neither a philosophical approach nor a logical approach is sufficient in isolation. Instead, the aletheic paradoxes demand a combination—a logical approach to carefully specify the principles truth predicates obey, along with a philosophical approach that identifies the relevant features of natural language truth predicates. In the next chapter I consider the combinations that have so far been proposed.

Once one concedes that paradoxical sentences are syntactically well-formed and meaningful, one's options for logical approaches to the liar are fairly limited. One can reject one or both of the primary aletheic principles involved in the paradoxical reasoning; I call these *classical approaches*, not because they *require* classical logic, but because they are *compatible with* classical logic. The other choice is to reject one of the logical principles involved in the paradoxical reasoning; I call these *non-classical approaches*. The classical option requires giving up one or more very plausible principles

involving truth; often these principles are so well-entrenched that to give them up is tantamount to giving up the concept of truth along with them. The non-classical options require giving up one or more very plausible principles involving logical constants or logical consequence; often these principles are so well-entrenched that to give them up is tantamount to giving up the logical concepts in question along with them. So, there are no good options here, just less bad ones. Whatever approach one chooses, an honest thinker will echo Winston Churchill by admitting that it is the worst approach to the aletheic paradoxes except for all the others that have been tried from time to time.

There are several preliminary issues that need to be addressed before we can commence with the classification. First, an artificial language is a mathematical construction—it has a set of symbols (the lexicon), an algorithm for how they can be legitimately combined into strings, which specifies the well-formed formulas of the language (the syntax), and an algorithm for deciding, for any given string, whether it is well-formed. Sentential languages (containing only variables for sentences and logical connectives for negation, conjunction, disjunction, conditional, and biconditional) and first order languages (containing only variables for individuals, names for individuals, function letters, predicate letters, quantifiers, and logical connectives) are examples of artificial languages. The artificial languages are almost always treated as having their logic "built in", so one often speaks of "a first order classical language". Thus, which arguments are valid is usually considered to be essential to the identity of the language. Again, this makes sense if we think of logical principles as constitutive of the logical expressions in question. The semantic features of the symbol strings are often given by model theory, which considers models of the language based on a subject matter for the language (a domain). In simple cases, a model assigns set theoretic items (e.g., sets, ordered pairs, or functions) to certain symbols and strings of symbols so that truth values can be assigned to the formulas of the language. For example, a monadic predicate letter might be assigned a set of

individuals from the domain and an individual constant might be assigned an element of the domain; if the element assigned to the individual constant is in the set assigned to the predicate letter, then the formula consisting of that predicate letter and that individual constant is true in that model.

Second, the vast majority of logical approaches are designed for language-specific truth predicates (e.g., true-in-L). I mentioned these in Chapter One in connection with disquotationalism and in the final section on polyadic truth predicates. They serve at least two purposes for logical approaches to the paradoxes: they permit one to focus on a single artificial language and ignore inter-linguistic truth attributions, and they help alleviate the problems caused by revenge paradoxes. We will see more on the latter issue in Chapter Eight.

Third, all the logical approaches specify a theory of truth, but here I am a bit more careful about how to use the term 'theory'. Here, a theory is a set of sentences closed under logical consequence; endorsing the theory of truth is accepting that truth predicates obey the principles in the theory. Also in this chapter, we encounter different theories of logical consequence, which we will have to track carefully. Many approaches to the liar and other paradoxes require a weakening of classical logic, which consists of the most widely accepted principles of reasoning (at least among analytic philosophers).

There are two primary kinds of theories: axiomatic theories and semantic theories. An *axiomatic theory* proposes a set of axioms and characterizes the theory of truth as the set of sentences that follow from those axioms (usually by classical logic). A *semantic theory* uses the techniques of set theory to define a set of sentences. There are two options for semantic theories for truth—(i) the defined set of sentences is the theory of truth, or (ii) the defined set of sentences is the extension of the truth predicate and the theory follows from this stipulation. The former is said to be the *inner theory* of the defined set, while the latter is the *outer theory* of that set. These might seem to the uninitiated to coincide, but they need not, and confusing them is a recipe for disaster. One

advantage that semantic theories have over axiomatic theories is that some sets of sentences are too complex to be axiomatizable; if the principles governing truth constitute such a set, then a semantic theory is the only way to go.[1] Also, an axiomatic theory for an expression (e.g., 'true-in-L') can only be given in a language that contains that expression. However, with a semantic theory, one can provide a theory of some expression where the theory is formulated in a language that does not have that expression. That fact gives semantic theories much more flexibility, which comes in handy when dealing with revenge paradoxes.

Fourth, we will be paying special attention to logic in this section. *Logic* is (roughly) the study of proper reasoning, whereas *a logic* is a theory of logical terms. A logic specifies the principles these words obey; for example, a conjunction is true if and only if both conjuncts are true. There are two major ways to present the principles for logical words—proof-theoretically as a collection of axioms and inference rules to be used in proofs, or model-theoretically as principles validated by classes of models. We can think of these as theories of the logical consequence relation that obtains between a sentence and a set of sentences by virtue of the logical words in those sentences. A *proof theory* is a set of axioms and inference rules that can be used to construct proofs using sentences with logical terms. One can think of these as characterizing the proof theoretic notion of logical consequence (to be denoted by '$\vdash$').[2] A *model theory* is a class of models such that the principles governing logical terms are true in every model in the class. One can think of these as characterizing a model-theoretic notion of logical consequence (to be denoted by '$\models$').[3] Proof-theoretic and model-theoretic consequence need not coincide.

---

[1] This can be seen as a bug instead of a feature since on some views, we cannot ever come to possess concepts that have extensions this complex; see McGee (1997) and Gupta (1997b) for a discussion of this point in connection with the revision theory.

[2] See Troelestra and Schwichtenberg (2000), Negri and von Plato (2001), and Restall (forthcoming).

[3] See Hodges (1997),

The identification of logical terms is a particularly thorny issue—I am partial to Tarski's view that lots of things can be treated as logical terms. Classical first order logic treats conjunction, disjunction, negation, the conditional, the biconditional, the existential quantifier, the universal quantifier, and identity as logical terms. Modal logic adds possibility operators and necessity operators to that list; deontic logics add permissibility operators and obligatory operators to the original list, and so on. We can use the term 'aletheic logics' for those that treat 'true' as a logical term (Field 2008a is explicit about this point, but does not use this expression).[4]

One way of thinking about the aletheic paradoxes is that they show that intuitive theories of truth are incompatible with intuitive theories of logical terms and intuitive theories of syntax. Since syntax is off the table, we have to choose between principles for our truth predicate and principles for our logical terms.[5] Classical logic is the theory that certain inference rules are valid; non-classical logics reject one or more of these rules (see Information Box 8 details). For our purposes in this chapter, we consider classical logics, weakly classical logics (which have restrictions on the so-called "meta-rules"), paracomplete logics (which reject the law of excluded middle), and paraconsistent logics (which reject the rule *ex falso quodlibet*—i.e., that any sentence follows from a contradiction).

[4] See also Tarski (1936), Peacocke (1976, 1987), McGee (1996), Feferman (1999), Sher (2003), and MacFarlane (2005b).
[5] Syntax is off the table because it is interdefinable with arithmetic via Gödel's technique of arithmetization.

Classical Logic | Information Box 8

Proof theory:

If $p \vdash q \wedge \sim q$, then $\vdash \sim p$

$\sim\sim p \vdash p$

$p, q \vdash p \wedge q$

$p \wedge q \vdash p$

$p \wedge q \vdash q$

$p \vdash p \vee q$

$q \vdash p \vee q$

If $\vdash p \vee q$ and $p \vdash r$ and $q \vdash r$, then $\vdash r$

If $p \vdash q$, then $\vdash p \rightarrow q$

$p, p \rightarrow q \vdash q$

Model theory:

$\vDash \sim p$ if and only if $\nvDash p$

$\vDash p \wedge q$ if and only if $\vDash p$ and $\vDash q$

$\vDash p \vee q$ if and only if $\vDash p$ or $\vDash q$

$\vDash p \rightarrow q$ if and only if either $\vDash q$ or $\nvDash p$

Fifth, all the approaches below are designed for languages that contain their own truth predicates (e.g., an artificial language L that contains a predicate 'true-in-L' that is intended to be true of all and only the true sentences of L). That choice reflects something of a sea change in this tradition that has occurred over the past forty years. Prior to that and since the work of Tarski, the conventional wisdom was that no language could contain its own truth predicate. That judgment was based on the assumption that truth predicates obey (T-In) and (T-Out) and all languages are classical. Given those constraints, one can have a truth predicate, 'true-in-L', for a language L in a different language, often called the *metalanguage*, but L could not contain its own truth predicate.

In 1975 two papers challenged the received view: Kripke's "Outline of a Theory of Truth" and Martin and Woodruff's "Defining 'true-in-L' in L".[6] They both had the same idea—a language can contain its own truth predicate if we drop the assumption that the truth predicate obeys (T-In) and (T-Out) or we drop that assumption that the language obeys classical logic. Since then, the vast majority of the work on logical approaches to the aletheic paradoxes has been on languages that contain their own (language-specific) truth predicates.

The classification scheme I use in this chapter is similar to the one found in Hartry Field's *Saving Truth From Paradox*.[7] It is based on non-trivial combinations of aletheic principles and logical principles. The aletheic principles in question are:

(T-In) If p, then $\langle \mathsf{p} \rangle$ is true.

(T-Out) If $\langle \mathsf{p} \rangle$ is true, then p.

(T-Intro) $\mathsf{p} \vdash \langle \mathsf{p} \rangle$ is true.

(T-Elim) $\langle \mathsf{p} \rangle$ is true $\vdash \mathsf{p}$.

Recall that if a theory accepts (T-In), then it accepts (T-Intro), and if a theory accepts (T-Out), then it accepts (T-Elim); however, some theories accept (T-Intro) without accepting (T-In) and some accept (T-Elim) without accepting (T-Out).[8] No classical theory can accept both (T-Intro) and (T-Elim), so no classical theory can accept both (T-In) and (T-Out), but weakly classical theories accept both (T-Intro) and (T-Elim).[9] Based on these principles, we have three classical options:

(Classical Glut) Accept (T-In), but reject (T-Out) and (T-Elim)

---

[6] See also Kindt (1978).
[7] Field (2008).
[8] To accept (T-Intro) without accepting (T-In), a theory must reject the standard rule of conditional proof (i.e., if one can derive p from q, then one can derive $p \rightarrow q$).
[9] Moreover, no classical theory or weakly classical theory can accept (T-In) and (T-Elim) together or (T-Out) and (T-Intro) together.

(Classical Gap) Accept (T-Out), but reject (T-In) and (T-Intro)

(Classical Symmetric) Reject (T-In), (T-Out), (T-Intro), and (T-Elim)

If one is willing to consider weakening classical logic by giving up the rules for hypothetical reasoning (e.g., conditional proof and reductio), then there is another option:

(Weakly Classical) Accept (T-Intro) and (T-Elim), but reject (T-In) and (T-Out).

Weakly classical approaches reject the conditionals (T-In) and (T-Out), but they accept the weaker inference rules (T-Intro) and (T-Elim); since hypothetical reasoning is invalid in weakly classical logics, it is impossible to derive the respective conditionals from the inference rules. If one is willing to consider a serious weakening of classical logic for even categorical reasoning, then there are two main options:

(Paracomplete) Accept (T-In) and (T-Out) and use a paracomplete logic.

(Paraconsistent) Accept (T-In) and (T-Out) and use a paraconsistent logic.

These are the six main families of logical approaches to the aletheic paradoxes.[10]

Before describing the individual logical approaches, I want to mention some details about the logics involved.

Classical logic, described above, is the received view on the principles of logical consequence and the logical connectives.[11] Familiar alternatives to classical logic are intutionistic logic (**I**) and relevance logics (of which there are many, but **R** is probably the most well-known). However, neither **I** nor **R** has been endorsed as part of an approach to the aletheic paradoxes because introducing a truth predicate into either one that obeys (T-Intro) and (T-Elim) results in triviality.[12]

---

[10] This way of categorizing logical approaches is most explicit in Hartry Field's comprehensive survey—Field (2008a)—but he is not careful to distinguish philosophical from logical approaches.

[11] Over half of the respondants in a recent poll accepted or "leaned toward" classical logic. See http://philpapers.org/surveys/results.pl.

[12] Again, a trivial system is one where every sentence follows from every set of sentences.

Instead, the non-classical logics that play a role in approaches to the aletheic paradoxes are less familiar to non-experts, and they diverge in more radical ways from classical logic than **I** or **R**.

Non-classical logics are often grouped according to interesting properties.[13] For our purposes, we consider five families of non-classical logics: paracomplete, paraconsistent, intuitionistic, relevant, and many-valued. The first three have easy definitions. A logic is *paracomplete* iff the law of excluded middle ($\vdash p \vee \sim p$) is not deducible in it.[14] A logic is *paraconsistent* iff the rule of explosion (also called ex falso quodlibet) ($p, \sim p \vdash q$) is not deducible in it.[15] A logic is *intuitionistic* iff the rule of double negation elimination (i.e., $\sim\sim p \vdash p$) is not deducible in it.[16] *Relevance logic* has a standard definition: its conditional has the *variable-sharing property*, which means that in any theorem whose primary connective is the conditional, the antecedent and consequent share at least one propositional variable); however, this definition is often ignored (e.g., when people say that **RM** or $\mathbf{RM_3}$ are relevance logics).[17] Even so, all the logics that are usually called relevance logics are paraconsistent; however, not all paraconsistent logics are relevant.[18] *Many-valued logic* has less of a standard definition, and is more of a conventional kind. Ostensibly, these are logics that have semantics that appeal to more than two truth values.[19] However, in reality, there are lots of logics that have semantics like this but are not usually called many-valued logics. For example, the relevance logic **RM** is sound and complete with respect to the Sugihara matrix (which treats all the integers as truth

---

[13] For overviews of non-classical logics, see Gobel (2001), Priest (2001), and Schechter (2005), and Burgess (2009). For advanced treatments, see Vigano (2000), Restall (2002), and Dunn and Bimbo (2008). See also Dunn and Hardegree (2001) for background on algebraic methods, which are used extensively to investigate non-classical logics.

[14] See Blamey (2001) and Field (2008a) for discussion of paracomplete logics.

[15] See Priest (2001), Priest (2006a), and Beall (2009) for more on paraconsistent logic.

[16] For more on intuitionistic logic, see Dummett (2000) and van Dalen (2001).

[17] See Priest (2001), Dunn and Restall (2002). RM and RM_3 have '$(A \wedge \sim A) \rightarrow (B \vee \sim B)$' as a theorem, which does not have the variable sharing property.

[18] See Anderson and Belnap (1975), Routley (1982), Read (1988), Anderson, Belnap, and Dunn (1992), Brady (2003), and Mares (2006) for discussion of relevance logics (also called *relevant* logics).

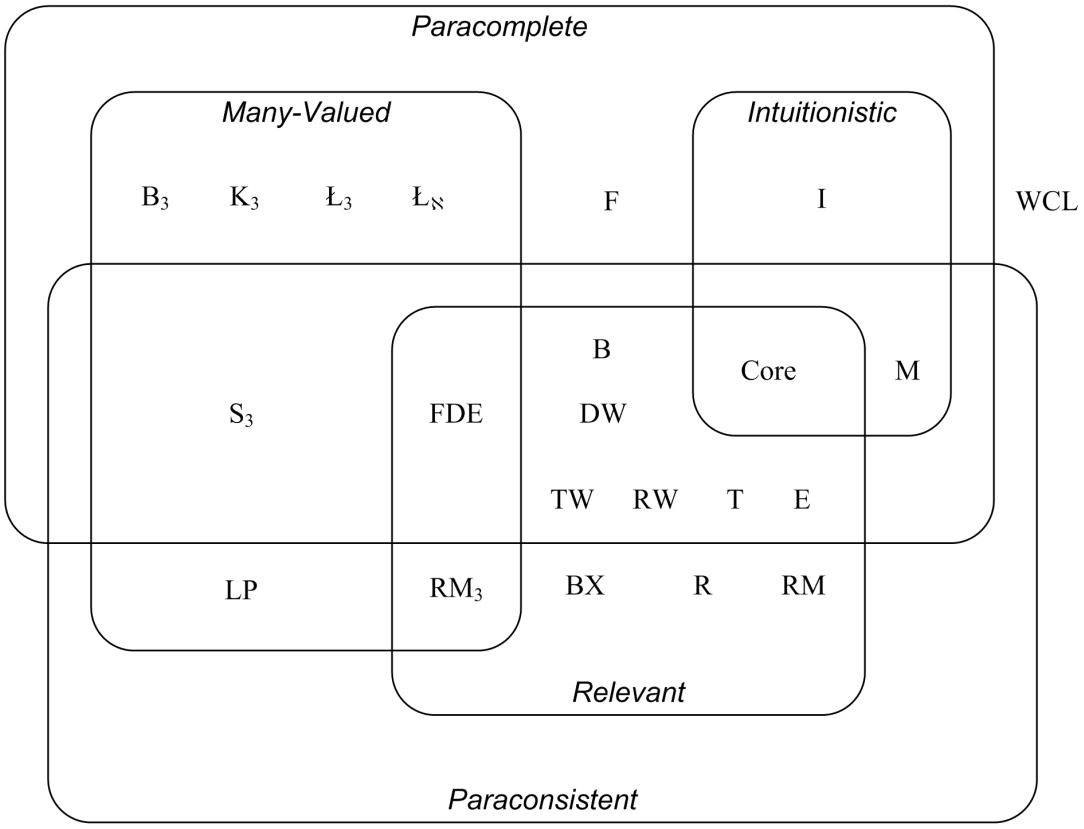
[19] For discussion of many-valued logics, see Gottwald (2001), Urquhart (2001), Hähnle (2001), and Beall and van Fraassen (2003).

values), but **RM** is rarely mentioned as a many-valued logic. Moreover, one can think of just about any logic as a many-valued logic with infinitely many values.[20] It makes more sense to think of many-valued logics as those inspired by particularly elegant or interesting many-valued semantics. That is obviously not a technical definition, but it should serve my purposes. Information Box 9 contains a diagram of these families of logics with many specific logics as well.

[20] That is, any logic that is closed under uniform substitution is weakly complete with respect to a many-valued semantics; see Priest (2001: 133-136) for an accessible proof.

Non-Classical Logics

Information Box 9



Paracomplete: $\nvdash A \vee \sim A$
Paraconsistent: $A, \sim A \nvdash B$
Intutionistic: $\sim\sim A \nvdash A$
Relevant: variable-sharing property
Many-Valued: extensional semantics

B_3: Weak Kleene Logic
K_3: Strong Kleene Logic
$Ł_3$: Łukasiewicz 3-valued Logic
$Ł_\aleph$: Łukasiewicz continuum-valued Logic
F: Field's Logic
I: Intuitionistic Logic
WCL: Weakly Classical Logic
S_3: sublogic of LP and K_3
FDE: First-Degree Entailments
B: Basic Relevant Logic
Core: Tennant's Intutionistic Relevant Logic
M: Minimal Logic
LP: Priest's Logic of Paradox
RM_3: 3-valued Relevance Logic
BX: Extension of B
DW: Extension of B
TW: Extension of DW
RW: Extension of TW
R: Logic of Relevant Implication
T: Ticket Entailments
E: Logic of Relevant Entailment
RM: Extension of R

In what follows, there are seven logics that play a role: **CL** (Classical Logic), **WCL** (Weakly Classical Logic), $\mathbf{K_3}$(Strong Kleene Logic), **F** (Field's Logic, which is an expansion of K_3), **LP** (The Logic of Paradox), **BX** (an extension of basic relevance logic), and **Core** (a substructural intutionistic relevance logic). All but the first are non-classical. $\mathbf{K_3}$ and **F** are paracomplete; that is, they deny the law of excluded middle. **LP** and **BX** are paraconsistent; that is, they deny the rule ex falso quodlibet. **Core** is both paracomplete and paraconsistent. **WCL** does not fit neatly into any of these categories; it has all the same theorems and inference rules as classical logic, but it does not have the same meta-rules. All of this will be explained in due course. For those who are interested, some features of these logics are detailed in Information Box 10, along with the details for Intutionistic Logic (**I**), the most common relevance logic (**R**), and a very basic, but commonly studied relevance logic (**FDE**), for comparison.[21, 22]

[21] Note that the names in Information Box 10 are frequently used for slightly different logics. For example, Bimbo (2007: 774-775) takes the relevance logic **B** to have excluded middle as a theorem, but Beall (2009: 32) does not. Moreover, different proof theoretic formulations of "the same logic" can yield different results. For example, one can provide an axiomatic formulation of **R** on which all the structural rules (listed on the top three lines) are valid, but on some natural deduction formulations of **R**, the structural rule of weakening (line 2) fails.

[22] One must take care in reading the table, since in some of the logics (e.g., **FDE** and $\mathbf{K_3}$), the conditional is defined as $\sim p \vee q$, whereas in others, (e.g., **BX** and **F**) it is not.

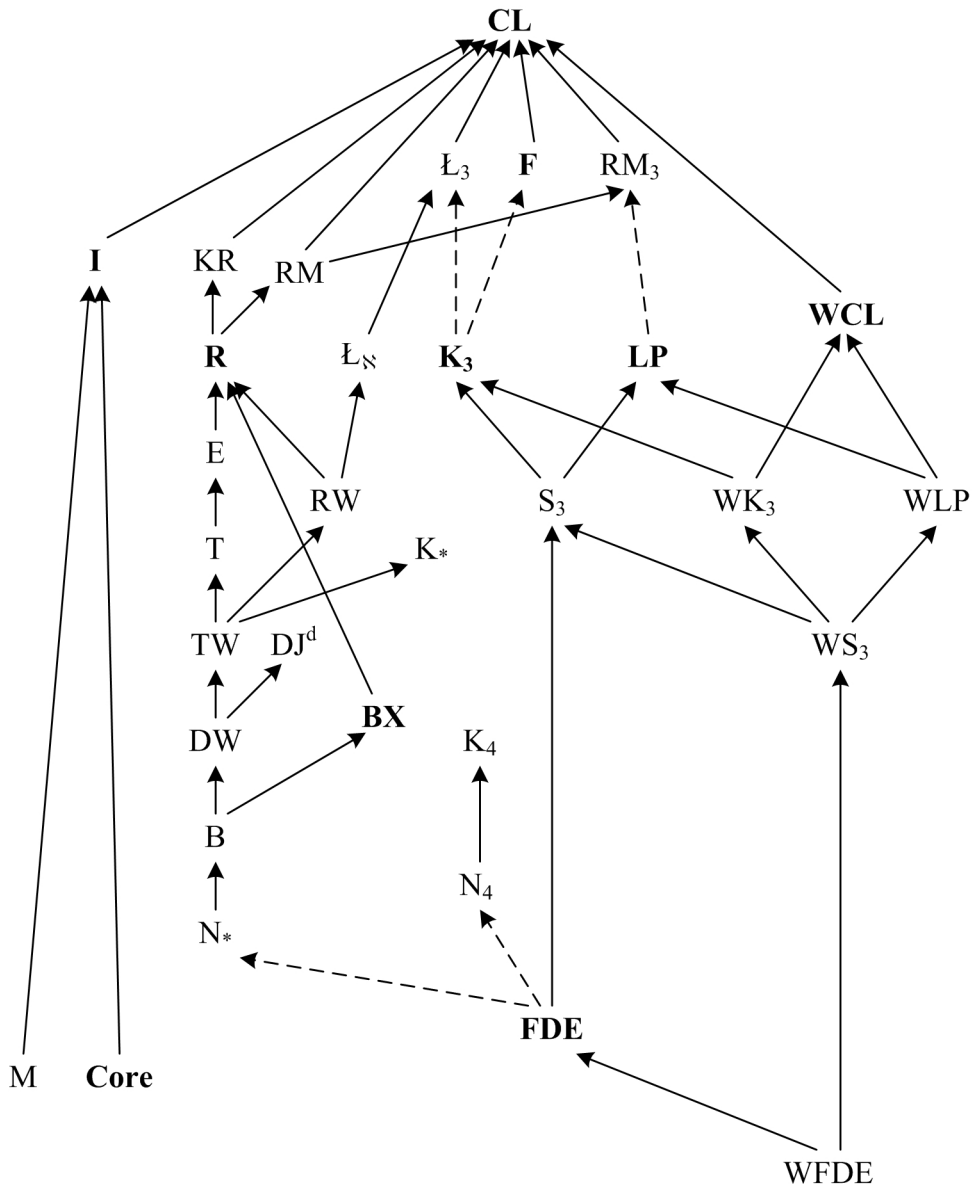
Table of Logics — Information Box 10

		Logics									
		CL	*WCL*	*I*	*Core*	*FDE*	*BX*	*R*	*LP*	K_3	*F*
Rules	If A∈Γ, then Γ ⊢ A	Y	Y	Y	N	Y	Y	Y	Y	Y	Y
	If Γ ⊢ A, then Γ∪Γ′ ⊢ A	Y	Y	Y	N	Y	Y	Y	Y	Y	Y
	If Γ ⊢ A and Γ, A ⊢ B, then Γ ⊢ B	Y	Y	Y	N	Y	Y	Y	Y	Y	Y
	A, B ⊢ A&B	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	A&B ⊢ A	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	A ⊢ A∨B	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	If A ⊢ C and B ⊢ C, then A∨B ⊢ C	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
	If A ⊢ B&~B, then ⊢ ~A	Y	N	Y	Y	N	N	Y	N	N	N
	If ~A ⊢ B&~B, then ⊢ A	Y	N	N	N	N	N	Y	N	N	N
	A ⊢ ~~A	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	~~A ⊢ A	Y	Y	N	N	Y	Y	Y	Y	Y	Y
	A, ~A ⊢ B	Y	Y	Y	N	N	N	N	N	Y	Y
	A, A→B ⊢ B	Y	Y	Y	Y	N	Y	Y	N	Y	Y
	If A ⊢ B, then ⊢ A→B	Y	N	Y	Y	N	N	N	Y	N	N
	A→B, B→C ⊢ A→C	Y	Y	Y	Y	N	Y	Y	N	Y	Y
	A→(B→C) ⊢ B→(A→C)	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
	A→(A→B) ⊢ A→B	Y	Y	Y	Y	Y	N	Y	Y	Y	N
	A→B ⊢ ~A∨B	Y	Y	N	N	Y	N	N	Y	Y	N
	~A∨B ⊢ A→B	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	A→B ⊢ ~B→~A	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	~B→~A ⊢ A→B	Y	Y	N	N	Y	Y	Y	Y	Y	Y
	A→(B→C) ⊢ A&B→C	Y	Y	Y	Y	Y	N	N	Y	Y	N
	A&B→C ⊢ A→(B→C)	Y	Y	Y	Y	Y	N	N	Y	Y	N
	A∨B, ~A ⊢ B	Y	Y	Y	Y	N	N	N	N	Y	Y
	A→C, B→C ⊢ A∨B→C	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	A→B, A→C ⊢ A→B&C	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Theorems	⊢ A→A	Y	Y	Y	Y	N	Y	Y	Y	N	Y
	⊢ A∨~A	Y	Y	N	N	N	Y	Y	Y	N	N
	⊢ ~(A&~A)	Y	Y	Y	Y	N	Y	Y	Y	N	Y
	⊢ A&(A→B)→B	Y	Y	Y	Y	N	N	N	Y	N	N

Also for those interested, Information Box 11 depicts some relationships between the more well-known non-classical logics; the seven non-classical logics listed above that we consider in this chapter are in bold. Definitions of all the logics mentioned in this chapter are given in an appendix.

Relations Between Non-Classical Logics

Information Box 11



3.1 Classical Glut Theories

Classical glut theories accept (T-In), reject (T-Out) and (T-Elim), and are compatible with classical logic. They are so named because it is common to call sentences that are both true and false *gluts*. The term arose as an antonym for 'gaps'. Gluts are sentences that are neither true nor false. As far as I know, there are no classical glut theorists; there are several theorists who accept that some sentences are both true and false, but none of them accept classical logic.[23]

Why does accepting (T-In) and classical logic gives rise to gluts? Let L be a liar sentence like (1) above, so L = 'L is not true'. Using (T-In), we get 'If L is not true, then 'L is not true' is true', which is equivalent to 'If 'L is not true' is not true, then L is true' by (Sub), which is equivalent to 'L is true' by classical logic. So the classical glut theory implies that L is true. Using (T-In) once more we get that the classical glut theory implies that 'L is true' is true. So the classical glut theory implies that L is true and that L's negation is true. Recall that there are several distinct ways of defining falsity, but the one accepted by those in this tradition is truth of negation.[24] Given this definition, we get that the classical glut theory implies that L is true and L is false (i.e., ~L is true); hence, it implies that L is a glut. Of course, on an alternative definition of falsity (e.g., negation of truth), L does not count as a glut and, indeed, there can be no classical glut theory at all.

So, if one is going to accept classical logic and keep either (T-In) or (T-Out), then the choice comes down to a theory that implies that some sentences and their negations are true (gluts) or a theory that implies that some sentences and their negations are not true (gaps). Faced with this choice, everyone picks the latter.

---

[23] Friedman and Sheard discuss (but do not endorse) theory A, which is a classical glut theory; see Friedman and Sheard (1987) and Leigh and Rathjen (2010).

[24] The two major accounts of falsity are: (i) ⟨p⟩ is false iff ⟨~p⟩ is true (i.e., truth of negation), and (ii) ⟨p⟩ is false iff ⟨p⟩ is not true (i.e., negation of truth).

3.2 Classical Gap Theories

Classical gap theories accept (T-Out), and reject (T-In), (T-Intro), and (T-Enter); they are compatible with classical logic. Again, a truth-value gap is a sentence that is neither true nor false, but since we have multiple ways of defining falsity, we end up with multiple definitions of 'gap'. On one definition (truth of negation), a gap is a sentence such that neither it nor its negation is true. That is the sense of 'gap' at work in this subsection.

Now that we see how to understand a classical gap theory, let us see why accepting (T-Out) and classical logic gives rise to gaps. Again, let L be a liar sentence, so L= 'L is not true'. The instance of (T-Out) for L is 'if 'L is not true' is true, then L is not true', which is equivalent to 'If L is true, then L is not true' by (Sub). It follows from classical logic that the classical gap theory implies that L is not true. Notice that L's negation is equivalent to 'L is true'. So if L's negation is true, then L is true, so the classical gap theory implies that L's negation is not true as well. Thus, we get a sentence such that it and its negation are not true (i.e., it is neither true nor false).

By far the most influential theory in this category is what I have (following Kripke) called the Orthodox view, which constructs a hierarchy of very restricted truth predicates using Tarski's techniques. A major innovation in Tarski's work is a model theory for sentences containing quantifiers (e.g., 'all' and 'some'). Tarski defined a relation, satisfaction, that holds between a formula and a sequence of objects from the domain. Using satisfaction, one can specify exactly the conditions under which a quantified formula is true. This, together with clauses for the other logical terms, allows one to give a recursive definition of truth for first order classical languages (again, this definition is formulated in some other language); see Information Box 12 for details on recursive definitions.

Recursive/Inductive Definitions Information Box 12

A recursive definition (sometimes called an *inductive definition*) defines some term by way of a set of rules. Some of the rules specify a base case explicitly and the recursive rules extend the definition and can be applied over and over to arrive at a complete definition.

For example, a recursive definition of 'Stu's ancestors' might be:

(Base) Stu's parents are among Stu's ancestors.

(Recursion) The parents of Stu's ancestors are among Stu's ancestors.

(Final) No one besides those specified in (Base) and (Recursion) are among Stu's ancestors.

Using these rules, one can determine, for any given person, whether that person is one of Stu's ancestors.

Tarski's method for defining truth can be used to indirectly give an account for a language that contains its own truth predicate. The key is defining a hierarchy of truth predicates. Consider a classical first order language L_0 that does not contain a truth predicate. We can define a truth in that language in some other language, M. We can extend L_0 to L_1 by adding a truth predicate, 'true_0' that behaves according to our definition; essentially, we treat the sentences of L_1 that do not contain 'true_0' as if they constituted their own language, L_0. 'true_0' is stipulated to apply only to sentences among those. Now we can define a truth in L_1, again in another language (perhaps M again); according to this definition, although L_1 has the capacity to construct liar type sentences (e.g., L_0 = 'L_0 is not true_0'), they are false since 'true_0' applies only to sentences of L_1 that do not contain 'true_0'. Again, we extend L_1 to a new language L_2 by adding a new truth predicate, 'true_1' to L_1, which behaves according to our definition and is stipulated to apply only to sentences of L_2 that do not

contain 'true$_1$'. Following this procedure, we can extend the language so that it contains a hierarchy of truth predicates of any finite order.[25]

The Orthodox approach is to treat natural language truth predicates as if they are ambiguous and they can express any one of the predicates in the hierarchy on an occasion of use. This counts as a classical gap approach because no matter which notion of Tarskian truth is expressed by the natural language truth predicate, (T-In) will have exceptions. If p is a sentence of natural language that contains the truth predicate and it gets interpreted as meaning *true$_n$* for some n, then (T-In) will fail when applied to p if the truth predicate in the instance of (T-In) is interpreted as meaning *true$_m$*, where m≤n. For example, 'if p is true$_1$, then 'p is true$_1$' is true$_1$' is an instance of (T-In) with the subscripts made explicit, and it is false; there is no reading of (T-In) on which all of its instances are true, according to the Orthodox approach.

I mention two axiomatic theories that fall into this category. They are called KF (for Kripke-Feferman) and VF (for van Fraassen); they are listed in Information Boxes 13 and 14.[26] They are both inspired by Kripke's work on truth, to which I turn now.

---

[25] The procedure can be extended into the transfinite; see Halbach (1996, 1997) for details.

[26] There are several variants of these theories. The ones used here are those discussed in Michael Sheard's overview, Sheard (1994). It might seem odd to include KF here, but (T-Out) is derivable from its axioms; see Field (2008a: 121-122) for discussion. Friedman and Sheard's system H is another axiomatic theory in the "classical gap" category.

Axiomatic Theory KF | Information Box 13

Solomon Feferman proposed this axiomatic theory that is inspired by the outer theory of Kripke's Strong Kleene inductive construction.

$$
\begin{array}{rcll}
T(\sim\sim p) & \leftrightarrow & T(p) & \\
T(p\vee q) & \leftrightarrow & T(p)\vee T(q) & \\
T(\sim(p\vee q)) & \leftrightarrow & T(\sim p)\vee T(\sim q) & \\
T(p\wedge q) & \leftrightarrow & T(p)\wedge T(q) & \\
T(\sim(p\wedge q)) & \leftrightarrow & T(\sim p)\wedge T(\sim q) & \\
T(p\rightarrow q) & \leftrightarrow & T(\sim p)\vee T(q) & \\
T(\sim(p\rightarrow q)) & \leftrightarrow & T(p)\vee T(\sim q) & \\
T(p\leftrightarrow q) & \leftrightarrow & T(p)\leftrightarrow T(q) & \\
T(\sim(p\leftrightarrow q)) & \leftrightarrow & T(\sim p)\leftrightarrow T(\sim q) & \\
T(\forall x\phi(x)) & \leftrightarrow & \forall x T(\phi(x)) & \\
T(\sim\forall x\phi(x)) & \leftrightarrow & \forall x T(\sim\phi(x)) & \\
T(\exists x\phi(x)) & \leftrightarrow & \exists x T(\phi(x)) & \\
T(\sim\exists x\phi(x)) & \leftrightarrow & \exists x T(\sim\phi(x)) & \\
T(T(p)) & \leftrightarrow & T(p) & \\
T(T(\sim p)) & \leftrightarrow & T(\sim p) & \\
T(\sim T(p)) & \leftrightarrow & T(\sim p) & \\
T(\sim T(\sim p)) & \leftrightarrow & T(p) & \\
T(p) & \leftrightarrow & p & \text{(for p atomic and arithmetic)} \\
T(\sim p) & \leftrightarrow & \sim p & \text{(for p atomic and arithmetic)} \\
 & \sim(T(p)\wedge T(\sim p)) & &
\end{array}
$$

Axiomatic Theory VF | Information Box 14

Andrea Cantini proposed this axiomatic theory that is inspired by the outer theory of Kripke's supervaluation inductive construction.

$T(p)$ (for p tautology)
$T(p)\wedge T(p\rightarrow q)\rightarrow T(q)$
$p\rightarrow T(p)$*
$T(p)\rightarrow p$
$\sim(T(p)\wedge T(\sim p))$
$\forall x T(\phi(x))\rightarrow T(\forall x \phi(x))$
$T(p)\rightarrow T(T(p))$
$T(\sim T(p))\rightarrow T(\sim p)$

*for p atomic and arithmetic and
for p negated atomic and arithmetic

I discussed Kripke's constructions above under the "Indeterminacy" heading because that is often how they are interpreted, but there are classical gap approaches that are based on Kripke's constructions and they do not seem to have much to do with indeterminacy.

Kripke showed how to begin with a classical first order language and extend it by adding a truth predicate to it. The truth predicate is interpreted in stages according to a model-theoretic recursive definition, but the key is that its extension and its anti-extension, though disjoint, are not exhaustive. The base case is usually taken to be a model where the extension of the truth predicate has all the true sentences that do not contain the truth predicate and the anti-extension contains all the false sentences that do not contain the truth predicate; all the sentences containing truth predicates are in neither set. For the recursion step, the sentences attributing truth to the sentences in either the extension or the anti-extension are placed in either the extension or the anti-extension of truth depending on whether they are true or false. This step requires a scheme for handling logically compound sentences. Here Kripke considers three options—the weak Kleene, the Strong Kleene,

and supervaluations. Supervaluations I have discussed above; see Information Boxes 15 and 16 for details on the Weak Kleene and Strong Kleene schemes. The major difference between them is that on the Weak Kleene scheme, if a sentence is in neither the extension nor the anti-extension, then every compound in which it is a component is in neither of them as well, whereas on the Strong Kleene scheme, sentences in neither can occur in compounds that are in one or the other. For example, if p is in neither and q is in the extension at a certain stage, then if the recursive step uses the Weak Kleene, [p∨q] will be in neither, whereas if it uses the Strong Kleene, [p∨q] will be in the extension. The Strong Kleene scheme is much more popular than the Weak Kleene scheme in approaches to the aletheic paradoxes.

Weak Kleene Logic | Information Box 15

The following are truth tables for the logical terms in Weak Kleene Logic ('T', 'F', and 'G' represent truth, falsity, and gaphood, respectively).

p	~p
T	F
G	G
F	T

∧	T	G	F
T	T	G	F
G	G	G	G
F	F	G	F

∨	T	G	F
T	T	G	T
G	G	G	G
F	T	G	F

→	T	G	F
T	T	G	F
G	G	G	G
F	T	G	T

Strong Kleene Logic | Information Box 16

The following are truth tables for the logical terms in Strong Kleene Logic ('T', 'F', and 'G' represent truth, falsity, and gaphood, respectively).

p	~p
T	F
G	G
F	T

∧	T	G	F
T	T	G	F
G	G	G	F
F	F	F	F

∨	T	G	F
T	T	T	T
G	T	G	G
F	T	G	F

→	T	G	F
T	T	G	F
G	T	G	G
F	T	T	T

Kripke proves that the process of adding sentences to the extension and to the anti-extension reaches a point where no more are added to either one. This is called a fixed point. To understand this aspect of the construction, one needs to know a little bit about the infinite.

At the end of the nineteenth century, in response to some serious mathematical puzzles involving the concept of infinity, Georg Cantor offered a new theory that has come to be essential to modern mathematics. Cantor suggested that there are many kinds of infinity and some are larger than others.[27] These are measured by *cardinal numbers*. The smallest one describes the size of the set of natural numbers (i.e., the set consisting of 1, 2, 3, and so on); it is $\aleph_0$ (pronounced *aleph-naught*, *aleph-zero*, or *aleph-null*), and the sets it describes are sometimes called *countably infinite*. One can prove that the set of all sets of natural numbers is strictly larger than the set of all natural numbers, so the former is described by a larger cardinal number (exactly which one depends on some delicate issues we do not need to get into). In fact, there are many cardinal numbers and so, many kinds of infinity. One might wonder just how many there are and how they are ordered. To investigate these matters, we need to talk about ordinal numbers as well.

Ordinal numbers describe kinds of orderings; indeed, one can think of the familiar finite numbers as ordinals—1 describes the basic ordering of just one thing, 2 describes the ordering of two things (e.g., <0, 1>, the numbers that come before it), 3 describes the ordering of three things (e.g., <0, 1, 2>, the numbers that come before it), and so on. The ordinal that describes the ordering of all the natural numbers is ω. The one that comes after it describes the ordering <0, 1, 2, …, ω> where the '…' is all the natural numbers; it is ω+1. 2ω is <0, 1, 2, …, ω, ω+1, ω+2, …>. We can continue defining kinds of orders in this way for a long time—we get 3ω, 4ω, and so on.

---

[27] Talking about kinds of infinity can be grating on the ears (or eyes) of those who know transfinite set theory, so please keep in mind that this chapter is supposed to be very accessible. See Halmos (1974), Lavine (1994), Hrbacek and Jech (1999) for more information.

The ordinal that describes the whole ordering is ω^2. One starts to see that this is a ridiculously complex process—it eventually reaches ω^3, ω^4, and so on; after all those is ω^ω. Eventually we get ω^{ω^ω}, $\omega^{\omega^{\omega^\omega}}$, and so on. After all of these is $\omega^{\omega^{\cdot^{\cdot^{\cdot}}}}$, where the '…' is ω ωs (we could describe this ordinal as "ω to the ω ω times"). This one has the special name 'ε_0'. The ordinals, of course, keep going and going, and we have only scratched the surface (even ε_0 has only a countable number of predecessors, so the set consisting of all the ordinals that come before it has cardinality $\aleph_0$), but this is good enough for our purposes.

The study of ordinal and cardinal numbers is a branch of mathematics called *transfinite set theory*. I have introduced it here to explain Kripke's constructions, which rely on transfinite recursion. Recall that each construction has a base step and a recursive rule for generating further steps. A transfinite recursive definition is just a recursive definition where the steps go up the ordinal numbers past ω. That is what Kripke offers, and he proves that at some point along the ordinals, no more sentences of our language are added to either the extension or the anti-extension of the truth predicate (there are only countably many sentences in the language so the process is guaranteed to stop at some point).[28] Again, the interpretation of the language constructed by the ordinal point at which the process stops is called a fixed point.

Kripke shows how to define a whole system of fixed points by varying the base step in the recursion and varying the logical scheme used in the recursion step. As I described it, the base step involves evaluating all the non-aletheic sentences of the language (i.e., those with no truth predicates), but there are lots of other options. I mentioned two logical schemes, Weak Kleene and Strong Kleene, but Kripke's construction works with supervaluations as well, and there are several

[28] Note that there are uncountable ordinals (i.e., ordinals with uncountably many predecessors).

ways to do the supervaluations by varying the conditions on acceptable valuations. I have listed some of the options in Information Box 17.[29]

Kripke Constructions | Information Box 17

All theories are based on transfinite recursive definition of 'true-in-L'

Base:	Recursive Clause:
Non aletheic sentences	Weak Kleene
Logical truths and falsehoods	Strong Kleene
Aletheic principles	Simple Supervaluation
Truth tellers	Weak Supervaluation
	Medium Supervaluation
	Strong Supervaluation

There are *inner theory* and *outer theory* interpretations for each combination

The final decision to be made when using Kripke-style recursive definition is how to interpret the sets of sentences one arrives at. I have followed Kripke in describing them as the extension and the anti-extension of the truth predicate for the language. One can, however, interpret the extension set as the theory of truth itself rather than as the extension of the truth predicate. Thinking of it as the extension of the truth predicate is the *outer theory* interpretation, whereas thinking of it as the theory of truth is the *inner theory* interpretation. On the outer theory interpretation, one constructs

[29] Weak, Medium, and Strong Supervaluation (listed in Information Box 16) are the following: (Weak) acceptable models assign classically consistent extension to 'true', (Medium) acceptable models assign deductively closed extension to 'true', and (Strong) acceptable models assign maximally consistent extension to 'true'. See Field (2008a: 176-186) for details.

the set of sentences of the language that are true; on the inner theory interpretation, one constructs the set of principles that govern the truth predicate.

The classical gap approaches based on Kripke constructions are the outer theories of the fixed points where the truth predicate is totally defined (i.e., anything not in the extension of the truth predicate is stipulated to be in its anti-extension). The Weak Kleene classical gap approach, the Strong Kleene classical gap approach, and the various supervaluation classical gap approaches differ slightly in their details (e.g., the law of excluded middle—$p \vee \sim p$—is true on the supervaluation classical gap approach but not on the Strong Kleene classical gap approach, and the supervaluation classical gap approach allows disjunctions to be true even though neither disjunct is true but the Strong Kleene classical gap approach does not).[30]

The major criticism of these theories is that they allow exceptions to (T-In), which, it is argued, is needed for truth predicates to perform their expressive role (see Chapter Six). Also, there are revenge paradox worries associated with most of the theories that result from Kripke constructions.

3.3 Classical Symmetric Theories

A classical symmetric theory denies both (T-In) and (T-Out) and accepts classical logic. In a classical framework, one can derive (T-In) using (T-Intro) and one can derive (T-Out) using (T-Elim), so these two inference rules must be rejected as well. Some classical symmetric theories

[30] In addition to the classical gap theories discussed so far, there are what Field calls *stratified classical gap theories*; these theories have a hierarchy of truth predicates (somewhat like the Tarski hierarchy), where the language at each step has its own truth predicate that obeys a classical gap theory; see Field (2008a: 214-225). There is also Stephen Read's version of Bradwardine's solution to the liar, which has a unique philosophical motivation, but is formally just like the other classical gap approaches. See Read (2007, 2008a, 2008b, 2009, 2010) and the papers in Rahman, Tulenheimo, and Genot (2008); see also Mills (1998) for a similar view. For other Medieval views on the aletheic paradoxes, see Spade (1988) and Sanson and Alwishah (2009).

accept (T-Enter) and (T-Exit), the categorical rules, since one cannot derive a contradiction from them alone.[31, 32]

Notice the argument for the existence of classical gluts for the classical glut theorist used (T-In) and the argument for the existence of classical gaps for the classical gap theorist used (T-Out). Thus, if we reject both of these, then we are no longer guaranteed to get gaps or gluts. Of course, this comes at a price—we no longer have anything to say about sentences that are exceptions to (T-In) and (T-Out).

The most prominent axiomatic theory in this category is called FS for 'Friedman' and 'Sheard'; see Information Box 18 for details.[33, 34]

---

[31] See McGee (1991), which contains a consistent theory of truth on which the truth rules are valid in categorical reasoning.

[32] Some classical symmetric theories also include the rules:

(~T-Enter) If $\vdash \sim\mathsf{p}$, then $\vdash \sim\mathrm{T}\langle\mathsf{p}\rangle$

(~T-Exit) If $\vdash \sim\mathrm{T}(\langle\mathsf{p}\rangle)$, then $\vdash \sim\mathsf{p}$

as well, but not all do.

[33] Friedman and Sheard's original proposal (theory D) is a bit different from this presentation of FS in that the original contained (T-Inf) and (E-Inf) as well (these say, roughly, that if all the instances of a universal generalization are true, then the generalization is true, and that if an existential generalization is true, then some instance of it is true, respectively). One drawback of theory D is that it is ω-inconsistent.

[34] Other axiomatic theories in this category are theories B, C, E, F, G, and I of Friedman and Sheard (1987). See also Leigh and Rathjen (2010) for discussion.

Axiomatic Theory FS | Information Box 18

Harvey Friedman and Michael Sheard proposed this axiomatic theory that is inspired by the outer theory of the revision construction for all finite levels.

$T(p)$ (for p tautology)
$T(p)\wedge T(p\rightarrow q)\rightarrow T(q)$
$p\rightarrow T(p)$*
$\sim(T(p)\wedge T(\sim p))$
$T(p)\vee T(\sim p)$
If $\vdash p$, then $\vdash T(p)$
If $\vdash T(p)$, then $\vdash p$
If $\vdash \sim p$, then $\vdash \sim T(p)$
If $\vdash \sim T(p)$, then $\vdash \sim p$

*for p atomic and arithmetic and
for p negated atomic and arithmetic

The major semantic theory of truth in this category is due to Vann McGee. McGee claims that 'true' is vague and uses a supervaluation-based theory of vagueness to arrive at a theory of truth. Like many who work on vagueness, McGee introduces a 'definite' operator to distinguish unproblematic from problematic cases of application. For example, one might say that someone is definitely bald, in which case the person does not fall within the borderline between baldness and nonbaldness. For someone who is in the borderline, we can say they are *unsettled*. Accordingly, McGee distinguishes between truth and definite truth. On McGee's theory, (T-In) and (T-Out) preserve definite truth, not truth. That is, the following principles hold:

(i) If p is definitely true, then 'p is true' is definitely true.

(ii) If p is definitely not true, then 'p is true' is definitely not true.

(iii) If p is unsettled, then 'p is true' is unsettled.

For his theory of definite truth, McGee uses the strong Kleene version of Kripke's fixed-point theory. By appealing to the notion of a partially interpreted language, McGee proves that both his supervaluation semantic theory for truth and his fixed point semantic theory for definite truth apply to the language in which they are formulated. One of the keys to this result is that the formal language in which his theories are formulated does not contain sentences that pose revenge paradoxes. That is, the sentence:

(2) (2) is false or unsettled,

is unsettled, and so not definitely true, but it is not *definitely* unsettled, thus, no paradox results from it.[35] Furthermore, because (2) is not a consequence of either theory, the usual argument is not available to show that it is not true.

In addition, McGee proves that both theories are expressible in the formal language in which they are formulated. He achieves this result by setting up his theories so that 'if p is definitely true, then p is true' is not definitely true. This is a counterintuitive result, but he needs it to ensure that (2) does not pose a revenge paradox. He also has to deny that the vague concept of truth he presents can be made more precise. Attempts at precisification result in revenge paradoxes.[36,37] [38]

In Part III, I offer a theory of truth that fits in this category as well.

3.4 Weakly Classical Theories

---

[35] McGee (1991, ch. 9).

[36] See Yablo (1989), Simmons (1993), Priest (1994a), Tappenden (1994), and Mills (1995) for discussion of this aspect of McGee's theory.

[37] Other classical symmetric theorists include Patrick Greenough and Thomas Hofweber. Greenough argues that it is illegitimate to suppose paradoxical sentences, which could be interpreted in different ways. I think the most plausible is restricting the aletheic principles to categorical contexts; see Greenough (2001). Hofweber argues that all the aletheic principles and indeed all logical principles have exceptions; see Hofweber (2007, forthcoming).

[38] Several axiomatic theories fit into this category; namely theories B, C, and I from Friedman and Sheard (1987); no one advocates one of these as far as I know.

Instead of choosing between accepting (T-In) or (T-Out), some theorists choose to accept weakened forms of them as (T-Intro) and (T-Elim). These are weakly classical theories.[39] The term 'weakly' is important because all the theories in this section accept **WCL** (described above and in the appendix to this chapter). **WCL** has all the inference rules and theorems of classical logic (**CL**), but **WCL** does not have some of the "meta-rules" of classical logic (e.g., conditional proof, reductio, and reasoning by cases). It is not hard to see why—imagine we allow conditional proof, which is the introduction rule for the conditional. We begin by supposing that p is true. Using (T-Elim) we arrive at p, and by conditional proof, we get 'if p is true, then p', which is (T-Out). Analogous reasoning holds for (T-Intro).

One can show that all weakly classical theories have the following counterintuitive consequence: they imply that for some paradoxical sentence $\langle p \rangle$, either (p and $\langle p \rangle$ is not true) or ($\langle p \rangle$ is true and $\sim p$), but they do not imply either disjunct. Recall that, for some paradoxical sentence $\langle q \rangle$, classical glut theories imply that $\langle q \rangle$ is true and $\sim q$, and classical gap theories imply that for some paradoxical sentence $\langle r \rangle$, ($\langle r \rangle$ is not true and r). Thus, weakly classical theories imply the disjunction of these two counterintuitive claims, but they do not imply either claim by itself. Again, this result stems from the fact that **WCL** does not allow reasoning by cases.

There are two major families of semantic theories in this category. One of the families of semantic theories consists of the internal theories for the recursive supervaluation constructions—that is, the theories that result from treating the set one defines using Kripke's construction with

---

[39] Some weakly classical theories also include the rules:

(~T-Intro) $\sim p \vdash \sim T\langle p \rangle$

(~T-Elim) $\sim T(\langle p \rangle) \vdash \sim p$

as well. Note that these are the contrapositives of (T-Elim) and (T-Intro), respectively. If we were dealing with conditional statements like (T-In) and (T-Out), then we would get the contrapositives as logical consequences, but since weakly classical theories accept only the inference rules and reject the corresponding conditional statements, they need not endorse the contrapositives.

supervaluations in the recursive rule as the theory of truth. These theories are weakly classical and have some interesting properties.[40]

The other family of semantic theories come from a very different kind of construction that we have not yet seen—revision constructions. Anil Gupta and Hans Herzberger (independently) proposed revision constructions in the early 1980s, but Gupta and Nuel Belnap substantially refined them in the mid 1990s. The idea is that we begin with a hypothesis about the extension (and perhaps the anti-extension as well) of the truth predicate and we use a revision rule to improve (or at least change) that hypothesis over and over. Using the initial hypothesis and the revision rule, we construct a revision sequence of hypotheses about the extension of the truth predicate. Revision sequences can be finite in length, but more powerful theories result from revision sequences that extend into the transfinite ordinals (discussed above). One extra complication for transfinite revision sequences is that the revision rule needs to work for both limit and non-limit ordinals. Limit ordinals are those that have no immediate predecessor (e.g., ω, 2ω, …), whereas non-limit ordinals are all the rest. Revision rules at non-limit ordinals are pretty straightforward, but they are tricky at limit ordinals. A key difference between Gupta and Belnap's revision constructions and Kripke's recursive constructions is the former need not reach fixed points in order to arrive at a theory, whereas the latter do; moreover, in the recursive constructions, once a sentence goes into either set, it stays there (this property is called *monotonicity*), whereas in revision constructions, sentences can go in or drop out of the sets from step to step.

Gupta and Belnap define the notion of stability—a sentence is *stably true* in a revision sequence if and only if there is an ordinal at which it is in the extension of truth and it stays in the extension of truth for every later ordinal, and a sentence is *stably false* in a revision sequence if and only if there is

---

[40] See van Fraassen (1968, 1970a, 1970b); see Field (2008a: chs. 10-12) for discussion. For recent discussions of supervaluations that are not focused on approaches to the aletheic paradoxes, see Varzi (2007), R. Williams (2008), and Asher, Dever, and Pappas (2009).

an ordinal at which it is in the anti-extension of truth and it stays there for every later ordinal. A revision theory is the theory one arrives at by basing the revision rule on (T-In) and (T-Out)—the resulting revision sequence does not reach a fixed point, so (T-In) and (T-Out) are not principles of any revision theory, although all of them accept (T-Intro) and (T-Elim), which is what makes them weakly classical.

Much of the criticism of weakly classical theories stems from the fact that they do not accept (T-In) or (T-Out), and these seem to be crucial for truth predicates to perform their expressive role. Also, revenge paradoxes are a problem for these theories. For example, a language that contains its own truth predicate and for which one gives a revision theory cannot contain its own stable truth predicate or the revision theory turns out to be inconsistent. In addition, some of the theories in this category are ω-inconsistent (some revision theories have this property), which is a pretty serious problem (see Chapter Two). Finally, the loss of the meta-rules is a major cost that, in my opinion, has not been fully appreciated.[41]

3.5 Paracomplete Theories

Now we turn to non-classical theories—those that reject some of the classically valid principles taken to govern logical terms in categorical contexts. Paracomplete theories accept both (T-In) and (T-Out), but reject central principles of classical logic, most notably the law of excluded middle (i.e., $p \vee \sim p$) and principles governing the conditional.

---

[41] This cost is more of aTimothy Williamson remarks about supervaluations as an approach to the sorites paradox that "Conditional proof, argument by cases, and reductio ad absurdum play a vital role in systems of natural deduction, the formal systems closest to our informal deductions. They are the rules by which premises are discharged, i.e. by which categorical conclusions can be drawn on the basis of hypothetical reasoning. … Thus supervaluations invalidate our natural mode of deductive thinking," (Williamson (1994: 152).

The inner theory of the Strong Kleene recursive construction is in this category, but Strong Kleene logic ($\mathbf{K_3}$) is very weak owing to its conditional. The real star of this category is Hartry Field's theory, which seems to me like the premier logical approach to the aletheic paradoxes right now. Field begins with the inner Strong Kleene *recursive* theory, but sets out to define a well-behaved conditional for it. He uses a *revision* construction to define an adequate conditional for his logic (to arrive at **F**). So his theory combines elements of Kripke's recursive constructions and Gupta and Belnap's revision constructions. The resulting theory of truth is just the T-sentences but the biconditional in them is based on Field's new conditional. Field's conditional acts just like a material conditional when the law of excluded middle is assumed. Moreover, it allows him to define a determinateness operator, D, which can be used to classify the liar sentences in the object language as not determinately true and not determinately false.[42] Furthermore, the determinateness operator iterates non-trivially so it can even be used to classify liar-type sentences that contain occurrences of the determinateness operator. For example, a sentence Q that is provably equivalent to 'Q is not determinately true' is not determinately determinately true. Indeed, by iterating the determinateness operator, one can generate a transfinite hierarchy of determinateness operators (i.e., a hierarchy that extends into the ordinals past ω). It is a delicate issue just how far this hierarchy extends since the language in question also contains a truth predicate, which can be used to generalize over the determinateness operators. Field argues that the hierarchy eventually breaks down, but in the interesting cases, the point of breakdown is indeterminate.[43] Thus, according to Field, the determinateness operators serve the purpose of classifying liar-type sentences without giving rise to pesky revenge paradoxes that plague other solutions (I discuss this issue in Chapter Eight). For

[42] 'Dp' is defined as '$p \wedge \sim(p \rightarrow \sim p)$'; Field (2008a: 236).
[43] Field (2008a: chs. 15, 17, and 22).

Field, the mistake in the reasoning that leads to the liar paradox is assuming that the law of excluded middle holds of truth claims in general.

Although Field's theory is new, there is already some secondary literature on it.[44] Some problems are that the conditional is still counter-intuitive especially when it is embedded in other conditionals, and Field offers no proof theory for his paracomplete logic. Also, there are problems with how to interpret the determinateness operators. His theory also has difficulties with revenge paradoxes. I argue in Chapters Seven, Eight, and Nine that it faces additional problems.[45]

3.6 Paraconsistent Theories

Paraconsistent theories also accept (T-In) and (T-Out) but reject key principles of classical logic, most notably the rule that anything follows from a contradiction. Paraconsistent theories are dialetheic in that they accept some contradictions, but they are not trivial (i.e., it is not the case that any sentence follows from any set of sentences).[46] Graham Priest is the person most responsible for popularizing them. Paraconsistent theories are so named because they are paired with paraconsistent logics. Priest's preferred logic is called **LP** (after the "logic of paradox"). Like paracomplete logics, basic paraconsistent logics have very weak conditionals; Priest has made a couple of suggestions over the years for introducing better-behaved conditionals into paraconsistent logics.[47]

Jc Beall is another paraconsistent dialetheist and his theory differs a bit from Priest's. Beall is a deflationist and accepts the principle of intersubstitutability, which states that for any sentence p, p

---

[44] See Priest (2005, 2007, 2010), Leitgeb (2007), Rayo and Welch (2007), Scharp (2007a, 2009), Shapiro (2010), Restall (2010), and McGee (2010).
[45] Another paracomplete approach is suggested by Hintikka (1996) by pairing what he calls independence-friendly (IF) logic with (T-In) and (T-Out). There is growing literature on IF logic, but qua approach to the aletheic paradoxes, it has not had much impact, probably because it seems to be inferior to the other paracomplete theories.
[46] For a discussion of trivialism, see Priest (1998, 2000b, 2006b: ch. 3), Azzouni (2003, 2007), and Bueno (2007).
[47] See Priest (2006a).

and 'p is true' are intersubstitutable *salva veritate* (preserving truth value) in any extensional context. Field and several other deflationists accept intersubstitutability as well. However, in a paraconsistent framework, one can accept all the T-sentences without accepting intersubstitutability; indeed, that is Priest's view. On the other hand, in a paracomplete framework, one can accept intersubstitutability without accepting all the T-sentences. Beall accepts the logic **BX**, which is a weak relevance logic (and hence, a paraconsistent logic), and he uses its conditional to formulate the T-sentences that constitute his disquotational theory of truth.[48]

Neil Tennant is a third paraconsistent dialetheist, but is almost never classified as such. Tennant accepts the primary aletheic principles, but he does not accept **LP** or **BX**. Instead, Tennant has proposed a completely different kind of logic, **Core**. It is the common core of intuitionistic logic (which differs from classical logic on how it treats negation and the conditional) and relevance logic (which differs from classical logic on how it treats disjunction and the conditional) with a twist. Whereas standard relevance logics reject the rule disjunctive syllogism (i.e., $p \vee q$, $\sim p \vdash q$), Tennant keeps it and restricts one of the so called *structural* rules that almost all logics take for granted; in particular, in **Core**, it is not the case that one may always chain two valid arguments together to make a third.[49] This kind of logic is usually called a *substructural logic*.[50] In **Core**, one can prove that the liar sentence is true and one can prove that the liar sentence is not true, but one cannot combine these to derive a contradiction.[51]

I distinguish these three flavors of paraconsistent approaches by using the terms 'opaque', 'transparent', and 'substructural'. Priest's theory is opaque in that it rejects intersubstitutability (for

---

[48] See also Ross Brady's theory in Brady (2006). He uses the logic $\mathbf{DJ}^{\mathbf{d}}$, which is also a weak relevance logic but it differs slightly from **BX**.

[49] That is, he restricts transitivity.

[50] See Restall (2002) for more on substructural logics.

[51] There are some subtlties here that I am slopping over for the sake of brevity. See Tennant (1982, 1995, 1997, forthcoming, MS1, MS2) for details.

example, p and 'p is true' are not intersubstitutable inside a negation), whereas Beall's is transparent because it accepts intersubstitutability. Tennant's is a substructural logic that is unlike any of the others I have considered.

Criticisms of dialetheism are legion, but probably the most common is that it is simply irrational to accept contradictions for any reason. It also has problems with revenge paradoxes (despite the assurances you hear from Priest).[52]

There are two appendices to this chapter. One is a partial list of aletheic principles—principles that, pretheoretically at least, truth seems to obey. The second is a presentation of some logics mentioned in this chapter.

[52] See Priest, Beall, and Armour-Garb (2005) for discussion.

Appendix 1: Aletheic Principles

The following is a list of principles that truth predicates (at least pretheoretically) seem to obey (it is not meant to be exhaustive).

Primary Principles

(T-Out) $T(\langle p\rangle)\rightarrow p$

(T-In) $p \rightarrow T(\langle p\rangle)$

(T-Elim) $T(\langle p\rangle) \vdash p$

(T-Intro) $p \vdash T(\langle p\rangle)$

(~T-Elim) $\sim T(\langle p\rangle) \vdash \sim p$

(~T-Intro) $\sim p \vdash \sim T(\langle p\rangle)$

(T-Entry) $\vdash p \rightarrow \ \vdash T(\langle p\rangle)$

(T-Exit) $\vdash T(\langle p\rangle) \rightarrow \ \vdash p$

Truth-functional Principles

(~-Imb) $\sim T(\langle p\rangle) \rightarrow T(\langle \sim p\rangle)$

(~-Exc) $T(\langle \sim p\rangle) \rightarrow \sim T(\langle p\rangle)$

($\wedge$-Imb) $T(\langle p\rangle)\wedge T(\langle q\rangle) \rightarrow T(\langle p\wedge q\rangle)$

($\wedge$-Exc) $T(\langle p\wedge q\rangle) \rightarrow T(\langle p\rangle)\wedge T(\langle q\rangle)$

($\vee$-Imb) $T(\langle p\rangle)\vee T(\langle q\rangle) \rightarrow T(\langle p\vee q\rangle)$

($\vee$-Exc) $T(\langle p\vee q\rangle) \rightarrow T(\langle p\rangle)\vee T(\langle q\rangle)$

($\rightarrow$-Imb) $(T(\langle p\rangle)\rightarrow T(\langle q\rangle)) \rightarrow T(\langle p\rightarrow q\rangle)$

($\rightarrow$-Exc) $T(\langle p\rightarrow q\rangle) \rightarrow (T(\langle p\rangle)\rightarrow T(\langle q\rangle))$

Quantificational Principles

($\forall$-Imb) $(\forall x)T(\langle \phi(x)\rangle) \rightarrow T(\langle (\forall x)\phi(x)\rangle)$

($\forall$-Exc) $T(\langle (\forall x)\phi(x)\rangle) \rightarrow (\forall x)T(\langle \phi(x)\rangle)$

($\exists$-Imb) $(\exists x)T(\langle \phi(x)\rangle) \rightarrow T(\langle (\exists x)\phi(x)\rangle)$

($\exists$-Exc) $T(\langle (\exists x)\phi(x)\rangle) \rightarrow (\exists x)T(\langle \phi(x)\rangle)$

Implication Principles

(MPC) $(p_1 \wedge \ldots \wedge p_n \rightarrow q) \rightarrow (T(\langle p_1\rangle) \wedge \ldots \wedge T(\langle p_n\rangle) \rightarrow T(\langle q\rangle))$

(SPC) $(p \rightarrow q) \rightarrow (T(\langle p\rangle) \rightarrow T(\langle q\rangle))$

(Sub-in) $p \leftrightarrow q \rightarrow T(\langle p\rangle) \leftrightarrow T(\langle q\rangle)$

(MPT) $(T(\langle p_1 \rangle) \wedge \ldots \wedge T(\langle p_n \rangle) \rightarrow T(\langle q \rangle)) \rightarrow (p_1 \wedge \ldots \wedge p_n \rightarrow q)$

(SPT) $(T(\langle p \rangle) \rightarrow T(\langle q \rangle)) \rightarrow (p \rightarrow q)$

(Sub-out) $T(\langle p \rangle) \leftrightarrow T(\langle q \rangle) \rightarrow (p \rightarrow q)$

Misc. Principles

(Non-Empty) $(\exists x)T(x)$

(Non-Full) $(\exists x)\sim T(x)$

(Taut) $T(\langle \top \rangle)$

(Contra) $\sim T(\langle \bot \rangle)$

(T-Del) $T(\langle T(p) \rangle) \rightarrow T(\langle p \rangle)$

(T-Rep) $T(\langle p \rangle) \rightarrow T(\langle T(\langle p \rangle) \rangle)$

(Reductio) $(T(\langle p \rangle) \rightarrow T(\bot)) \rightarrow T(\langle \sim p \rangle)$

(DN) $T(\langle \sim \sim p \rangle) \rightarrow T(\langle p \rangle)$

(TT) $T(\langle T(\langle p \rangle) \rightarrow p \rangle)$

Appendix 2: Logics

In this appendix, I present all the logics mentioned in this chapter. When possible, I give proof-theoretic formulations axiomatically (i.e., as Hilbert systems). When helpful, I offer model-theoretic formulations as well. In some cases, natural deduction systems are used instead of Hilbert systems. Occasionally, I use a model-theoretic formulation alone. The reader should look back at Information Boxes 9, 10, and 11 for more information on these particular logics.

Let $\mathfrak{L}$ be a sentential language with $\wedge$, $\vee$, and $\sim$ as logical operators with the usual syntax. In all the logics for $\mathfrak{L}$, one can define a conditional, $\rightarrow$, in the usual way: $A \rightarrow B \equiv_{df} \sim A \vee B$.

Structural rules

Most of the logics presented here obey the following structural rules (where Γ and Γ' are sets of sentences of $\mathfrak{L}$ and A and B are sentences of $\mathfrak{L}$:

(Reflexivity) If $A \in \Gamma$, then $\Gamma \vdash A$

(Weakening) If $\Gamma \vdash A$, then $\Gamma \cup \Gamma' \vdash A$

(Transitivity) If $\Gamma \vdash A$ and $\Gamma, A \vdash B$, then $\Gamma \vdash B$

Notice that the structural rules make no mention of the logical connectives of $\mathfrak{L}$. They are rules that govern the behavior of '$\vdash$' alone.

WFDE

We begin with a very weak logic, WFDE, which consists of the structural rules and the following rules:

($\wedge$-Intro) $A, B \vdash A \wedge B$

($\wedge$-Elim1) $A \wedge B \vdash A$

($\wedge$-Elim2) $A \wedge B \vdash B$

($\vee$-Intro1) $A \vdash A \vee B$

($\vee$-Intro2) $B \vdash A \vee B$

($\sim\wedge$-Intro1) $\sim A \vdash \sim(A \wedge B)$

($\sim\wedge$-Intro2) $\sim B \vdash \sim(A \wedge B)$

($\sim\vee$-Intro) $\sim A, \sim B \vdash \sim(A \vee B)$

($\sim\vee$-Elim1) $\sim(A \vee B) \vdash \sim A$

($\sim\vee$-Elim2) $\sim(A \vee B) \vdash \sim B$

($\sim\sim$-Intro) $A \vdash \sim\sim A$

(~~-Elim) $\sim\sim A \vdash A$

FDE

FDE is a very commonly studied logic. It results from adding the following two *meta-rules* to WFDE:

(~∧-Intro)
$$\frac{\sim A \vdash C \quad \sim B \vdash C}{\sim(A \wedge B) \vdash C}$$

(∨-Intro)
$$\frac{A \vdash C \quad B \vdash C}{A \vee B \vdash C}$$

A popular semantics for FDE uses the set, {1, b, n, 0}, of truth values, where 1 and b are designated. The clauses for the logical connectives can be read off the following tables:

A	~A
1	0
b	b
n	n
0	1

∧	1	b	n	0
1	1	b	n	0
b	b	b	0	0
n	n	0	n	0
0	0	0	0	0

∨	1	b	n	0
1	1	1	1	1
b	1	b	1	b
n	1	1	n	n
0	1	b	n	0

$\mathbf{S_3}$ and $\mathbf{WS_3}$

S_3 results from adding the following rule to FDE:

(UE) $A \wedge \sim A \vdash B \vee \sim B$

Likewise, WS_3 results from adding (UE) to WFDE.

$\mathbf{K_3}$ and $\mathbf{WK_3}$

K_3 is the Strong Kleene Logic. It results from adding the following rule to FDE:

(EFQ) $A \wedge \sim A \vdash B$

Likewise, WK_3 results from adding (EFQ) to WFDE.

A popular semantics for K_3 uses the set, {1, n, 0}, of truth values, where 1 is designated. The clauses for the logical connectives can be read off the tables for FDE. That is, the tables for K_3 result from deleting the rows and columns for the truth value b from the tables for FDE.

LP and WLP

LP is Priest's Logic of Paradox. It results from adding the following axiom to FDE:
(LEM) $\vdash A \vee \sim A$
Likewise, WLP results from adding (LEM) to WFDE.

A popular semantics for LP uses the set, {1, b, 0}, of truth values, where 1 and b are designated. The clauses for the logical connectives can be read off the tables for FDE. That is, the tables for LP result from deleting the rows and columns for the truth value n from the tables for FDE.

CL and **WCL**

CL is classical logic, which results from adding (EFQ) and (LEM) to FDE.
Likewise, WCL (weakly classical logic) results from adding (EFQ) and (LEM) to WFDE.

In all the logics that follow, we break the classical equivalence between $A \rightarrow B$ and $\sim A \vee B$. Many of these logics can be thought of as expansions of the logics above, where a conditional is added to $\mathfrak{L}$. Let $\mathfrak{L}^+$ be a sentential language with $\wedge$, $\vee$, $\sim$, and $\rightarrow$ as logical operators with the usual syntax.

K_4, **N_4**, **K_***, and **N_***

All four of these logics expansions of FDE—they differ from FDE only in the treatment of the conditional. I present them model-theoretically.[53]
Let a model $\mathfrak{M}=\langle W, \mathfrak{R}\rangle$, where W is a set of worlds and $\mathfrak{R}$ is a relation between pairs of sentences of $\mathfrak{L}^+$ and worlds and members of the set of truth values {1, 0}. We use '$|A|_{\mathfrak{R}(w)}$' to denote the truth value assigned to sentence A at world w by $\mathfrak{R}$. We assume that $\mathfrak{R}$ is given for the propositional variables of $\mathfrak{L}^+$, and the logic is specified by a set of clauses that determine $\mathfrak{R}$ for the sentences of $\mathfrak{L}^+$ that contain logical connectives and the definition of validity.
The first two logics agree on the following clauses:
$|\sim A|_{\mathfrak{R}(w)} = 1$ iff $|A|_{\mathfrak{R}(w)} = 0$
$|\sim A|_{\mathfrak{R}(w)} = 0$ iff $|A|_{\mathfrak{R}(w)} = 1$[54]
$|A \wedge B|_{\mathfrak{R}(w)} = 1$ iff $|A|_{\mathfrak{R}(w)} = 1$ and $|B|_{\mathfrak{R}(w)} = 1$
$|A \wedge B|_{\mathfrak{R}(w)} = 0$ iff $|A|_{\mathfrak{R}(w)} = 0$ or $|B|_{\mathfrak{R}(w)} = 0$
$|A \vee B|_{\mathfrak{R}(w)} = 1$ iff $|A|_{\mathfrak{R}(w)} = 1$ or $|B|_{\mathfrak{R}(w)} = 1$
$|A \vee B|_{\mathfrak{R}(w)} = 0$ iff $|A|_{\mathfrak{R}(w)} = 0$ or $|B|_{\mathfrak{R}(w)} = 0$

K_4 is the logic that results when the conditional is obeys the following clauses:
$|A \rightarrow B|_{\mathfrak{R}(w)} = 1$ iff for all $w' \in W$, if $|A|_{\mathfrak{R}(w')} = 1$ then $|B|_{\mathfrak{R}(w')} = 1$
$|A \rightarrow B|_{\mathfrak{R}(w)} = 0$ iff for all $w' \in W$, if $|A|_{\mathfrak{R}(w')} = 1$ and $|B|_{\mathfrak{R}(w')} = 0$
and *validity* is defined as:
$\Gamma \vdash A$ iff for all models $\langle W, \mathfrak{R}\rangle$ and for all $w \in W$, if for all $B \in \Gamma$, $|B|_{\mathfrak{R}(w)} = 1$, then $|A|_{\mathfrak{R}(w)} = 1$.

[53] See Priest (2001: ch. 9).
[54] Remember that R is a relation, not a function, so these clauses need to be given independently for 1 and 0.

N_4 is just like K_4, except that the models have "non-normal" worlds. That is, let a model $\mathfrak{M}=\langle W, N, \mathfrak{R}\rangle$, where W and $\mathfrak{R}$ are as above and $N \subseteq W$ (N is the set of normal worlds). Here, $\mathfrak{R}$ is given for all the propositional variables of L+ and all the conditionals at non-normal worlds (i.e., $w \notin N$). The clauses for the conditional are the same, and validity is defined as:
$\Gamma \vdash A$ iff for all models $\langle W, N, \mathfrak{R}\rangle$ and for all $w \in N$, if for all $B \in \Gamma$, $|B|_{\mathfrak{R}(w)} = 1$, then $|A|_{\mathfrak{R}(w)} = 1$.

For the next two logics, let a model $\mathfrak{M}=\langle W, *, \mathfrak{I}\rangle$, where W is a set of worlds, * is a function for worlds to worlds such that $w^{**}=w$, and $\mathfrak{I}$ is a function from pairs of sentences of $\mathfrak{L}^+$ and worlds to truth values {1, 0}. We assume that $\mathfrak{I}$ is given for the propositional variables of $\mathfrak{L}^+$, and the logic is specified by a set of clauses that determine $\mathfrak{I}$ for the sentences of $\mathfrak{L}^+$ that contain logical connectives and the definition of validity. The next two logics agree on the following clauses:
$|\sim A|_{\mathfrak{I}(w)} = 1$ iff $|A|_{\mathfrak{I}(w^*)} = 0$
$|A \wedge B|_{\mathfrak{I}(w)} = 1$ iff $|A|_{\mathfrak{I}(w)} = 1$ and $|B|_{\mathfrak{I}(w)} = 1$
$|A \vee B|_{\mathfrak{I}(w)} = 1$ iff $|A|_{\mathfrak{I}(w)} = 1$ or $|B|_{\mathfrak{I}(w)} = 1$

K_* is the logic that results from adding the following clause for the conditional and the definition of validity for K_4 (above).
$|A \rightarrow B|_{\mathfrak{I}(w)} = 1$ iff for all $w' \in W$, if $|A|_{\mathfrak{I}(w')} = 1$ then $|B|_{\mathfrak{I}(w')} = 1$.

N_* is the logic that results from allowing non-normal worlds (as above), the clause for the conditional as in K_*, and the definition of validity as in N_4.

It should be obvious that N_4 is a sublogic of K_4, and that N_* is a sublogic of K_*.

B

B is the weakest mainstream relevance logic. It consists of the following axioms and rules:

(MP)	$A, A \rightarrow B \vdash B$
(ADJ)	$A, B \vdash A \wedge B$
(PRE)	$A \rightarrow B \vdash (C \rightarrow A) \rightarrow (C \rightarrow B)$
(SUF)	$A \rightarrow B \vdash (B \rightarrow C) \rightarrow (A \rightarrow C)$
(CON)	$A \rightarrow \sim B \vdash B \rightarrow \sim A$
(ID)	$\vdash A \rightarrow A$
($\wedge$-Elim1)	$\vdash (A \wedge B) \rightarrow A$
($\wedge$-Elim2)	$\vdash (A \wedge B) \rightarrow B$
($\vee$-Intro1)	$\vdash A \rightarrow (A \vee B)$
($\vee$-Intro2)	$\vdash B \rightarrow (A \vee B)$
(DN)	$\vdash \sim\sim A \rightarrow A$
(Dist)	$\vdash A \wedge (B \vee C) \rightarrow ((A \wedge B) \vee (A \wedge C))$
(M1)	$\vdash ((A \rightarrow B) \wedge (A \rightarrow C)) \rightarrow (A \rightarrow (B \wedge C))$

(M2) $\vdash ((A\rightarrow C)\wedge(B\rightarrow C))\rightarrow((A\vee B)\rightarrow C)$

As for semantics for B, let a model $\mathfrak{M}=\langle W, N, R, *, \mathfrak{I}\rangle$ where W is a set of worlds, $N\subseteq W$, R is a ternary relation on W, * is a binary relation on W such that $w^{**}=w$, and $\mathfrak{I}$ is a function from pairs of sentences of $\mathfrak{L}^+$ and worlds to the set of truth values $\{1, 0\}$.

We assume that $\mathfrak{I}$ is given for the propositional variables of $\mathfrak{L}^+$, and the logic is specified by a set of clauses that determine $\mathfrak{I}$ for the sentences of $\mathfrak{L}^+$ that contain logical connectives and the definition of validity. The following clauses extend $\mathfrak{I}$:
$|\sim A|_{\mathfrak{I}(w)}=1$ iff $|A|_{\mathfrak{I}(w^*)}=0$
$|A\wedge B|_{\mathfrak{I}(w)}=1$ iff $|A|_{\mathfrak{I}(w)}=1$ and $|B|_{\mathfrak{I}(w)}=1$
$|A\vee B|_{\mathfrak{I}(w)}=1$ iff $|A|_{\mathfrak{I}(w)}=1$ or $|B|_{\mathfrak{I}(w)}=1$
$|A\rightarrow B|_{\mathfrak{I}(w)}=1$ iff $w\in N$ and for all $w'\in W$, if $|A|_{\mathfrak{I}(w')}=1$ then $|B|_{\mathfrak{I}(w')}=1$.
$|A\rightarrow B|_{\mathfrak{I}(w)}=1$ iff $w\notin N$ and for all $w', w''\in W$, if $Rww'w''$ $|A|_{\mathfrak{I}(w')}=1$ then $|B|_{\mathfrak{I}(w'')}=1$.
Validity is defined as:
$\Gamma\vdash A$ iff for all models $\langle W, N, R, *, \mathfrak{I}\rangle$ and for all $w\in N$, if for all $B\in\Gamma$, $|B|_{\mathfrak{I}(w)}=1$, then $|A|_{\mathfrak{I}(w)}=1$.

BX, **DW**, **DJ**$^{\mathbf{d}}$, **TW**, **RW**, **T**, **E**, and **R**

All eight of these logics are mainstream relevance logics, and they are extensions of B.

BX results from adding (LEM) to B.
DW results from adding the contraposition axiom to B:
(Con) $\vdash (A\rightarrow\sim B)\rightarrow(B\rightarrow\sim A)$

DJ^d results from adding the transitivity rule and a special metarule to DW:
(Trans) $\vdash ((A\rightarrow B)\wedge(B\rightarrow C))\rightarrow(A\rightarrow C)$

(Special) $\dfrac{A\vdash B}{C\vee A\vdash C\vee B}$

TW results from adding axioms for prefixing and suffixing to DW:
(Pre) $\vdash (A\rightarrow B)\rightarrow((C\rightarrow A)\rightarrow(C\rightarrow B))$
(Suf) $\vdash (A\rightarrow B)\rightarrow((B\rightarrow C)\rightarrow(A\rightarrow C))$

RW results from adding an axiom corresponding to modus ponens to TW:
(IMP) $\vdash A\rightarrow((A\rightarrow B)\rightarrow B)$

T (called the Logic of "Ticket Entailment") results from adding the following axioms to TW:
(Con) $\vdash (A\rightarrow(A\rightarrow B))\rightarrow(A\rightarrow B)$

E (called the Logic of Entailment) results from adding the following two axioms to T:

(E) $\vdash ((A\rightarrow A)\rightarrow B)\rightarrow B$

R (called the Logic of Relevant Implication) results either from adding (Con) to RW or from adding (IMP) to E.

All of these logics have semantics that are variations on the semantics for B (I omit the details).

KR, **RM**, and **RM$_3$**

These three logics are extensions of R and are often called relevance logics. However, they lack the characteristic feature of relevance logics (i.e., a conditional having the variable-sharing property). Nevertheless, they are weaker than classical logic.

KR results from adding the axiom that corresponds to the ex falso rule to R:

(EFQ) $\vdash A\wedge\sim A\rightarrow B$

RM results from adding what is called the mingle axiom to R:

(Min) $\vdash (A\rightarrow A)\rightarrow A$

RM_3 results from adding the following axiom to RM:

(3) $\vdash A\vee(A\rightarrow B)$

All three of these logics have semantics that are variations on the semantics for B (I omit the details). RM_3, however, has a semantics that is just like the semantics for LP (above) except that that the table for the conditional is:

$\rightarrow$	1	b	0
1	1	0	0
b	1	b	0
0	1	1	1

Thus, RM_3 can be thought of as a three-valued logic.

Ł$_\aleph$ and **Ł$_3$**

Ł$_\aleph$ (pronounced "wook-aleph") is Łukasiewicz's continuum-valued logic. It results from adding the following axioms to RW (above):

(PP) $\vdash A\rightarrow(B\rightarrow A)$

(CD) $\vdash ((A\rightarrow B)\rightarrow B)\rightarrow(A\vee B)$

The semantics for Ł$_\aleph$ take all the real numbers between 0 and 1 to be truth values and 1 to be designated. Let $\mathfrak{I}$ be a function from the sentences of L+ to the set [0, 1] of real numbers. The clauses for the logical connectives are:

$|\sim A|_{\mathfrak{I}} = 1-|A|_{\mathfrak{I}}$

$|A\wedge B|_{\mathfrak{I}} = \text{Min}(|A|_{\mathfrak{I}},|B|_{\mathfrak{I}})$

$|A \vee B|_{\mathfrak{I}} = \text{Max}(|A|_{\mathfrak{I}}, |B|_{\mathfrak{I}})$
$|A \rightarrow B|_{\mathfrak{I}} = 1$ if $|A|_{\mathfrak{I}} \leq |B|_{\mathfrak{I}}$
$|A \rightarrow B|_{\mathfrak{I}} = 1 - (|A|_{\mathfrak{I}} - |B|_{\mathfrak{I}})$ if $|A|_{\mathfrak{I}} > |B|_{\mathfrak{I}}$

$Ł_3$ (pronounced "wook-three") is Łukasiewicz's three-valued logic. Its semantics are the same as above except with {0, .5, 1} as the set of truth values where 1 is designated. A different semantics for $Ł_3$ is just like that for K_3 (above) except the table for the conditional is:

$\rightarrow$	1	n	0
1	1	n	0
n	1	1	n
0	1	1	1

F

F is Field's Logic. It is an expansion of K_3 with a conditional. Its semantics are a bit different than what we have seen so far. Let a model $M=\langle W, F, \mathfrak{I}\rangle$, where W is a set of worlds, F is a function from W to the set of subsets of W, and $\mathfrak{I}$ is a function from pairs of sentences of $\mathfrak{L}^+$ and worlds to the set of truth values {0, .5, 1}, where 1 is designated. Let $@ \in W$ such that for all $X \in F(@)$, $@ \in X$ and $\{@\} \notin F(@)$. Assume that for all $w \in W$ and for all X, $Y \in F(w)$, there is a $Z \in F(w)$ such that $Z \subseteq X \cap Y$.
The clauses for the logical connectives are as follows:
$|\sim A|_{\mathfrak{I}(w)} = 1 - |A|_{\mathfrak{I}(w)}$
$|A \wedge B|_{\mathfrak{I}(w)} = \text{Min}(|A|_{\mathfrak{I}(w)}, |B|_{\mathfrak{I}(w)})$
$|A \vee B|_{\mathfrak{I}(w)} = \text{Max}(|A|_{\mathfrak{I}(w)}, |B|_{\mathfrak{I}(w)})$
$|A \rightarrow B|_{\mathfrak{I}(w)} = 1$ if for some $X \in F(w)$ and for all $w' \in X$, $|A|_{\mathfrak{I}(w')} \leq |B|_{\mathfrak{I}(w')}$
$|A \rightarrow B|_{\mathfrak{I}(w)} = 0$ if for some $X \in F(w)$ and for all $w' \in X$, $|A|_{\mathfrak{I}(w')} > |B|_{\mathfrak{I}(w')}$
$|A \rightarrow B|_{\mathfrak{I}(w)} = .5$ otherwise
Validity is defined as if @ is the only normal world:
$\Gamma \vdash A$ iff for all models $\langle W, F, \mathfrak{I}\rangle$, if for all $B \in \Gamma$, $|B|_{\mathfrak{I}(@)} = 1$, then $|A|_{\mathfrak{I}(@)} = 1$.

M and **I**

M is known as Minimal Logic and I as Intuitionistic Logic.

M has the following rules and axioms:

(MP) $A, A \rightarrow B \vdash B$
($\wedge$-Elim1) $\vdash (A \wedge B) \rightarrow A$
($\wedge$-Elim2) $\vdash (A \wedge B) \rightarrow B$
($\vee$-Intro1) $\vdash A \rightarrow (A \vee B)$
($\vee$-Intro2) $\vdash B \rightarrow (A \vee B)$
(ADJ) $\vdash A \rightarrow (B \rightarrow (A \wedge B))$
(PP) $\vdash A \rightarrow (B \rightarrow A)$

(M3) $\vdash ((A\rightarrow C)\rightarrow((B\rightarrow C))\rightarrow((A\vee B)\rightarrow C))$

(I) $\vdash (A\rightarrow(B\rightarrow C))\rightarrow((A\rightarrow B)\rightarrow(A\rightarrow C))$

I results from adding (EFQ) to M.

Core

Core is Tennant's Logic that incorporates aspects of relevance logic and intutionistic logic; however, unlike other relevance logics, disjunctive syllogism:

(DS) $A, \sim A\vee B \vdash B$

is valid in Core. Moreover, all three structural rules (above) are invalid in Core. Below is the natural deduction formulation of Core:

(¬-I)
$$\dfrac{\begin{array}{c}\Box\!\!-\!\!-\!\!-(i)\\ \varphi\\ \vdots\\ \bot\end{array}}{\neg\varphi}(i)$$

(¬-E)
$$\dfrac{\neg\varphi \qquad \begin{array}{c}\vdots\\ \varphi\end{array}}{\bot}$$

(∧-I)
$$\dfrac{\begin{array}{c}\vdots\\ \varphi\end{array} \qquad \begin{array}{c}\vdots\\ \psi\end{array}}{\varphi\wedge\psi}$$

(∧-E)
$$\dfrac{\varphi\wedge\psi \qquad \begin{array}{c}(i)\!\!-\!\!-\!\!\Box\!\!-\!\!-(i)\\ \underbrace{\varphi\,,\psi}\\ \vdots\\ \theta\end{array}}{\theta}(i)$$

(∨-I)

$$\frac{\begin{array}{c}\vdots\\ \varphi\end{array}}{\varphi \vee \psi} \qquad \frac{\begin{array}{c}\vdots\\ \psi\end{array}}{\varphi \vee \psi}$$

(∨-E)

$$\frac{\varphi \vee \psi \quad \begin{array}{c}\Box\!\!-\!\!(i)\\ \varphi\\ \vdots\\ \theta\end{array} \quad \begin{array}{c}\Box\!\!-\!\!(i)\\ \psi\\ \vdots\\ \theta\end{array}}{\theta}(i) \qquad \frac{\varphi \vee \psi \quad \begin{array}{c}\Box\!\!-\!\!(i)\\ \varphi\\ \vdots\\ \bot\end{array} \quad \begin{array}{c}\Box\!\!-\!\!(i)\\ \psi\\ \vdots\\ \theta\end{array}}{\theta}(i) \qquad \frac{\varphi \vee \psi \quad \begin{array}{c}\Box\!\!-\!\!(i)\\ \varphi\\ \vdots\\ \theta\end{array} \quad \begin{array}{c}\Box\!\!-\!\!(i)\\ \psi\\ \vdots\\ \bot\end{array}}{\theta}(i)$$

(→-I)

$$\frac{\begin{array}{c}\Box\!\!-\!\!(i)\\ \varphi\\ \vdots\\ \bot\end{array}}{\varphi \to \psi}(i) \qquad \frac{\begin{array}{c}\Diamond\!\!-\!\!(i)\\ \varphi\\ \vdots\\ \psi\end{array}}{\varphi \to \psi}(i)$$

(→-E)

$$\frac{\varphi \to \psi \quad \varphi \quad \begin{array}{c}\Box\!\!-\!\!(i)\\ \psi\\ \vdots\\ \theta\end{array}}{\theta}(i)$$

Casual readers should not expect to understand this formulation—it is included for the experts.

Chapter 4

Unified Theories of Truth

The first three chapters presented views on the nature of truth, philosophical approaches to the aletheic paradoxes, and logical approaches to the aletheic paradoxes. In this chapter we assemble these parts into more comprehensive wholes. First, I consider combinations of philosophical approaches and logical approaches to the paradoxes. Then I turn to what I call unified theories of truth, which incorporate all three elements.

4.1 Combinations of Philosophical and Logical Approaches

Significant features of the presentation in this work include the distinction between descriptive and prescriptive projects and the distinction between philosophical and logical approaches to the paradoxes. These are often ignored, which makes the space of alternatives confusing and muddled. I organized the presentation around philosophical vs. logical approaches because I think this distinction is blurred more often and to worse effect. One downside to this choice is that it masks important connections between certain philosophical approaches and certain logical approaches that are designed to work together. As part of the remedy, I want to offer a view on the relation between these kinds of approaches and mention some of the portions of logical space that have been spoken for. Information Boxes 19 and 20 contain overviews of the philosophical approaches and the logical approaches.

Philosophical Approaches | Information Box 19

Grammaticality

(Jorgensen, Kattsoff)

Meaningfulness

(Ushenko, Grover, Brandom)

Assertibility

(Martinich, Goldstein, Kearns)

Intensionality

(Skyrms)

Epistemicism

(Horwich)

Ambiguity

(Orthodox, Williamson)

Context Dependence

(Parsons, Burge, Barwise & Etchemendy, Gaifman, Simmons, Glanzberg)

Circularity

(Gupta, Herzberger, Belnap, Yaqūb, Kremer)

Indeterminacy

(van Fraassen, Kripke, McGee, Soames, Maudlin, Field, Tennant)

Inconsistency

(Chihara, Priest, Yablo, Eklund, Beall, Patterson, Ludwig, Scharp)

Logical Approaches | Information Box 20

Classical Glut

Classical Gap

(Orthodox, Burge, Parsons, Feferman, Barwise & Etchemendy, Simmons, Cantini, Glanzberg, Maudlin)

Classical Symmetric

(Friedman & Sheard, McGee, Hofweber, Scharp)

Weakly Classical

(van Fraassen, Herzberger, Gupta & Belnap, Yaqūb, Yablo, Kremer, L. Shapiro)

Paracomplete

(Kripke, Soames, Field)

Paraconsistent

(Priest, Beall, Tennant)

4.1.1 Measurement Theory

It seems to me that a good way to understand the relation between philosophical and logical approaches is in terms of measurement theory. Measurement theory is the study of how formal systems apply to the physical world; as I said, I like to think of it as somewhat analogous to set theory, but for science—it serves as an all-purpose background theory for science in the way that set theory serves as an all-purpose background theory for mathematics. Patrick Suppes, who has probably done more than anyone to develop measurement theory, writes:

> A procedure of measurement is needed in any area of science when we desire to pass from simple qualitative observations to the quantitative observations necessary for the precise prediction or control of phenomena. To justify this transition we need an algebra of empirically realizable operations and relations which can be shown to be isomorphic to an appropriately chosen numerical algebra. Satisfying this requirement is the fundamental problem of measurement representation.[1]

Any time we use mathematics to describe, explain, predict, or control the physical world, we are implicitly using measurement theory. It gives a single coherent method of applying mathematics and formal theories to empirical phenomena.

According to measurement theory, the process of measurement involves three structures:

(i) A *physical structure*, which consists of physical entities, their properties, and relations.

(ii) A *relational structure*, which consists of a set of (idealized) objects defined by principles specifying their properties and relations.

(iii) A *mathematical structure*, which consists of a set of mathematical entities with mathematical properties and relations.

In addition to these three structures, there is a connection between the physical structure and the relational structure, and there is a connection between the relational structure and the mathematical structure. I use the term 'measurement system' for something consisting of these three structures and the connections between them. *Measurement* is a two-step process: (i) specifying the connection between the physical objects and physical relations in the *physical structure* to the idealized objects and relations defined by the axioms of the idealizing *relational structure*, and (ii) constructing mathematical functions to connect the relations of the idealizing *relational structure* to relations of the *mathematical structure* so that the mathematical entities can represent the properties of the physical objects via the properties of the idealized objects.

Consider, as an example, the measurement system for length. In this case, the physical structure might be a group of straight rigid rods. The relational structure is a set whose members represent

---

[1] Suppes (1998: 244).

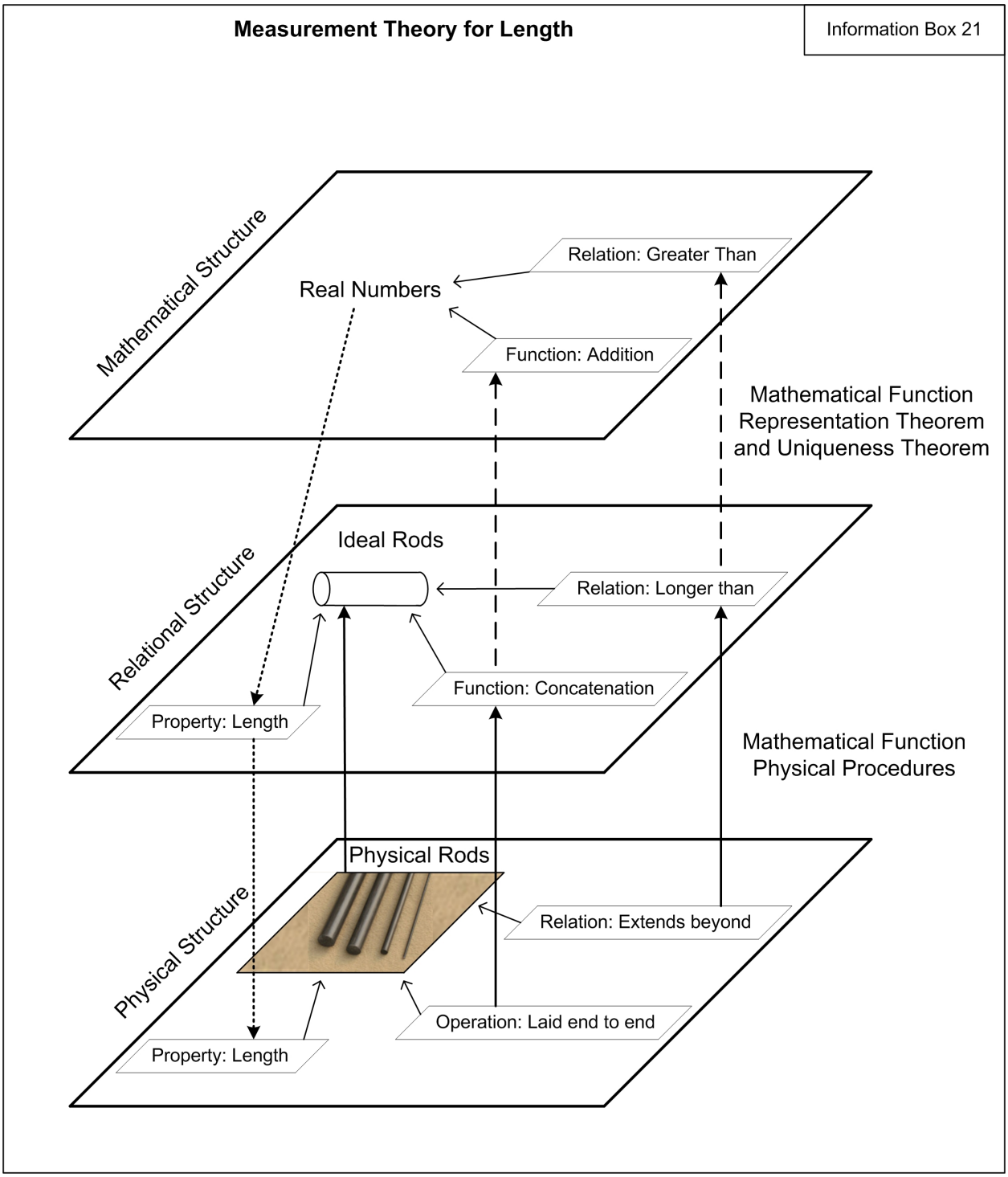
the rods in the physical structure, and the relations 'longer than', and 'concatenation' (i.e., the result of laying two rods end to end). The axioms defining the relational structure include "longer than' is transitive', 'concatenation is associative', 'if A is longer than B, then the concatenation of A and C is longer than B'. The mathematical structure is the set of real numbers with the addition function and the greater than relation.

The connection between the physical structure of rods and the relational structure of idealized rods is as follows. Each real rod is assigned an ideal rod. One lines up rods A and B and determines whether A extends beyond B or B extends beyond A to determine whether the relation 'A is longer than B' holds in the relational structure between the ideal rods assigned to A and to B. One arranges rods A and B end-to-end to arrive at the concatenation of A and B, which is assigned a new ideal rod. One lines up rod A with the concatenation of B and C to determine whether the relation 'A is longer than the concatenation of B and C' holds in the relational structure between the ideal rods assigned to A and to the concatenation of B and C.[2]

The connection between the relational structure and the mathematical structure consists of mathematical functions. For example, a function ϕ assigns real numbers to the ideal rods in such a way that if A is longer than B then $\phi(A) > \phi(B)$, and ϕ (the concatenation of A and B) = $\phi(A) + \phi(B)$. The function ϕ assigns numbers to the members of the relational structure so that the "greater than" relation can represent 'longer than' and 'the concatenation of' can be represented by addition. In the case of length, numbers are assigned to the physical rods as their lengths by combining the connection between the physical rods and the ideal rods with the connection between the ideal rods and the real numbers. The real number assigned to an ideal rod represents the length of whatever

[2] Note that there are more ideal rods than physical rods—there is an ideal rod for the concatenation of any number of physical rods.

the ideal rod models—either a physical rod or some concatenation of physical rods. Information Box 21 has a diagram of the entire measurement system for length.



There are two crucial results for any measurement system: a representation theorem and a uniqueness theorem. A *representation theorem* says that there is a function from the relational structure to the mathematical structure that preserves the relations between the elements of the relational structure; e.g., a function ϕ from the relational structure of rods, 'longer than', and concatenation to the real numbers so that rod x is longer than rod y iff[3] $\phi(x) > \phi(y)$, and rod z is the concatenation of rods x and y iff $\phi(z) = \phi(x) + \phi(y)$. A *uniqueness theorem* specifies the class of relation-preserving functions from the relational structure to the mathematical structure; e.g., if ϕ is one such function, then $\phi' = \alpha\phi$ is another ($\alpha>0$).[4]

4.1.2 Combination Approaches as Measurement Systems for Truth

Philosophical approaches to the aletheic paradoxes do two things: they tell us something about the truth predicate that is relevant to solving the aletheic paradoxes (i.e., to pursuing one or more of the projects and solving one or more of the problems) and they tell us something about the paradoxical truth bearers and the paradoxical reasoning.

Logical Approaches specify principles truth predicates obey and logics that are compatible with these principles. The theories of truth offered by logical approaches apply to certain artificial languages and these theorists use techniques from mathematical logic to investigate the properties of these theories and to prove things about them (e.g., consistency relative to a background mathematical theory).

---

[3] I use 'iff' as short for 'if and only if' from here on.

[4] For more information on measurement theory, see Suppes (1998) and Narens (2007) for an introduction and see Suppes et al (1971, 1989, 1990) and Narens (2002) for more advanced topics. See also Suppes (2002), which applies the framework of measurement theory to many scientific topics.

In measurement-theoretic terms, the first few chapters have been dedicated to investigating: (i) a physical structure—our practice of using the concept of truth, which includes how we use truth predicates of natural languages, (ii) relational structures—various precise principles truth predicates obey and other relevant principles (e.g., logical principles), which are studied for artificial languages, and (iii) mathematical structures—the mathematical models for artificial languages, truth predicates, principles of truth, and logics. Once we have this structure in view, it becomes clear that *philosophical approaches* to the paradoxes state conditions on relations between the physical structure and the relational structure; occasionally, they contain full-on specifications of the connection, other times we are left to guess. Either way, they tell us something about items in the physical structure (e.g., the syntactic, semantic, and pragmatic features of the truth predicate and the sentences that contain it) and its connection to the relational structure (e.g., how idealized truth predicates work in artificial languages that are meant to model natural languages). It also becomes clear that logical approaches specify (at least partial) relational structures, mathematical structures, and connections between them. Thus, the link between philosophical approaches and logical approaches is the link between the Physical – Relational connection and the Relational – Mathematical connection. Furthermore, a combination approach to the aletheic paradoxes would be a full-on measurement system for truth.[5]

Finally, recall the distinction between the descriptive project (i.e., a theory of our aletheic practice as it is) and the prescriptive project (i.e., a theory of what our aletheic practice should be). Depending on one's diagnosis of the paradoxes, one might need two theories—a complete measurement theory for truth as we currently use it (i.e., a descriptive project) and a complete measurement theory for truth as we should use it (or related notions) (i.e., a prescriptive project). This is exactly what I offer in Part III of the book.

---

[5] I do not think these ideas are new, but I have not seen this literature described in quite this way before.

4.1.3 Examples of Combinations

Take Field's combination approach for example. He specifies a class of artificial languages that contain their own truth predicates and obey the intersubstitutability principle (which says that substituting p for 'p is true' or vice versa in extensional contexts preserves truth value), but the theory of the logical terms of this language is paracomplete—not all instances of the law of excluded middle are valid. However, many familiar logical principles are valid (e.g., double negation elimination, modus ponens, etc.). These artificial languages, these principles for truth, and these logical principles together constitute the relational structure. Field presents a mathematical structure (actually, he presents several—a neighborhood structure, an algebraic structure, and a topological structure), which is to serve as a model for his relational structure.[6] He claims that natural language truth predicates display indeterminacy in the sense that one ought not accept or assert all the instances of excluded middle involving truth predicates. In particular, excluded middle fails for ungrounded sentences of natural languages. However, all the T-sentences for sentences of natural language are true (at least when 'if…then…' constructions are treated as expressing Field's conditional). Paradoxical sentences are indeterminate, which means, for Field, that they are neither determinately true nor determinately not true. Indeed, Field explains 'determinately' in these natural language sentences by appeal to the determinateness operator defined for the relational structure. The relational structure models the physical structure and helps us explain many important features of the physical structure (e.g., what semantic features truth predicates have, what is wrong with the reasoning in the aletheic paradoxes, what status paradoxical sentences have); the mathematical

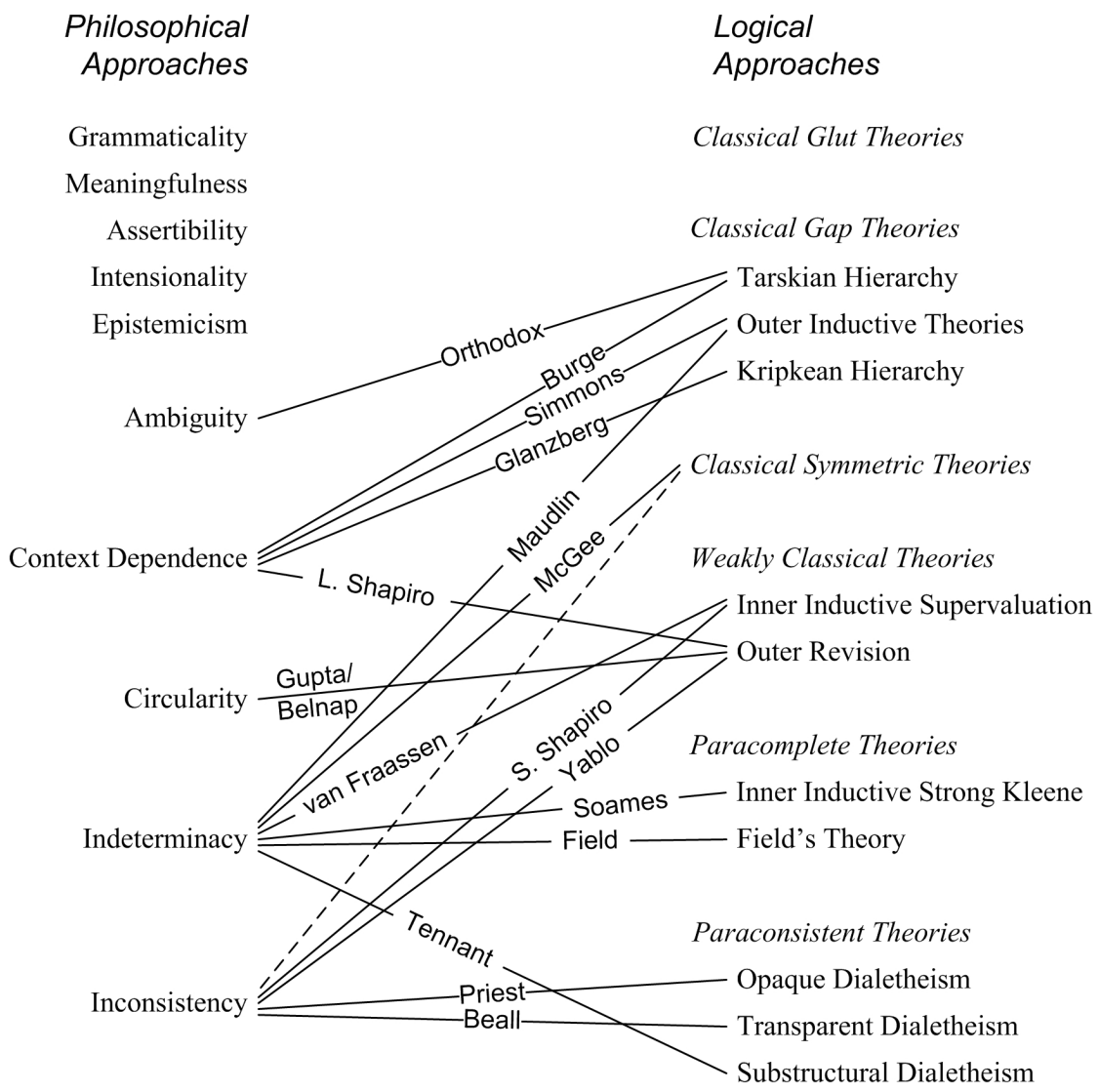
---

[6] Field (2008a: 259-266). If Field had offered a proof theory for his logic (he only gives a model theory), then he could have proven several theorems about how deduction in the artificial language (relational structure) and validity (mathematical structure) relate: a soundness theorem shows that anything provable in the theory of truth and logic is valid in the mathematical structure, and a completeness theorem shows that anything valid in the mathematical structure is provable in the theory of truth and logic. That is, soundness theorems and completeness theorems are just special cases of representation theorems and uniqueness theorems.

structure models the relational structure and helps explain many important features of the relational structure (e.g., the consistency of the theory of truth, the soundness and completeness of the logic). As a team, Field's philosophical and logical approaches to the paradoxes specify (or at least suggest) a total measurement system for truth. Anyone else who defends a combination approach can be interpreted measurement-theoretically in the same way.

Describing each combination would be tedious and not add much to this chapter. Instead, I have depicted them in Information Box 22, which contains a diagram showing philosophical approaches on one side and logical approaches on the other, with important connections between them and the theorists who work on them. The broken line indicates my view, which is developed in Part III.

Philosophical and Logical Approaches Information Box 22



There are several trends to notice. First, most of the first few philosophical approaches have no connections to the logical approaches—that is because the first few try to find some problem with paradoxical truth bearers, whether it is that they are syntactically defective, semantically defective, or pragmatically defective. Approaches like these do not need logical approaches to handle paradoxical sentences because they imply that there is no problem to be handled.

Second, a single philosophical approach might be paired with distinct logical approaches. For example, Field, Tim Maudlin, and Vann McGee champion the indeterminacy approach. Field pairs it with a paracomplete approach, McGee with a classical symmetric approach[7], and Maudlin with a classical gap approach.[8] In each of these combinations, the notion of indeterminacy is interpreted differently. The same point holds of context-dependence views as well.

Third, a single logical approach might be paired with distinct philosophical approaches. For example, the orthodox approach pairs a Tarskian hierarchy (a classical gap approach) with an ambiguity approach—it interprets 'true' as ambiguous so that it can have the meaning of any one of the predicates in the Tarskian hierarchy. Tyler Burge uses the same logical approach (i.e., the Tarskian hierarchy), but he pairs it with a context-dependence philosophical view.[9] He claims that 'true' is an indexical, which has an invariant meaning (i.e., character) and variable content; in any given context, 'true' can have the content of any of the predicates in the Tarskian hierarchy. These are very different interpretations (i.e., ambiguity and context dependence) of the same formal structure (i.e., the Tarskian hierarchy). Another example is the revision theory, which is a weakly classical approach and was initially designed by Gupta and Belnap to be paired with a circularity approach (a philosophical approach).[10] However, Lionel Shapiro suggests that the context dependence approach is a better fit for revision theories, while Stephen Yablo combines a revision theory with an inconsistency view.[11] Again we have several distinct philosophical interpretations of the same mathematical structure.

Fourth, there is a trade off between thinking of natural language truth predicates as univocal and invariant (having a single meaning across uses) and accepting classical logic. Philosophical

---

[7] McGee (1991).
[8] Maudlin (2004).
[9] Burge (1979a, 1982a, 1982b).
[10] Gupta and Belnap (1993).
[11] L. Shapiro (2006); Yablo (1993a, 1993b).

approaches that posit multiple contents for truth predicates because of ambiguity or context dependence tend to go with classical or weakly classical logical approaches; these theorists retain classical logic by *fragmenting* the concept of truth. On the other hand, indeterminacy approaches and inconsistency approaches tend to get paired with weakly classical or non-classical logical approaches; these theorists prize the unity of truth, but pay for it by losing cherished logical principles.

Fifth, even if we ignore the variations internal to each kind of logical approach (so we have six logical approaches) and we set aside the five "dismissive" philosophical approaches at the top of the diagram (so we have five philosophical approaches), that gives us thirty possible combinations, of which only about a third are spoken for. At least with respect to combinations of approaches to the paradoxes, we have a sparsely populated logical space; or, to accentuate the positive: there are many opportunities for new work in this area.

4.2 Unified Projects

Recall from Chapter One that we can classify many theories of the nature of truth along two dimensions, first by that in terms of which they explain truth, and second by the strength of the explanation. The options are Correspondence, Coherence, Pragmatic, Epistemic, Deflationary, Modest, and Pluralist (see Information Box 23 for a summary).

Theories of the Nature of Truth Information Box 23

Correspondence
(Russell, Austin, Field (1972), Armstrong, Devitt, Candlish, Dodd, Fumerton, Vision, Englebretsen)

Coherence
(Joachim, Blanchard, Young)

Pragmatic
(Peirce, James, Dewey, Rorty)

Epistemic
(Dummett, Habermas, Putnam, Tennant, Alcoff, Misak)

Deflationist
(Ramsey, Ayer, Strawson, Quine, Grover, Camp, Belnap, Leeds, Williams, Brandom, Horwich, Kraut, McGee, Field (1994), Soames, Hill, Halbach, Beall, Ebbs)

Modest
(Carnap, Kneale, Mackie, Gupta & Belnap, Künne)

Pluralist
(Wright, McGrath, Lynch, Kölbel)

Non-Reductive
(Davidson, Patterson, Scharp)

Each of these offers a central biconditional:

x is true iff F(x)

Künne gave us five strengths of the central biconditional offered by a theory of truth:

(i) Same sense – analytic biconditional

(ii) Same intension and self-evident – self-evident biconditional

(iii) Same intension and apriori – apriori biconditional

(iv) Same intension – necessary biconditional

(v) Same extension – true biconditional

We can classify theories of the nature of truth along these two lines (not all fit, e.g., Davidson). As I mentioned, Correspondence, Coherence, Pragmatic, Epistemic and Modest all offer general central biconditionals (universally quantified). Deflationism offers a central biconditional for each truth bearer instead of a general one. Pluralism offers as many biconditionals as there are truth properties. Correspondence, Coherence, Pragmatic, Epistemic and Modest all offer analytic biconditionals. In section 1.10, I considered some alternatives to these. Deflationism offers at least necessary biconditionals—it does not seem that a deflationist could accept that there is a possible world where some truth bearer p is true but the claim that p is true is false (or vice versa). Pluralism certainly does not offer self-evident biconditionals, and it seems to me that they are probably not even apriori—one could find out, from empirical evidence, that a certain discourse has a certain truth property. Anyone who accepts the T-biconditionals accepts that the central biconditional for each truth bearer is at least true.

I suggested in Chapter Two that there were several projects one might engage in when addressing the liar paradox, which include the descriptive project (describing our current aletheic practice) and the prescriptive project (describing how our aletheic practice should be). There is also a philosophical aspect and a logical aspect to any project. I also claimed in the last section that a combined philosophical and logical approach would be a complete measurement theory for truth, and that one might need two of these—one for the descriptive project and one for the prescriptive project.

The descriptive/prescriptive distinction is going to affect one's views on the nature of truth—it is entirely possible to pair a descriptive approach to the paradoxes with a view on the nature of truth

(as it is currently used) and pair a prescriptive approach to the paradoxes with a different view on the nature of truth (as it should be used). Each view on the nature of truth can be thought of as a combination of a central biconditional(s) plus its status.

When we put all these pieces together, we arrive at the following idea. Let us call a *unified theory of truth* a combination of a view on the nature of truth together with a combined approach to the paradoxes. That could involve: (i) a descriptive central biconditional(s) (or an alternative to them), (ii) the reading of the descriptive central biconditional (or its alternative), (iii) a philosophical approach to the descriptive project, (iv) a logical approach to the descriptive project, (v) a prescriptive central biconditional (or alternative to them), (iv) the reading of the prescriptive central biconditional (or its alternative), (vii) a philosophical approach to the prescriptive project, and (viii) a logical approach to the prescriptive project. It might be that a unified theory of truth says that we need not change our practice in any way, in which case it will not have items (v)-(viii).

Instead of focusing on the central biconditional, one could adopt Davidson's view of truth and combine it with a philosophical approach to the paradoxes and a logical approach to the paradoxes. I presented Davidson's theory of truth and suggested that it could be cast in terms of measurement theory in Chapter One. I characterized the philosophical/logical distinction between approaches to the paradoxes in terms of measurement theory in the first section of this chapter. It should not come as a surprise that these two components could be combined. The combination would take a measurement system for truth (as described above), which includes a philosophical and logical approach to the paradoxes, and embed it in a broader Davidsonian theory of rationality. The theory of rationality would ideally be formulated as a measurement system as well. The overall theory would combine Davidson's view on the nature of truth with a philosophical and a logical approach to the paradoxes. No one has yet attempted such a thing, but this is exactly what I do in Chapters Thirteen, Fourteen, and Fifteen.

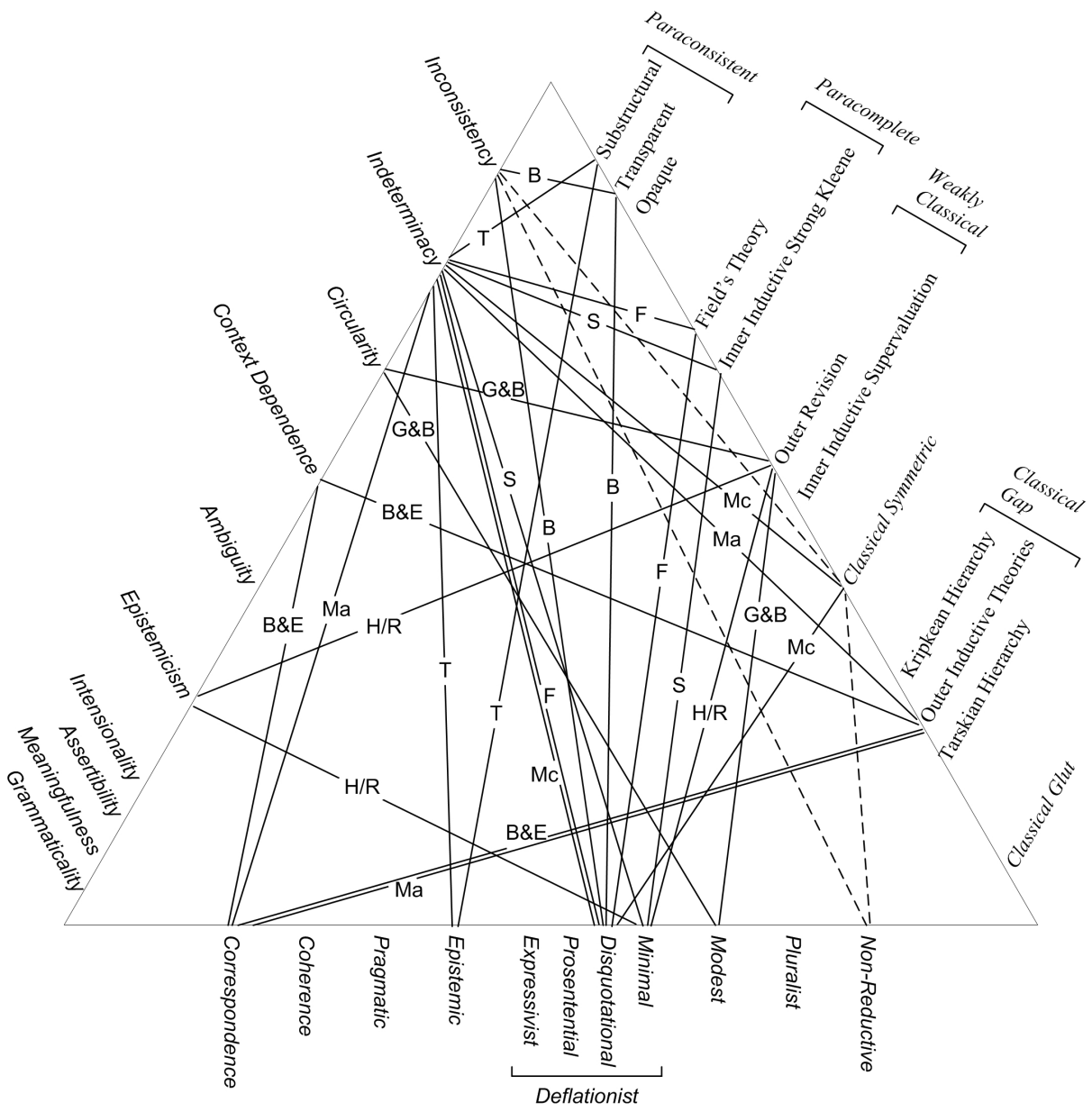
4.3 Unified Theories of Truth

We have seen combinations of philosophical and logical approaches to the paradoxes and been reminded of the views on the nature of truth. Information Box 24 displays the connections between the philosophical approaches to the paradoxes (left), the logical approaches to the paradoxes (right), and the theories of the nature of truth (bottom). One thing that jumps immediately off the page is that deflationists are the primary bridge builders between work on the nature of truth and work on the paradoxes. This is probably owing to two things: deflationism has long been thought to have significantly more trouble with the paradoxes than other views on the nature of truth so deflationists have been especially eager to dispel this myth,[12] and deflationism is more focused on the principles truth predicates obey than are other views on the nature of truth. Another feature of this diagram is that only unified theories of truth are represented; i.e., those that offer a combination of views on the nature of truth, a philosophical approach to the liar, and a logical approach to the liar. These unified theories show up as triangles in this diagram.[13] The broken line represents the unified theory of truth I offer in Part III. This section is a brief tour through the work of these authors, organized chronologically.

[12] One can find this view in Dummett (1959) and Simmons (1999); see also Gupta (2005).
[13] Note that 'B&E' and 'G&B' stand for 'Barwise and Etchemendy' and 'Gupta and Belnap', respectively. These pairs of authors worked together on their respective unified theories of truth. 'H/R' stands for 'Horwich / Restall'; these two authors worked independently on the theory of truth in question.

Unified Theories of Truth

Information Box 24



B&E	Barwise & Etchemendy	Ma	Maudlin
G&B	Gupta & Belnap	F	Field
H/R	Horwich / Restall	B	Beall
Mc	McGee	T	Tennant
S	Soames		

4.3.1 Barwise and Etchemendy

Barwise and Etchemendy offer a unified theory that incorporates a correspondence view.[14] They begin by presenting an innovative theory of self-referential propositions based on Peter Aczel's work on non-well-founded sets.[15] This is technically beyond the scope of this chapter, but it treats propositions as sets that can contain themselves in a fully rigorous and consistent way. These propositions are the primary truth bearers in their theory.

Barwise and Etchemendy are decidedly inflationist on the nature of truth and they seem to endorse a kind of correspondence theory. Theirs is an update of J. L. Austin's correspondence theory from the 1950s. Austin claimed that descriptive conventions link sentences to types of situations in the world, and demonstrative conventions link propositions with particular situations in the world. A proposition is true, for Austin, when the particular situation linked to it by demonstrative conventions is of the type linked to the sentence expressing that proposition by descriptive conventions.[16] One consequence of Austin's view is that there is always a contextual element to the proposition expressed by a sentence—any sentence could express distinct propositions in different contexts if the demonstrative conventions differed. Barwise and Etchemendy exploit this feature so that paradoxical sentences express different propositions than they seem to express. Indeed, the situations linked to paradoxical sentences shift around to avoid the paradoxical consequences. The result is that they combine a context-dependence philosophical approach with a classical gap approach for the paradoxes. Sentences containing truth predicates express different propositions in different situations and they allow that (T-In) has exceptions. It

[14] Barwise and Etchemendy (1987).
[15] Aczel (1987). See also Barwise and Moss (2004).
[16] Austin (1950).

seems to me that the correspondence/context-dependence/classical gap combination is a natural one.[17]

4.3.2 Gupta and Belnap

Anil Gupta and Nuel Belnap offer a unified theory of truth that is not deflationary.[18] They do not say much about the nature of truth except that all T-sentences for a language constitute a definition of truth for that language.[19] One could treat this as a kind of modest theory of the nature of truth or one could treat it as a primitive theory that says p and 'p is true' have the same intension (i.e., same truth value in all possible worlds), but denies that one can give an analysis of truth. The latter reading seems harder to square with their text than the former. Either way, they claim that once one takes the T-sentences to define truth, one has to treat truth as a circularly defined concept, which is their philosophical approach to the paradoxes. Their theory of circularly defined concepts, based on revision sequences, yields a weakly classical logical approach. So it seems as though they offer a modesty/circularity/weakly classical combination as their unified theory of truth.[20]

4.3.3 Horwich / Restall

Paul Horwich focuses mostly on the nature of truth—he is responsible for proposing and defending minimalism, which is a deflationary theory that takes propositions as primary truth bearers.[21] However, one can draw conclusions on the basis of some of his other commitments for approaches

---

[17] See Koons (1992), McLarty (1993), and Priest (1993) for discussion.
[18] At least, it is not disquotational or minimalist. They are non-committal on whether it is compatible with prosentential theories, which makes sense because Belnap once defended a prosentential theory; see Grover, Camp, and Belnap (1976).
[19] Gupta and Belnap (1994: ch. 4).
[20] For discussion see P. Kremer (1993), Yaqub (1994), Antonelli (1994, 2000), Chapuis (1996), McGee (1997), Martin (1997), Gupta (1997a, 1997b), Welch (2001), Cook, (2002, 2003), and M. Kremer (2002) for discussion.
[21] Horwich (1998, 1999, 2001a, 2004, 2006, 2008)

to the paradoxes. Both Greg Restall and (independently) Bradley Armour-Garb and Jc Beall suggest that Horwich has committed himself to an epistemicist approach to the paradoxes (a philosophical approach).[22] That is, either (T-In) or (T-Out) fails for paradoxical propositions, but it is impossible to know which one. If this is accurate, then Horwich pairs an epistemicist approach to the paradoxes with a variety of deflationism. Restall argues as well that the logical approach best suited to the minimalist/epistemicist combination is the outer theory of the revision procedure.[23]

4.3.4 McGee

Vann McGee's views on truth are so novel and innovative that they are hard to classify—he thinks that our ordinary conception of truth is hopelessly incoherent and he sets out to propose two concepts that will replace it.[24] This places him squarely in the prescriptive project category. He does not say much that would constitute a descriptive project at all. He defends disquotationalism, but it is unclear whether he thinks it holds of our incoherent everyday notion of truth or whether he thinks that one of his replacement notions is disquotationalist.[25] It seems to me that the answer is both, but I am not confident about that.

As for the replacements, McGee recommends a vague concept of truth (i.e., one that admits of borderline cases) and a supervaluation theory of vagueness. As with most vague concepts, one can distinguish between truth and definite truth, which, together, are to replace our incoherent notion. McGee suggests the outer strong Kleene recursive theory for definite truth and a modified supervaluation theory for the vague concept of truth. That results in a classical symmetric logical approach to the paradoxes. Overall, he seems to offer a disquotational/inconsistency pairing for his

---

[22] See Restall (2005) and Armour-Garb and Beall (2005).
[23] See Gupta (1993), Davidson (1996), Dodd (1997), and Ludwig (2004) for discussion.
[24] McGee (1989, 1991).
[25] McGee (1993, 2000, 2005a, 2006).

descriptive project (without a logical approach, so I am not counting it as a unified theory) and a disquotational/indeterminacy/classical symmetric combination for his prescriptive project.[26]

There are several avenues that McGee does not discuss, most notably a theory of our incoherent everyday notion of truth he seeks to replace. It seems as though he belongs in the inconsistency camp here, but he does not say much beyond this. It should not come as a surprise that the view I end up presenting in Part III owes much to McGee's work; in particular, we agree that our ordinary concept of truth is incoherent (I prefer the term 'inconsistent'), and that it should be replaced. However, we disagree when it comes to the choice of replacements. McGee suggests vague truth and definite truth, whereas I strike out in a wholly different direction.[27]

4.3.5 Soames

Scott Soames deserves the credit for being one of the first philosophers to attempt to integrate the literature on the nature of truth with the literature on the paradoxes for a descriptive project. Soames offers a minimalist theory of truth (a deflationary on that takes propositions as primary), a philosophical approach to the paradoxes based on indeterminacy, and a paracomplete logical approach (based on the logic K_3).[28] Soames follows Kripke in most of this and sees himself as filling in the details of Kripke's views. Soames' paracomplete theory is a strong Kleene theory, but he does not attempt to deal with the problem of a seriously weak conditional. His minimalism/indeterminacy/paracomplete combination seems to do a good job of capturing intuitions.[29]

---

[26] McGee (1991).
[27] See Yablo (1989), Priest (1994a), and Tappenden (1994) for discussion.
[28] Soames (1997, 1999, 2002a, 2002b).
[29] See Gupta (2002), Tappenden (2002), Williamson (2002), and McGrath (2002) for discussion.

4.3.6 Maudlin

Tim Maudlin combines correspondence intuitions about the nature of truth with an indeterminacy (philosophical) approach and a classical gap (logical) approach. Although Maudlin does not explicitly endorse a correspondence theory of truth, he does claim that truth values of aletheic truth bearers (i.e., those containing truth predicates) supervene on the truth values of non-aletheic truth bearers and he offers a procedure for determining exactly how the truth values of aletheic truth bearers are determined.[30] The can easily be parlayed into a theory of truthmakers and truthmaking that would serve as the basis for a correspondence theory of truth; indeed, Maudlin even offers a suggestion for how this is to be done.[31] This view on the determination of truth values fits nicely with the idea that paradoxical sentences are indeterminate and the classical gap logical approach. Together they offer an enticing alternative to Barwise and Etchemendy's package for the correspondence theorist.[32]

4.3.7 Field

Hartry Field began his career defending a version of correspondence, but ended up abandoning it for disquotationalism in the early 1990s.[33] After a brief flirtation with paraconsistent dialetheism,[34] he developed a paracomplete logical approach to the paradoxes that is paired with an indeterminacy approach. The result is probably the most thoroughly cohesive unified theory where each component is formulated in detail, vigorously defended, and supports the other two. He is probably the most prominent defender of each component.

---

[30] Maudlin (2004: ch. 2). Given his comments there and in Maudlin (2004: ch. 9), it is clear that he takes himself to be offering a theory of the nature of truth.
[31] Maudlin (2004: 181-187).
[32] See Field (2006c), Gupta (2006), Belnap (2006), and Maudlin (2006) for discussion.
[33] Field (1972, 1986, 1994a, 1994b, 2001c, 2001d, 2001e, 2001f, 2005c, 2005d, 2006a, 2006b).
[34] Field (2001h).

Field's approach to the aletheic paradoxes is part of a much larger project that focuses on partially defined expressions. From this account of partially defined expressions he derives a theory of truth and indeterminacy, a theory of vagueness, and a theory of properties.[35] To accompany his account, he presents a new formulation of disquotationalism (within the paracomplete framework), a new paracomplete logic, and a non-standard probability calculus that allows him to explain degrees of belief in propositions that display indeterminacy.[36]

He adds a new conditional to the internal theory of Kripke's strong Kleene minimal fixed point, and he uses it to define a determinacy operator, which can be used to classify all the paradoxical sentences of the language in question (even those that contain the determinacy operator). He uses this conditional to define a paracomplete biconditional that features in the formulation of the T-sentences for his disquotational theory.[37] In sum, he offers a disquotational/indeterminacy/paracomplete unified theory.

4.3.8 Beall

Jc Beall rounds out our unified theory menu with the only inconsistency view. Graham Priest, who introduced paraconsistent dialetheism (the idea that paradoxical sentences are both true and not true, together with a non-classical logic that rejects the explosion rule), has not offered a theory of the nature of truth, and there are good reasons to think that Priest's approach to the paradoxes is

---

[35] See Field (2002, 2003a, 2003b, 2003c, 2004, 2005a, 2005b, 2007, 2008a, 2010a, 2010b) for the theory of truth and indeterminacy, Field (2003b, 2003c) for the theory of vagueness, and Field (2003c, 2004) for the theory of properties. Associated with this project is the work in Field's John Locke Lectures, which focus on the rational revisability of logic; Field (2008b).

[36] See Field (1998, 2000, 2001b, 2001g, 2001h, 2003b, Forthcoming) for the non-standard probability calculus. One should be aware that Field presents two non-standard probability calculi, one that is classical and the other non-classical. He now endorses only the non-classical version; see Field (2003c: 462).

[37] See Priest (2005), Leitgeb (2007), Priest (2007, 2010), Rayo and Welch (2007), Scharp (2007, 2009), and Shapiro (2010) for discussion.

not compatible with deflationism.[38] The problem is that, although Priest's theory of truth contains all the T-sentences (for a given language), it does not validate even a very weak intersubstitutability principle; in particular, the paraconsistent logic Priest adopts allows that 'p is not true' and '~p' have different truth values. Beall, on the other hand, is a disquotationalist. He offers a paraconsistent logic, BX, that is compatible with the intersubstitutability principle.[39] The result is the first unified theory of truth to feature the disquotationalism/inconsistency/paraconsistent combination.

4.3.9 Tennant

Neil Tennant has long endorsed an epistemic view of the nature of truth.[40] He has also defended a proof-theoretic diagnosis of what goes wrong in the reasoning of the aletheic paradoxes; namely, in each case, the argument to contradiction (e.g., 'the liar sentence is true and the liar sentence is not true') cannot be put into normal form. Very roughly, an argument is in normal form if it does not have any extraneous steps.[41] Although each "part" of the reasoning (e.g., the argument to 'the liar sentence is true' and the argument to 'the liar sentence is not true'), is normalizable, they cannot be put together in a normal-form proof. Tennant takes normalizability to be a necessary condition on valid arguments. As such, in his preferred logic, Core (discussed in Chapter Three), all valid arguments are normalizable. However, this feature comes at a cost—the rejection of transitivity.[42] That is, in Core logic, it can happen that there is a valid argument from some set of premises Γ to a conclusion A, and there is a valid argument from Γ and A together to a conclusion B, but there is no valid argument from Γ alone to B. This is the kind of problem that occurs in the liar reasoning.

---

[38] See Field (2008a: chs. 24-27).
[39] Beall (2009).
[40] Tennant (1997) contains the definitive statement of his view.
[41] See Tennant (1982, 1995, MS1, MS2).
[42] Tennant proves that Core logic obeys a weaker form of transitivity, but it is not relevant to our discussion.

Although the two "parts" of the reasoning are valid in Core logic, they cannot be combined to arrive at a valid argument whose conclusion is the contradiction. Tennant suggests that in cases like this, neither "part"s conclusion is assertible (e.g., one may not assert that the liar is true and one may not assert that the liar is not true). Tennant's philosophical interpretation of this state of affairs is that languages that contain their own truth predicates display indeterminacy, but this indeterminacy has its source in the logical consequence relation—it can happen that all the steps of an argument are locally acceptable, but when combined, the result is unacceptable. Overall, Tennant advocates a combination of an epistemic theory of the nature of truth, and indeterminacy philosophical approach and Core logic, which is a paraconsistent logical approach.

4.4 Summary

Information Box 25 contains a summary of the nine unified theories in this section. In Part III, I offer my own unified theory, which has a descriptive project and a prescriptive project. The descriptive project combines a Davidsonian theory on the nature of truth with an inconsistency philosophical approach to the paradoxes and a classical symmetric logical approach that treats truth predicates as assessment-sensitive (i.e., they have invariant content, but their extensions shift from one context of assessment to another). The prescriptive project features two new concepts, ascending truth and descending truth; it uses a Davidsonian theory for their nature and it is fully classical.

Nine Unified Theories of Truth Information Box 25

(i)	Correspondence	Contextual	Classical Gap	(Barwise and Etchemendy)
(ii)	Modest	Circularity	Weakly Classical	(Gupta and Belnap)
(iii)	Minimal	Epistemicist	Classical	(Horwich / Restall)
(v)	Disquotational	Indeterminacy	Classical Symmetric	(McGee)
(iv)	Minimal	Indeterminacy	Paracomplete	(Soames)
(vi)	Correspondence	Indeterminacy	Classical Gap	(Maudlin)
(vii)	Disquotational	Indeterminacy	Paracomplete	(Field)
(viii)	Disquotational	Inconsistency	Paraconsistent	(Beall)
(ix)	Epistemic	Indeterminacy	Paraconsistent	(Tennant)

Chapter 5

The Aletheic Penumbra

Truth occupies a central place in our conceptual scheme—it is without question one of the most versatile and explanatorily saturated concepts we have. It features in all kinds of philosophical explanations, and it is linked to many other fundamental concepts.[1] This section lays out some of these connections and considers the prospects for unified theories of truth to do justice to them.

5.1 Penumbral Connections

Kit Fine gave the term 'penumbral' a philosophical usage in his celebrated 1975 paper "Vagueness, Truth, and Logic." It has several non-philosophical uses, the most common of which designates the part of a shadow where the object casting the shadow only partially obscures the light source in question. Fine uses it to designate conceptual links that are to be preserved through conceptual or linguistic changes. For example, we might decide to sharpen the meaning of 'poor' so that someone is poor only if he or she makes less than $11,000 per year. This change should not affect the claim that if Gil makes less than Frank, then Frank is poor only if Gil is poor. The link between being poor and the "makes less money" relation is an example of a penumbral connection. Fine offers the following description:

> If language is like a tree, then penumbral connection is the seed from which the tree grows. For it provides an initial repository of truths that are to be retained throughout all growth. Some of the connections are internal. They concern the different borderline cases of a given predicate: if Herbert is to be bald, then so is the man with fewer hairs on his head. But many

[1]Deflationists will, of course, deny that these are genuine explanations, but no one can deny that truth figures in a tremendous number of purported explanations.

> other of the connections are external. They concern the common borderline cases of different predicates: if the blob is to be red, it is not to be pink; if ceremonies are to be games, then so are rituals; if sociology is to be a science, then so is psychology. Thus penumbral connection results in a web that stretches across the whole of language. The language itself must grow like a balloon, with the expansion of each part pulling the other parts into shape.[2]

For what it is worth, I agree that, other things being equal, penumbral connections ought to be preserved through conceptual or linguistic changes, but they are not sacrosanct. Sometimes the utility of a given change outweighs the fact that it requires giving up a cherished penumbral connection; for example, the link between physical space and Euclidean space was severed at the beginning of the twentieth century with the advent of general relativity.

5.1.1 Being

The idea that truth depends on being goes back at least to Aristotle, who writes:

> [W]hereas the true statement is in no way the cause of the actual thing's existence, the actual thing does seem in some way the cause of the statement's being true: it is because the actual thing exists or does not exist that the statement is called true or false.[3]

These days, philosophers often speak of the truthmaking relation (discussed in Chapter One in connection with correspondence theories of truth). For example, J. L. Austin writes: "When a statement is true, there is, *of course*, a state of affairs which makes it true."[4] Austin is endorsing a truthmaker principle.[5] Often, the view that every true truth bearer has a truthmaker is called *truthmaker maximalism*. Some truthmaker theorists prefer to focus on truthmakers for certain discourses and remain agnostic about truthmaker maximalism. There is a large literature on the

---

[2] Fine (1975: 275-276).
[3] Aristotle (1941)
[4] Austin (1979: 123).
[5] See also Fine (1982), Alston (1996), and Armstrong (1997, 2004).

candidate truthmakers for various kinds of claims (e.g., ethical, modal, negative, etc.), the precise formulation of truthmaker principles, and the extent to which they hold.[6]

Truthmakers and the relation of truthmaking have come to be a well-worn instrument in the toolbox of contemporary analytic metaphysics. Recall that in the discussion of philosophical methods in the Introduction (section 0.1.5), I mentioned reductive explanation, which can appeal to translation or apriori entailment as the relation between claims about the phenomena in question and claims in the reductive base. However, if one prefers a weaker relation between the two sets of claims, one can use truthmaking. That is, whatever the reductive base describes provides truthmakers for all the claims about the phenomena in question. Used in this way, the truthmaking relation becomes an all-purpose tool for philosophical theorizing. It seems to me that many of those working in the truthmaker debates more or less explicitly see themselves as engaged in this sort of project. For example, one might reject the Quinean construal of ontological commitment in terms of entities quantified over for a truthmaker view. On the former, if a theory T quantifies over entities of some type, then anyone who accepts T is committed to the existence of those entities.[7] On the truthmaker view, if some entities serve as truthmakers for a theory T, then anyone who accepts T is committed to the existence of those entities.[8] One reason to prefer the truthmaker view is that it is less strict—it allows one to accept theories that quantify over certain entities even if one denies the existence of those entities as long as one can provide suitable truthmakers for the theory that do not embed those entities. For example, if one denies the existence of compound objects (e.g., tables), one can still accept that there are tables as long as one is committed to the simple constituents that can serve as truthmakers for 'there are tables'.

---

[6] See the papers in Beebe and Dodd (2005) and Lowe and Rami (2009) for discussion.
[7] Quine (1948).
[8] Heil (2003), Armstrong (2004), and Cameron (2008).

As I mentioned in Chapter One, defenders of correspondence theories of truth are engaged in truthmaker disputes. The correspondence theorist, like the truthmaker theorist who regards truth as dependent upon being, must specify, for each truth bearer, some item(s) in a world, where the item(s) *make the truth bearer true*. The difference between the correspondence theorist and the truthmaker theorist is that the former typically offers an analysis of truth, whereas the latter need not do so, and the former is committed to explaining the truth of all true truth bearers, whereas the latter need not be.[9]

There are two major views on the nature of the truthmaking relation:

(Necessity) For all truth bearers b and for all worlds w, if b is true at w, then there exists an x such that x is in w and for all worlds w′ if x is in w′, then p is true at w′ (i.e., x *necessitates* the truth of b in any world where x exists).[10]

(Supervenience) For all truth bearers b and for all worlds w, if b is true at w then there exist some objects $x_1, \ldots, x_n$ and some properties or relations $F_1, \ldots, F_m$ such that $F_1 \ldots F_m$ hold of $x_1 \ldots x_n$ in w , and for every world w′ where $x_1 \ldots x_n$ exist and $F_1 \ldots F_m$ hold of $x_1 \ldots x_n$, b is true in w′ (i.e., the truth of b *supervenes* on what there is and how it is).[11]

These ways of understanding the truthmaking relation result in quite different theories of truthmaking and truthmakers. Furthermore, when implemented in a correspondence theory of truth, they result in different theories. I present a problem for the connection between truth and being later in this chapter.

5.1.2 Realism and Objectivity

To be a *realist* about something, it is often said, is to believe that it exists independently of our beliefs about it, our way of conceptualizing it, etc. And, for something to be *objective* is for it to be a certain

---

[9] See David (1994), Vision (2004), and Merricks (2007).
[10] Bigelow (1988) and Armstrong (1997); see Merricks (2007) and Schaffer (2008a, 2008b) for discussion.
[11] Lewis (2001a) and Bricker (2006); see Merricks (2007) and Schaffer (2008a, 2008b) for discussion.

way independently of our beliefs about it, our way of conceptualizing it, etc. Usually the term 'realist' applies either to people by virtue of their beliefs or behavior (e.g., Michael Devitt is a scientific realist) or to theories (e.g., Alston offers a realist theory of truth), and the term 'objective' applies to phenomena or discourses (e.g., propositional attitudes are objective, or mathematical discourse is objective). Philosophers have often explained both realism and objectivity in terms of truth: to be a realist about X is to endorse a correspondence theory of truth for discourse about X, and for discourse about X to be objective is for it to admit of correspondence truth.[12]

The opponents of realism have changed over the years. In the nineteenth and early twentieth centuries, idealism was the main competitor, and it was usually characterized by the acceptance of a coherence theory of truth (described in Chapter One). During the twentieth century, idealism faded away and various forms of what has come to be known as anti-realism took over as opponents to realism. One kind of anti-realism, championed by Michael Dummett, holds that truth cannot transcend our capacity to recognize or verify it; consequently, Dummett denies bivalence (i.e., that every truth bearer is either true or false) and the law of excluded middle (i.e., $p \vee \sim p$) along with it.[13] This view of truth is a variety of epistemic theory.[14] Other anti-realisms include error theories (i.e., all the claims of the discourse in question are false)[15], fictionalism (which is just an error theory with the claim that the discourse is still useful for some purposes)[16], and expressivism (i.e., all the claims of the discourse in question express some stance or non-cognitive attitude of the speaker, and so are neither true nor false).[17]

---

[12] See Boghossian (1990a, 1990b), Devitt (1991), Taylor (2006), and the papers in Greenough and Lynch (2006).
[13] Note that, as an intuitionist, Dummett does not accept the negations of these claims either.
[14] See Dummett (1991) and Tennant (1997).
[15] See Mackie (1977).
[16] See Field (1980), Yablo (1998, 2001), and Eklund (2005b).
[17] See Ayer (1939), Blackburn (1984), Gibbard (1990), and Schroeder (2007).

However, the entire debate about realism and objectivity has undergone immense changes in the last few decades, mostly because of deflationists on the one hand, and Crispin Wright on the other. Paul Horwich spearheaded this project for the deflationists in the early 1980s, and he deserves much of the credit for this trend.[18] Horwich is a deflationist (minimalist), and he argues that one can be a realist about some phenomenon without endorsing a correspondence theory of truth. For Horwich, there is no significant difference between asserting a claim of some discourse and asserting that that claim is true, so truth cannot be used to distinguish discourses that really purport to describe an independent reality and those that do not. However, Horwich does not offer an alternative way of characterizing realism and objectivity—he thinks that these debates are not worthwhile because once they have been unhinged from discussions of truth, there is nothing substantive to discuss any more.[19] Most deflationists accept Horwich's decoupling of realism and objectivity issues from accounts of truth, and many share his view about the futility of realism and objectivity debates.[20]

Crispin Wright agrees with Horwich and the other deflationists that realism and objectivity should not be explained in terms of the acceptance of a correspondence theory of truth; however, Wright thinks that the debates about realism and objectivity are legitimate. He completely reoriented them with the publication of his *Truth and Objectivity* in 1992. Wright's views are intricate and subtle, but the rough idea is as follows. He suggests that there are several independent and orthogonal ways of being objective; for example, not being constrained by what is knowable in principle (being epistemically unconstrained), not being dependent on the responses of rational entities (being response independent), being such that substantive disagreement must be due to cognitive shortcoming (displaying cognitive command), and featuring in multiple independent explanations of diverse phenomena (having wide cosmological role). In addition, Wright rejects

---

[18] Horwich (1982, 1998, 2004); see also Soames (1984).
[19] Horwich (1998: 52-67).
[20] Field (1994a, 1994b), Brandom (1994: ch. 5), and Williams (2002).

deflationism (his "inflationary argument" is taken up in Chapter Six); instead, he endorses a pluralist theory on which there are several distinct properties that can serve as truth properties (e.g., correspondence to facts and superassertibility, which is the property of being justified under any improvement of information). According to Wright, being a realist or anti-realist about some kind of discourse is a matter of how one regards the kinds of objectivity that arise for that discourse by virtue of the kind of truth property particular to that discourse; for example, one kind of anti-realism consists in holding that a discourse is epistemically constrained, which results from having superassertibility as its truth property.[21] Wright's view is currently reverberating throughout the various specific realism debates and having a significant impact on how they are conducted.[22, 23]

The old link between realism and the correspondence theory of truth has led some to call the correspondence theory of truth a *realist* theory of truth; epistemic, pragmatic, and coherence theories of truth were called anti-realist theories of truth.[24] The term 'realist theory of truth' is ambiguous; in one sense it is synonymous with 'theory of truth that can be used in conjunction with realism about some subject', while in another, it is synonymous with 'theory of truth that implies realism about truth'. A correspondence theory of truth could be a realist theory of truth in the first sense because one way of understanding realism is that it is the doctrine that a correspondence theory of truth is appropriate for some linguistic items. It could be a realist theory of truth in the second sense because many correspondence theorists are realists about truth (i.e., they think that truth is a property that sentences have independently of what humans believe). Sometimes deflationism is labeled as an anti-realist or irrealist theory of truth (Paul Boghossian uses this unfortunate

---

[21] See Wright (1989, 1992, 1998, 2003). For alternative takes on the relation between realism and truth, see Engel (2002), Kölbel (2002), and Taylor (2006). See Chapter One for a discussion of superassertibility.
[22] For example, see Shapiro (2007).
[23] For more recent developments on the notion of objectivity, see Nozick (2001), Debs and Redhead (2007), Daston and Galison (2007), van Fraassen (2008), Peacocke (2009), and Burge (2009, 2010).
[24] See Kirkham (1995) and Alston (1996).

terminology).[25] This is probably a response to the tendency of some deflationists to characterize their doctrine by asserting that truth is not a property. As a result, they seem to be anti-realists about truth. Accordingly, a realist about truth would be one who takes truth to be a real property. However, many deflationists (disquotationalists and minimalists in particular) do not qualify as anti-realists about truth in any of the respects listed above (e.g., they reject epistemic theories, error theories, fictionalism, and expressivism). I find it best to avoid the phrase 'realist theory of truth' since it generates so much confusion.

5.1.3 Meaning

It is commonplace to think that a sentence or belief is true or false by virtue of its meaning or content *and* the way the world is. In particular, the truth principles (T-In) and (T-Out) seem to hold by virtue of the meaning of the sentence in question. For example, 'snow is white' is a consequence of "snow is white' is true' because 'snow is white' means that snow is white. The dependence on meaning is more obvious when considering sentences of other languages; e.g., if 'Schnee ist weiss' means that snow is white, then 'Schnee ist weiss' is true if and only if snow is white.

The idea that one can exploit the connection between truth and meaning so as to explain meaning in terms of truth has a long history, but in the 1960s, Donald Davidson cast it in the form most familiar to contemporary philosophers.[26] Davidson is frequently interpreted as saying that the meaning of a sentence is its truth conditions (i.e., the conditions under which it is true). This is, at best, an oversimplification. Davidson claims that a Tarskian truth definition (discussed in Chapters One and Three) for a given language will serve as a meaning theory for that language. A *meaning theory* for a language provides a systematic specification of the meaning of each sentence in that

---

[25] Boghossian (1990a).
[26] Davidson (1968, 1973, 1974, 1990, 2005).

language. He also claims that a Tarskian truth definition for a given language serves as an *interpretation theory* for that language, where an interpretation theory for a language is a theory such that, if an interpreter knows it, she will be able to understand the utterances of a speaker of that language. Both of these views are defended in a fairly non-standard way. Davidson begins with a Tarskian theory of truth for a given language and asks how to give empirical content to it (or, alternatively, how one could test it). His answer is in the form of a procedure that a somewhat idealized person (called the *radical interpreter*) could use to construct a Tarskian truth definition for a person's language without knowing any of that person's propositional attitudes or understanding his language in advance (discussed in Chapter One). Davidson thinks of this procedure as similar to the procedures used in the field known as measurement theory (discussed in Chapter Four) to give empirical content to many kinds of formal theories (e.g., theories of weight, length, subjective probability, utility, etc.).[27] It is common for contemporary philosophers to ignore this aspect of Davidson's work, and instead concentrate on the idea that it is a necessary condition on any theory of meaning that the meaning of a sentence determines its truth conditions.

Davidson's ideas about meaning and truth conditions have been highly influential on formal semantics, which sits at the intersection of linguistics, philosophy of language, and logic. Formal semantics uses mathematical techniques from mathematical logic to model the meanings of words, phrases, and sentences of natural languages and artificial languages.[28]

5.1.4 Validity

[27] See Davidson (1968, 1973, 1990, 1995).
[28] See Heim and Kratzer (1998), Chierchia and McConnell-Ginet (2000), Portner (2005) and the papers in Portner and Partee (2002).

Validity is supposed to track good reasoning; roughly, if an argument is valid then one ought not believe its premises without also believing its conclusion.[29] So validity is a property of reasoning that adheres to the canons of good reasoning. It is very common to explain validity in terms of truth preservation: an argument is *valid* iff it is necessarily truth-preserving; i.e., an argument $\langle\Gamma, \phi\rangle$ is valid iff necessarily, if all the members of Γ are true, then ϕ is true. Indeed, this is usually taken to be the definition of validity. There is a connection between validity, consequence, and entailment, which links truth to the latter two as well: an argument $\langle\Gamma, \phi\rangle$ is valid iff ϕ is a *consequence* of Γ iff Γ *entails* ϕ. So the above definition of validity in terms of truth turns into a definition of consequence and entailment in terms of truth as well.

5.1.5 Inquiry

The concept of inquiry covers a lot of ground, and I cannot summarize it all here. Suffice it to say that many philosophers think that truth is a goal of inquiry. When one engages in inquiry, of whatever kind, one is aiming at, among other things, arriving at truths. Moreover, truth is used as a standard by which to evaluate an inquiry—an inquiry is successful only if it arrives at truths.[30]

5.1.6 Belief

Belief is often taken to be a kind of propositional attitude—a mental state constituted by a relation between a person's mind and a proposition, mental content, or potential state of affairs (other propositional attitudes include desiring, hoping, and intending). Truth is often invoked to distinguish belief from other propositional attitudes by saying that believing that p is taking the

[29] There are well-known disputes about the relationship between logic and rational belief; see Harman (1986, 1999, 2009), Hanna (2006), and Field (2009) for discussion.
[30] See Stalnaker (1987), Wright (1992), and Lynch (2004, 2009) for discussion.

proposition that p to be true. Truth is also used to explain successful beliefs. For example, it seems that, other things being equal, acting on true beliefs is more likely to satisfy one's desires or accomplish one's goals.[31] Finally, truth is also used as a standard for belief—it is correct to have a belief only if that belief is true.[32]

5.1.7 Assertion

Assertions are a particular type of speech act in which one utters a sentence. Of course, asserting is more than just the uttering of a sentence, but how to specify what more it consists in is difficult. Many think that a speaker's assertion purports to express that speaker's belief(s). Some have argued that an assertion is an utterance of a sentence with the intention of saying something true. That is, truth is the goal of assertion or the point of assertion.[33] Some philosophers have argued that assertion is the kind of speech act whose constitutive rule is to assert only what one believes is true.[34] Truth is also used as a standard by which to evaluate assertions—it is correct to assert a sentence p only if p is true; otherwise, it is incorrect and that asserter may be sanctioned for it.[35]

Huw Price has recently written on the connection between truth and assertion. Price claims that there are at least three norms governing assertion:

> *Subjective assertibility*: A speaker is incorrect to assert that p if she does not believe that p.; to assert that p in those circumstances provides prima facie grounds for censure or disapprobation.
>
> *Personal warranted assertibility*: A speaker is incorrect to assert that p if she does not have adequate (personal) grounds for believing that p; to assert that p in these circumstances provides prima facie grounds for censure.

---

[31] See Millikan (1984), Dretske (1988), and Fodor (1990).
[32] See Williamson (2000) for discussion.
[33] Dummett (1959); see Wright (1992), Williamson (2000a), Simmons and Bar-On (2006), and Lynch (2009) for discussion.
[34] See Beaver (2001), Soames (2004, 2007), and MacFarlane (forthcoming c) for discussion.
[35] Price (1988, 1998, 2003).

> *Truth*: If p is not true, then it is incorrect to assert that p; if p is not true, there are prima facie grounds for censure of an assertion that p.[36]

Price argues that our linguistic practice would be unrecognizable without all three norms. The third norm in particular is crucial for participants in a conversation to count as disagreeing with one another. Without it, Price claims, our assertions would be akin to ordering an entre at a restaurant. That is, without it, Roy's assertion that grass is green and Poochie's assertion that grass is not green would be no more incompatible than if one ordered shrimp and the other ordered chips and salsa.[37] Moreover, this kind of incompatibility is essential to our assertional practice. I examine assertion in detail in Chapter Six.

5.1.8 Knowledge

Going back to Plato, the standard definition of knowledge is justified true belief (i.e., an agent S knows that p iff S believes that p, S is justified in believing that p, and S's belief that p is true).[38] Although the twentieth century saw a flurry of criticism and alternatives to this definition, most of it concerns the justification component, not the truth component. Virtually all philosophical views of knowledge imply that if S knows that p, then $\langle\mathsf{p}\rangle$ is true.

5.1.9 Analyticity

The standard definition of analyticity is truth in virtue of meaning, so the bearers of analyticity (if there are any) have to be the kinds of things that can be truth bearers and the kinds of things that can have meanings (e.g., sentences). W. V. Quine launched an attack on analyticity in the 1950s that

---

[36] I have altered this third norm by substituting 'not true' for 'not-p' in Price's formulation. His discussion makes clear that he intends the third norm to appeal to the concept of truth.
[37] Price (2003).
[38] Plato (1961)

was very successful; however, more and more philosophers have been trying to resurrect the notion recently.[39]

Typically, the sentences taken to be analytic are those that are constitutive of the concepts involved, but, as I have mentioned (in Chapter Two), there are at least two ways to understand constitutive principles: as things that must be true for a word to express a given concept, or as things that must be believed for a person to possess a given concept. Quine's attack targets the first reading, but not the second. It is unclear whether the second is to be explained in terms of truth; I argue in Chapter Twelve that it need not be.

5.1.10 Necessity

Necessity is often called a modal notion because it was thought of as a mode of truth. That is, being necessary is a way of being true. Given this connection it is not a surprise that most philosophers agree that all necessary sentences are true. Moreover, just as with validity, truth is used in the definition of necessity.

There are several notions of necessity that philosophers currently discuss, including logical, metaphysical, and physical. A sentence is logically necessary iff it is true by virtue of its logical form. A sentence is metaphysically necessary if and only if it is true in all possible worlds. A sentence is physically necessary if and only if it is true in all possible worlds that have the same natural laws as the actual world. Thus, an account of truth will have consequences for an account of necessity.[40]

5.1.11 Proof

---

[39] Quine (1951), Boghossian (1997), Soames (2002c), Russell (2008), Williamson (2008) and Juhl and Loomis (2009).
[40] Both McGee (1991) and Gupta and Belnap (1993) present theories of necessity that are compatible with their respective theories of truth.

There are many technical issues associated with the notion of proof, but the one I want to emphasize is that Gödel showed that one cannot define truth in certain interpreted languages in terms of deduction in a formal system because he provided a method of constructing true sentences of such languages that are not deducible.[41] It is not that proof is explained in terms of truth, although the soundness of a deductive system is frequently demonstrated by appealing to the notion of truth-in-a-model. Instead, truth is used to explain what is special or important about these undecidable sentences. A theory of truth ought to explain why undecidable sentences are true even though they are unprovable.[42]

5.2 Paradoxes and the Explanatory Role

The penumbral connections between truth and other concepts can be treated in several distinct ways. First, one can treat truth as more fundamental than the other concept in question and assume that a theory of truth ought to be able to explain the penumbral connection. Second, one can treat the other concept as more fundamental and assume that the penumbral connection ought to be explained by a theory of it. Third, one can treat truth and the other concept as equi-fundamental and offer a theory that explains both at once in addition to the penumbral connection between them.

It is in this arena that so many of the disputes about theories of truth take place. In what remains of this chapter, I consider the impact that the aletheic paradoxes have on truth's penumbral connections. It turns out that, although this issue is almost completely unexplored, it is a rich and fruitful domain, but I can only cover a few topics here.

---

[41] Gödel (1931); see Mendelson (2001), Franzen (2004, 2005), and Boolos, Rosen, and Jeffrey (2007) for technical treatments and see Smith (2007) for an accessible presentation.
[42] For discussion see Shapiro (1998) and Tennant (2002).

5.2.1 Paradox and Being

Contrary to the claims of correspondence theorists and truthmaker maximalists, considerations arising in discussions of the aletheic paradoxes suggest that *having a truthmaker* or *corresponding to the world* are not necessary conditions for truth. We can formulate the principal target as:

(N) If p is true, then p has a truthmaker.

Indeed, I show that (N) is not only false, it is self-refuting.

Consider the sentence, 'snow is white'.[43] A truthmaker theorist would say that the truthmaker for 'snow is white' is snow, or snow's being white, or the fact that snow is white, or perhaps something else that is more fundamental and on which snow's being white supervenes (e.g., the world). The choice between these options for truthmakers depends on one's particular truthmaker theory, but they do not matter for my purposes.

What is the truthmaker for the sentence, "snow is white' is true'? Before answering this question, we need to formulate a constraint on truthmaker theories (and for correspondence theories of truth as well). Since the point of truthmaker theory is specifying that on which the truth of any given truth bearer depends, it would be illicit to say that the truthmaker for some truth bearer p is the truth of p. As Jonathan Schaffer says, "[T]he core of truthmaking theory [is] the idea that truth is a derivative aspect of reality, and thus needs grounding."[44] We can formulate this constraint as:

(*) The truthmaker for p is not the truth of p or anything that depends on the truth of p.

Every truthmaker theorist and correspondence theorist accepts (*); I formulate it explicitly to motivate the choice point below.

---

[43] I use sentences as truth bearers throughout for ease of exposition, but nothing hangs on this point.
[44] Schaffer (2010: 319).

Now back to ''snow is white' is true'. When specifying a truthmaker for this truth bearer, the truthmaker theorist has a choice:

(i) the *redundant option*: "snow is white' is true' has the same truthmaker as 'snow is white'. In general, whatever grounds the truth of p grounds any further attributions of truth to p.

(ii) the *recursive option*: the truthmaker for "snow is white' is true' is 'snow is white' or the sentence 'snow is white', or the truth of the sentence 'snow is white', or the fact that 'snow is white' is true, or something else that is more fundamental and on which 'snow is white's being true supervenes (e.g., the world). In general, the truthmaker theorist first specifies truthmakers for all the truth bearers that do not involve truth, then specifies truthmakers for those that attribute truth to those truth bearers, and then to those that attribute truth to *those* truth bearers, and so on. The truthmakers for truth bearers involving truth are built up recursively.

Although these two strategies are distinct in that they assign different truthmakers to aletheic truth bearers, it is what they have in common that interests me. They both imply that *only grounded truth bearers have truthmakers*. Recall (from Chapters Two and Three) that Kripke defined groundedness as part of his recursive constructions for handling the aletheic paradoxes. A sentence is grounded iff its truth value depends entirely on the truth values of sentences that do not contain 'true'.

Regardless of whether one chooses the redundancy option or the recursive option to assign truthmakers to aletheic truth bearers, the result will be that only those truth bearers whose truth values ground out in the non-semantic facts will be assigned truthmakers.[45] Correspondence theorists and truthmaker maximalists should accept this fundamental principle: only grounded truth bearers can correspond to something in the world, or only grounded truth bearers have truthmakers.

The problem with (N) is that there are many true ungrounded truthbearers. For example, the following sentences are ungrounded:

(1) For all x, x is a true universal generalization iff all the instances of x are true.

(2) No truth bearer is both true and not true.

---

[45] For a discussion of this point in a formal setting, see Leitgeb (2005).

(1) is a core entry in any respectable semantic theory that handles quantifiers. Since (1) is a universal generalization, its truth value is not to be determined by the truthvalues of all the non-aletheic universal generalizations (i.e., the universal generalizations that do not have either a truth predicate or the concept of truth as a constituent). That is, it could be that all universal generalizations except (1) are true iff all their instances are true, but (1) is not. The result is that truthmaker maximalism and correspondence theories of truth must deny that (1) is true because of their commitment to (N). Since (1) is part of just about any semantic theory, it follows that a commitment to (N) is incompatible with acceptance of contemporary semantic theory.

Still, the truthmaker theorist might bite the bullet on this point. However, (2) is more problematic since it should be a consequence of any non-dialetheic truthmaker theory or correspondence theory of truth. Note that since (2) is itself a truth bearer, its truth value is not determined by the non-aletheic truth bearers; it could be that all non-aletheic truth bearers are either true or false, but (2) fails to be true (only) because it is both true and not true. Thus, any non-dialetheic truthmaker theory or correspondence theory of truth will have a consequence, (2), that it deems untrue by virtue of (N). Nevetheless, the truthmaker theorist might find some way to avoid having (2) as a consequence and so accept the counterintuitive consequence that (2) is not true.

What about (N) itself? It is easy to see that it is ungrounded as well. That is, the truth value of 'if p is true, then p has a truthmaker' is not determined by the truth values of the non-alethiec truth bearers. Hence, (N) is ungrounded and accepting (N) is tantamount to accepting that only grounded truth bearers are true. Thus, according to (N), (N) is not true. It follows that (N), and with it truthmaker maximalism and correspondence theories of truth, are self-refuting.

Let us consider how a correspondence theorist or truthmaker maximalist might respond. Ross Cameron and Jonathan Shaffer have each offered truthmaker theories that allow entire worlds as truthmakers; call these *promiscuous truthmakers*. They purport to solve the perennial problems of

negative truthmakers, truthmakers for general claims, etc. Could the friend of (N) avoid the objection by appealing to about a promiscuous truthmaker?

If the world does not have aletheic constituents (e.g., facts about truth), then it will not ground the truth of (1), (2), or (N). If the world does have aletheic constituents, but these are built up recursively as described above, then, again, (1), (2), and (N) remain ungrounded. However, if the world has aletheic constituents that are not built up recursively, then the truthmaker theorist denies (*) and the whole theory is malignantly circular. Either way, no one who endorses a correspondence theory or a truthmaker theory thinks that the world has (non-recursive) aletheic constituents anyway. For example, Schaffer's definition of the world is: the instantiation of the conjunction of all natural properties by the fusion of all thin particulars.[46] Obviously truth does not count as a natural property or else there would be no point to formulating a truthmaker theory in the first place (no one thinks we need a whitemaker theory or a humanmaker theory unless they deny that being white or being human are natural properties). Thus, recourse to promiscuous truthmakers is of no help.

Another possibility is to allow truthmak*ing* without truthmak*ers*. That is, a truthmaker theorist can deny that there is any thing that makes a true truthbearer true, but accept that truth supervenes on the non-aletheic truths. That is, if p is true in world w and not true in world v, then there is some non-alethic difference between w and v.[47] However, it is easy to see that this will not help the truthmaker theorist. Because (1), (2), and (N) are ungrounded, there are worlds w and v that are alike in all non-aletheic respects, but (1), (2), and (N) are true in w and not true in v. How can this be so? If (1) is an exception to (1), (2) is an exception to (2), and (N) is an exception to (N) in v, then these are the only differences between w and v. That is just what it means for there to be

[46] Schaffer (2010: 309).
[47] See Lewis (2001a), Sorenson (2001: ch. 11), and Melia (2005) for discussion of this option.

ungrounded truths—that truth does not supervene on what there is an how it is (in non-aletheic respects).

Instead, could the truthmaker theorist deny (*) by positing brute truthmakers? That is, could the truthmaker theorist say that the truthmakers for ungrounded truthbearers like (1), (2), and (N) be the very facts that (1) is true, that (2) is true, and that (N) is true?

Sure, the truthmaker theorist could say this, but only at the expense of robbing truthmaker theory of any interest whatsoever. After all, truthmakers are supposed to perform metaphysical work by showing how truth is dependent on being. If the truthmaker for each truthbearer were just the fact that that truthbearer is true, then truth would no longer be shown to depend on being in the way supposed by all proponents of truthmaker theory. Moreover, correspondence theorists take themselves to be offering analyses of the concept of truth (or perhaps reductive explanations of the property of truth). If the correspondence theorist holds that 'p is true' corresponds to the fact that p is true, then neither of these projects is remotely tenable.[48]

What if the truthmaker theorist reinterprets (1), (2), and (N) as the following:

(1′) for all x, x is a true grounded universal generalization iff all instances of x are true.

(2′) no grounded truthbearer is both true and not true.

(N′) if p is grounded and true, then p has a truthmaker.

I grant that the truthmaker theorist could treat ungrounded truthbearers as if they have an implicit domain restriction to grounded truthbearers, but this is a desperate move. For example, a consequence would be that 'every truth has a truthmaker' (when read with the implicit domain restriction) and '(2) is true and (2) has no truthmaker' are consistent, which would be absurd. Moreover, to offer this reading of (N) would be to give up truthmaker maximalism, which is just to

[48] Any attempt to use something like Kripke's maximal intrinsic fixed point to give ungrounded sentences truth values is essentially the same move. It requires stipulating that some sentences have brute truthmakers.

grant my point. Thus, appealing to some kind of implicit domain restriction is merely pretending to be a truthmaker maximalist.

Finally, it is an option for the truthmaker theorist to give up truthmaker maximalism and endorse a variant of (N) for some restricted class of truths, but this would be to grant my objection. Moreover, it comes at a serious cost. If the whole point of truthmaker theory is to show how truth is dependent on being, then it simply does not make sense to say that some truths (i.e., the ones the truthmaker theorist can handle) cry out for grounding in the world, while others (i.e., the ones the truthmaker theorist cannot handle) do not. Without some principled distinction between the truths that are dependent on being and those that are not, being a truthmaker theorist while rejecting truthmaker maximalism seems ad hoc at best, but it is closer to incoherent. Moreover, this option is unavailable to the correspondence theorist, who, by definition, endorses a theory about all truths.

Other moves for the correspondence theorist or truthmaker maximalist might be: (i) come up with some way of assigning truthmakers for truths involving truth (instead of the redundant option or the recursive option) that gives ungrounded truths truthmakers, or (ii) embrace a self-refuting theory. I see no way of implementing the first strategy within the strictures of the correspondence theory or truthmaker maximalism. As for the second, self-refutation has usually been taken to be a fatal flaw in a theory.[49] However, one theorist, Tim Maudlin, has recently defended a self-refuting theory of truth—I present an objection to it in Chapter Eight.

5.2.2 Paradox and Objectivity

[49] See Fitch (1946) for a nice discussion.

I mentioned above that pluralists about truth typically select a few principles to serve as platitudes or truisms that characterize the role a property must play for it to count as a truth property. Wright and Lynch both select principles that are inconsistent, as evidenced by the paradoxes.[50] If one follows Wright in using a pluralist theory of truth as a basis for a theory of objectivity and realism, then the paradoxes do affect the extent to which truth can explain them. If there is no way to characterize the truth role without using principles that the paradoxes show to be inconsistent, then it seems that this way of developing a truth-based theory of objectivity and realism is a non-starter.

Consider Lynch's view. He argues that several distinct properties can play the truth rule, depending on the discourse on considers, where the truth rule is constituted by a set of platitudes. Lynch lists among these platitudes the following:

(Objectivity) The belief that p is true iff with respect to the belief that p, things are as they are believed to be.

(Norm of Belief) It is prima facie correct to believe that p iff the proposition that p is true.

(End of Inquiry) Other things being equal, true beliefs are a worthy goal of inquiry.

Lynch calls these core truisms and says of them that "denying *many* or *all* would mean that you would be regarded by other users of the concept as changing the subject."[51] Lynch then offers the following criterion for a theory of truth:

> A theory counts as a theory of *truth* (as opposed to a theory of something else) only if it incorporates the core truisms about truth. As noted, there may be disagreement amongst philosophers about just what those core truisms are. But in this book, I will take them to include, at least, the truisms that truths are objective, correct to believe, and an end of inquiry. To incorporate a truism into a theory is to either list it among the principles of the theory or endorse a principle that entails it.[52]

---

[50] See Shapiro (forthcoming).
[51] Lynch (2009: 13).
[52] Lynch (2009: 17).

So a theory of truth must, by definition, include the core truisms. In addition, Lynch correctly notes that it is very straightforward to derive versions of (T-In) and (T-Out) from (Objectivity) using simple platitudes like 'with respect to my belief that p, things are as they are believed to be iff p'.[53]

Putting these claims together, we can see that Lynch is saying that a theory of truth must validate (T-In) and (T-Out). But we already know that these principles lead to contradiction by way of the reasoning in the aletheic paradoxes. This reasoning can be avoided, but only at the cost of giving up classical logic. So, it seems that Lynch thinks that every theory of truth is inconsistent or committed to non-classical logic.

I agree with Lynch that the above principles are truisms in the sense of constitutive principles that explain concept possession and linguistic competence (i.e., a concept possessor must bear some relation to that concept's constitutive principles—e.g., belief, a disposition to accept, etc.); indeed, I turn to this topic in Chapter Eleven. However, we part ways when he insists that any theory of truth must incorporate them.

Here is a way that the pluralist can retain the idea of a truth role and avoid the above problem. Instead of insisting that any theory worthy of being called a theory of truth should incorporate the core truisms, the pluralist should say that any theory of truth must imply *that* the above principles are truisms about truth (or, to use terminology I prefer, that they are constitutive of the concept of truth). With that small change, the pluralist can avoid the untoward consequence exposed above.

Moreover, the pluralist's emphasis on platitudes or truisms regarding truth makes it a very nice fit for inconsistency approaches to the paradoxes. Inconsistency approaches too insist that the concept of truth has certain constitutive principles that any theory of truth should respect—however, respecting a constitutive principle is different from incorporating it into one's theory since inconsistency theorists think the constitutive principles for truth are inconsistent. The dialetheic

[53] Lynch (2009: 8-10; my formatting.

inconsistency theorists (e.g., Priest and Beall) think that it follows that paradoxical sentences are both true and not true, but the pluralist need not follow them. Instead, the other group of inconsistency theorists (e.g., Chihara, Yablo, Eklund, and myself) seems a better fit with pluralism. Of course, such a pluralist must accept that properties that count as manifestations of truth need not obey all the truisms, since no property obeys all of them. Instead, the pluralist might claim that manifestations of truth are those that fit best with all the truisms or satisfy a weighted majority of all the truisms. It is unclear to me just how problematic this move might be.[54]

5.2.3 Paradox and Meaning

There are reasons to think that the paradoxes threaten truth-conditional theories of meaning, and so threaten explanations of meaning in terms of truth. The problem comes in specifying truth conditions for sentences containing truth predicates. Since paradoxical sentences are meaningful, any truth-conditional theory of meaning ought to be able to specify truth conditions for them. However, to avoid inconsistency, the truth-conditional theory of meaning ought to incorporate some approach to the paradoxes.

As I have mentioned, most approaches to the paradoxes face revenge paradoxes and respond to them by restricting the class of languages to which they apply so that it does not include those with the capacity to formulate revenge-paradoxical sentences. Call the languages to which a theory applies its *target languages*. There are at least four kinds of these restrictions:

(i) No target language L can express its own truth predicate ('true-in-L'). Example: Tarski's theory.

(ii) Each target language L can express its own truth predicate ('true-in-L'), but it cannot express the theory of truth-in-L. Examples: Kripke's theory (inner strong Kleene) and Gupta & Belnap's theory (outer revision).

---

[54] Thanks to Michael Lynch for discussions on this issue.

(iii) Each target language L can express its own truth predicate ('true-in-L') and the theory of truth-in-L, but it cannot express other linguistic resources associated with revenge paradoxes (e.g., exclusion negation). Examples: McGee's theory (classical symmetric), Maudlin's theory (outer strong Kleene), Field's theory (paracomplete), and Beall's theory (transparent paraconsistent).[55]

(iv) No expressive restrictions on target languages. Examples: none.

For example, Davidson, the prototypical truth-conditional theorist, accepts a type (i) theory (Tarski's). However, there is good reason to think that this is unacceptable.[56]

Assume Sherri and Terri are interpreting one another and Sherri speaks language S, while Terri speaks language T. Sherri uses S to construct a truth-conditional theory of meaning for T (in the form of a Tarskian truth definition in S for T), while Terri uses T to construct a truth-conditional theory of meaning for S (in the form of a Tarskian truth definition in T for S). Obviously, if S and T are natural languages, then they are going to contain their own truth predicates. However, a Tarskian theory of truth is not acceptable for such a language. Hence, neither Sherri nor Terri can succeed.

A truth conditional meaning theorist might respond to this objection in the following way. Let S* be the sublanguage of S that does not express truth-in-S and let T* be the sublanguage of T that does not express truth-in-T. Now Sherri can construct in S a definition of truth-in-T* and Terri can construct in T a definition of truth-in-S*. These theories will allow Sherri and Terri to specify the truth conditions for all the sentences of each other's languages that do not contain truth predicates.

Fair enough, but the sentences of S and the sentences of T that contain truth predicates are meaningful, and meaning is supposed to be explained in terms of truth conditions. These sentences

---

[55] Note that there are additional constraints these theories would have to meet. For example, Field's theory and Beall's theory are disquotational, which will not work as a meaning theory. See Chapter Six for discussion.
[56] See Chihara (1976) and Lycan (forthcoming) for discussion of this point.

have truth conditions. What are they? A theory of meaning that cannot be used on obviously meaningful sentences is not worth calling a theory of *meaning*.

It seems to me that this argument pushes a truth-conditional meaning theorist from accepting a type (i) theory of truth to a type (ii) theory of truth. However, one can run the same argument in terms of theories of truth to push the truth-conditional meaning theorist from accepting a type (ii) theory of truth to accepting a type (iii) theory of truth. That is, the sentences left out of the target languages are meaningful, so a truth-conditional theory of meaning ought to be able to specify their meanings. One can then turn the crank again, this time by appealing to the other linguistic resources (e.g., exclusion negation) to push a truth-conditional meaning theorist from accepting a type (iii) theory of truth to accepting a type (iv) theory of truth. However, there are no type (iv) theories of truth.

This argument can be resisted at a couple of points. First, one might deny that the additional linguistic resources (e.g., exclusion negation) are meaningful or coherent.[57] If that is the case, then the motivation to move from a type (iii) theory to a type (iv) theory disappears. I discuss this move later in the book (Chapters Six and Twelve). Second, one might follow Kirk Ludwig in denying that commitment to a truth-conditional meaning theory requires commitment to a theory of truth.[58] For Ludwig, one can use an inconsistent truth theory to arrive at a consistent and true meaning theory for a language.

Here is Ludwig's suggestion. Let $\mathcal{L}$ be the target language and let T be an *interpretive* truth theory for $\mathcal{L}$, which means that the expressions contained in the axioms of T (which is formulated in a metalanguage $\mathcal{M}$ for $\mathcal{L}$) are synonymous with the associated expressions of $\mathcal{L}$. For example, an

---

[57] Maudlin and Field both make this move.
[58] Ludwig (2002); see also Lepore and Ludwig (2004) and Badici and Ludwig (2007). See Patterson (2009) for discussion.

axiom of T might be 'for any sentences ϕ and ψ of $\mathcal{L}$, $[\phi \wedge \psi]$ is true iff ϕ is true and ψ is true'; the expression 'and' of $\mathcal{M}$ is used in giving the truth conditions for the expression '$\wedge$' of $\mathcal{L}$, and these two expressions are synonymous. A meaning theory, M, for $\mathcal{L}$ consists of the following components:

(i) T is an interpretive truth theory for L.

(ii) The axioms of T are A_1, …, A_n, …

(iii) A_1 means in L that …; A_2 means in L that …; …

(iv) CP is a canonical proof procedure for T.

(v) For any sentence s of $\mathcal{L}$, if s is the last line of a canonical proof in T, then the corresponding M-sentence is true in $\mathcal{M}$.

Ludwig does not say much about canonical proof procedures other than they result in "proofs that draw solely on the content of the axioms" of T.[59] I consider several ways of understanding them below.

The key to Ludwig's suggestion is that even if $\mathcal{L}$ contains its own truth predicate (i.e., 'true-in-$\mathcal{L}$'), which results in T being inconsistent, M (the meaning theory for $\mathcal{L}$) need not be inconsistent. From the canonical T-sentences (i.e., those having canonical proofs), one can infer the corresponding M-sentence:

s in $\mathcal{L}$ means that p,

where s is a structural description of a sentence of $\mathcal{L}$ and p is its translation into $\mathcal{M}$. For example, if $\mathcal{L}$ contains its own truth predicate and the means to refer to its own sentences, then it also contains a liar sentence like:

---

[59] Ludwig (2002: 149).

(1) (1) is not true-in-$\mathcal{L}$.

Via a canonical proof, we get a T-sentence for (1): '(1) is true-in-L iff (1) is not true-in-$\mathcal{L}$'; from it, the corresponding M-sentence, "(1) is not true-in-$\mathcal{L}$' means that (1) is not true-in-$\mathcal{L}$' follows. Thus, even though T is inconsistent, we can use it to generate a consistent meaning theory for $\mathcal{L}$. If this strategy works, then there is no reason to pair a truth-conditional theory of meaning with an approach to the aletheic paradoxes because the aletheic paradoxes do not infect the theory of meaning.

It should be clear that everything turns on Ludwig's account of canonical proof. Remember, in the cases we care about, the truth theory is inconsistent and the background logic is classical, so the truth theory is the whole language. That is, every sentence of $\mathcal{L}$ is a theorem of T. Thus, every sentence of the form 's is true-in-$\mathcal{L}$ iff p' (where s is a structural description of a sentence of $\mathcal{L}$ and p is a sentence of $\mathcal{M}$) is a theorem of the theory. Ludwig tries to pick out the canonical T-sentences (i.e., those whose right-hand-side is a *translation* of the sentence described on the left-hand-side) by appeal to the notion of canonical proof. For Ludwig, a canonical proof is "a proof meeting certain constraints that ensure that only the content of the axioms is drawn on in proving it. This can be accomplished by restricting the rules we can appeal to in proofs and what we can apply them to. We can call proofs that satisfy the constraints *canonical proofs*."[60] So, which rules are we allowed to use in a canonical proof? Ludwig writes: "For any given theory and logic, it would be straightforward, if somewhat tedious, to write out what restrictions were required. Once we had a characterization of the restrictions required in some logical system, we could in fact weaken the system so that it consisted of only the moves so allowed. In this case every T-theorem of the theory would also be a

[60] Ludwig (2002: 148).

T-sentence."[61] Ludwig does offer a sample canonical proof procedure, which consists of the following rules:

(UQI) For all sentences σ, all variables υ, and all singular terms τ, Inst(σ, υ, τ) may be inferred from UQuant (σ, υ).

(RPL) For all sentences σ_1, σ_2, S(σ_2) may be inferred from Eq(σ_1, σ_2) and S(σ_1).

(SUB) For all singular terms τ_1, τ_2, S(τ_2) may be inferred from S(τ_1) and Ident(σ_1, σ_2).

In these rules, the following terms are use: 'UQuant(σ, υ)' means *the universal quantification of σ with respect to υ*; 'Inst(σ, υ, τ)' means *the result of replacing all instances of the free variable υ in σ with the singular term τ*; 'Eq(σ_1, σ_2) means *the biconditional linking σ_1 with σ_2 (in that order)*; 'Ident(τ_1, τ_2)' means *the identity sentence linking σ_1 with σ_2 (in that order)*; 'S(x)' stands for a sentence containing the grammatical unit x, which may be a word, phrase, or sentence.[62]

As an example, Ludwig gives us the following canonical proof:

1. 'Fa $\wedge$ Ra' is true-in-$\mathcal{L}$ iff 'Fa' is true-in-$\mathcal{L}$ and 'Ra' is true-in-$\mathcal{L}$. (by UQI from the axiom defining conjunction)

2. 'Fa' is true-in-$\mathcal{L}$ iff the referent of 'a' is red. (by UQI from the axiom defining the predicate 'F')

3. 'Ra' is true-in-$\mathcal{L}$ iff the referent of 'a' is round. (by UQI from the axiom defining the predicate 'R')

4. 'Fa' is true-in-$\mathcal{L}$ iff Alfred is red (by SUB from 2 and the axiom defining 'a')

5. 'Ra' is true-in-$\mathcal{L}$ iff Alfred is round (by SUB from 3 and the axiom defining 'a')

6. 'Fa $\wedge$ Ra' is true-in-$\mathcal{L}$ iff Alfred is red and Alfred is round (by RPL—twice—from 1, 4, 5)

---

[61] Ludwig (2002: 148).
[62] Ludwig (2002: 157-158). These rules and definitions are not carefully formulated and should be attributed to Ludwig, not me.

The last step, line 6, contains the T-sentence for 'Fa ∧ Ra'. Since this the above is a canonical proof of this T-sentence, we can infer that 'Fa ∧ Ra' in $\mathcal{L}$ means that Alfred is red and Alfred is round.

It should be clear that Ludwig's proposal is not workable. The most obvious problem is that line 1 contains a T-sentence that is arrived at by canonical proof, so we should be able to infer that 'Fa ∧ Ra' in $\mathcal{L}$ means that 'Fa' is true-in-$\mathcal{L}$ and 'Ra' is true-in-$\mathcal{L}$. So we have two M-sentences that are clearly not equivalent for a single sentence of $\mathcal{L}$.[63] Another problem is that we can derive multiple incompatible M-sentences for liar sentences of $\mathcal{L}$. Let 'T' be the truth predicate in $\mathcal{L}$, and let 'b' be a singular term of $\mathcal{L}$ that refers to the sentence '~Tb', which is also a sentence of $\mathcal{L}$. We have the following canonical derivation for '~Tb':

1. '~Tb' is true-in-$\mathcal{L}$ iff it is not the case that ('Tb' is true-in-$\mathcal{L}$) (by UQI from the axiom defining negation)

2. 'Tb' is true-in-$\mathcal{L}$ iff the referent of 'b' is true-in-$\mathcal{L}$. (by UQI from the axiom defining the predicate 'T')

3. 'Tb' is true-in-$\mathcal{L}$ iff '~Tb' is true-in-$\mathcal{L}$. (by SUB from 2 and the axiom defining the singular term 'b')

4. '~Tb' is true-in-$\mathcal{L}$ iff it is not the case that ('~Tb' is true-in-$\mathcal{L}$) (by RPL from 1 and 3).

So far so good—we have a canonical proof of a T-sentence for '~Tb'. However, we can continue:

5. '~Tb' is true-in-$\mathcal{L}$ iff it is not the case that (it is not the case that ('~Tb' is true-in-$\mathcal{L}$)) (by RPL from 3 and 4).

Now we have a problem since we now have a canonical proof of a different T-sentence for '~Tb'. Following Ludwig's strategy, we can derive the following M-sentences:

(M1) '~Tb' in $\mathcal{L}$ means that it is not the case that ('~Tb' is true-in-$\mathcal{L}$).

---

[63] Ludwig stipulates that the T-sentences in this example contain no semantic terms, but for languages that contain their own truth predicates, this stipulation would prevent one from deriving canonical T-sentences for sentences containing the truth predicate.

(M2) '~Tb' in $\mathcal{L}$ means that it is not the case that (it is not the case that ('~Tb' is true-in-$\mathcal{L}$).

So, we end up with two M-sentences that attribute incompatible (indeed contradictory) meanings to a single sentence of $\mathcal{L}$. I conclude that Ludwig has not given us a workable account of canonical proof. Thus, Ludwig's suggestion for how to use an inconsistent truth theory to arrive at a consistent meaning theory fails.

5.2.4 Paradox and Validity

Hartry Field has recently argued that the standard definition of validity is untenable in light of the aletheic paradoxes because it is incompatible with every logical approach to the paradoxes. Recall the six categories of logical approaches: (i) classical gap, (ii) classical glut, (iii) classical symmetric, (iv) weakly classical, (v) paracomplete, and (vi) paraconsistent. Field only considers five of these (he leaves out (iii)). Let T be the theory composed of the logic in question plus the principle(s) of truth in question (e.g., in case (i), T is classical logic plus (T-In)).[64]

Case (i): T implies that some principles of classical logic are not truth-preserving.[65]

Case (ii): T implies that some principles of truth are not truth-preserving.[66]

Case (vi): Either T implies that some principles of truth are not truth preserving or T implies that some principles of (weakly) classical logic are not truth-preserving.[67]

Case (v): T is inconsistent with the claim that all principles of paracomplete logic are truth-preserving.[68]

---

[64] In what follows, 'implies' should be read in the proof-theoretic sense.

[65] For a liar sentence λ (i.e., λ='λ is not true'), T implies that λ is true, T implies that $\lambda \rightarrow (\sim\lambda \rightarrow \bot)$ is true, and T implies that $\sim\lambda \rightarrow \bot$ is not true. Thus, T implies that an instance of modus ponens is not truth preserving. See Field (2006a)

[66] T implies that 'λ is true $\rightarrow \lambda$' is not true, but 'λ is true $\rightarrow \lambda$' is an instance of (T-Out). Thus, T implies that an instance of (T-Out) is not truth-preserving. See Field (2006a).

[67] See Friedman and Sheard (1987) for the argument; see also Field (2006a).

[68] See Field (2006a) for the argument.

> Case (vi): Either T implies that some principles of paraconsistent logic are not truth preserving or T trivializes when conjoined with the claim that all principles of paraconsistent logic are truth preserving.[69]

Here I am including under the heading of 'truth preserving' the condition that the theory implies that all its axioms are true—the instance of an axiom is like a trivial inference rule. Although option (iii), classical symmetric theories, are left out of Field's treatment, there is good reason to think that they cannot treat validity as truth preservation since they do not even allow the move from p to 'p is true' or vice versa in hypothetical contexts. The upshot is that no matter which of the five options one chooses, one should not accept that validity is necessary truth-preservation.

In general, we have two prominent ways of thinking about validity: as the property of canons of good reasoning and as necessary truth preservation. The lesson of Field's argument can be put as: given any combination of a theory of truth and a logic, it is unacceptable that the canons of good reasoning preserve truth.[70]

This argument of Field's is relatively new, and it is buried in a much more complex discussion of Gödel's Second Incompleteness Theorem and formal theories of truth, so it has yet to generate much literature. However, I find it convincing, and this conclusion has an effect on my response to some difficult problems that arise in Part III.

This brings Part I of the book to a close. Part II attempts a realignment of the philosophical discussion about truth by emphasizing four key ideas about truth, one (truth's expressive role) from the discussions of the nature of truth, two (empirical paradoxes and revenge) from the discussions of the paradoxes, and one (internalizability) of my own creation. My contention is that, together,

---

[69] See Field (2006a) for the argument.

[70] The effects of this split can be seen all over the literature on logical approaches to the paradoxes. For example, Maudlin defines validity in terms of truth preservation, and that leads him to claim that (T-In) and (T-Out) are valid on his theory. Well, they are truth-preserving according to his theory, but they are not canons of good reasoning according to his theory. See Field (2006c) for discussion.

these ideas have dramatic consequences for many of the theories of truth and approaches to the liar outlined in Part I. Indeed, it is by reflecting on these key ideas that one appreciates the need for a new kind of theory and a new kind of approach to the paradoxes. That is the goal of Part III.

Part II

The Realignment

> Have patience with everything unresolved in your heart and to try to love the questions themselves as if they were locked rooms or books written in a very foreign language. Don't search for the answers, which could not be given to you now, because you would not be able to live them. And the point is, to live everything. Live the questions now. Perhaps then, someday far in the future, you will gradually, without even noticing it, live your way into the answer.
>
> --Rainer Rilke, *Letters to a Young Poet*, pp. 34-35

Chapter 6

What is the Use?

The first five chapters summarized work on the nature of truth and on approaches to the aletheic paradoxes. Now we turn in Part II to the four key ideas that, in my opinion, anyone concerned with the concept of truth ought to address. For these key ideas, when brought together, have significant consequences for our views on the nature of truth and how to approach the aletheic paradoxes, and by reflecting on these consequences, we are drawn to a theory of the kind presented in Part III. Each of the first four chapters in Part II is devoted to explaining one of these key ideas and to drawing out some of the more important consequences.

The first key idea is the expressive role that truth plays in our linguistic and cognitive lives. Having a truth predicate allows us to say and think things we could not otherwise. It does this in several distinct ways. Moreover, there are some important features of any linguistic practice that are crucial to keep in mind in evaluating descriptive theories of truth.

This chapter has four sections. The first is on truth's expressive role, while the second explains communication and pragmatics in general—it defends a condition that plays an important role in the rest of the book. The third section lays out some consequences of the first key idea, and the fourth summarizes the impact on the views found in the literature.

6.1 The Expressive Role

In addition to its explanatory role in philosophy, linguistics, and logic, a truth predicate serves an important expressive role. These two roles feature prominently in disputes about the nature of truth, where deflationists claim that truth's expressive role exhausts its nature, while inflationists

claim that it plays an explanatory role as well. Although the expressive role gets plenty of lip service, it is rarely the target of in-depth study. It turns out that truth plays at least two different expressive roles (as a device of endorsement and as a device of generalization) and they are not equivalent.

6.1.1 Endorsement

The uses that get the most attention in philosophical discussions are often called *lazy uses*. In a lazy use of a truth predicate, a speaker asserts a sentence containing a truth predicate (e.g., Doris asserts "snow is white' is true'), but she could instead have just asserted the sentence or sentences to which truth is being attributed (e.g., 'snow is white'). Notice that in this example, Doris used quotation marks to form a name of the sentence to which she was attributing truth. In what follows I call the object or objects to which truth is attributed the *target* or *targets* of the truth attribution.

A lazy use of a truth predicate need not involve a quoted sentence.[1] Instead, Doris could have asserted 'Church's thesis is true'. If Doris could just have easily asserted Church's thesis (i.e., 'every effectively computable function is recursive'), then this counts as a lazy use as well. Thus, the way of picking out the target of a truth attribution is irrelevant to whether it counts as a lazy use. If Doris happened to know that the first sentence uttered by a nun in Capital City on April Fools' Day of 1978 was a formulation of Church's thesis, then she could have used this description in a lazy use of the truth predicate (again, assuming she could have assured Church's thesis directly). Furthermore, lazy uses can have multiple targets (e.g., all the axioms of Robinson arithmetic—Q—are true). All that matters when it comes to lazy uses, is whether the speaker could have asserted all the targets of the truth attribution instead of asserting the truth attribution itself.

[1] It seems to me that the received view is that only truth attributions containing quote-names for their targets are considered lazy uses.

Not all our uses of 'true' are lazy. In fact, those that are not lazy are much more interesting and important for the purposes of this study. For it is the fact that we can use 'true' in cases where we cannot simply assert the targets of our truth attribution that gives the truth predicate a crucial expressive power. For example, imagine that Doris wanted to attribute truth to Church's thesis but she was unable at that moment to recall how to formulate Church's thesis. Although she might perform the same assertion (i.e., asserting 'Church's thesis is true'), in this case, her use of the truth predicate would not be lazy because she could not have simply asserted Church's thesis. There is no accepted term for non-lazy uses of truth predicates, so I call them *indispensable uses*.

There are several distinct kinds of indispensable uses. In the example just given, Doris has some sense of Church's thesis, but cannot remember exactly how to formulate it. If, instead, she did not know what Church's thesis says but had it on good authority that she should accept Church's thesis, then she might perform the same assertion. Cases where a person does not know the content of the targets of one's truth attribution are sometimes called *blind* truth attributions.[2] I am not fond of this term, but it seems to be fairly entrenched. Notice a particularly interesting kind of indispensible use occurs when the speaker does not know the number of targets to which he wishes to attribute truth. For example, Doris might assert 'everything Prof. Frink said in yesterday's lecture is true' even if she was not in attendance. Doris does not know which sentences are the targets of her truth attribution, she does not know their meanings, and she does not know how many there are; she might not even know what language he was speaking. However, this is a perfectly legitimate use of 'true', and an important one.

One of the most significant kinds of indispensible uses involves truth attributions with an infinite number of targets. These uses are unavoidably indispensible since there is no way a speaker can assert an infinite number of sentences. For example, Doris asserts 'all the axioms of Peano

[2] Schantz (2002: 6).

Arithmetic are true'. Peano Arithmetic (first order) is a formal theory of natural numbers, and it has infinitely many axioms. Thus, there is no way for Doris to assert all of them. If she wants publically to endorse Peano Arithmetic, then she has to use some indirect method like using a truth predicate.

6.1.2 Generalization

It is sometimes said that a truth predicate is a "device of generalization", which means that the presence of a truth predicate in our language permits us to formulate general claims we would not otherwise be able to formulate.[3] For example, one might think that only if monkeys do not grow on trees should a rational agent believe that monkeys do not grow on trees. Moreover, only if brown dogs bark should a rational agent believe that brown dogs bark. Indeed, the particular content of the sentences 'monkeys do not grow on trees' and 'brown dogs bark' is irrelevant. To generalize from these two cases, we might say that a rational agent should believe a proposition only if that proposition is … What goes in the blank? Here is where the truth predicate comes in: a rational agent should believe a proposition only if that proposition is *true*. Without a truth predicate, it is not obvious how we might state the general claim.

One might wonder about the relation between device-of-generalization use and the other uses discussed above. First, notice that the former is not a direct truth attribution since the truth attribution is embedded as the consequent of a conditional. The point of the claim is not that some particular propositions are true; rather, the point is to give a condition on what rational agents ought to believe, and that seems like a legitimate difference. Second, we do not need a truth predicate in order to formulate Peano Arithmetic—instead, logicians and mathematicians often use an axiom

[3] See Horwich (1998) and Field (1994a).

scheme, which is a formal device for characterizing infinitely many axioms.[4] Natural languages do not have this sort of formal device; instead, we use truth predicates. Our more familiar generalization devices (e.g., quantifiers—'all', 'some', 'none', etc.) do not work in these circumstances because they bind terms, not sentences. Instead of ordinary quantification, we can use sentential quantification in order to formulate the above theory about what rational agents should believe; we could say: for all p, a rational agent should believe that p only if p. Here 'p' ranges over propositions (it stands in for a sentence). Instead of this sort of sentential quantification, we use our everyday quantification and a truth predicate.[5] To sum up, we use truth as a device of generalization when we embed a truth attribution in a more complex sentence in order to arrive at a theory of some phenomena that generalizes over some collection of truth bearers. Without a truth predicate (or something equivalent), our language would not have these sentences.

There is an important connection between 'true' as a device of endorsement and 'true' as a device of generalization. Consider again the claim that a rational agent should believe a proposition only if that proposition is true. It follows from this generalization and the claim that a rational agent should believe the proposition that monkeys do not grow on trees, that the proposition that monkeys do not grow on trees is true. However, recall the original point of the generalization; it was to generalize a host of instances, which included: if a rational agent should believe that monkeys do not grow on trees, then monkeys do not grow on trees. In order to recover this result, we need it to follow from the proposition that monkeys do not grow on trees is true that monkeys do not grow on trees. This is exactly what is required for 'true' to function as a device of endorsement. Thus, in order for 'true' to function as a device of generalization, it has to function as a device of

---

[4] Of course, one could use a truth predicate to formulate Peano Arithmetic; see Feferman (1991) and Field (2006b) on axiom schemes and schematic variables.

[5] See Künne (2003: 357-365) for an argument that natural languages have the capacity for sentential quantification.

endorsement (or, more carefully, it has to obey the principles that allow it to function as a device of endorsement).

What about the converse? Do devices of endorsement also serve as devices of generalization? To explore this question, let 'x is endo' be a predicate that is synonymous with 'x is a sentence that expresses a proposition I endorse'. Clearly, 'endo' is a device of endorsement. Imagine Cletus asserts 'Everything Brandine said in lecture yesterday is endo'. Assume that during the lecture in question, Brandine said that grass is green. Moreover, let us assume that Cletus does not know what Brandine said. Does 'grass is green' follow from Cletus's claim? No. Does "Grass is green' is endo' follow? Yes. So Cletus has succeeded in endorsing 'grass is green' even if he doesn't know what Brandine said. So 'endo' does serve as a device of endorsement.

Despite the fact that 'endo' is a device of endorsement, there is a big difference between it and 'true'. For one, 'endo' does not obey convention T—⟨⟨p⟩ is endo if and only if p⟩ is false since ⟨if ~p then ⟨p⟩ is not endo⟩ is false as is ⟨if p then ⟨p⟩ is endo⟩. Moreover, if Cletus says 'p is endo' and Brandine says 'p is not endo', they have not disagreed; Cletus has said that he endorses p and Brandine has said that she does not. Hence, there is an important element of context dependence in a mere device of endorsement that one does not find in a truth predicate (this should be obvious from the indexical in the definition of 'endo'). That is, a mere device of endorsement is going to be context dependent, but we do not think 'true' displays this kind of context dependence. If Cletus says 'p is true' and Brandine says 'p is not true', they have disagreed; 'p is true' means the same thing in Cletus's mouth as it means in Brandine's. Therefore, there has to be more to truth than its just being a device of endorsement.

Does 'endo' serve as a device of generalization? Imagine a list of sentences:

A rational agent should believe that snow is white only if snow is white.

A rational agent should believe that grass is green only if grass is green.

Etc.

In an attempt to endorse all sentences of the form 'a rational agent should believe that p if and only if p', Cletus asserts 'a rational agent should believe a proposition only if that proposition is endo'. Has Cletus succeeded in generalizing? What follows from his claim? 'A rational agent should believe that grass is green only if 'grass is green' is endo' follows. Does it follow that a rational agent should believe that grass is green only if grass is green? No. Thus, 'endo' does not work as a device of generalization. Therefore, devices of endorsement need not serve as devices of generalization. The upshot is that truth predicates serve two distinct expressive roles and these roles are not equivalent. Philosophers routinely disregard this difference (I discuss examples below).

6.1.3 Infinite Conjunction and Disjunction?

For a while, deflationists were fond of saying that truth predicates serve as a device of infinite conjunction and disjunction. Here is a quote from Field:

> [T]he word 'true' has an important logical role: it allows us to formulate certain infinite conjunctions and disjunctions that can't be formulated otherwise. There are some very mundane examples of this, for instance, where we remember that someone said something false yesterday but can't remember what it was. What we are remembering is equivalent to the infinite disjunction of all sentences of the form 'She said 'p', but not-p'.[6]

Anil Gupta does a good job of exposing the problems with this way of characterizing truth's expressive role:

> A universal statement (e.g., [For all sentences x: (x is true and snow is white)]) does not have the same sense as the conjunction of its instances (e.g., [('Sky is blue' is true and snow is white) and ('Chicago is blue' is true and snow is white]). The two typically do not even imply the same things; they are equivalent only in a much weaker sense. I think that the proponents of the disquotational theory have gone astray because they have ignored the

[6] Field (1994a: 263-4).

> difference between wanting to affirm a generalization and wanting to affirm each of its instances.[7]

To be fair, Field does mention this point in a footnote:

> [W]hat I've given allows statements that are really a bit stronger than infinite conjunctions, in the same way that first order quantifications are stronger than the totality of their instances even when every object has a name.[8]

So, what is this dispute all about?

Gupta and Field are right that a universal generalization is stronger than even an infinite conjunction, and an existential generalization is weaker than even an infinite disjunction. It is consistent to deny a universal generalization and affirm an infinite conjunction of all its instances; likewise, it is consistent to affirm an existential generalization and deny an infinite disjunction of its instances. This is just a technical point about quantification theory.[9] The real question is: does it matter?

Take an example like the one Field gives: Elizabeth asserts 'Janie said something true yesterday'. We can think of this as an existential generalization: $\exists x(Jx \,\&\, Tx)$, where 'J' is a predicate for 'is a sentence Janie uttered yesterday' and 'T' is a truth predicate. Let s_1, s_2, … be a list of all the sentences of whatever language Janie speaks. Imagine we have some device for formulating infinite disjunctions, 'OR'. Does $OR_i\, Js_i \,\&\, Ts_i$ follow from the claim Elizabeth asserted? No, as I indicated above. So, even if we assume both (T-In) and (T-Out), it does not follow from the claim Elizabeth asserted that $OR_i\, Js_i \,\&\, s_i$, contra Field's suggestion.

Should this cause concern to deflationists? I do not see that it should. This entire line of characterizing truth's expressive role by appealing to infinite conjunctions and disjunctions is misguided—I see it as a holdover from Frank Ramsey's idea that truth predicates could be

---

[7] Gupta (1993a: 63).
[8] Field (1994a: 264n17).
[9] For more on this see Smith (2007: ch. 16). See also Halbach (1997).

systematically eliminated from any language.[10] The idea is supposed to be that generalizations containing 'true' could be rephrased as infinite conjunctions of truth attributions, which could then be reformulated as infinite conjunctions without truth predicates. The difference between generalizations on the one hand and infinite conjunctions and disjunctions on the other suggests otherwise. Indeed, the point about infinite conjunctions and disjunctions illustrates the essential expressive power we derive from truth predicates.

Gupta's paper is probably the most influential critical piece on deflationism, but it has not proven to be a decisive objection to deflationism; instead, its effect has been that deflationists stopped talking about infinite conjunctions and disjunctions, and focused instead on the idea that a truth predicate allows one to construct a sentence to assert in a given situation where one would otherwise be unable to assert anything, as in the examples of generalization above. The talk of infinite conjunctions and disjunctions seems as though it was just an unfortunate mistake in the way deflationists formulated the expressive role of truth predicates rather than a substantive problem with the view itself.

6.1.4 Intersubstitutability

There are two interrelated issues associated with explaining these expressive uses of truth predicates. The first is the status of the T-sentences (i.e., sentences of the form: $\langle \mathsf{p} \rangle$ is true iff p). Just about everyone investigating the nature of truth agrees that a theory of truth should have all the T-sentences as consequences (with possible exceptions for the paradoxical sentences). However, what is controversial is the status of the T-sentences. Is the connective in them just a material biconditional? Are they necessary? Apriori? Conceptual truths? The question confronting us in

[10] Ramsey (1926).

this section is: what status must the T-sentences have for truth predicates to perform their expressive roles?

A very closely related issue is whether truth predicates obey intersubstitutability principles. Intersubstitutabilty principles all have the following form:

(Intersub) Substitution of p for 'p is true' or 'p is true' for p in extensional contexts preserves X.

The weakest of these principles has 'truth value' for X; i.e., intersubstitution preserves truth value. In other words, if one substitutes one or more occurrences of p for 'p is true' or vice versa in some sentence q to arrive at a new sentence q′, then q and q′ have the same truth value.

Stronger intersubstitutability principles arise when X is replaced with 'intension' or 'aboutness' or 'sense'. On the first, q and q′ have the same truth value in every possible world. On the second, q and q′ are both about the same thing; e.g., "snow is white' is true' and 'snow is white' would both be about snow. On the third, q and q′ have the same meaning or content. I take it that this sequence is ordered by strength—e.g., intersubstitutability preserving sense is sufficient, but not necessary for intersubstitutability preserving aboutness.

There is a connection between the status of the T-sentences and the intersubstitutability principles. If one has a suitable conditional in one's language (e.g., a conditional for which 'if p, then p' is a necessary truth), then one can derive the T-sentences from the intersubstitutability principle. Moreover, the reading one gives of the connective of the T-sentences (material biconditional, strict biconditional, definition) is tied to the feature one takes intersubstitutability to preserve. If intersubstitutability preserves truth value, then one gets the material biconditional T-sentences. If intersubstitutability preserves intension, then one gets the strict biconditional T-sentences.

One should note that even the weakest intersubstitutability principle (i.e., preservation of truth value) is very strong—it is inconsistent in classical or even weakly classical logic. One would need something weaker like a paracomplete logic or a paraconsistent logic to endorse it.[11]

Notice also that one can explain the endorsing use and the generalizing use of truth predicates with the intersubstitution principle. Indeed, Field argues that acceptance of a particularly strong intersubstitution principle is required to explain truth's expressive role.

> Suppose I can't remember exactly what was in the Conyers report on the 2004 election, but I say
>
> (1) If everything that the Conyers report says is true, then the 2004 election was stolen.
>
> Suppose that what the Conyers report says is A_1, …, A_n. Then relative to this last supposition, (1) better be equivalent to
>
> (2) If A_1 and … and A_n, then the 2004 election was stolen.
>
> And this requires True(<A>) to be intersubstitutable with A even when A is the antecedent of a conditional.[12]

The key to understanding this passage is grasping what Field means by 'equivalent'. Throughout the paper, he uses 'equivalent' as short for 'cognitively equivalent', which he defines in a footnote: "to call two sentences that a person understands 'cognitively equivalent' for that person is to say that the person's inferential procedures license a fairly direct inference from any sentence containing an occurrence of one to the corresponding sentence with an occurrence of the other substituted for it. … I would also take the claim of cognitive equivalence to imply that the inferences are more or les indefeasible."[13] With this reading, we see that Field is appealing to at least an intensional intersubstitutability principle (i.e., one that says p and 'p is true' have the same truth value in all possible worlds). Perhaps he wants something even stronger, but the intensional version is enough for our purposes. He says in the first quotation that if the T-sentences are not necessary or we do

---

[11] See Chapter Three for discussions of these kinds of logics.
[12] Field (2008a: 210).
[13] Field (1994a: n2).

not have an intensional intersubstitutability principle, then truth predicates cannot perform their expressive role.

6.2 Communication and the Gricean Condition

In this section, we switch gears and focus on how we use language in general, not just truth predicates. We will arrive at what seems like an innocent principle governing language use in general, but it turns out that it has far reaching consequences for views on the nature of truth and approaches to the paradoxes.

6.2.1 Communication

It is a truism that a primary use of language is for communication. Some philosophers claim that we use language to think as well, but either way, communication is one of the major reasons we have language.[14] The syntactic, semantic, and pragmatic features of words and sentences are geared toward facilitating conversation. Although there is no accepted way of demarcating these realms, the following is a rough sketch. *Syntax* is the study of the basic words of a language and how they combine into more complex phrases and sentences. *Semantics* is the study of the meanings of basic words and how they combine into meanings of more complex phrases and sentences. *Pragmatics* is the study of the features linguistic expressions have by virtue of being used and the features of linguistic acts.

One assumption linguists and many philosophers of language make regarding communication is that it is a rational enterprise. That is, the participants in conversation have certain goals in mind and perform linguistic actions in the interest of achieving those goals. Paul Grice is one of the

[14] See Sellars (1969).

major figures in this area—he claimed that participants in conversation follow a cooperative principle, which states: "Make your contribution such as required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged."[15] Accordingly, when a speaker utters a sentence in a conversation, that speaker is trying to accomplish some goal (and frequently has several goals in mind), and has chosen that sentence to utter because by uttering it, as opposed to uttering any other sentence or not uttering anything at all, the speaker believes she is most likely to achieve her goal. The audience too uses this principle in interpreting the speaker. From this one principle, Grice derived a list of maxims that govern conversation. One can violate these maxims in conversation as long as one makes it clear to the audience that that is what is happening; indeed, many familiar linguistic phenomena arise in this way. For example, one of the maxims is that one should assert only what one believes to be true. However, sarcasm occurs when one violates this maxim by saying the opposite of what one believes; nevertheless, communication with the sarcastic can still be successful as long as the speaker gives the audience some indication that he is being sarcastic.

One of the points to come out of Grice's work is the extremely complicated web of phenomena associated with communication. There is far more to it than just the speaker picking a sentence to utter and the audience figuring out what the sentence means. Most of those phenomena fall under the heading of pragmatics.

6.2.2 Pragmatics

It is not easy to say exactly what the topic of pragmatics is. The problem is that there are many distinct but interconnected phenomena associated with pragmatics, and there are disputes about the

[15] Grice (1975: 45).

legitimacy of these phenomena and how best to explain them. Right now there are two dominant traditions—neo-Griceans and Relevance theorists.[16] Neo-Griceans usually accept Grice's framework, but attempt to replace his complicated list of conversational maxims with just two or three general principles. Relevance theorists, on the other hand, attempt to explain all conversational phenomena by appeal to the principle that linguistic utterances carry a presumption of their own maximal relevance. Relevance, for these theorists, is the optimal combination of most information conveyed and least cognitive effort. Relevance theorists borrow heavily from cognitive science in their theorizing, whereas neo-Griceans stick closer to mainstream linguistics.

Both neo-Griceans and relevance theorists accept Grice's distinction between what a speaker says and what a speaker means; the terms 'sentence meaning' and 'speaker meaning' are also used to mark the distinction. That there is a need for this distinction should be evident from examples like the one above where a person uses sarcasm; in a sarcastic utterance, the speaker meaning is frequently the opposite of the sentence meaning, and audiences have no trouble tracking this difference as long as the speaker gives them enough clues.

One might think that sentence meaning is a matter for semantics, whereas speaker meaning is a pragmatic phenomenon, but, in fact, there are several pragmatic issues that can come into play before one determines sentence meaning. These are sometimes called *near-side* pragmatic features since they occur prior to the determination of sentence meaning. They include any information the audience gleans from the context of utterance to determine sentence meaning; e.g., disambiguation, context dependence, reference determination, and anaphora are all near-side pragmatic phenomena. On the other hand, everything that figures into speaker meaning beyond the sentence meaning is

[16] Work by neo-Griceans includes Horn (1984, 1989, 2004, 2005, 2007a, 2007b), Levinson (1983, 2000), Atlas (1989, 2005), Bach (1987a, 1987b, 1994, 2001), Bach and Harnish (1979), Roberts (1996, 2002, 2003, 2004) and the papers in Horn, Birner, and Ward (2006); work by relevance theorists includes Sperber and Wilson (1986, 1993, 1996, 1997, 1998 2002a, 2002b, 2004, 2008), Carston (1998, 1999, 2002, 2004, 2005, 2007, 2008, 2009), the papers in Carston and Ushida (1998), and the papers in Novek and Sperber (2004). For overviews, see Sperber and Wilson (2006), Bach (2006), Carston and Powell (2006), Szabo (2006b), and the papers in Horn and Ward (2004).

often called *far-side* pragmatics. It includes phenomena like presuppositions, and conventional and conversational implicatures.

Theories of illocutionary force also belong in the realm of pragmatics, but they are difficult to place. Frequently one needs to determine illocutionary force before doing farside pragmatics. Depending on how one draws the line between semantics and pragmatics, felicity determinations and presuppositions could play a role in phenomena called *impliciture* or *explicature*, which occur when one has to use aspects of context to determine the truth-conditional content of an assertion.[17] So, on some views, some aspects of pragmatic presupposition could be placed on the nearside.

To illustrate some semantic and pragmatic phenomena, consider the following conversation between Ned and Lenny:

Ned: Do you and Carl like the First Church of Springfield?

Lenny: I stopped going, but Carl likes it.

Consider Lenny's utterance (in what follows, technical terms explained above are in italics). The sentence he utters is *syntactically well-formed* and *meaningful*, and it has the mood of a declarative sentence. Lenny's utterance seems to have the *illocutionary force* of an *assertion*, which suggests that it has a *truth-conditional meaning*.[18]

Focus now on how Ned interprets Lenny's utterance. To determine its meaning, Ned decodes the *conventional meanings* of each word in the sentence (e.g., to like something is to have a positive attitude toward it) and uses the syntactic structure of the sentence to put these meanings together (the explanation for how these words get their conventional meanings and how Ned comes to know these meanings is a matter for a theory of meaning, which I do not address here).

---

[17] Do not confuse the terms 'implicature' and 'impliciture'. See Carston (1988), Bach (1994, MS), and Cappelen and Lepore (2005) for discussion.

[18] If Ned was unable to find a suitable truth-conditional meaning for Lenny's sentence, then he might reconsider the illocutionary force of Lenny's utterance.

However, the structure of the sentence and the conventional meanings of its parts alone are not enough. To arrive at a truth-conditional meaning for it, Ned needs to go though several near-side pragmatic processes. In particular, he needs to assign a value to the *context dependent* element, 'I'; to do this he can follow the conventional rule of letting 'I' refer to the speaker of the sentence—in this case, Lenny. Ned also needs to assign a value to the pronoun 'it'. In this case, it is an *anaphoric dependent* that inherits its value from the definite description 'the First Church of Springfield' used in Ned's sentence. In addition Ned needs to *fix the referent* of the name 'Carl', since many people have that name. In this case it is obvious from the context that 'Carl' refers to the same person Ned referred to when he used that name. Furthermore, depending on one's views about the kinds of truth-conditional meanings sentences have, Ned might need to make explicit exactly what Lenny stopped going *to*, namely the First Church of Springfield. Ned draws this inference because it is the only relevant one in the context. This pragmatic feature has been called *explicature* by relevance theorists and *impliciture* by some neo-Griceans (e.g., Bach).[19]

The result of these considerations leaves us with the proposition that Lenny stopped going to the First Church of Springfield, and Carl likes the First Church of Springfield as the truth conditional content of Lenny's sentence. Following Grice, linguists frequently take this to be *what is said* by Lenny's utterance (i.e., *sentence meaning*). Notice that it has a simple conjunction, 'and', instead of the contrastive 'but'. The fact that Lenny is contrasting his behavior with that of Carl is typically not taken to be part of the truth conditional meaning of the sentence; instead, it is treated as a *conventional implicature*. This is a far-side pragmatic phenomenon.[20] That is, it is not part of the sentence meaning or what Lenny's utterance says. Instead, for the Gricean, it is part of what Lenny

---

[19] See Carsten (1988, 2004) and Bach (1994, MS). Others (e.g., Capellen and Lepore) would say that the first conjunct just means that *Lenny stopped going*; where the information about what Lenny stopped going to is a far-side pragmatic phenomenon (conversational implicature). See Capellen and Lepore (2005).
[20] See Potts (2005) for a new view on the nature and function of conventional implicature.

conveys to his audience in this particular circumstance by uttering a sentence with this particular meaning.

Other far-side pragmatic phenomena include the *presupposition* that Lenny used to go to the First Church of Springfield. That is not usually taken to be part of the meaning of the sentence. Lenny's utterance is felicitous in this context (in part) because it is common knowledge between Ned and him that Lenny used to go to this church. If Lenny had never gone to that church, then it might have been infelicitous for him to utter this sentence.[21] Another is the *conversational implicature* that Lenny did not like the First Church of Springfield.[22] Ned figures this out either by using a conversational rule (according to neo-Griceans) or by relevance considerations (according to relevance theorists). Either way, Ned takes Lenny to be cooperating in the conversation by providing enough information to answer his question. Ned reasons that, in the absence of any further information, if Lenny quit going to the church, then he probably did not like it. The conventional implicature from the contrastive 'but' helps here—Carl likes the church, and this attitude is contrasted with that of Lenny. Following Grice, some linguists and philosophers of language would say that the proposition that Lenny does not like the First Church of Springfield and Carl does like the First Church of Springfield is what Lenny meant (i.e., *speaker meaning*). Linguists typically contrast far-side pragmatic phenomena like implicatures and presuppositions with semantic phenomena like *entailments*.[23] For example, Lenny's sentence entails that someone likes it. Entailments are taken to be part of the truth-conditional meaning of the sentence.

One might wonder how to decide whether some information the hearer comes to through this linguistic exchange counts as truth-conditional meaning, an entailment, an explicature, an implicature

[21] Presuppositions can sometimes be accommodated, which means that the common knowledge of the conversation is updated to include the information presupposed; see Lewis (1979a), Simons (2003), and von Fintel (ms) for discussion.
[22] Conventional implicatures attach to particular words and are present in each context of utterance, whereas conversational implicatures are generated by the interplay of the sentence uttered and the context of utterance.
[23] One often hears linguists saying that entailments, presuppositions, and implicatures are *kinds* of meaning. That way of using 'meaning' usually sounds odd to philosophers of language.

or a presupposition. These are very delicate issues and linguists and philosophers of language have developed an interconnected set of diagnostics for determining what kind of phenomenon one is dealing with (Information Box 26 has some further details[24]). Very often, the details of these diagnostics depend on controversial theoretical commitments, and so are a matter of dispute. We are lucky that we do not need to delve into these issues for our purposes.

[24] Information Box 26 is based on Partee (MS).

Semantic and Pragmatic Consequences of an Utterance Information Box 26

Assume that an agent utters a sentence, *s*, and *s* expresses the proposition P in this context.

An *entailment* of *s* is a proposition Q such that it is impossible that P is true and Q is not true.

A *semantic presupposition* of *s* is a proposition that must be true for P to be either true or false.

A *pragmatic presupposition* of the utterance of *s* is a proposition that must be in the common ground of the conversation in question for the utterance of *s* to be felicitous.

A *conventional implicature* of *s* is a proposition that is part of the conventional meaning of *s*, but not part of the truth conditions of *s*, and arises from the particular words that compose *s*.

A *conversational implicature* of the utterance of *s* is a proposition that is implied by the utterance by virtue of conversational maxims.

A phenomenon is *conventional* just in case it is generated by the conventional meaning alone.

A phenomenon is *cancelable* iff further utterances can deactivate it.

A phenomenon is *backgrounded* iff it should be part of the common ground for an utterance associated with it to be felicitous.

A phenomenon *detaches* iff it need not be present in an utterance of a truth-conditionally equivalent sentence

A phenomenon *embeds* if it is preserved when the sentence uttered is embedded under truth-functional operators.

	Property of	*Conventional?*	*Cancelable?*	*Backgrounded?*	*Detaches?*	*Embeds?*
Entailment	Sentences	Y	N	N	N	N
Semantic Presupposition	Sentences	Y	N	Y	Y	Y
Pragmatic Presupposition	Utterances	Y/N	Y/N	Y	Y	Y
Conventional Implicature	Sentences	Y	N	N	Y	Y
Conversational Implicature	Utterances	N	Y	N	N	Y/N

Y: Yes N: No Y/N: Sometimes Yes, sometimes No

6.2.3 Pragmatic Theories

There are several proposals for how to understand the way conversations develop. Here, I review Robert Stalnaker's notion of common ground, David Lewis's scorekeeping theory, and Craige

Roberts' model; these form a progression in the sense that Lewis incorporates Stalnaker's model and Roberts incorporates Lewis' model.

Stalnaker's theory of discourse focuses on assertion and presupposition. The main idea is that when a speaker makes an assertion in a conversation, the content of the sentence asserted furthers the conversation in a certain way. In particular, the content rules out ways the world might be that were previously live options in the conversation. If the content asserted is accepted by everyone in the conversation, then the potential ways the world might be have been narrowed, and that is one of the central goals of conversation.

To model this idea, assume that we have a conversation consisting of several people. Each person has many beliefs. Stalnaker defines a participant's *presupposition* as a purportedly shared belief in the conversation. It requires that the participant believes it, the participant believes that everyone else believes it, the participant believes that everyone else believes that everyone else believes it, and so on. Since beliefs are often taken to be attitudes toward propositions, and propositions are often taken to determine a set of possible worlds in which they are true, we can simplify matters by talking about propositions. A participant's presupposition will divide the class of possible worlds into two—those in which the presupposed proposition is true, and the rest. If a participant's presupposition is also a presupposition of all the other participants, then it is a *shared presupposition*. The set of shared presuppositions is called the *common ground*—it is what everyone in the conversation agrees on, agrees they agree on, and so on. The crucial notion for Stalnaker's view is the set of possible worlds in which all the propositions in the common ground are true; call this the *context set*. As the conversation develops, the common ground expands and the context set shrinks. When a participant in the conversation makes an assertion, the proposition asserted should not be entailed by the common ground; that is, it should be false in some worlds in the context set prior to

the assertion.[25] That way, if everyone in the conversation accepts the assertion, it narrows the context set. Stalnaker's model of conversation has been extremely influential, and offers a powerful explanation for a variety of pragmatic phenomena.[26]

Lewis's views on conversations are similar to Stalnaker's in that it treats conversations as rule-governed, and Lewis' model incorporates Stalnaker's ideas of presupposed propositions and common ground, but is more general. Lewis begins with an analogy between the score in a baseball game and the score in a conversation. One can model the score in a baseball game as a septuple with entries for visiting team runs, home team runs, half of the inning, inning, strikes, balls, and outs.[27] Rules of baseball then come in four kinds:

(i) *Specifications of the kinematics of score*: rules that specify how the score changes over time in response to the behavior of players (e.g., a home team runner crossing home plate without being tagged out as the result of a hit or steal increases the home team runs by one).

(ii) *Specifications of correct play*: rules that specify what is permissible and obligatory behavior for the players as determined by the score (e.g., if in the top half of an inning, the outs reaches three, then the home team players leave the field, while the visiting team players take the field).

(iii) *Directive requiring correct play*: all players ought to obey the specifications of correct play at all times.

(iv) *Directives concerning score*: players try to make the score change in certain ways (e.g., visiting team players try to increase the visiting team runs, visiting team players try to prevent home team runs from increasing, and in the bottom half of innings, visiting team members try to increase the outs).[28]

Lewis suggests that conversations can be usefully modeled along the same lines. The conversational score consists of a mathematical structure that includes "sets of presupposed propositions,

---

[25] Obviously, this feature of Stalnaker's model is an idealization since it would rule out asserting necessary propositions (e.g., in mathematical conversations).
[26] See Stalnaker (1970, 1973, 1974, 1978, 1987, 1998, 1999, 2002, 2004, 2009).
[27] Notice that Lewis' formulation is incomplete since it does not account for runners, batting order, pinch hitters, etc., all of which might affect specifications of correct play and directives concerning score.
[28] Lewis (1979a: 236).

boundaries between permissible and impermissible courses of action, and the like."[29] The four types of rules carry over. (i) Conversational score changes in a rule-governed way in response to the behavior of participants (e.g., when an assertion is accepted by everyone, the proposition asserted gets added to the common ground). (ii) Acceptable behavior for the participants at any stage in the conversation is determined by the score (e.g., it is unacceptable to assert something that has already been accepted by everyone). (iii) Participants are expected to cooperate by following the rules for acceptable behavior. (iv) Participants try to change the score in certain ways (e.g., a speaker attempts to get others to accept what she believes by making assertions in the hopes that they are accepted and added to the common ground). The beauty of Lewis' model is that it is not restricted to assertions—it is designed to handle commands, questions, suppositions, challenges, promises, and a wide range of other discourse actions; it also allows information to come off the record through retraction or accommodation. Although not as influential as Stalnaker's model, it too has been used to explain a variety of pragmatic phenomena.[30]

Roberts expands Lewis's model and fills in many of the details for dealing with non-assertoric utterances. According to Roberts, the conversational score consists of the following structures (these should be thought of as relativized to a time t):

---

[29] Lewis (1979a: 238).

[30] Robert Brandom offers a novel variant of Lewis's model, which takes as primitives the notions of *deontic status* and *deontic attitude*. Statuses come in two flavors: *commitments* and *entitlements*. The former are similar to responsibilities and the latter are similar to permissions. There are three types of attitudes: *attributing*, *undertaking*, and *acknowledging*. One may attribute, undertake, and acknowledge various commitments and entitlements. There are several different kinds of commitments that correspond to aspects of discursive practice. *Doxastic commitments* correspond to assertions and beliefs, *inferential commitments* correspond to reasons, and *practical commitments* correspond to intentions. The members of a discursive practice keep track of each other's commitments and entitlements. At a given moment in a conversation, the score is just the set of commitments and entitlements associated with each participant. Each member of the conversation keeps score on all the participants (including herself). Every time one of the participants undertakes (implicitly adopts), acknowledges (explicitly adopts), or attributes (takes another as if he adopts) a commitment or entitlement, it changes the score. Moreover, each participant keeps two sets of books on the other participants—one for the commitments and entitlements of that participant according to what that participant accepts, and one for the commitments and entitlements of that participant according to what the scorekeeper accepts. See Brandom (1994, 2001) for details; see also Lance and Kremer (1994, 1996), Lance (2001), Lance and Kukla (2009), Restall (2008, 2009), and John MacFarlane's program GOGAR (Game Of Giving and Asking for Reasons) at http://johnmacfarlane.net:9094/.

(i) I: A set of *interlocutors* at time t.

(ii) G: A function from pairs of individuals in I and times t to sets of *goals* in effect at t such that for each i∈ I and each t, there is a set, G(<i, t>), which is i's set of goals at t.

(iii) $G_{com,}$: the set of *common goals* at t; i.e., {g| for all i∈ I, g∈ G(<i, t>)}.

(iv) M: The set of *moves* made by interlocutors up to t with the following distinguished subsets—A, the set of assertions; Q, the set of questions; R, the set of requests; and Acc, the set of accepted moves.

(v) <: a total order on M that reflects the chronological order of moves.

(vi) CG: The *common ground*; i.e., the set of shared presupposed propositions at t.

(vii) DR: The set of *discourse referents*; i.e., the ontological commitments of the claims in CG.[31]

(viii) QUD: The set of *questions under discussion* at t; i.e., a subset of Q∩Acc such that for all q∈ QUD, CG does not entail an answer to q and the goal of answering q is a common goal.

On Roberts' model, the conversational score is updated in the following ways:

(i) Assertion: if an assertion is accepted by all the interlocutors, then the proposition asserted is added to CG.

(ii) Question: if a question is accepted by all the interlocutors, then the set of propositions associated with the question is added to QUD. A question is removed from QUD iff either its answer is entailed by CG or it is determined to be unanswerable.

(iii) Request: If a request is accepted by an interlocutor, i, then the goal associated with the request is added to G_i, and the proposition that i intends to comply with the request is added to CG.

One nice aspect of Roberts' model is that it relates the conversational score back to Grice's original insight that participating in a conversation is a rational enterprise—each participant has certain beliefs and desires, and each participant engages in the conversation to rationally further his ends. The common goals of the conversation and the question under discussion are meant to help explain why the participants are engaging in a conversation at all, and why they pursue their own particular

[31] This element plays a role in modeling anaphora.

strategies in the conversation. These structures also allow Roberts' model to explain the pragmatic significance of questions and commands. Roberts' model is relatively new, but she uses it to explain a number of recalcitrant data pertaining to demonstratives, anaphora, definite descriptions, ellipsis, and prosodic deaccentuation.[32, 33]

For the rest of the book, I refer back to these discourse models when discussing various views on the nature of truth and approaches to the aletheic paradoxes. It turns out that many theories of truth are incompatible with basic assumptions about conversation built into these models. Given some methodological considerations discussed below, these incompatibilities lead me to reject the theories of truth in question.

6.2.4 The Gricean Condition

In this subsection I am going to argue for a condition on sentence meanings. It should not be a controversial condition; indeed, if you ask any linguist about it, you will probably get a blank stare in response, as if to say "who in their right mind would ever doubt that?!". Nevertheless, it turns out that many theories of truth (including views on the nature of truth and philosophical approaches to the paradoxes) violate this condition. The Gricean Condition together with the fact that truth predicates have expressive uses (e.g., endorsement) make a powerful combination. Later, when we add empirical paradoxes and revenge paradoxes to the mix, we arrive at a seemingly insurmountable challenge to any unified theory of truth. The goal of Part III is to show that there is, however, at least one unified theory of truth that can meet the challenge.

---

[32] See Roberts (1996, 1998, 2002, 2003, 2004, 2005, 2010).

[33] I regret that I do not have the space to discuss dynamic semantic theories, which would fit well here. Although these have had a huge impact in linguistics, they have been almost completely ignored by philosophers of language; see Kamp (1981), Heim (1983), Groenendijk and Stakhof (1991), Muskens (1996), Beaver (2001), Dekker (2010), and van Eijck and Visser (forthcoming) for details. For what it is worth, I think that when philosophers catch on to this trend, there will be a revolution in philosophy of language similar to the one that has occurred in linguistics.

In accordance with the idea that communication is a rational enterprise, and the claim that linguistic expressions and actions have their syntactic, semantic, and pragmatic properties in order to facilitate communication, we can conclude that the syntactic, semantic, and pragmatic features of linguistic expressions and actions are *available* or *retrievable* by the participants in a conversation.[34] This idea is a consequence of each of the three models of communication described above. Here is Roberts' formulation:

> (Retrievability) In order for an utterance to be rationally cooperative in a discourse interaction D, it must be reasonable for the speaker to expect that the addressee can grasp the speaker's intended meaning in so-uttering in D.[35]

Roberts focuses on semantic features, but the same point holds for syntactic and pragmatic features. That is, the participants in a typical conversation can figure out the syntactic features of linguistic expressions, the semantic features of linguistic expressions, and the pragmatic features of linguistic actions. For example, it would be very odd to say that some particular word of a particular natural language is an adjective even though no one knows that about it. Likewise, it would be hard to make sense of the claim that some particular sentence of a particular natural language has such and such meaning even though no one knows that it does. The same goes for pragmatic features—I would not know how to respond to some one who said that one of my utterances is an assertion even though no one could ever tell.

Why do these claims sound so strange? I think the answer is that, because a primary use for our linguistic expressions and actions is communication, it would not make sense if they had communication-relevant features that were inaccessible to conversational participants. In addition, it seems as though our linguistic expressions and actions have these features because of how we treat them and what we do with them. It would not make sense to say a word had some meaning even

---

[34] Stalnaker uses the term 'available' in Stalnaker (1999) and Roberts uses the term 'retrievable' in Roberts (1996, 2010).
[35] Roberts (2010).

though no one ever took it to have that meaning since there does not seem to be any other explanation for how it came to have that meaning in the first place. However, this explanatory claim brings us into complex and controversial issues in the theory of meaning, and I do not intend to defend it here.

A careful formulation of what I will call the Gricean Condition is:

(GC) If S is a competent speaker of a natural language L, and S understands all the linguistic expressions in a lexicon Λ of L, then S can (easily)[36] come to know the syntactic, semantic, and pragmatic properties of expressions composed of elements of Λ when they are used felicitously in a conversational context.

Again, this should not be a controversial principle. Nevertheless, here are some reasons to accept it. Assume a speaker performs a felicitous utterance and the syntactic or semantic features of the sentence uttered, or the pragmatic features of the utterance are unavailable to the audience. By 'felicitous' we mean that the utterance conforms to the standards set by the conventions of conversation nicely summarized by Grice as "be cooperative" but explicated in multiple ways, the most familiar of which is the four maxims: quality (provide information that you have good reason to believe is accurate), quantity (provide the right amount of information), relevance (provide information that is relevant to the topic at hand), and manner (provide information in an accessible way). In following these maxims, the speaker intends to convey some information to the audience and intends that the audience recognizes this intention.[37] If the hearer is unable to recover the proposition expressed, then either the hearer is incompetent in some way or the speaker has been negligent in performing the utterance.

This argument does little but remind the reader that the Gricean condition is constitutive of communication as we understand it today. The key is in the reflective nature of the speaker's intention. This element of communication insures that if the hearer is competent and the utterance

[36] By 'easily', I mean with minimal effort, without consulting a dictionary, etc.
[37] Grice (1989).

is felicitous, then the syntactic, semantic, and pragmatic features are retrievable. More generally, the idea that linguistic communication is a variety of *strategic interaction* underwrites the Gricean condition. That is, we do not analyze linguistic communication as the interaction between a rational agent and its merely physical environment. Of course, some aspects of communication are explained at this level (e.g., phonetics), but not all of it is. Instead, we analyze linguistic communication as an interaction between a rational agent and other rational agents, where each agent recognizes that the others are rational and reasons practically about its own behavior by considering what the other rational agents might do. That is what is known as strategic interaction. Unless the syntactic, semantic, and pragmatic features of a felicitous utterance are retrievable, it is impossible to make sense of why the rational agents in question are engaging in communication.

There is much more to say about what it is for a proposition to be retrievable: what is it to grasp a proposition? What is it for a proposition to be grasp*able*? What constitutes competence on behalf of the audience? These are complex questions but we need not go into the suggested answers to them because these debates all take something like the Gricean condition for granted—if one proposed a theory of what it is for a proposition to be graspable by a competent hearer that violated the Gricean condition, it would be inadequate.[38]

Many philosophers of language are onto the fact that the Gricean Condition is an accepted tenet of linguistics. For example, in their paper on quantifier domain restriction, Jason Stanley and Zoltan Szabo confine their attention to "typical assertions" which, they assume, are "successful just in case the hearer can identify the proposition the speaker intends to communicate."[39] Stalnaker too indicates his acceptance of the Gricean condition when he makes the following claim about context dependence:

---

[38] For discussion of these issues, see Chomsky (1986, 1995), Pettit (2002, 2006), Gross (2005, 2006), Longworth (2008a, 2008b, 2009), Devitt (2006), and the papers in Barber (2003).
[39] Stanley (2007: 75).

> It is a substantive claim that the information relevant to determining the content of context-dependent speech acts is presumed to be available to the participants of a conversation—that it is included in the presuppositions of the context—but it is a claim that is motivated by natural assumptions about the kind of action one performs in speaking. It is not unreasonable to suppose that speakers, in speaking, are normally aiming to communicate—at least to have the addressees understand what is being said. Succeeding in this aim requires that the information relevant to determining content be available to the addressee.[40]

Stalnaker is right—the Gricean Condition is a substantive claim; however, without it, the sense in which communication is a rational enterprise would be lost to us.

Notice that the Gricean condition is an assumption of many of the most well-known tenets of semantics. Take, for example, compositionality: the meanings of complex expressions are be determined by their structure and the meanings of their constituents. Compositionality is supposed to explain an amazing feature of linguistic interaction; namely, that natural language users have the ability to produce and understand an unlimited number of sentences from a finite set of resources. Unless one thinks that the meanings of complex expressions are retrievable in normal circumstances, there is no reason to treat compositionality as a condition on a semantic theory. Of course, these issues are complex and subtle, but the details need not concern us here. The big idea is that the Gricean condition is a central tenet of linguistics.

For illustration, consider some potential counterexamples. Assume that, in a conversation with Ned, Lenny utters 'Carl is at the bank', but Ned is unable to disambiguate 'bank' to determine whether Lenny is saying that Carl is at the $\text{bank}_{\text{financial}}$ or the $\text{bank}_{\text{river}}$. I am not saying that that could not happen—just that if it did happen, then Carl would not be using 'bank' felicitously. When language is used properly, competent audience members can understand it.

Consider another potential counterexample. 'here' can have its content determined automatically by the location in which a given sentence is uttered. However, if features unavailable to the conversational participants determine the content of a sentence containing 'here', then it

---

[40] Stalnaker (1999: 6).

violates Grice's cooperation principle to utter that sentence in that conversational context. Imagine Apu is hiking when he loses his way and becomes lost. He does not know how to get back or how to find the trail. None of the geographical features is familiar. He does, however, have cell-phone service. So, he calls his partner, Manjula, in the hope that she will know how to help him. He tells her that he is lost, and she asks him to explain. Then he asserts 'oh, wait, I've figured out where I am', followed by asserting 'I am here'.

The sentence Apu asserted, 'I am here', is syntactically well-formed and it has a meaning. The context in which he asserts it determines the proposition it expresses. Moreover, the sentence is true, he believe the proposition it expresses, and he has good reason to believe it. However, his assertion is infelicitous in the sense that in asserting it, he violates conversational rules—he has not conveyed any information to Manjula. She would, of course, respond with "that isn't helpful," or "why are you telling me that?", or "at least you haven't lost your sense of humor", or some other utterance that indicates his assertion is not accepted as legitimate.

Again, according to Grice, every time one performs a speech act, one has to intend that one's audience recognize one's intention to convey a certain meaning. It simply does not make sense for a speaker to intend that his audience recognize his intention to convey a certain meaning even though neither of them is capable of determining the meaning of the sentence in question.

Apu could, however, say 'It's cold here'. So it cannot be that it is always wrong to use a context-dependent term in this way. Neither of them knows where Apu is, so it might seem that neither of them knows the content of that claim. Is this a counterexample to the Gricean Condition? Not necessarily. Although the topic of the vocabulary that may be used by a semantic theory to pick out the contents of context-dependent expressions is relatively unexplored, there is no reason to think that 'Apu's current location' could not be used for this purpose. If so, then the content of Apu's sentence would be the same as that of 'it is cold at Apu's current location'. There is no problem

with him asserting a sentence with that content in the context in question—it is retrievable by Manjula. If it turns out that Apu is on Mt. Useful, then the content of the sentence has the same truth conditions as 'It is cold on Mt. Useful'. However, if Apu intended to convey *this* information to Manjula, then, again, his assertion would be improper.

The lesson is that every major pragmatic theory endorsed by linguists has the Gricean Condition as a consequence.[41] Moreover, it follows from the Gricean characterization of communication as a rational endeavor, and it is a feature of all the most widely accepted models of discourse.

If one's favored view of truth or approach to the aletheic paradoxes conflicts with the Gricean Condition, then it is the philosophical view that should go. Thomas Hofweber, who has recently explored this relationship between philosophy and the sciences, writes:

> The *modest attitude* towards the relationship between the sciences and philosophy (modest from the point of view of philosophy) holds that the sciences don't need philosophy for their final vindication, nor does philosophy have the authority to overrule the results of the sciences. They are just fine without us. Collectively, that is. Individual philosophers can of course fruitfully join in on the scientific enterprise, and help out in ways that their philosophical training has especially prepared them for. What is at issue is not that, but how the results of philosophy and metaphysics, the disciplines, relate to those of the sciences. To have the modest attitude is not to have science worship. One can have the modest attitude and be critical of various sciences. One might hold that a particular science overstates its claims, or hasn't gathered enough evidence to be accepted as true, or the like. But what one can't do, with the modest attitude, is to hold that there is an open philosophical question whether p is the case even though one of the acceptable sciences has shown something that immediately implies p.[42]

I think this is exactly right. To endorse a philosophical theory of truth that conflicts with the science of linguistics is just as condemnable as being a creationist or a flat-Earther or a proponent of any other non-empirical superstition. We philosophers should be past this by now.[43]

---

[41] One might worry whether relevance theorists accept it. Not only is it an easy consequence of relevance theory, relevance theorists explicitly endorse it; see Sperber and Wilson (2002a).
[42] Hofweber (2009: 263).
[43] For discussion of linguistics and scientific methodology, see Devitt (2006a, 2006b, 2009, 2010), Culbertson and Gross (2009), Textor (2009), Fitzgerald (2010), Slezak (MS), and Sprouse and Almeda (MS), and the papers in Katz (1985) and Everaert et al (2010).

I imagine that some readers want to protest that the modest attitude makes philosophy impotent to overturn findings in the sciences. This objection is off the mark. If a philosophical position implies that an empirically supported tenet of a science is false, then there are several courses of action that might transpire. Imagine that I give a philosophy talk in which I present a new theory of physical objects and it comes out that a consequence of my theory is that there are no gauge bosons (e.g., photons and gluons). Of course, if there are no gauge bosons, then the standard model of particle physics is false.[44] In this case, it seems perfectly legitimate to reject my theory of physical objects out of hand. However, if I present an alternative to the standard model of particle physics that is both compatible with my theory of physical objects and as empirically verified as the standard model, then I am back on firm ground. The modest attitude implies that if a philosophical theory is incompatible with a tenet of one of the sciences, then empirical confirmation should be the deciding factor.

Another might objection to my use of the modest attitude might be: linguists assume that propositions are sets of possible worlds, and this is obviously false since it would make all true mathematical sentences synonymous. So it seems that philosophical considerations can trump the assumptions of scientists. The modest attitude would require that philosophers reject any theory that takes propositions to be something other than sets of possible worlds.

My reply: yes, linguists do assume that the Gricean condition is true and they assume that propositions are sets of possible worlds. However, these two assumptions play very different roles in linguists' theories. If linguists give up the claim that propositions are sets of possible worlds and accept that propositions are individuated much more finely, their semantic and pragmatic theories will still deliver the same empirical results. All that is needed for the theories to be empirically significant is that propositions *determine* sets of possible worlds, not that they are identical to sets of

[44] I am ignoring issues associated with anti-realist interpretations of scientific theories.

possible worlds. When asked about the problem, linguists will freely admit that the identification of the two is an *idealization*. Of course, with the idealization in force, semantic theories deliver false results with respect to mathematical discourse, but dropping the idealization does not affect the empirical predications and explanations offered by semantic theories for the other areas of discourse on which these theories have done such a good job. On the other hand, if we look at what happens to these theories when we drop the Gricean condition, then we see that they lose their explanatory power and empirical predications. If there is no reason to think that the propositions expressed by sentences uttered in communication are available to the participants of the conversation, then we no can no longer make sense of why these participants are engaging in conversation at all. Communication stripped of strategic interaction is just making noise. In sum, not all assumptions are created equal. The modest attitude pertains to tenets of the sciences, not idealizations. The claim that propositions are sets of possible worlds is an idealization linguists make, while the Gricean Condition is a tenet of linguistics.

6.2.5 Anti-descriptivism and the Gricean Condition

Before moving on to the consequences of the first key idea, I want to consider a particular philosophical challenge to the Gricean Condition. My guess is that two potential problems might jump to mind for readers familiar with contemporary philosophy of language. The first is semantic externalism and the second is Millianism (or direct reference theories). Both arise out of the anti-descriptivist revolution that has occurred over the last forty years. A brief foray into history is necessary to place these views in context.

Contemporary descriptivism originated in the work of Frege and Russell; in particular, it arose out of their solutions to several outstanding problems in the explanation of language. One of the most famous is how it can be that one identity claim (e.g., 'Hesperus = Hesperus') is uninformative

and can be known apriori, while another identity claim that results from substituting a co-referring name in the first (e.g., 'Hesperus = Phosphorus') is informative and is known only aposteriori. Frege and Russell solved these puzzles by assuming that linguistic expressions have two semantically relevant features: meaning and reference. The *meaning* of a linguistic expression is what a speaker grasps when she understands that expression, while *reference* is a relation between the expression and one or more objects. In the case of proper names (e.g., 'London'), the meaning is identical to the meaning of a definite description (e.g., 'the largest city in England'). If some unique object satisfies the description, then it is the referent of the name; otherwise, the name has a meaning, but no referent. One can give a similar analysis of natural kind terms (e.g., 'cat') and other predicates by treating their meanings as descriptive conditions (e.g., 'domesticated feline') that determine their extensions (e.g., the set of cats).[45]

When combined with other intuitive views on the nature of language and the mind, this account of the semantic features of linguistic expressions constitutes a powerful theory with far-reaching consequences. The resulting picture of language has come to be known as *descriptivism*. The following are five tenets of descriptivism as explicated by Soames:

(i) One must distinguish between the meaning of a linguistic expression and its referent; for almost any linguistic expression (including proper names), its meaning is given by a description, which determines its referent.

(ii) Understanding a linguistic expression consists in mentally grasping its meaning and associating this meaning with the expression.

(iii) Meaning is transparent; that is, if two linguistic expressions have the same meaning, then anyone who understands them can tell that this is the case. (Because anyone who understands an expression mentally grasps its meaning and associates that meaning with the expression, a person who understands two expressions can tell whether he has mentally grasped the same meaning and associated it with each of them.)

(iv) The meaning of a person's linguistic expression and the content of the person's mental state it expresses are determined entirely by internal features of the person in question.

---

[45] See Frege (1892), Russell (1905, 1910); see Soames (2002c) for discussion.

(Because the meaning of an expression is something that is mentally grasped by someone who comprehends the language in question, a person's physical and social environments have no direct impact on the meanings of her expressions.)

(v) A proposition is apriori if and only if it is necessary; both apriority and necessity are explained in terms of meaning. (Because the meaning of an expression is something that is mentally grasped by someone who comprehends the language in question, simply comprehending a language enables one to know certain truths that are grounded in the meanings of the expressions of that language.)[46]

Although descriptivists differ on the details of how these principles are to be worked out, and it is not the case that all descriptivists accept all of them, the general picture of how linguistic expressions function and how they relate to both the minds of those who comprehend them and the objects in the world was the received view in analytic philosophy from the beginning of the twentieth century until the late 1960s.

Kripke is perhaps the most famous opponent of descriptivism—the force and clarity of his criticisms have been immensely influential. Kripke argues that if names had descriptive meanings, then sentences containing names (or the propositions expressed by them) would have modal and epistemic properties that are different from the ones they actually have. Moreover, he denies that the referent of a name is determined by a definite description (or cluster of definite descriptions). For example, if a name, 'Clancy', is synonymous with a definite description, 'the chief of the Springfield police department', then the proposition expressed by 'if Clancy exists, then Clancy is the chief of the Springfield police department' is necessary and apriori.[47] However, Clancy might not have been the chief of the Springfield police department. Hence, the proposition in question is not necessary. Moreover, a person's justification for the belief that if Clancy exists, then he is the chief of the Springfield police department will certainly depend on empirical evidence; hence, the

---

[46] Soames (2005: 1-2). Soames actually lists seven tenants (including anti-essentialism and the claim that the aim of philosophy is conceptual analysis), but I am not concerned with these issues in this paper.

[47] I assume that modal properties are properties of propositions. There are several popular theories of propositions, but for my purposes, it does not matter which one is correct; see Soames (2002c, 2010) and Schiffer (2003) for discussion.

proposition in question is not known apriori. In addition, if the referent of 'Clancy' is whatever satisfies 'the chief of the Springfield police department', then understanding 'Clancy' would require knowing that its referent is fixed by this definite description, which is clearly not correct.[48]

In place of descriptivism with respect to names, Kripke suggests that names are *rigid designators*. That is, a name refers to the same object in all possible worlds in which that object exists, and the name never refers to anything else. Moreover, he offers an alternative account of how the referents of names are fixed, on which the referent of a name is the object that initiated a chain of reference transmissions. The chains usually begin with a person proposing a name for an object; these people use the name to refer to that object without associating any particular description or cluster of descriptions with the name. Other people can learn to use the name too; the name refers to the original object so long as a person intends to use it with the same reference as the person from whom he learned the name. In this way, the name comes to be used by people "further down the chain" without the help of definite descriptions.

David Kaplan proposes similar objections to the descriptivist theory of indexicals and demonstratives. He argues that indexicals and demonstratives are not synonymous with descriptions and that their referents are not determined by descriptions. In place of the descriptivist theory, he offers an account of indexicals and demonstratives on which they are rigid designators. He goes beyond Kripke's views by endorsing a direct reference theory of indexicals and demonstratives. For Kaplan, the content of an indexical or demonstrative just is its referent.[49] Kaplan offers an alternative account of the meaning and content of indexicals and demonstratives in addition to an alternative account of their reference. Kripke presented only an account of the referents of proper names; he is silent about their meanings. However, other anti-descriptivists, including Nathan

---

[48] Kripke (1972); see Soames (2002c) for discussion.
[49] Kaplan (1978, 1989); see also Perry (1977, 1979, 2001). I am distinguishing between the meaning and the content of a context dependent expression; its meaning remains constant throughout changes in context, but its content changes.

Salmon and Scott Soames, have offered direct reference theories for the meanings of proper names.[50]

The attacks on descriptivism extend beyond its consequences for names and indexicals. Indeed, Kripke suggests that natural kind terms are rigid designators and that they are not synonymous with descriptions or clusters of descriptions.[51] Thus, the descriptivist account of natural kind terms comes under attack as well. At around the same time, Hilary Putnam presented a sequence of papers arguing that natural kind terms are not synonymous with descriptions or clusters of descriptions. Moreover, Putnam argued, the meanings of natural kind terms are determined in part by the physical environment in which they are used. Thus, the meaning of a natural kind term is not determined entirely by features internal to the mind of a person who uses it. According to Putnam, it is possible for there to be two people with qualitatively identical mental states using the same word, yet with the word having one meaning when used by one person and a different meaning when used by the other. This theory has come to be known as *semantic externalism*. Putnam also offers a non-descriptivist account of the meaning of natural kind terms, which is based on the notion of a stereotype. The stereotype of a tiger, for example, is something like the cluster of properties that a normal tiger should have.[52]

Tyler Burge proposes several versions of semantic externalism. He argues that the meanings of natural kind terms depend not only on the physical environment, but on the social environment as well. That is, a linguistic expression used by two people with the same mental states in the same physical environment can have different meanings for them because they are members of linguistic communities that have different linguistic norms.[53] Burge also argues that semantic externalism (or

---

[50] Salmon (1986) and Soames (2002c).
[51] Kripke (1972).
[52] Putnam (1975); see also Burge (1986).
[53] Burge (1979b).

anti-individualism as he sometimes calls it) is true of many other types of linguistic expressions as well.[54] Furthermore, Burge claims that semantic externalism should apply not only to the meanings of linguistic expressions, but also to the contents of mental states and perceptual experiences as well.[55]

Although the anti-descriptivists disagree on many issues, one can draw several broad conclusions from their attacks on descriptivism. First, names, indexicals, and natural kind terms are not synonymous with definite descriptions, and definite descriptions do not determine the referents of these linguistic expressions. Second, understanding a name, an indexical, or a natural kind term is not simply a matter of mentally grasping a meaning and associating this meaning with it. Third, one can understand two synonymous expressions without knowing that they are synonymous; hence, meaning is not transparent. Fourth, the meanings of many linguistic expressions and the contents of many mental states are determined in part by the physical or social environment in which they are used or occur.

One might worry that on semantic externalist views, the Gricean Condition fails, since participants in a conversation on Earth where someone utters 'the glass contains water' and participants in a conversation on Twin-Earth where someone utters the same sentence would treat them as saying the same thing about the contents of the glasses in question. I want to make two points. First, since all of the conversations involving 'true' take place on Earth (or at least, all the ones I am concerned with), I do not need to take a stand on this issue. Participants in actual conversations occupy the same physical and social environment, so there is no problem with the Gricean Condition as a principle governing actual human communication on Earth. Second, almost from the start, semantic externalists have been criticized for being unable to explain how we know

---

[54] Burge (1986).
[55] Burge (1979b, 1986b); see also McDowell (1992).

what our words mean. This criticism has turned into *the* major topic of discussion in debates over semantic externalism.[56] Whatever turns out to be the solution, if there is one, will no doubt also explain how semantic externalism could be compatible with the Gricean Condition. That is, if one can reconcile semantic externalism with the claim that competent speakers know the meanings of their own words (even if they do not know the relevant facts about their environment), then one can reconcile semantic externalism with the Gricean Condition. For, all the Gricean Condition says is that in normal circumstances, speakers and hearers know the meanings of the words that are used. Thus, if there is a solution to this problem affecting semantic externalism, then semantic externalism is compatible with the Gricean Condition, but if there is not, then so much the worse for semantic externalism.[57]

Millianism (or direct reference theories) might also pose a problem for the Gricean Condition. Here is one way of bringing out the problem. If the meaning of the word 'water' is just the property of being H_20, then someone who felicitously asserts 'the glass is full of water' asserts the proposition that the glass is full of H_20; so it seems that the proposition expressed by this sentence is not retrievable by a hearer who is ignorant of contemporary chemistry despite the fact that the hearer might be a competent user of 'water' and the utterance was felicitous.

Is this a problem for the Gricean Condition? I am not convinced that it is. Stefano Predelli has argued recently that Millianism does not have these sorts of counterintuitive consequences. According to Predelli, these consequences disappear once one keeps track of the distinction between a semantic theory (what Predelli calls an *interpretive system*) and a linguistic practice. In particular, one needs to distinguish between the mathematical model of semantic content (what Predelli calls a *t*-

[56] See Davidson (1987, 1988), Burge (1988), Boghossian (1989), McKinsey (1991), Kobes (1996), Sawyer (1999), Brown (2004), and the papers in the papers in Ludlow and Martin (1998) and Nuccetelli (2003).
[57] For what it is worth, it seems to me that one could satisfy the intuitions that power Putnam's and Burge's thought experiments and avoid the entire problem of self-knowledge by treating semantic externalism as a variety of what has come to be known as *non-indexical contextualism* (discussed in detail in Chapter Fourteen), but I do not defend that claim here.

distribution) in the interpretive system and the intuitive truth conditions of utterances in the linguistic practice. Millianism, properly understood, is a view about the mathematical model of semantic content, not a view about intuitive truth conditions. I do not have the space to explain in detail how this helps Millianism avoid its seemingly counterintuitive consequences, but it seems to me that Predelli has an important insight here.[58] Moreover, this insight could show how Millians may keep their philosophical position along with a modest attitude toward the sciences (in particular, toward the Gricean Condition as a tenet of linguistics).

Even if Predelli's diagnosis is ultimately untenable, there are other ways to square Millianism with the Gricean Condition. Some contemporary Millians follow Scott Soames in saying that when a speaker performs an utterance, the speaker thereby utters many propositions simultaneously. Some of these propositions are picked up by the audience, and some are not.[59] Although Soames does not emphasize this point, one can see that this makes it easier to square Millianism with the Gricean Condition since only one of these propositions, the one intended by the speaker to be picked up by the addressee, needs to be available to the audience to satisfy the Gricean Condition.

On the other hand, some Millians happily say "ordinary practice be damned." For example, Ted Sider and David Braun write:

> Soames's stand on intuitions about particular sentences, roughly speaking, is that they are correct about *something*, namely, asserted content. … We think that the correct stand is rather that, in some cases, speakers' intuitions about particular sentences are correct about *nothing*. … Particular intuitions are best taken as concerning semantic content. Thus taken, some of them are simply mistaken. Speakers fundamentally misunderstand the rules that govern language use. In a sense, then, we are reformers in a way that Soames is not. Speakers regularly utter such sentences as "Lois Lane does not believe Clark Kent can fly". We think they should stop—such utterances violate the rules of use of English.[60]

---

[58] See Predelli (2005: ch. 5).
[59] Soames (2002c).
[60] Sider and Braun (2006: 681).

The problem is, of course, that the science of linguistics is at odds with this attitude. No self-respecting linguist would be caught dead saying that English speakers should stop uttering sentences that do not square with Sider and Braun's version of Millianism. So, before going on about "the rules of use of English", it might be prudent to check with the experts on this topic. In sum, the challenge for Sider and Braun (and any other philosopher whose pet theory is incompatible with the results of established sciences) is to explain why their view is not just as unacceptable as creationism or alchemy.

6.3 Consequences

Each of the primary four chapters of Part II is divided into a discussion of one key idea, the consequences of that discussion for theories of truth, and a summary of its impact. Accordingly, this section is devoted to four consequences of the above discussion of truth's expressive role and the Gricean Condition.

6.3.1 Deflationism and Sense Identity

Some philosophers (mostly deflationists) advance theories on which p and 'p is true' have the same meaning or express the same proposition.[61] The fact that 'true' has expressive uses, together with the Gricean condition, rules out these theories. For example, imagine Leopold has it on good authority that Goldbach's conjecture is true, but he does not know what Goldbach's conjecture is. He asserts 'Goldbach's conjecture is true' in a conversation where no one else knows what it is either. According to the sense-identity view, Leopold's sentence means that every number greater than two is the sum of two primes, but no one in the conversation knows this. Thus, these theories

[61] Frege (1918), Ramsey (1926), Ayer (1936), Grover, Camp and Belnap (1976), Brandom (1994, 2002), and Lance (1997); see Künne (2003: 34-52) for discussion.

either make bad predictions about truth's expressive role (e.g., that Leopold's utterance is infelicitous) or they violate the Gricean Condition. Either way, they are unacceptable.

6.3.2 Deflationism and Truth's Explanatory Role

A quick perusal of the stock objections to deflationism (listed in the Appendix to Chapter One) shows that many of the criticisms of deflationism come in the form of explanatory challenges—deflationists are unable to explain such and such connection between truth and some other concept. Indeed, many take it as definitive of deflationism that truth cannot play an explanatory role in any philosophical theory; most, perhaps all, deflationists seem to agree. However, they disagree about what to say about such theories. Some deflationists reject them completely and offer alternative views, whereas others deflationists suggest that truth predicates are playing a merely expressive role in the theories in question.

For example, it is often said that deflationists cannot accept a truth-conditional theory of meaning since it would cast truth in an explanatory role; to accept that truth plays an explanatory role is to accept an inflationary notion of truth. Many prominent deflationists, including Hartry Field, Vann McGee, Paul Horwich, and Robert Brandom hold this view.[62] One must be careful since it is important to realize that even a deflationist can accept that sentences have truth conditions, as Field notes in the following passage:

> If I understand 'Snow is white', and if I also understand a notion of disquotational truth as explained above, then I will understand "snow is white' is true', since it will be equivalent to snow is white. … For the cognitive equivalence of 'snow is white' is true' and 'snow is white' will lead to the (more or less indefeasible) acceptance of the biconditional 'snow is white' is true iff snow is white'; and a natural way to put this (more or less indefeasible) acceptance is to say 'snow is white' has the truth conditions that snow is white'. A pure disquotational notion of truth gives rise to a purely disquotational way of talking about truth conditions.[63]

---

[62] Field (1994a), McGee (1993), Horwich (1998), and Brandom (2002).
[63] Field (1994a: 251).

So, a deflationist can accept that 'snow is white' has the truth conditions that snow is white. However, what a deflationist cannot accept is that the meaning of 'snow is white' is its truth conditions, or that to specify the meaning of 'snow is white' is to specify the conditions under which it is true.

Michael Williams is one of the only deflationists to go against this trend. He argues that deflationists can accept a Davidsonian truth-conditional theory of meaning because every occurrence of a truth predicate in such a theory is serving a merely expressive role, not an explanatory role. According to Williams, Davidson's theory is split into two parts, the theory of meaning, which specifies the meanings of the sentences in a particular language, and the theory of interpretation, which states conditions on theories of meaning by way of the radical interpreter. However, 'true' as it occurs in the theory of meaning is used to generalize over specifications of the meanings of sentences for some language, whereas 'true' as it occurs in the theory of interpretation is used to generalize over sentences to state compositionality, consistency, agreement, and distal requirements on a theory of meaning. So despite Davidson's protestations to the contrary, his theory of meaning does not explain meaning in terms of an inflationary notion of truth. Rather, it merely employs the truth predicate to achieve generality.[64]

One problem with Williams' argument is that a deflationary theory of truth is typically just a list of T-sentences for a given language, but this is inadequate for Davidson's theory of meaning. Davidson is quite explicit that a theory of meaning for a language L should take the form of a Tarskian truth definition for L, and a theory of meaning should be compositional (i.e., it should show how the meaning of each complex sentence is determined by its structure and the meanings of its parts). Moreover, Davidson, like Tarski before him, emphasizes that a list of T-sentences is not

---

[64] Williams (1999).

an adequate basis for this kind of theory of meaning because the resulting theory would not be compositional.[65] The theory that Davidson uses has compositional clauses for the logically complex sentences (e.g., a conjunction is true iff both conjuncts are true). Deflationary theories do not include principles like these. As such, they are not suitable to serve as the basis for a Davidsonian theory of meaning. Although Williams might be right that the truth predicates that occur in the formulation of a Davidsonian theory of meaning and a Davidsonian theory of interpretation are playing expressive roles, the deflationist's theory of truth is not capable of playing the role a Davidsonian theory of meaning demands of it. So, a Davidsonian about meaning cannot accept a deflationary theory of truth, at least, if a deflationary theory of truth is just a list of T-sentences.

Let us look at the issue in a bit more depth. Let the following serve as an example of a truth-conditional theory of meaning:

(TCTM) To specify the meaning of a sentence is to state the conditions under which it is true.

To avoid problems with what meanings are, I have formulated it as a theory of what it is to specify the meaning of a sentence instead of as a theory of what the meaning of a sentence is. Imagine a deflationist who wants to endorse TCTM. Assume that this deflationist accepts the T-sentence for 'snow is white' (i.e., "snow is white' is true iff snow is white'). From (TCTM) it follows that to specify the meaning of 'snow is white' is to state the conditions under which 'snow is white' is true. Since the deflationist accepts truth-value intersubstitutability, acceptance of (TCTM) brings with it the claim that to specify the meaning of 'snow is white' is to state the conditions under which snow is white. Note that we have substituted 'snow is white' for "snow is white' is true'.

What are the conditions under which snow is white? The standard explanation is that snow is frozen H_20, and the crystals that form when H_20 freezes have the right shape and structure to ensure that when light hits them, they reflect all wavelengths equally well, but at different angles so the

[65] Davidson (1968, 1973, 1990).

crystals appear a translucent white, instead of transparent or some particular color. So, it seems that at least one rather plausible way of stating the conditions under which snow is white is to say that H_20 has the molecular properties that render frozen crystals of it capable of reflecting light translucently at all wavelengths. However, let us reconsider the claim above in light of this explanation: to specify the meaning of 'snow is white' is to say that H_20 has the molecular properties that render frozen crystals of it capable of reflecting light translucently at all wavelengths. But that does not sound right at all! The claim about the molecular properties of H_20 need not be what one means when one says that snow is white. If this is the right way to think of the phrase 'the conditions under which snow is white', then the deflationist's reading of this instance of (TCTM) ought to be rejected.

From these reflections it seems clear that the problem with deflationists accepting truth-conditional theories of meaning is not that these theories appeal to an inflationary notion of truth; or, at least, that is not the most perspicuous way of putting the problem. Rather, when combined with the deflationist's principles involving truth, the truth conditional theory of meaning has consequences that are radically implausible.

Another example confirms these findings. Consider the explanatory role of truth in a theory of assertion. Keith Simmons and Dorit Bar-On put forward the following theory:

(TTA) One asserts that p iff one presents $\langle \mathsf{p} \rangle$ as true.[66]

Again, consider a particular instance of it: one asserts that snow is white iff one presents 'snow is white' as true. Although the phrase 'p is true' does not occur in here, it seems plausible that a deflationist would just read this as: one asserts that snow is white iff one presents snow as being white. However, that does not seem right either. One can present snow as being white without

[66] See Simmons and Bar-On (2006).

asserting anything, say, by holding up a handful of snow for inspection in the right circumstances. Again, the consequences a deflationist draws from the theory of assertion in question are false.

Yet another example, however, has a different result. Consider the explanatory role of truth as a standard for belief:

(TSB) A rational agent ought to believe that p only if ⟨p⟩ is true.

Can a deflationist accept this principle? I do not see why not. When we use the intersubstitution principle to arrive at an instance like 'a rational agent ought to believe that snow is white only if snow is white', that seems like a perfectly legitimate thing to accept, as do the other instances. Thus, in this case, the consequences of the theory (TSB) are not problematic.

It seems to me that the lesson here is that the explanatory challenges to deflationism have been misunderstood. It is not that a deflationist cannot accept a theory of meaning or assertion or anything else that is formulated using a truth predicate. They can. However, given the deflationist's strong principles linking p and 'p is true', the deflationist is going to be forced to accept counterintuitive consequences from some of these theories.

It seems to me that it is implausible to think that one can tell whether a given use of a truth predicate is expressive or explanatory (or even that this distinction is ultimately intelligible). Instead, one should see what follows given other principles one accepts. These considerations support the idea that deflationists should be very careful in accepting theories that are formulated with truth predicates, not because such theories appeal to some notion of truth that betrays the deflationist's view on the nature of truth, but instead because the deflationist's own principles about truth when combined with the theory in question often (but not inevitably) lead to unhappy consequences about the topic in question.

6.3.3 Inflationary Arguments

Crispin Wright's book *Truth and Objectivity* begins with a much-discussed criticism of deflationism, which usually goes by the name 'the inflationary argument'.[67] It seems to be a very powerful objection to deflationism since it depends only on the deflationist's claim that truth is a device of endorsement. Below is a reconstruction of Wright's argument (augmented by his discussion in Wright 1999):

1. If S asserts ⟨Fa⟩ where ⟨F⟩ is a device of endorsement, then S asserts that the referent of ⟨a⟩ meets some doxastic standard.[68]
2. 'true' is nothing more than a device of endorsement.
3. When S asserts 'p is true', S asserts (at least in part) that p meets some doxastic standard.
4. When S asserts 'p is not true', S asserts (at least in part) that p does not meet some doxastic standard.
5. It is true that **p** if and only if **p**.[69]
6. It is not true that **p** if and only if ~**p**.
7. It is not true that **p** if and only if it is true that ~**p**.
8. ~(p does not meet doxastic standard D if and only if '~p' meets doxastic standard D).
9. ⊥

Wright takes this argument to show that step 2 should be rejected and deflationist theories along with it. Given what has been said in this chapter already, we can see a major flaw in Wright's argument—premise 2 is false—no one should think that 'true' is no more than a device of endorsement. As I argued above, 'true' is also a device of generalization, which is a role that is much more demanding.[70]

---

[67] Wright (1992: ch. 1).
[68] A doxastic standard is a standard a belief must meet for it to be rational for someone to hold it—this usually involves having a certain amount of evidence for the belief.
[69] Wright routinely slips between treating 'true' as a predicate and treating it as an operator. Although this is sloppy, it could be cleaned up so as not to affect the validity of his argument.
[70] Huw Price makes the same mistake in Price (2003).

However, there is another major problem here. Claim 3 is false as well. When one asserts that p is true, one does not thereby assert that p meets some doxastic standard. As everyone admits, there could be true propositions that no one is justified in believing. For example, either there was an even number of living people on Earth at midnight December 31, 2000, or there was an odd number. Although someone might believe one or the other of these propositions, neither of them meets a doxastic standard—i.e., no one is remotely justified in believing either one. However, one of them is true. So, there are propositions that are true, but do not meet a doxastic standard. No reasonable deflationist would ever deny this.

I can imagine a defender of Wright saying "Hold on. When a person asserts that p is true, that person is thereby endorsing p. Everyone in the audience is justified in assuming that the person not only believes that p, but also takes his or her belief to meet some doxastic standard. Otherwise, why would they have endorsed it? So it does seem that when one asserts that p is true, one asserts that p meets some doxastic standard." It is correct that when a person asserts that p is true, the audience is justified in believing that that person believes that p and believes that p meets some doxastic standard. However, the claim that p meets some doxastic standard is not part of the *content* of the assertion. Rather, it is *conversationally implicated* by the assertion. Wright has confused the content of the assertion with one of its implicatures.

Why should we take this to be a conversational implicature rather than part of the meaning? First, the conversational implicature reading makes more sense than the entailment reading because, when a speaker asserts that p is true, audiences infer that p meets some doxastic standard *not* from the content of 'p is true' alone, but from the fact that it was asserted (as evidenced by the question above—"why would they have endorsed it?"). That is a paradigmatic feature of conversational implicature. In addition, a standard test for conversational implicature is that it is cancelable (see above). That is, the speaker can say or do something that would prevent the audience from drawing

that conclusion. For example, if I assert 'the proposition that my car is right now being broken into is true, but I have absolutely no reason to believe it' I have thereby cancelled the conversational implicature. I have successfully attributed truth to the proposition in question, but it would be wrong for my audience to take me to believe that it meets some doxastic standard. Think for a moment about how ridiculous it would be for an audience member to respond by saying "well, if you have no reason to believe it, it can't be true," which would make sense if Wright were correct in thinking that meeting a doxastic standard is part of the content of a device of endorsement (e.g., that 'p is true' entails 'p meets some doxastic standard').

Moreover, taking 'p is true' to *entail* 'p meets some doxastic standard' is incompatible with one of the constitutive principles of truth:

(T-In) If p, then p is true.

Take any unknown claim, say, P=NP.[71] An instance of (T-In) is 'if P=NP, then 'P=NP' is true'. If 'p is true' entails 'p meets some doxastic standard', then we also get 'if 'P=NP' is true, then 'P=NP' meets some doxastic standard'. By transitivity, we then get 'if P=NP, then 'P=NP' meets some doxastic standard. The same argument form shows that any claim meets some doxastic standard, which is absurd. Since (T-IN) and the fact that truth is a device of endorsement are on much firmer ground than Wright's principle (1) (in the argument above), it should be obvious that the latter is unacceptable. Note also that Wright accepts (T-In) and that truth is a device of endorsement; his inflationary argument is supposed to show that truth is *more than* a mere device of endorsement. So, by his lights, every claim meets some doxastic standard.

I conclude that if the argument above is an accurate reconstruction of Wright's inflationary argument, then it has two insurmountable defects.

---

[71] This is probably the most famous unsolved problem in theoretical computer science.

6.3.4 Language-Specific Truth Predicates

Language specific truth predicates (e.g., 'true-in-English') have come up several times already in discussions of deflationism and logical approaches to the aletheic paradoxes. In this subsection, I want to offer an objection to any theory that explains natural language truth predicates in terms of language-specific truth predicates. I call these *language-specific theories* (or LS theories).

The first problem for LS theories is that speakers of natural languages routinely apply their truth predicates to foreign sentences. For example, one can attribute truth to a German sentence by using the truth predicate of English (e.g., "Schnee ist weiss' is true' is a true sentence of English). However, "Schnee ist weiss' is true-in-English' is not a true sentence of English. Thus, the simple LS theory on which 'true' just means *true-in-English* has false consequences.

A more complex LS theory avoids this problem. For example, an LS theory on which 'true' is ambiguous and can have the meaning of any of the LS truth predicates is untouched by this objection. On this view, in the sentence "snow is white' is true', 'true' means true-in-English. In the sentence "Schnee ist weiss' is true', 'true' means *true-in-German*. Call this the *ambiguity LS theory*.

There is a substantial body of literature in linguistics on tests one can perform to determine whether a linguistic expression is ambiguous.[72] For example, one cannot express the claim that Carl went to the financial institution and Lenny went to the edge of a river by asserting 'Carl and Lenny each went to the bank' because 'bank' is ambiguous; in addition, sentences like 'Carl went to the bank, but he did not go to the bank' have non-contradictory readings because 'bank' is ambiguous. However, one can express the claim that 'Schnee ist weiss' is true and 'snow is white' is true by asserting "Schnee ist weiss' and 'snow is white' are true'; in addition, the sentence "Schnee ist weiss'

[72] See Zwicky and Sadock (1975), Cruse (1988), Atlas (1989), Gillon (2004), and Kennedy (2010) for discussion.

is true, but 'Schnee ist weiss' is not true' is a contradiction. Thus, 'true' (as conceived by the ambiguity LS theorist) fails the standard ambiguity tests.[73]

In addition, the ambiguity LS theory faces another problem. Consider the following example. Ned and Maude are at a bar, having a conversation in English. Maude is a distinguished expert on ring-tailed lemurs, and Ned is aware of this fact. Maude tells Ned that on Monday she was at a talk given by Helen, another expert on ring-tailed lemurs. Maude informs Ned that Helen argued for a certain thesis, but Maude does not tell Ned what the thesis is because the complexities do not matter for her purposes in the conversation. Maude simply refers to it as *Helen's thesis*. Maude remarks that Helen's thesis implies that a theory Maude recently published is false, and she tells Ned that she now agrees with Helen.[74] Later that morning, Ned bumps into Tim at the library. Tim is writing a paper on ring-tailed lemurs, and he informs Ned that he is planning to rely on Maude's recently published theory. Ned tells Tim that Maude's theory is false. Tim knows that Ned is usually sincere and trustworthy, but that Ned does not know much about the literature on ring-tailed lemurs; accordingly, Tim challenges Ned on his assertion. Ned responds by asserting 'if Helen's thesis is true, then Maude's theory is false' and 'Helen's thesis is true'. Ned, of course, explains to Tim that Maude informed him of these facts. After hearing this, Tim scurries off to the bar to find Maude so that he can find out what Helen's thesis is.

Ned's assertion of 'Helen's thesis is true' is an expressive use of a truth predicate since Ned is using it to endorse Helen's thesis even though he is unable to assert Helen's thesis directly (i.e., it is a blind use). Moreover, it is felicitous. A problem for the ambiguity LS view is that the Gricean Condition says its meaning is retrievable. However, by the ambiguity LS view, 'true' is ambiguous. So, what is the meaning of 'true' in Ned's sentence? If it meant *true-in-English* and Helen's thesis is a

---

[73] It seems to me that these results cast considerable doubt on Kölbel's claim that 'true' is ambiguous; see Kölbel (2006).
[74] I assume that Maude is right about the truth of Helen's thesis and its consequence.

sentence of some other language, then Ned's sentence would be false. The problem is, Ned does not know which language Helen was speaking, so he does not know which meaning to pick. Thus, the ambiguity LS theory requires too much of speakers using 'true' in situations like this. It implies (given the Gricean Condition) that Ned should know what language Helen was speaking in order for his utterance to be felicitous.

An ambiguity LS theorist might claim that 'true' just gets whatever meaning is appropriate independently of what Ned intends. This idea flies in the face of evidence from linguistics about ambiguous terms, but so be it. Assume Helen was speaking French. Now, according to the LS-ambiguity theory, the truth predicate in Ned's sentence means *true-in-French*, and his sentence means that Helen's thesis is true-in-French. However, now the problem is that the meaning of Ned's sentence is not available to anyone in his conversation with Tim. Neither Ned nor Tim knows what Ned's sentence means and there is no way for them to recover its meaning from what has been said or the context of their conversation. Thus, the ambiguity LS theorist who follows this path runs afoul of the Gricean Condition. The Gricean Condition and the expressive uses of truth predicates form a powerful combination.

Instead of defending the ambiguity LS theory, some deflationists who advocate LS theories suggest that natural language truth predicates are synonymous with 'translatable into a sentence of L that is true-in-L'. Call this a *translational LS theory*.[75] For example, on this view the English sentence "Schnee ist weiss' is true' means that 'Schnee ist weiss' is translatable into an English sentence that is true-in-English.

---

[75] The move is familiar in the face of other criticisms leveled against LS approaches; for such criticisms, see Blackburn (1984), David (1989, 1994), Richard (1996), Soames (1997), Brendel (2000), Horwich (2001), Künne (2002), and Shapiro (2003). These philosophers all address deflationists who advocate the LS approach. Some philosophers who work on the logic of truth have criticized Tarski's commitment to the LS approach; see Field (1972), Dummett (1978: introduction), and Putnam (1985). See also Davidson (1990, 2005) for a discussion of this issue.

There are at least two options for how a translational LS theory interprets an English sentence like "Schnee ist weiss' is true': the *quantificational version*, which treats this sentence as '(∃x)(x is a sentence of English and x is a translation of 'Schnee ist weiss' and x is true-in-English)', and the *constant version*, which treats it as 'p is a sentence of English and p is a translation of 'Schnee ist weiss' and p is true-in-English', where 'p' is a constant. When interpreting multiple-target truth attributions (e.g., 'all the sentences Carl asserted yesterday are true'), the quantificational version of the translational LS theory is the only acceptable option.[76] Thus, one might as well endorse it in general.

Although the translational LS approach seems to explain the English word 'true' with only a single LS truth predicate ('true-in-English'), this is an illusion. Consider the sentence '"Schnee ist weiss' ist wahr' is true'. This should come out as a true sentence of English. However, according to the translational LS theory, it means "Schnee ist weiss' ist wahr' is translatable into a sentence of English that is true-in-English. What sentence could this be? Perhaps "Schnee ist weiss' is true-in-English'? No, this sentence will not work since it is not true-in-English. The only option for an LS theorist is to use "Schnee ist weiss' is true-in-German' as the translation into English. But, of course, that would require English to have 'true-in-German' as well as 'true-in-English'. Thus, the translational LS theory cannot get away with using a single language-specific truth predicate. It needs just as many as the ambiguity LS theory.

The translational LS approach is plausible only for accounts of language and accounts of translation on which all languages are intertranslatable. Otherwise, it faces an obvious criticism.

---

[76] The quantificational version can render this claim as: 'for all x, if x is a sentence Carl asserted yesterday, then for some y, y is a sentence of English and y is a translation of x and y is true-in-English'. How should the constant version treat this sentence? Perhaps 'for all x, if x is a sentence Carl asserted yesterday, then p is a sentence of English and p is a translation of x and p is true-in-English'? This cannot be right because it implies that p is a translation of all the sentences Carl asserted. Another option might be: 'for all x, if x is a sentence Carl asserted yesterday, then p and q are sentences of English, and either p or q is a translation of x, and p and q are true-in-English'. This suggestion works only if Carl asserted at most two sentences on the day in question. A supporter of the constant version might suggest that the logical form of the truth attribution depends on the number of its targets, but this hardly seems plausible. Moreover, it abandons the view that 'true' means 'translatable into a sentence of English that is true-in-English'.

Pick a true sentence p of a language L that is not translatable into English. The sentence 'p is true' is a true sentence of English, but the translational LS approach implies that it is false (on this view, 'p is true' means that for some x, x is a sentence of English and x is a translation of p and x is true-in-English—but, by stipulation, there is no such sentence of English). Thus, I assume that, given the notions of language and translation employed by the translational LS theorist, all languages are intertranslatable.

This concession does not save the translational LS approach.[77] Consider again the example story with Ned, Maude, Helen, and Tim. Maude tells Ned about Helen's thesis, but she does not tell him which language Helen was speaking when Helen asserted it. We can alter the example so that Maude tells Ned that Helen's thesis cannot be translated into English because it involves technical jargon that currently belongs only to the language Helen was speaking when she asserted it.

Again, Ned's use of the truth predicate is expressive—he is using it to endorse a proposition that he cannot assert directly. Furthermore, Ned believes that Helen's thesis is not translatable into English; consequently, Ned believes that there is no sentence of English that is both a translation of Helen's thesis and true-in-English. However, according to the translational LS theory, Ned's sentence means that there is a sentence of English that is both true-in-English and a translation of Helen's thesis. So, again, this theory imposes too strict a requirement on speakers using 'true' in situations like this. It implies (again, given the Gricean Condition) that Ned should know that Helen's thesis is translatable into English in order for his utterance to be felicitous.

---

[77] However, it does save the translational LS approach from other criticisms; for example, Shapiro (2003) argues that if it is not the case that all languages are intertranslatable, then the translational LS theorist has to accept a notion of logical consequence that is not acceptable to a deflationist. See Field (2001e), and Shapiro (2005) for comment.

In sum, I have argued that the basic LS theory, the ambiguity LS theory, and the translational LS theory all conflict with the combination of truth's expressive role and the Gricean Condition.[78] The right conclusion to draw is that truth predicates are not language-specific in any way.

6.4 Impact

The impact of the first key idea concerns two areas of study: deflationism and logical approaches to the paradoxes. The results for deflationism are mixed. While the inflationary argument, which is one of the most influential objections to deflationism, has been defanged in section 6.3.3, the prosentential theory (a version of deflationism) is, in my view, a non-starter given the considerations regarding sense identity in section 6.3.1. I have suggested that the explanatory challenge to deflationism has been misconstrued, but the proper understanding of issues as presented in section 6.3.2 offers a mixed verdict for deflationism—it seems as if deflationism is compatible with some theories of other concepts that invoke truth, but incompatible with others. The objections to language-specific truth predicates in section 6.3.4 are a serious blow to disquotationalism. That makes two deflationist theories maimed by considerations of truth's expressive role, which the deflationists have fought so hard to emphasize. In addition, since almost all logical approaches to the paradoxes are formulated only for language-specific truth predicates, these considerations threaten the relevance of those approaches. At the very least, anyone who formulates an approach to the paradoxes focusing on language-specific truth predicates owes an explanation of how the approach is to be scaled up for an unrestricted truth predicate like the one we find in natural language.

---

[78] Gary Ebbs tries to deal with the problem by offering a new account of words. However, if successful, it would work only for attributions of truth (in English) to sentences that contain only words that have the same extensions as English words. Sentences like "Nelson fühlt Schadenfreude' is true-in-English' still pose a problem for Ebbs since there is no English equivalent of 'Schadenfreude'. See Ebbs (2009: 141).

Chapter 7

Risky Business

We saw in the last chapter that truth predicates endow natural languages with a certain expressive power. In this chapter, we explore a feature of the aletheic paradoxes. Some aletheic paradoxes are *empirical* in the sense that the occurrence of the paradox depends on certain empirical facts. Had these facts not obtained, no paradox would have resulted. It is not clear whether all paradoxes have this feature, but the fact that some aletheic paradoxes do is significant. Again, the chapter is broken into several sections on the key issue and then a section on its consequences.

7.1 Empirical Paradoxicality

It is common knowledge among those who work on the liar paradox that one can construct paradoxical sentences with the use of empirical predicates. These sentences are paradoxical because of some empirical facts; if the facts had been different, they would not have been paradoxical. We can say that such sentences are *empirically paradoxical*.

Philosophers have known of empirical versions of the liar paradox since it first became an object of study over two millennia ago. For example, the predicate 'is a complete sentence in the first section of Chapter Seven of Scharp's *Replacing Truth* whose first letter is an 'E", can be used to construct a version of the liar paradox.

> Every complete sentence in the first section of Chapter Seven of Scharp's *Replacing Truth* whose first letter is an 'E' is false.

The fact that the previous sentence is the only complete sentence in section one of this chapter to begin with an 'E' is an empirical fact about that sentence. If I had chosen to place it in a different section or if I had included some other sentences in this section, then it might not have uniquely

satisfied that empirical predicate and, thus, it would not have been paradoxical. Nevertheless, it seems obvious that this change would not have altered the sentence's syntactic or semantic features.[1]

7.2 Kripke on Riskiness

Although empirical versions of the liar paradox are as old as the paradox itself, they have not received much attention in contemporary discussions. Certainly the most influential examination of empirically paradoxical sentences is found in Saul Kripke's paper on truth.[2] Kripke makes a striking remark about empirical versions of the liar paradox:

> The versions of the Liar paradox which used empirical predicates already point up one major aspect of the problem: *many, probably most, of our ordinary assertions about truth and falsity are liable, if the empirical facts are extremely unfavorable, to exhibit paradoxical features*.[3]

Presumably, lazy uses where the target does not contain a truth predicate (described in Chapter Six—e.g., "snow is white' is true') are the least risky. Any truth attribution to a truth bearer that contains a truth predicate (e.g., "Helen's thesis is true' is true') is liable to be much more risky since the target could turn out to be paradoxical if *its* targets are paradoxical. Generalizing uses are very risky as well since they often quantify over many truth bearers.

To illustrate, Kripke provides an example in which Nixon and Jones assert sentences that, owing to their satisfaction of empirical predicates, are paradoxical. In his example Jones asserts:

(1) Most (i.e., a majority) of Nixon's assertions about Watergate are false.

And one of Nixon's assertions is:

(2) Everything Jones says about Watergate is true.

---

[1] Did you look at every sentence in this section to confirm? If so then you now have a great example of knowing the syntactic and semantic features of a sentence without knowing whether it is paradoxical. Think for a moment how bizarre it would be to say that while you were perusing the other sentences of this section looking for 'E's, you were learning about the syntax or the meaning of that one sentence.

[2] Kripke (1975).

[3] Kripke (1975: 691); italics in original.

If, excluding (2), Nixon uttered the same number of true claims about Watergate as false ones, then (1) and (2) are both paradoxical; otherwise, they are not.

Kripke draws some remarkable conclusions from this example:

> The example of (1) points up an important lesson: it would be fruitless to look for an *intrinsic* criterion that will enable us to sieve out—as meaningless, or ill-formed—those sentences which lead to paradox. (1) is, indeed, the paradigm of an ordinary assertion involving the notion of falsity ; just such assertions were characteristic of our recent political debate. Yet no syntactic or semantic feature of (1) guarantees that it is unparadoxical. Under the assumptions of the previous paragraph, (1) leads to paradox. Whether such assumptions hold depends on the empirical facts about Nixon's (and other) utterances, not on anything intrinsic to the syntax and semantics of (1). … The moral: an adequate theory must allow our statements involving the notion of truth to be *risky:* they risk being paradoxical if the empirical facts are extremely (and unexpectedly) unfavorable. There can be no syntactic or semantic "sieve" that will winnow out the "bad" cases while preserving the "good" ones.[4]

These considerations receive plenty of lip service from philosophers who write on the liar paradox, but they are rarely given the attention they deserve.[5] As we will see, they are very powerful when combined with the considerations on the expressive role of truth from the previous chapter.

Kripke demonstrates the power of these considerations in a couple of highly influential objections to the orthodox approach to the aletheic paradoxes. Recall that the orthodox approach specifies a hierarchy of type-restricted truth predicates defined using Tarski's methods and it implies that natural language truth predicates are ambiguous—they can be synonymous with any of the predicates in the hierarchy. According to the orthodox approach, when a speaker utters a sentence containing a natural language truth predicate, the speaker must determine which concept of truth from the hierarchy is to be expressed. In the following passage, Kripke describes how each truth predicate in the hierarchy is associated with sentences of a particular level:

> The notion of differing truth predicates, each with its own level, seems to correspond to the following intuitive idea … First, we make various utterances, such

---

[4] Kripke (1975: 692)

[5] See Church (1946), Cohen (1957, 1961), Prior (1958, 1961), van Fraassen (1968), Burge (1979a), Gupta (1982), Yablo (1982), Martinich (1983), Parsons (1984), Barwise and Etchemendy (1987), Kremer (1988), Stebbins (1992), Gaifman (1992), Simmons (1993: ch. 8), Visser (2001), and Goldstein (2001) for remarks on empirical paradoxicality.

> as 'snow is white', which do not involve the notion of truth. We then attribute truth values to these, using a predicate 'true$_1$'. ('True$_1$' means—roughly—"is a true statement not itself involving truth or allied notions.") We can then form a predicate 'true$_2$' applying to sentences involving 'true$_1$', and so on.[6]

Since it is customary to distinguish the predicates in the hierarchy offered by the orthodox approach by using subscripts (e.g., 'true$_1$'), Kripke sometimes talks about speakers *attaching* subscripts to natural language truth predicates—all that this means is that the speaker intends the truth predicate to express one or another of the concepts in the hierarchy.

The following is a portion of Kripke's remarks on the orthodox approach:

> If someone makes such an utterance as (1), he does *not* attach a subscript, explicit or implicit, to his utterance of 'false', which determines the "level of language" on which he speaks. An implicit subscript would cause no trouble if we were sure of the "level" of *Nixon's* utterances; we could then cover them all, in the utterance of (1) or even of the stronger
>
> (4) All of Nixon's utterances about Watergate are false.
>
> simply by choosing a subscript higher than the levels of any involved in Nixon's Wategate-related utterances. Ordinarily, however, a speaker *has no way of knowing the "levels" of Nixon's relevant utterances*. Thus Nixon may have said, "Dean is a liar," or "Haldeman told the truth when he said that Dean lied," etc., and the "levels" of these may yet depend on the levels of Dean's utterances, and so on. If the speaker is forced to assign a "level" to (4) in advance [or to the word 'false' in (4)], he may be unsure how high a level to choose; if, in ignorance of the "level" of Nixon's utterances, he chooses too low, his utterance of (4) will fail of its purpose. The idea that a statement such as (4) should, in its normal uses, have a "level" is intuitively convincing. It is, however, equally intuitively obvious that the "level" of (4) should not depend on the form of (4) alone (as would be the case if 'false'—or, perhaps, 'utterances'—were assigned explicit subscripts), nor should it be assigned in advance by the speaker, but rather its level should depend on the empirical facts about what Nixon has uttered.[7]

This is Kripke's first objection and it contributed substantially to the decline in popularity of the orthodox approach to the liar. However, the criticism is not exactly clear; I present my preferred reading of it below. Here is the second objection:

> Another situation is even harder to accommodate within the confines of the orthodox approach. Suppose Dean asserts (4), while Nixon in turn asserts
>
> (5) Everything Dean says about Watergate is false.

---

[6] Kripke (1975: 695). Presumably, if a sentence in class C is false, then 'all sentences in class C are true' has level n where n-1 is the highest level of a false sentence in C.

[7] Kripke (1975: 695-696); bracketed text is in the original.

> Dean, in asserting the sweeping (4), wishes to include Nixon's assertion (5) within its scope (as one of the Nixonian assertions about Watergate which is said to be false); and Nixon, in asserting (5), wishes to do the same with Dean's (4). Now on any theory that assigns intrinsic "levels" to such statements, so that a statement of a given level can speak only of the truth or falsity of statements of lower levels, it is plainly impossible for both to succeed: if the two statements are on the same level, neither can talk about the truth or falsity of the other, while otherwise the higher can talk about the lower, but not conversely. Yet intuitively, we can often assign unambiguous truth values to (4) and (5). Suppose Dean has made at least one true statement about Watergate [other than (4)]. Then, independently of any assessment of (4), we can decide that Nixon's (5) is false. If all Nixon's other assertions about Watergate are false as well, Dean's (4) is true; if one of them is true, (4) is false. Note that in the latter case, we could have judged (4) to be false without assessing (5), but in the former case the assessment of (4) as true depended on a *prior* assessment of (5) as false. Under a different set of empirical assumptions about the veracity of Nixon and Dean, (5) would be true [and its assessment as true would depend on a prior assessment of (4) as false]. It seems difficult to accommodate these intuitions within the confines of the orthodox approach.[8]

There are several other objections in Kripke's paper as well, but they do not pertain to empirical paradoxicality.

The second objection is the easier of the two to interpret. To do so, one needs an account of falsity by default. A sentence that attributes truth_i or falsity_i to a sentence whose level is greater than or equal to i is *false by default*. It is possible that neither (4) nor (5) is false by default (indeed, Kripke describes situations in which this occurs). However, on the orthodox approach, one of them is false by default. If for some i, (4) is true_i, then (4) attributes falsity_{i-1} to its targets. (5) is among (4)'s targets. Thus, (5) must have level j where $j < i$. Of course, (4) is among the targets of (5). So, (5) is a level j sentence that attributes falsity_{j-1} to a sentence whose level is greater than j. Therefore (5) is false by default. Similar reasoning holds for the other direction. Therefore, the orthodox approach implies that either (4) and (5) are false by default, but there are situations in which neither one is false by default. Indeed, it is hard to accept the idea that any syntactically well-formed meaningful truth attributions of natural language are false by default.

---

[8] Kripke (1975: 696-697).

Notice that this point does not seem to have anything to do with the speaker having to pick a particular concept from the hierarchy in advance. Indeed, we could augment the orthodox approach by saying that the level of a truth predicate contained in a truth attribution is determined not by the intentions of the speaker, but rather by the levels of its targets, so that, if the highest level of a target of a truth attribution is i, then the truth predicate in that attribution has level i+1. This change would not avoid the second objection.

Kripke's first objection is more subtle. The problem here is no mere technical glitch. Instead, Kripke points out that we frequently attribute truth to truth bearers without knowing the levels of those truth bearers, and although it is perfectly legitimate for speakers to use truth predicates in this way, this fact is incompatible with the orthodox approach. These uses were a focus of Chapter Six and are integral to truth's expressive role. The reason we frequently do not know the levels of the targets when making truth attributions is that the level of a target cannot always be determined by its syntactic or semantic properties, and even these are often unknown when truth is used in its expressive role. The point Kripke makes in this first objection is a combination of truth's expressive role and the phenomenon at the root of empirical paradoxicality. That is, empirical paradoxicality is a special case of a more general phenomenon—the level of a sentence can depend on just about any fact, and so is often not determined by the sentence's syntactic or semantic features. We might call this phenomenon *empirical level-determination*. In many circumstances, speakers have no idea about the levels of the sentences to which they are attributing truth, so they cannot be expected to pick a concept from the hierarchy offered by the orthodox approach. There are plenty of cases where it is permissible for a speaker to assert a truth attribution even though she does not know the levels of its targets. Thus, if the orthodox approach were correct, then it would prohibit these uses of the truth predicate and, consequently, seriously limit truth's expressive role. In other words, the orthodox

approach is too demanding on speakers in much the same way as the language-specific views discussed in Chapter Six.

Once again, one might attempt to rescue the orthodox approach by stipulating that the subscript of the natural language truth predicate on an occasion of use is determined not by the intentions of the speaker (since the speaker often does not know which subscript to choose), but instead by the levels of the targets in question. That is, when a speaker asserts a truth attribution, the truth predicate gets a subscript that is one greater than the highest level of its targets. The problem with this suggestion is that it violates the Gricean Condition for which I argued in Chapter Six—it would imply that, in many cases, neither the speaker nor the audience has the information to determine what the speaker's sentence means. Thus, the combination of truth's expressive role, empirical level-determination, and the Gricean Condition sinks the orthodox approach to the aletheic paradoxes. This is how to read Kripke's first objection.

If a natural language truth predicate in a sentence is interpreted as being synonymous with one of the Tarskian truth predicates on the basis of the speaker's intentions, then a speaker should know the levels of the targets of a truth attribution before uttering it. But if speakers obey this condition on asserting truth attributions then they would be prevented from exploiting the truth predicate's expressive power. In particular, indispensable uses of truth predicates (i.e., where the speaker is incapable of directly asserting the propositions in question) would violate this condition.

7.3 Supervenience Theses

That there are empirically paradoxical sentences seems almost trivial when it is first presented, but Kripke was the first to realize that it could have far-reaching consequences.[9] He points out some of

[9] See also Prior (1961), which develops some of the problematic features of empirical paradoxes.

these and I mentioned them above: (i) paradoxicality is not determined by a sentence's syntactic features and (ii) paradoxicality is not determined by a sentence's semantic features. Moreover, his first objection depends on what I called *empirical level-determination* above, which has the following consequences: (i) a sentence's level is not determined by its syntactic features and (ii) a sentence's level is not determined by its semantic features.

We have already seen that these claims are powerful, especially when combined with the facts about truth's expressive role, and I use them substantially in what follows. It will be helpful to have a more precise formulation of them (and several others). Since they are the denials of dependence relations, and philosophers have found it illuminating to formulate dependence relations in terms of supervenience, we might consider thinking of them in those terms as well.

Supervenience came up in Chapter Five—it is a relation between kinds of properties. Philosophers have developed a large menu of supervenience relations to use in formulating philosophical theses, but only a few of these are relevant for our purposes.[10] All of them are explained as modal claims—that is, claims about relations between possible worlds. Here are the definitions of weak and strong supervenience:

> *A*-properties *weakly supervene* on *B*-properties if and only if for any possible world w and any individuals x and y in w, if x and y are *B*-indiscernible in w, then they are *A*-indiscernible in w.
>
> *A*-properties *strongly supervene* on *B*-properties if and only if for any possible worlds w_1 and w_2 and any individuals x in w_1 and y in w_2, if x in w_1 is *B*-indiscernible from y in w_2, then x in w_1 is *A*-indiscernible from y in w_2.

Individuals are A-indiscernible (or B-indiscernible) iff they have all the same A-properties (or B-properties) and lack all the same A-properties (or B-properties).[11] In our case, the individuals we

---

[10] For more on supervenience see McLaughlin and Bennett (2005), Leuenberger (2008), and the papers in Savellos and Yalçin (1995).

[11] These are both individual supervenience concepts (they pertain to the properties of individuals), but there are global ones as well, which are about entire possible worlds. Since we are interested in whether participants in a conversation

care about are sentences, the A-properties we care about are paradoxicality and having a certain level, and the B-properties we care about are syntactic and semantic properties, which should at least include the following:

> The *syntactic* properties of a sentence include: whether it is syntactically well-formed, its mood, its phrase structure, and the syntactic types of the words it contains.
>
> The *semantic* features of a sentence include: its sentential meaning, the proposition it expresses, the subsentential meanings of the words it contains, the contents of context-dependent terms it contains, and the referents of its singular terms.

Recall that in Chapter Six, I argued that in a conversation where the participants' utterances are felicitous, these properties are available to the participants.

The big question is whether to formulate the principles in terms of weak or strong supervenience. For our purposes, the denial of the strong supervenience claims is all that we need since we wish to consider different sets of empirical claims. So I stick with those:[12]

> *Syntactic Empirical Paradoxicality:* paradoxicality does not strongly supervene on the syntactic properties of a sentence; i.e., it is not the case that for any possible worlds w_1 and w_2 and any sentences x in w_1 and y in w_2, if x in w_1 and y in w_2 have the same syntactic properties, then they have the same paradoxicality status.
>
> *Semantic Empirical Paradoxicality*: paradoxicality does not strongly supervene on the semantic properties of a sentence; i.e., it is not the case that for any possible worlds w_1 and w_2 and any sentences x in w_1 and y in w_2, if x in w_1 and y in w_2 have the same semantic properties, then they have the same paradoxicality status.
>
> *Syntactic Empirical Level Determination*: the level of a sentence does not strongly supervene on the syntactic properties of the sentence; i.e., it is not the case that for any possible worlds w_1 and w_2 and any sentences x in w_1 and y in w_2, if x in w_1 and y in w_2 have the same syntactic properties, then they have the same level.
>
> *Semantic Empirical Level Determination*: the level of a sentence does not strongly supervene on the semantic properties of the sentence; i.e., it is not the case that for any possible worlds w_1 and w_2 and any sentences x in w_1 and y in w_2, if x in w_1 and y in w_2 have the same semantic properties, then they have the same level.

---

have enough information to determine either the level of some sentences or whether some sentences are paradoxical, it makes sense for us to focus on individual supervenience theses.

[12] It seems to me that the analogs of the syntactic theses for weak supervenience are correct as well, but I do not argue for that here.

Since we can think of being grounded as having a level, the claim that whether a sentence is grounded does not strongly supervene on its syntactic or semantic properties follows from the Empirical Level Determination theses (recall that groundedness was discussed in Chapters Three and Five).

Given the way Kripke introduces these ideas, it seems that the justification for these theses is that we need not recognize any relevant difference (syntactic or semantic) between the case where the sentence in question is non-paradoxical and the case where it is paradoxical; likewise we need not recognize a difference between cases where the sentence's level differs. These points are evident from Kripke's Nixon-Jones example. Consider two possible worlds, one in which, prior to uttering (2), Jones has uttered an equal number of true statements and false statements about Watergate, and the other in which they are unequal. In both worlds, Nixon's sentence, (1), has the same syntactic and semantic features, but in the first world it is paradoxical while in the second it is not. If, instead, we consider two possible worlds, one in which the maximum level of Jones' Watergate-related sentences uttered is 1 and the other where it is 2, then we get the similar results with respect to level determination.

It seems to me that the same considerations support extending these claims to pragmatic properties as well. With that we get:

> *Pragmatic Empirical Paradoxicality*: paradoxicality does not strongly supervene on the pragmatic properties of a sentence; i.e., it is not the case that for any possible worlds w_1 and w_2 and any sentences x in w_1 and y in w_2, if x in w_1 and y in w_2 have the same pragmatic properties, then they have the same paradoxicality status.

> *Pragmatic Empirical Level Determination*: the level of a sentence does not strongly supervene on the pragmatic properties of the sentence; i.e., it is not the case that for any possible worlds w_1 and w_2 and any sentences x in w_1 and y in w_2, if x in w_1 and y in w_2 have the same pragmatic properties, then they have the same level.

Under "pragmatic properties" one ought to include: the speaker's meaning of the sentence uttered, the conventional implicatures of the sentence uttered, the force of the utterance, the presuppositions of the utterance, and the conversational implicatures of the utterance.

The upshot is that the aletheic paradoxes are not a syntactic, semantic, or pragmatic phenomenon. Of course, for some paradoxical sentences, if one knows the syntactic, semantic, and pragmatic features of the sentence, then one can determine that it is paradoxical—no one denies that. However, the fact is that knowing this information is not sufficient to determine whether a sentence is paradoxical in general. The consequences are far-reaching—anyone who treats paradoxicality as a syntactic, semantic, or pragmatic feature seriously misunderstands the problem.

7.4 Consequences

The rest of this chapter is dedicated to consequences of the Empirical Paradoxicality theses and the Empirical Level Determination theses. Many of these consequences follow from these theses plus claims about truth's expressive role and the Gricean Condition (both presented in Chapter Six).

7.4.1 Disquotationalism, Minimalism, and Non-Classical Logic

This section covers what I take to be an important connection between the two traditions: if one accepts a disquotational theory of truth or a minimalist theory of truth, then it is very difficult to avoid adopting a non-classical logical approach to the paradoxes (either paracomplete or paraconsistent). This idea is not novel, but the argument I give in this section is new and appeals to empirical paradoxicality.[13]

[13] See Beall and Glanzberg (2008) for a precursor.

Begin with disquotationalism. If the disquotationalist accepts classical logic, then she must find some way of excluding the T-sentences for paradoxical sentences of L from her theory; otherwise, the disquotational theory for L is inconsistent. The alternative to excluding T-sentences for paradoxical sentences is to weaken one's logic to the point where a T-sentence for a paradoxical sentence does not engender inconsistency (or triviality). For the moment, I focus on the first option—excluding the troublesome T-sentences from the theory.

Vann McGee showed that the disquotationalist cannot simply stipulate that the theory of truth for L is just the maximally consistent set of T-sentences for L—there is no unique set that satisfies this criterion.[14] Instead, the disquotationalist must find some other way of specifying which T-sentences are parts of the theory and which ones are not.[15]

Jc Beall and Michael Glanzberg argue that a theory of transparent truth cannot exclude paradoxical T-sentences. A truth predicate, T, is *transparent* if and only if p and 'p is true' are intersubstitutable *salva veritate* in all extensional contexts. It is easy to show that a classical language with a transparent truth predicate and the expressive resources to formulate liar sentences is trivial (i.e., every sentence is a consequence of every set of sentences).[16] Thus, a disquotationalist who accepts classical logic has good reason to deny that the truth predicate in question is transparent. There is room for such a view since accepting all the T-sentences for a language and accepting that a truth predicate is transparent can come apart. For example, the inner theory of truth that results from Kripke's minimal fixed point with a strong Kleene scheme is transparent, but not all the T-sentences for that language are true; on the other hand, Priest's dialetheic theory of truth includes all

[14] McGee (1992).
[15] For discussion of this point and whether the disquotationalist can accommodate it, see Weir (1996), Gauker (2001), and Armour-Garb and Beall (2003).
[16] Beall and Glanzberg (2008).

the T-sentences for a language, but his truth predicate is not transparent.[17] Moreover, disquotationalism, as I have defined it, is free to reject intersubstitutability since it is free to reject paradoxical T-sentences in order to maintain consistency. The fact that there has been a fruitful debate about just how disquotationalists should pick out the T-sentences to reject suggests that this reading of disquotationalism is plausible.

One might wonder, will a non-transparent truth predicate still serve the purposes of the disquotationalist? Disquotationalists claim that truth predicates are devices of endorsement, rejection, and generalization. Assume we have a theory of truth T for a language L that consists of all and only the non-paradoxical T-sentences of L.[18] If p is not paradoxical then one can endorse p by asserting 'p is true' or reject p by asserting 'p is not true'. It does not seem like a major loss to not be able to indirectly endorse or reject paradoxical sentences. The generalizing function is a bit trickier. According to a disquotationalist, when one asserts 'a rational agent should believe something only if it is true', one uses 'true' as a device of generalization; one is generalizing over all sentences like 'a rational agent should believe that grass is green only if grass is green', and 'a rational agent should believe that monkeys grow on trees only if monkeys grow on trees', etc. Call these the *target claims*. Of course, the instances of the generalization are 'a rational agent should believe that grass is green only if 'grass is green' is true', and 'a rational agent should believe that monkeys grow on trees only if 'monkeys grow on trees' is true', etc. Given the T-sentences in T, when one asserts the generalization, one commits oneself to *almost* all the target claims. The only ones left out are ones pertaining to paradoxical sentences (e.g., L='L is false') like 'a rational agent should believe that L is false only if L is false'. It is not obvious that this limitation causes any trouble. Thus, if the disquotationalist excludes paradoxical T-sentences from the theory of truth, the truth predicate still

---

[17] See Kripke (1975) and Priest (2006a); see Field (2008a) for discussion.
[18] I am not concerned, at this point, with how the disquotationalist specifies this class of T-sentences.

serves as a device of endorsement, a device of rejection, and a device of generalization. It does not function in *exactly* the same way, but it does not seem that this difference really matters since it only leaves out paradoxical sentences.

However, there is good reason to think that the disquotationalist cannot make this move (i.e., eliminate the T-sentences for paradoxical sentences from her theory of truth). The argument is simple and depends on two points: (i) disquotational truth is a use-independent property, so the T-sentences of the theory are necessary, and (ii) there are contingently paradoxical sentences.

According to the disquotationalist, when one asserts "snow is white' is true', one attributes whiteness to snow; that is, for the disquotationalist, truth attributions pass straight through the sentences they seem to be about and talk about the world. This fact can seem perplexing, but the key to understanding it is the relativization of the truth predicate to a language. Usually we think that a sentence's meaning plus the world work together to determine that sentence's truth value, but for the disquotationalist, the meaning of the sentence is rendered irrelevant by the relativization of the truth predicate to a language. Disquotationalists differ on the details of this relativization, but it is essential to the theory. Field puts the point this way:

> [T]he use-independence of disquotational truth is *required* for the purposes just reviewed. For if 'All sentences of type Q are true' is to serve as an infinite conjunction of all sentences of type Q, then we want it to entail each such sentence, and be entailed by all of them together. This would fail to be so unless 'S is true' entailed and was entailed by S. But the only way that can be so is if 'true' doesn't ascribe a use-dependent feature to S. Suppose for instance that Euclidean geometry is true, and that we try to express its contingency by saying that the axioms together might have been false. Surely what we wanted to say wasn't simply that speakers might have used their words in such a way that the axioms weren't true, it is that space itself might have differed so as to make the *axioms as we understand them* not true. A use-independent notion of truth is precisely what we require.[19]

In order for disquotational truth to be use-independent, the T-sentences that make up the theory of disquotational truth for a language L must be necessary. If one of them is contingent, then for the

[19] Field (1994a: 266).

relevant sentence, p, of L, p and 'p is true' have different intensions since there would be a world in which they differ in truth value. This point is not controversial—all the disquotationalists accept this consequence of their theory, but it is relevant here for my argument.

The second point is the central topic of this chapter: there are contingently paradoxical sentences; i.e., sentences that are paradoxical but need not have been if empirical facts had turned out differently. Presumably, the label 'contingently paradoxical' should be applied only to a sentence that is actually paradoxical. We can use the term 'possibly paradoxical' for a sentence that is paradoxical in some worlds, but not in others; a possibly paradoxical sentence need not be paradoxical in the actual world. Given the extent to which we use a truth predicate as a device of endorsement and generalization, possibly paradoxical sentences abound. This fact is captured nicely in Kripke's quip, "*many, probably most, of our ordinary assertions about truth and falsity are liable, if the empirical facts are extremely unfavorable, to exhibit paradoxical features.*"[20]

Now let us put these two points together. If a T-sentence is in the disquotationalist's theory of truth for a language L, then that T-sentence is necessary. If q is a possibly paradoxical sentence of L, then there is a world, w, in which q is paradoxical. If q is paradoxical in w, then the T-sentence for q is also paradoxical in w; hence, the T-sentence for q is not true in w. Not only must the disquotationalist exclude the T-sentences for paradoxical sentences of L from her theory, she must exclude the T-sentences for *possibly* paradoxical sentences of L from her theory. Otherwise, she loses the necessity of the T-sentences, and with it, the use-independence of the disquotational truth predicate.

The major obstacle for the disquotationalist should now be clear—it is not that the disquotationalist must specify some way of excluding T-sentences for possibly paradoxical sentences of L; rather, it is that there is very little left of the disquotational theory once all the T-sentences for

[20] Kripke (1975: 691); italics in original.

possibly paradoxical sentences have been excluded. So far, discussions of disquotationalism and the liar have focused on how the disquotationalist can exclude the T-sentences for paradoxical sentences. My point is that since the disquotationalist is committed to use-independence, she must exclude not just T-sentences for paradoxical sentences of L, but T-sentences for *possibly* paradoxical sentences of L as well. Even if the disquotationalist can find a way to do that, the resulting theory tells us nothing about a substantial number of our sentences containing the truth predicate. Indeed, to put it in Kripke's terms, without the T-sentences for possibly paradoxical sentences, the disquotational theory of truth has nothing to say about "many, probably most, of our ordinary assertions about truth and falsity."[21] This result is damning—there is no viable disquotational theory of truth for L that excludes all the T-sentences for possibly paradoxical sentences of L.

Now consider minimalism. Unlike disquotationalists, minimalists treat propositions as primary truth bearers, so the issues of use-independence and language-specific truth predicates are avoided. However, the problem of empirical paradoxicality surfaces in a different way for the minimalist. Recall that minimalism about truth is the theory consisting of all and only instances of the following schema:

(Min) The proposition that **p** is true iff **p**.

Of course, if we let 'L' be a name for the proposition expressed by the sentence 'L is not true', then we can derive a contradiction from the minimalist theory using the liar reasoning. Paul Horwich, the founder of minimalism about truth, claims that paradoxical instances of (Min) should be eliminated from the theory; hence, the minimalist theory of truth consists of all and only non-paradoxical instances of (Min).[22] About the restriction to (Min), Horwich writes:

> Given our purpose, it suffices for us to concede that certain instances of the equivalence schema are not to be included as axioms of the minimal theory, and to note that the

[21] Kripke (1975: 691).
[22] Horwich (1998: 41-42).

> principles governing our selection of excluded instances are, in order of priority: (a) that the minimal theory not engender 'liar-type' contradictions; (b) that the set of excluded sentences be as small as possible; and—perhaps just as important as (b)—(c) that there be a constructive specification of the excluded instances that is as simple as possible.[23]

Like the disquotationalist described above, the minimalist needs to specify some principled way of excluding all the instances of (Min) from the theory of truth.

The points about empirical paradoxicality carry over to the minimalist's favored truth bearers. That is, just as some sentences are empirically paradoxical, some propositions are as well. For example, the propositions expressed by the sentences uttered in the exchange between Nixon and Jones in Kripke's example are empirically paradoxical. Those propositions would not have been paradoxical had the empirical facts (e.g., the number of claims Nixon made about Watergate) turned out differently. Just as above, let a proposition be *possibly paradoxical* iff it is paradoxical in some world.

The minimalist cannot eliminate all the instances of (Min) for possibly paradoxical propositions. Again, there are just too many of them. Instead, the minimalist needs some principled way of eliminating just the instances that are actually paradoxical. However, there does not seem to be any way to do it since whether a proposition is paradoxical might depend on any empirical fact. Consider the empirically paradoxical sentence in the first section—it expresses a paradoxical proposition, and so the instance of (Min) pertaining to it should not be included in the minimalist theory of truth. However, the only way to figure out that it should be eliminated is to go through all the sentences of the first section to see whether it is the only one that begins with an 'E'. The point can be generalized. Because of empirical paradoxicality, in order for the minimalist to specify which instances of (Min) are to be eliminated from the theory of truth, the minimalist would have to know all the empirical facts about the actual world. That is hardly a simple specification and it is not

---

[23] Horwich (1998: 42).

constructive. Moreover, only an omniscient being could figure out what the theory of truth is. For many sentences, the rest of us would be unable to determine whether their instances of (Min) belong in the theory or not.

A further problem is that this strategy for avoiding the aletheic paradoxes makes the theory of truth, and the concept of truth it implicitly defines beholden to the empirical facts about the actual world. If there had been another sentence beginning with 'E' in section one and that sentence had been true, then the sentence in question would not have been paradoxical. Thus, if there had been another sentence beginning with 'E' in section one and that sentence had been true, then the instance of (Min) in question would have been included in the minimalist theory. Consequently, in different possible worlds, we get different minimalist theories of truth. How troubling is this result? Consider what Horwich says about the relation between the minimalist theory and the concept of truth:

> The concept of truth (i.e., what is meant by the word 'true') is that constituent of belief states expressed in uses of the word by those who understand it—i.e., by those whose use of it is governed by the equivalence schema. And the theory of truth itself—specifying the explanatorily fundamental facts about truth—is made up of instances of that schema. Thus, the minimal theory of truth will provide the basis for accounts of the meaning and function of the truth predicate, of our understanding of it, of our grasp upon the concept of truth, and of the character of truth itself.[24]

Here Horwich posits a tight connection between the minimalist theory of truth and the concept of truth itself.

Consider again the possible world in which there is an additional true sentence in the first section of this chapter that begins with an 'E'. Call this world w′. As I said, the minimalist theory of truth for w′ is different from the minimalist theory of truth for the actual world (i.e., the former has the instance of (Min) for the proposition expressed by the inset sentence in section one, but the

[24] Horwich (1998: 37).

latter does not). However, there is no difference in anyone's conceptual competence, the explanatorily fundamental facts about truth, or anyone's understanding with respect to w′ and w; the only difference is in the sentences placed in section one of this book. Here is the crucial question: according to the minimalist, is the concept of truth different in those two worlds? If, on the one hand, it is, then minimalism retains a tight connection between the minimalist theory of truth and the concept of truth, but it makes the identity of that concept dependent on seemingly random empirical facts. It is hard to make sense of the idea that the concept of truth itself differs from world to world. If, on the other hand, it is not, then the minimalist avoids having to say that the concept of truth depends on seemingly random empirical facts, but this answer severs the tight connection between the minimalist theory of truth and the concept of truth. In w and w′ we would have the same concept of truth but different minimalist theories of truth, which would cast doubt on the explanatory efficacy of the minimalist theory.

To sum up: empirical paradoxicality poses serious problems for the disquotationalist and the minimalist insofar as these theorists intend to eliminate paradoxical instances of the T-schema from their theories. It seems to me that, at root, the problem is the same—in general, sentences or propositions, by themselves, are neither paradoxical nor non-paradoxical. Only when paired with an empirical description of the world can we say whether a truth bearer (of whatever kind) is paradoxical or not. Paradoxicality is a feature of truth bearer / world pairs, not a feature of truth bearers alone. However, by trying to eliminate instances of the T-schema for paradoxical sentences or propositions, the disquotationalist and the minimalist expose a fundamental misunderstanding of the problem posed by the aletheic paradoxes in general and empirical paradoxicality in particular.

The other alternative for the minimalist or disquotationalist is to abandon classical logic and use a weaker connective for the T-sentences than the material biconditional. However, pursuing this strategy requires solving a difficult *optimization problem*:

(i) the theorist needs a biconditional that is weak enough for one to accept the T-sentences even for paradoxical sentences or propositions, but

(ii) the biconditional needs to be strong enough to ensure that p and 'p is true' are intersubstitutable *salva veritate* in extensional contexts.

This is not an easy problem to solve. We find one workable solution from Hartry Field, and another from Jc Beall. As we have seen, Field's unified theory includes disquotationalism but abandons classical logic for paracomplete logic.[25] His disquotational theory includes all the T-sentences formulated with the paracomplete biconditional. Horwich could adopt this strategy as well for minimalism, but it would require giving up classical logic. Jc Beall's unified theory also includes disquotationalism, but abandons classical logic for paraconsistent logic. His disquotational theory includes all the T-sentences formulated with his paraconsistent biconditional. Again, this is a live option for Horwich as well.

The upshot of the argument in this section is that anyone committed to minimalism or disquotationalism is thereby committed to abandoning classical logic. Moreover, there are specific constraints on an acceptable logic; for example, intuitionistic logic is unacceptable. Field has shown that paracomplete logic *does* satisfy the disquotationalist's or minimalist's constraints and Beall has shown that paraconsistent logic does as well.

7.4.2 Syntax and Paradox

Some early approaches to the aletheic paradoxes imply that paradoxical sentences are not syntactically well formed.[26] I mentioned these briefly in Chapter Two. Although being inconsistent with established theories of syntax seems like a huge problem for these views, another problem

---

[25] Notice that another prominent disquotationalist, Vann McGee, does not fare so well. He advocates a classical symmetric logical approach (McGee 1991), which, in light of the point made in this section, is incompatible with his disquotationalism (McGee 1993).

[26] See Jorgensen (1953, 1955) and see Kattsoff (1955) for discussion.

stems from the Syntactic Empirical Paradoxicality Thesis. Namely, paradoxicality does not supervene on a sentence's syntactic features, and being syntactically well-formed is certainly a syntactic feature. Thus, it cannot be that all paradoxical sentences are not well-formed.

7.4.3 Meaningfulness and Paradox

Some contemporary philosophers still accept the meaningless approach to the paradoxes (Roy Sorensen endorsed it as recently as 2005).[27] However, the same considerations that show that paradoxical sentences are well-formed also show that they are meaningful. That is, given truth's expressive role and empirical paradoxicality, it would violate the Gricean Condition if all paradoxical sentences were meaningless.

A similar view that one sees occasionally is that paradoxical sentences do not express propositions. Indeed, some theorists hold that there are no paradoxical propositions whatsoever.[28] Hartry Field brings out the obvious problem with this view: "Beliefs can be paradoxical in classical logic, just as sentences can: I might believe that what the least intelligent man in the room is saying is untrue, on the mistaken assumption that the least intelligent man is someone other than me. In this case, the proposition believed is paradoxical in classical logic."[29] It would be highly counterintuitive to say that the person in question did not actually have a belief because there is no proposition to have the attitude of belief toward. Still, a proponent of the "no paradoxical propositions" view might bite the bullet on this point. However, the view also conflicts with the Gricean Condition. Consider a conversation in which someone asserts a sentence that turns out to be empirically paradoxical despite the fact that no one in the conversation is in a position to know that it is paradoxical. On the view in question, that sentence would not express a proposition, but no one in

[27] Sorensen (2005).
[28] See Glanzberg (2004) and Patterson (2010).
[29] Field (2008a: 296).

the conversation would be able to tell that it failed to express a proposition; this is impossible given the Gricean Condition. The lesson is that there are paradoxical propositions and one's approach to the aletheic paradoxes should be able to handle them.

7.4.4 Ambiguity and Paradox

Kripke attacked the orthodox approach, but it should be clear from the above discussion that Kripke's criticism generalizes to other philosophical approaches that appeal to ambiguity. The three characteristics of the orthodox approach that render it susceptible to Kripke's attack are: (i) a natural language truth predicate is interpreted as having multiple independent meanings, (ii) the appropriate meaning of the truth predicate on an occasion of use depends on features (i.e., the levels) of the sentences to which truth (of some kind or other) is being attributed, and (iii) the features of the sentences to which truth is being attributed might be unknown to a speaker felicitously uttering a truth attribution. Ambiguity approaches are almost always used in conjunction with a hierarchy of some sort (e.g., a hierarchy of truth predicates, determinacy operators, negations, conditionals, truth values). In each case, it is temping to treat the natural language term in question as if it is ambiguous, but the considerations offered in this chapter should prove decisive. In each case, the approach in question would satisfy the above three criteria and, thus, it would run afoul of empirical paradoxicality, empirical level seeking, truth's expressive role, or the Gricean condition.

These points against ambiguity approaches fit well with more general considerations attacking philosophical appeals to ambiguity. Paul Grice introduces what he calls Modified Occam's Razor: senses are not to be multiplied beyond necessity.[30] He uses this principle in a variety of objections

[30] Grice (1989: 47).

to those philosophers who posit hitherto unrecognized ambiguities to solve philosophical problems.[31] Kripke too rails against many philosophical uses of ambiguity in this oft-quoted passage:

> [I]t is very much the lazy man's approach in philosophy to posit ambiguities when in trouble. If we face a putative counterexample to our favorite philosophical thesis, it is always open to us to protest that some key term is being used in a special sense, different from its use in the thesis. We may be right, but the ease of the move should counsel a policy of caution: Do not posit an ambiguity unless you are really forced to, unless there are really compelling theoretical or intuitive grounds to suppose that an ambiguity is really present.[32]

Elizabeth Anscombe expresses a similar sentiment in the following passage: "where we are tempted to speak of 'different senses' of a word which is clearly not equivocal, we may infer that we are pretty much in the dark about the concept it represents."[33] The lesson for us is that those philosophers who appeal to ambiguity as part of an approach to the aletheic paradoxes are "pretty much in the dark" about the concept of truth.

7.4.5 Context Dependence and Paradox

Ambiguity approaches are not the only ones to fail in light of empirical paradoxicality; context-dependence approaches do as well, and for basically the same reason.

A context-dependence philosophical approach to the paradoxes claims that sentences containing natural language truth predicates are context dependent. However, there has been a tremendous amount of work recently on different kinds of context dependence. The most common view we can call *aletheic contextualism*, which holds that sentences containing 'true' express different propositions in different contexts. We can treat 'true' as having an invariant character (or meaning), which can be modeled as a function from contexts to contents, and a variable content, which can be modeled as a function from contexts to truth values.

---

[31] See Neale (1992) for discussion.
[32] Kripke (1977: 19).
[33] Anscombe (1957: 1). See also Atlas (1989).

7.4.5.1 Interpretive Systems

Let us be a bit more careful about the view. Following Stefano Predelli's recent treatment, let us distinguish between a *linguistic practice*, which consists of rational entities making noises and inscriptions in the course of their interactions with other rational entities, and an *interpretive system*, which is used as a tool by natural language semanticists to explain the semantic properties of those noises and inscriptions.[34] In all that follows, one must keep this distinction firmly in mind. Natural language semantics is typically taken to be the enterprise of systematically assigning truth conditions to sentences of natural language. However, as Predelli emphasizes, it is considerably more complex because the tool of natural language semantics, the interpretive system, is an abstract theory that does not take sentences of natural language as inputs and does not yield truth conditions as outputs. Instead, there is an extra layer of processing that occurs between the linguistic practice and the interpretive system. For example, most English sentences are ambiguous and must be disambiguated before they can be assigned truth conditions; 'e.g., Montgomery is at the bank' could mean that Montgomery is at the river bank or that Montgomery is at the financial institution. For this reason, interpretive systems do not take natural language sentences as input; instead, their inputs are complex structures that result from disambiguating sentences. I use Predelli's neutral term 'clause' for these items.

For context-dependent sentences, the interpretive system needs something in addition to the clause as input. Assume for example that Nelson is in fourth grade and Stu is an adult with a high

---

[34] Predelli (2005); 'interpretive system' is Predelli's term, 'linguistic practice' is my own. For those readers unfamiliar with formal semantics, Predelli is more careful about this distinction, but otherwise his presentation accords with the received view one finds in Lewis (1980), Dowty, Wall, and Peters (1981), Kaplan (1989), Chiercha and McConnell-Ginet (1990), Larson and Segal (1995), and Heim and Kratzer (1998). There are, of course, disputes about the details, but these do not matter for my purposes. Predelli's emphasis on this distinction also permits a defense of traditional formal semantics from its detractors (e.g., Travis (2008) and Recanati (2001, 2004, 2010)).

school diploma; if Nelson says 'I am in fourth grade' and Stu says 'I am in fourth grade', then they have uttered the same sentence (type), but Nelson has uttered something true, while Stu has not. Not only do the two utterances have different truth values, they express different propositions as well. The proposition Nelson uttered is true iff Nelson is in fourth grade, while the proposition Stu uttered is true iff Stu is in fourth grade. Clearly, the difference in propositions uttered should be traced to the word 'I' that occurs in the sentence. 'I' is a paradigmatic indexical, which has different semantic features in different contexts of use. In order for an interpretive system to handle cases like this, it needs more input than just a clause. It needs information about the environment in which these utterances occurred. Again, following Predelli, I use the term 'index' for the information that gets fed into the interpretive system, and 'context' for the concrete environment in which the utterance is performed. (Note that these two terms are used in so many different ways in philosophy of language and linguistics that it is a wonder anyone can follow.) For sentences like the one uttered by Nelson and Stu, the input for the interpretive system is a clause/index pair. For this example, the index needs to contain, at least, the speaker who uttered the sentence, since 'I' often (but not invariably) refers to the person who uttered or inscribed the sentence in which it occurs. Since there are other kinds of indexicals (e.g., 'here', 'now', 'that'), indexes need to contain more information (e.g., place, time, demonstratum).[35] Note that contexts in this sense can be modeled using the pragmatic theories described in Chapter Six (e.g., Stalnaker's theory, Lewis's theory, and Roberts' theory).

Just as interpretive systems accept only specific inputs, they produce special outputs. Recall that the goal is assigning truth conditions to sentences uttered in the linguistic practice; however, there is an additional level of complexity between the output of the interpretive system and the assignment of truth conditions. The truth conditions of a sentence are usually taken to be its truth value across

[35] Predelli (2005: ch. 1).

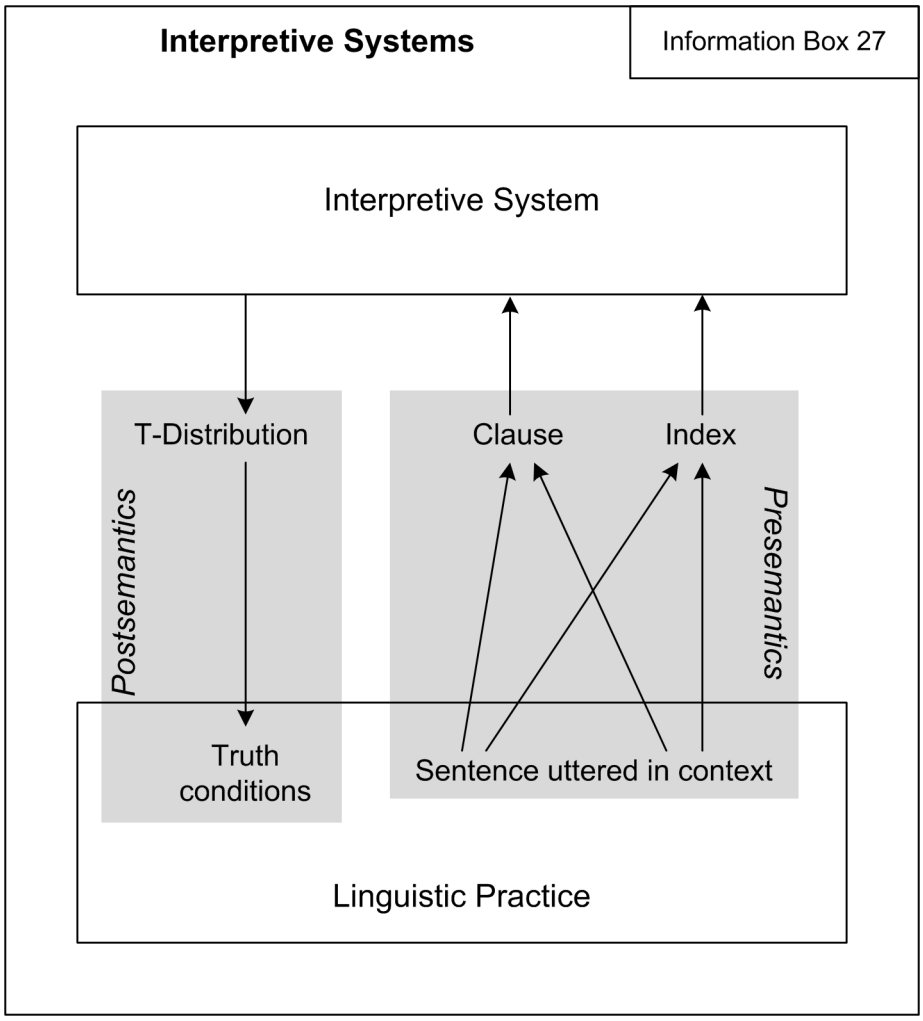
a variety of empirical descriptions (ways the world might be or have been). Predelli notices that this idea conflates the distinction between the interpretive system and the linguistic practice. Instead, utterances made in the linguistic practice have *truth conditions*, while the interpretive system outputs *t-distributions*, which are assignments of truth values to clause/index pairs relative to points of evaluation. Predelli writes:

> A crucial criterion for the assessment of an interpretive system's interest and tenability is its empirical adequacy: what is desired is that it yield results compatible with the intuitions of competent, intelligent speakers, or at least with those intuitions that we are willing to recognize as relevant. But such intuitions do not pertain to the mapping of abstract pairs with t-distributions: rather, they concern the conditions under which utterances of certain expressions on particular occasions turn out to be true. … Competent speakers' intuitions thus yield evaluations of particular utterances, while entertaining certain hypotheses about the way things may be: we may be invited to consider an utterance of 'Felix is on the mat' while imagining that Felix is on the table—that is, we may be invited to indulge in the thought-experiment involving that utterance and that (actual or merely imaginary) situation. Remaining deliberately neutral with respect to the metaphysical questions regarding the ontological status and structure of such 'ways things may be', I shall hereafter refer to the conditions of certain items, such as Felix's whereabouts, as a worldly condition. In this manner, competent speakers may be tested for their intuitions about the truth-conditions of an utterance: that is, about its truth-value with respect to alternative worldly conditions.[36]

The interpretive system takes a clause/index pair as input and assigns it truth values at each point of evaluation. The points of evaluation contain information like a possible world and a time. For each point of evaluation, the interpretive system assigns the clause/index pair a truth value. This array of truth values at points of evaluation is called a *t-distribution*. There is a gap between the t-distribution assigned to a clause/index pair and the truth conditions belonging to an utterance. That is, just as there is interpretive work to be done in moving from an utterance to a clause/index pair, there is interpretive work to be done in moving from a t-distribution to truth-conditions. A t-distribution is at the level of the interpretive system, whereas truth conditions are at the level of the linguistic practice. See Information Box 27 for a depiction of this relationship.

---

[36] Predelli (2005: 138-139).



So far, interpretive systems have been characterized as *black boxes*—that is, they have been described only in terms of their inputs and outputs. How exactly do they work? It should not come as a surprise that they are just mathematical theories familiar to anyone who has taken a class on symbolic logic. That is, they assign mathematical structures to the syntactic components of the clauses and a team of mathematical functions transform these mathematical structures into a t-distribution. To take a simple example, an interpretive system will have a domain (i.e., a set of entities); given a clause, for each point of evaluation, it assigns to each singular term in the clause an item from the domain, it assigns to each 1-place predicate in the clause a set of items from the domain, it assigns to each 2-place predicate in the clause a set of ordered pairs of items from the

domain, and so on. So, the clause 'the cat is on the mat' is broken into a singular term, 'the cat', a singular term, 'the mat', and a 2-place predicate, 'is on'. 'the cat' is assigned an entity from the domain at each point, 'the mat' is assigned an entity from the domain at each point, and 'is on' is assigned a set of ordered pairs of entities from the domain at each point.[37] For a given point, if the entity assigned to 'the cat' and the entity assigned to 'the mat' belong to an ordered pair assigned to 'is on', then 'the cat is on the mat' is assigned truth at that point (usually, the mathematical structure uses the number 1 for truth and the number 0 for falsity, but it does not really matter). For more complex linguistic phenomena (e.g., tense), the mathematical theory needs to be more complex, but this example should get across the basic idea.

Here is how an interpretive system handles indexicals. Consider again an example where Nelson asserts 'I am in fourth grade' in one conversation and Stu asserts 'I am in fourth grade' in a different conversation. We need to assign clause/index pairs to these two utterances. The clause is the same in each case—'I' is a singular term and 'am in fourth grade' is a 1-place predicate. The index is an n-tuple with entries for a speaker, a time, a location, a possible world, and possibly other information. For Nelson's utterance, the speaker entry in the index is Nelson, the time is the time at which Nelson utters the sentence, the location is Nelson's location when he utters the sentence, etc. Likewise, for Stu's utterance, the speaker entry in the index is Stu, the time is the time at which Stu utters the sentence, etc. Since 'I' is an indexical, the interpretive system assigns it a function from the index to the speaker entry in the index. Thus, when computing the t-distribution for the clause/index pair associated with Nelson's utterance, the interpretive system assigns Nelson as the semantic value (i.e., referent) of 'I' at each point. From this information, the interpretive system can assign a t-distribution to the clause/index pair associated with Nelson's utterance. For the point that

[37] For simplicity, I am treating clauses as sentences of a first order predicate calculus; a more subtle treatment would consider syntactic trees in categorical grammars. See Montague (1974), Dowty, Wall, and Peters (1981), Cherchia and McConnell-Ginet (1990), and van Benthem and ter Meulen (1997) for details.

represents the actual world and the current time, this clause/index pair is assigned truth because Nelson is a member of the set of things that are in fourth grade at that point of evaluation. For Stu's utterance, the interpretive system assigns Stu as the semantic value (i.e., the referent) of 'I' at each point, and it uses this information to compute the t-distribution for the clause/index pair associated with Stu's utterance. For the point that represents the actual world and the current time, this clause/index pair is assigned falsity because Stu is not a member of the set of things that are in fourth grade at that point of evaluation. The function assigned to 'I' in each case is the same—it picks out the speaker entry in the index; this is called the *character*. However, the object assigned to 'I' in the two cases is different; this is called the *content*. The content of 'I' differs from utterance to utterance, but the character is invariant. For whole sentences, one can think of the content as the t-distribution (i.e., truth values at points of evaluation), whereas one can think of the character as a function from indices to contents. The character of the sentence Nelson utters is the same as the character of the sentence Stu utters, but their contents are different—Nelson's utterance is true and Stu's is false.

7.4.5.2 Aletheic Contextualism

Context-dependence approaches to the aletheic paradoxes imply that 'true' is an indexical.[38] Since 'true' is a 1-place predicate, we need to see exactly how this view fits into the above description of interpretive systems. To do so, we use an analogy between 'true' and gradable adjectives like 'tall', 'flat', 'rich', 'old', etc. These display context dependence in the following way: if, in a conversation about fourth graders, Martin asserts 'Nelson is tall', but in a conversation about NBA stars, Kearney asserts 'Nelson is not tall', these two utterances could both be true despite the fact that Nelson does

[38] Here and throughout, I am using 'indexical' in its broad sense; Stanley (2007: 38) distinguishes between the two senses. This distinction does not matter for my purposes, and I flag it only to avoid misinterpretation.

not change in height between the two conversations. The reason is that the standards for what counts as tall change from conversation to conversation. We model this by including a slot in the index for tallness standards associated with utterances involving 'tall'.[39] In the first conversation, say, anyone who is over 160 cm counts as tall, while in the second conversation, anyone who is over 210 cm counts as tall. The interpretive system gives the character to 'tall' that picks out the standard entry in the index. If Nelson's height is 180 cm, then Martin's utterance is true because Nelson is a member of the set of entities greater than 160 cm in the point of evaluation that represents the actual world and time of Martin's utterance, and Kearney's utterance is also true because Nelson is not a member of the set of entities greater than 210 cm in the point of evaluation that represents the actual world and time of Kearney's utterance.

Advocates of context dependence approaches to the aletheic paradoxes (*aletheic contextualists* from here on) claim that truth predicates behave in the same way, except that instead of standards of truth, the index needs a slot that picks out a restricted concept of truth. Thus, on this view, clauses containing truth predicates have invariant characters, but their contents (i.e., t-distributions) differ depending on the index. At the linguistic practice level, utterances of sentences with truth predicates have different truth conditions in different contexts.

Tyler Burge proposes a kind of aletheic contextualism. He suggests that the content of 'true' in a context of use is the content of one of the truth predicates in the Tarskian hierarchy. That is, instead of ambiguity as the interface between natural language truth predicates and the Tarskian hierarchy (as the orthodox approach holds), Burge claims that indexicality is the interface.[40] Other aletheic contextualists disagree with Burge, but they all agree that: (i) there is a set of restricted concepts of truth, (ii) sentences containing natural language truth predicates have contents that

---

[39] See Kennedy (1999) for details on the semantics of gradable adjectives.
[40] Burge (1979a, 1982a, 1982b).

differ from context to context, and (iii) these contents are constituted in part by restricted concepts of truth. If we use the term 'aletheic standards' to mark the slot in the index that determines which restricted concept of truth is relevant for the utterance in question, then aletheic contextualists hold that the aletheic standard for a given clause/index pair is determined by context of the utterance in question.[41]

7.4.5.3 Semantic Blindness

There are some serious problems for aletheic contextualism and one of the worst is that we do not treat truth predicates as if they are context dependent in this way. Now, there are plenty of terms that might not seem as though they are context dependent, but upon reflection it is pretty clear that they are. For example, gradable adjectives mentioned above are context dependent, but might not seem so at first glance. However, a bit of reflection should clear this up. A second grade teacher says that her six-foot student is tall, and basketball coach says that his six-foot player is not tall. They are not disagreeing about whether six-foot people are tall. In the first context, the six-foot person is tall, while in the second, the six-foot person is not tall. Any competent English speaker ought to be able to identify this instance of context dependence with minimal prompting.

However, there are words that, even upon reflection, do not seem to be context dependent; nevertheless, that does not deter some philosophers from claiming that they are context dependent. One important discussion for my purposes pertains to contextual analyses of knowledge. For

[41] Note that some philosophers have proposed tests for context dependence. For example, Cappelen and Lepore (2003) suggest the Intercontextual Disquotation Test (IDT), which states that e is a context-dependent expression iff one can truly assert that, for some sentence $\langle p \rangle$ containing e, there are false utterances of $\langle p \rangle$ even though p. Obviously, in the case of truth, the IDT tests whether a truth predicate obeys (T-In). Thus, passing the IDT would prevent truth from playing its expressive role as described in Chapter Six. Another example is the disagreement test from Cappelen and Hawthorne (2009). The disagreement test (DT) states that if e is a context-dependent expression, then there will be cases where a person A asserts that p (where $\langle p \rangle$ contains e) in one context, a person B asserts $\langle \sim p \rangle$ in a different context, yet they cannot correctly be said to disagree by someone in a third context. Again, it seems as if 'true' fails this test as well. That might be bad news for the aletheic contextualist, but I do not put much stock in these kinds of tests since their outcomes are so susceptible to one's antecedent theoretical commitments.

centuries, philosophers have pondered the problem of epistemological skepticism, which threatens to undermine the extent of our knowledge because of our inability to rule out bizarre hypotheses like being deceived by an evil demon or being a brain in a vat.[42] It is very difficult to reconcile our intuition that we know many things with the problem of skepticism. One solution that has become very popular and the source of tremendous debate is that 'knows' is context dependent—in everyday contexts the extension of 'knows' is wide, but in contexts where skeptical hypotheses are salient, its extension is very narrow.[43] This view of 'knows' as context dependent has the potential to resolve the tension in our intuitions. However, it has come in for plenty of criticism. One of the most common objections is that the claim that 'knows' is context dependent is itself highly counterintuitive since speakers, even upon reflection, take it to be invariant. This worry and others like it have come to be known as *semantic blindness* objections (because contextualists about knowledge take speakers to be blind to the semantic features of 'knows').

Stephen Schiffer was one of the first to push this objection:

> For the speaker would not only have to be confounding the proposition she's saying; she'd also have to be totally ignorant of the sort of thing she's saying. One who implicitly says that it's raining in London in uttering "It's raining" knows full well what proposition she is asserting; if articulate, she can tell you that what she meant and was implicitly stating was that it was raining in London. But no ordinary person who utters "I know that p", however articulate, would dream of telling you that what he meant and was implicitly stating was that he knew that p relative to such-and-such standard.[44]

For Schiffer, the problem with epistemological contextualism is that speakers do not know which propositions they are asserting when they assert knowledge claims. Thomas Hofweber disagrees:

> According to the contextualist, it is not so that when one speaker utters "A knows that p" and another speaker utters the same sentence then the content of the two utterances will be the same. And it is not so that if one speaker utters "A knows that p" and another speaker utters "A does not know that p" then the contents of these two utterances are incompatible.

---

[42] Descartes (1641) and Putman (1981); for discussion see Stroud (1984) and Williams (1996).
[43] See Cohen (1986), DeRose (1992), and Lewis (1996) for examples; see Hawthorne (2004) and Stanley (2005) for discussion.
[44] Schiffer (1996: 326).

> Furthermore, according to contextualism, the speakers won't be aware of these facts about difference and compatibility of contents. This follows from the fact that ordinary speakers are not aware of the semantic context sensitivity of their knowledge ascriptions, and from the claim that lots of details of the context are relevant for what the content of a knowledge ascription is. … A contextualist will thus not only hold that speakers have no access (in the strong sense spelled out above) to the content of their utterances, but also no access to sameness, difference and incompatibility of the contents of their utterances. … I think that *it*, not hidden relativity per se, is the really problematic aspect of the philosophy of language part of a contextualist theory about knowledge ascriptions. It is one thing to deny that speakers have access to the content of their utterances in the strong sense spelled out above. … It is quite another thing to deny that sameness, difference and incompatibility of contents of utterances is inaccessible to ordinary speakers.[45]

According to Hofweber, that speakers do not know exactly which propositions they are asserting when they assert knowledge claims is not a problem; rather, the fact that they would be unable to determine when knowledge claims express the same or different or incompatible propositions that poses the real problem. John Hawthorne presents a more general version of this worry:

> Suppose that Joe says 'I know that p', and at the time of utterance he expresses the same relation as you do by 'know'. You accept what he says. The sentence 'Joe knows that p (plus a date index)' goes into your belief box. But now suppose that your standards for knowledge rise. Your belief about Joe's knowledge will now come out false—as will, presumably, hundreds of other once true beliefs—unless you somehow update the sentences in your belief box. Moreover, you will no longer have a cognitive hold on those true propositions that your belief box once truly expressed and that, as a result, you once truly believed.
>
> Similarly, if the standards for knowledge fall, many beliefs that deny knowledge will now come out false and much true information will be lost, unless updating occurs. Suppose further that we are semantically blind in the way suggested: the semantic content of 'knows' shifts, but our language organ does not supply us with a standards index—analogous to a dating method—with which to enrich knowledge ascriptions that are tokened in the belief box. Then shifting semantic values for 'know' would spell disaster—in the ways just outlined—for our belief set.[46]

Hawthorne takes the problem to be that if we were semantically blind in the way that epistemological contextualists suggest, then we would be unable to keep track of our own and others

[45] Hofweber (1999: 101)
[46] Hawthorne (2004: 110).

beliefs dynamically in the way that is a minimal requirement for participating in a conversation. That seems to be what bothers Hofweber as well.[47]

The semantic blindness objection has come to be one of the major issues in the debate over epistemological contextualism and contextualist views in philosophy more generally. The contextualists' responses to it have mostly been of the "learn to live with it" and "it is not that counterintuitive" variety.[48] I am not going to evaluate these replies, but I do want to draw some comparisons to aletheic contextualism.

Notice that as long as speakers and audience members are able to determine the epistemic standards that are relevant to a given knowledge attribution, epistemological contextualism is *compatible* with the Gricean Condition defended in Chapter Six, which requires that information for determining the contents of properly used context-dependent expressions be available to participants in a conversation.[49] Thus, the semantic blindness objection (or objections, since they seem to differ slightly from one another) appeal to a condition on language use that is *more* demanding than the Gricean Condition. Still, it seems to me that semantic blindness objections do pose serious problems for epistemological contextualism and for aletheic contextualism, which is one of the most popular philosophical approaches to the aletheic paradoxes.[50]

7.4.5.4 Aletheic Standards and the Gricean Condition

---

[47] See also Stanley (2005: 22, 115-122).
[48] See DeRose (2006) and Montminy (2009). See Capellen (2007), Brogaard (2008), and Weiner (MS) for discussion.
[49] Of course, if speakers do not know that 'knows' is context dependent, then they will probably not utilize this information, even though it could be available to them.
[50] Zoltan Szabo's comment seems apt: "[A]ppeals to context sensitivity have become 'cheap'—the twenty-first century version of ordinary language philosophy's rampant postulations of ambiguity. Not only is this 'the lazy man's approach to philosophy,' it undermines systematic theorizing about language. The more we believe context can influence semantic content, the more we will find ourselves at a loss when it comes to explaining how ordinary communication (let alone the transmission of knowledge through written texts) is possible," Szabo (2006a: 31).

Even if the semantic blindness objection is *not* decisive against epistemological contextualism and aletheic contextualism, there is a bigger problem for *aletheic* contextualists. That problem, as you must have guessed, is that when speakers use truth predicates in their expressive roles, it is often the case that neither the speaker nor the audience knows the levels or paradoxicality status of the targets of the truth attribution. Moreover, it is these features of the targets that would determine the content of the truth predicate according to aletheic contextualists. Thus, not only does aletheic contextualism face the semantic blindness objection, but also it confronts a serious violation of the Gricean Condition in situations where truth plays its expressive role.

Consider the problem in a bit more detail. When, in Kripke's example, Nixon asserts 'All of what Jones says about Watergate is true', what is happening according to the pragmatic theories covered in Chapter Six and the semantics of interpretive systems described above? Nixon is part of a conversation, and (following Stalnaker, Lewis, and Roberts) that conversation has a common ground, common goals, questions under discussion, etc. These elements of the conversation are the context of Jones' utterance. The aletheic contextualist holds that 'true' in the sentence Nixon asserts is an indexical.[51] Accordingly, the truth conditions of Nixon's utterance are determined by the t-distribution an interpretive system assigns to the clause/index pair that represents Nixon's utterance. Determining a clause is straightforward, but the problem comes when specifying an index; in particular, the aletheic standard entry in the index should be read off the context of Nixon's utterance. What is the aletheic standard in this case? If Nixon's utterance is felicitous, then the aletheic standard will have to be able to encompass all of Jones' claims about Watergate. If we are using the Tarskian hierarchy as our example for the set of restricted truth predicates, then the aletheic standard will have to be higher than the highest level of the sentences Jones uttered about

[51] Again, there are aletheic contextualists (e.g., Glanzberg) that think the context dependence should be traced to a quantifier domain rather than to the truth predicate. The objection carries over easily to these views.

Watergate. The problem is that there is no reason to think that this information will be part of the common ground in Nixon's conversation. Thus, the aletheic contextualist holds that the aletheic standard in the index is to be filled in by consulting the context, but when one looks to the context, at least in the vast majority of cases where truth is being used in its expressive role, that information is nowhere to be found.[52]

All that is needed for Nixon's utterance to be felicitous is that he is following the rules of the conversation—that is, he is cooperating, and so on. Given the description of the context, his utterance is felicitous. However, if the aletheic contextualist is right, then even though his utterance is felicitous, it cannot be assigned truth conditions by an interpretive system since the context in which it was made does not determine an aletheic standard. On the other hand, the aletheic contextualist might deny that Nixon's utterance is felicitous for these very reasons. If so, then this move would be tantamount to demanding that sentences containing truth predicates can be felicitously uttered only when the context contains enough information to determine an aletheic standard. This view is obviously empirically false—we often use truth predicates in contexts insufficient to determine an aletheic standard, and competent natural language users correctly regard these utterances as felicitous.

Instead, the aletheic contextualist might claim that the aletheic standard is specified in some other way that goes beyond the information available in the context. If so, then the content of the sentence Nixon utters is not available to anyone in the conversation. Thus, although Nixon's utterance is felicitous, it would violate the Gricean Condition defended in Chapter Six. Recall that the Gricean Condition is accepted by every major theory of pragmatics; thus, aletheic contextualism

---

[52] Curiously, Keith Simmons discusses Stalnaker's pragmatic theory in conjunction with Simmons' version of aletheic contextualism in Simmons (2003), but he does not recognize the problem.

is at odds with a fundamental result from the science of linguistics. Again, one might as well be a creationist or a flat-Earther.

To sum up: aletheic contextualism is a popular philosophical approach to the aletheic paradoxes, but when examined a bit more closely, it encounters what I take to be an insuperable difficulty; namely, it is incompatible with the combination of truth's expressive role, empirical paradoxicality, and the Gricean Condition.

7.4.6 Field on Truth and Determinate Truth

The last two subsections on ambiguity and context-dependence point out problems for approaches to the aletheic paradoxes that posit a group of more or less restricted truth predicates and attempt to link them to a natural language truth predicate via some hitherto unrecognized semantic feature. Earlier I called these *fragmentary theories of truth* since they try to explain a seemingly unitary notion like truth in terms of a group of more restricted notions (e.g., those in the Tarskian hierarchy). It should be pretty clear that any fragmentary theory of truth that appeals to something like ambiguity or context-dependence to relate natural language truth predicates to the group of restricted truth predicates is going confront a semantic blindness objection. Moreover, it is going to fall prey to the much more serious objection I have presented in this chapter based on Kripke's criticism of the orthodox approach; namely, it cannot accept the combination of the Gricean Condition, the fact that truth predicates have expressive uses, and the fact that there are empirically paradoxical sentences. In my opinion, that gives us good reason to avoid fragmentary theories of truth altogether.

What might not be so obvious is that these same problems can crop up even for someone who does not endorse a fragmentary theory of truth. In this subsection, I consider Hartry Field's combination of indeterminacy philosophical approach and paracomplete logical approach to the

aletheic paradoxes. Recall that Field offers an artificial language that has a paracomplete logic (K_3), but with a new conditional (I will call it a *paracomplete conditional*) that is defined via a revision sequence (he also gives a neighborhood semantics for it and an algebraic semantics for it, but we will not be concerned with those details—see Appendix Two of Chapter Three). Using the paracomplete conditional, Field defines a determinateness operator, D. The language Field considers also has its own truth predicate and it can express Field's preferred theory of truth, which is just the set of T-sentences for the sentences of the language, except that the T-sentences contain Field's paracomplete biconditional. The determinateness operator can be paired with the truth predicate to express determinate truth, and there is a crucial difference between saying that a sentence is true and saying that a sentence is determinately true.

Of course, Field's language has liar sentences (e.g., L= 'L is not true'), and his theory of truth contains T-sentences for all of them. However, the paracomplete biconditional and the background paracomplete logic are weak enough to prevent a proof of contradiction from them. As for their truth value, Field's theory neither implies that they are true nor that they are not true (remember, it is a paracomplete logic, so it does not validate the law of excluded middle); moreover it does not imply that they are neither true nor not true—it is silent about their truth value. However, one can use the determinateness operator to say something about them: liar sentences are not determinately true and not determinately not true; that is implied by the theory. So the determinateness operator comes in handy for classifying paradoxical sentences. Since it is in the language, though, one can formulate revenge liar sentences with it (e.g., R= 'R is not determinately true'). Here, the theory does not imply that they are true nor that they are not true, and it does not imply that they are neither true nor not true; in addition, it does not imply that they are not determinately true and not determinately not true either. However, it does imply that they are not determinately determinately true and not determinately determinately not true. That is, the determinateness operator iterates

non-trivially so that iterations of it can be used to classify revenge liars. Indeed, Field shows how to define a hierarchy of determinateness operators that stretches into the recursive ordinals. However, his view does not allow for a general notion of determinateness that encompasses all those in the hierarchy.[53]

One might complain that instead of fragmenting truth, Field has fragmented determinateness. His reply? Guilty as charged, but fragmenting determinateness is much less problematic:

> On the views considered here, we do have a unified notion of truth (and of satisfaction too). It is the notions of truth and satisfaction, not determinate truth, that we need to use as devices of generalization. … [T]he ability to use truth as a device of generalization isn't affected at all in the paracomplete theories now under discussion, since the truth predicate isn't even quasi-stratified.[54]

Field's point is that it makes much more sense to fragment determinateness than truth given the expressive role that truth plays. Let us see if he is right.

Consider three examples:

(i) Moe asserts 'The Riemann hypothesis is not true'.

(ii) Carl asserts 'The sentence on the blackboard is not true', where the sentence in question is empirically paradoxical (like sentence (1) of this chapter).

(iii) Lenny asserts 'The sentence on the whiteboard is not true', where the sentence in question is an empirically paradoxical revenge liar (e.g., 'Most of Jones' statements about Watergate are not determinately true).

These are all expressive uses of the truth predicate of English and these utterances are all felicitous (given normal conversational circumstances). Our job is to see how Field's theory handles them.

In (i), it is pretty easy; 'true' in Moe's sentence means *True* (which is the concept of truth expressed by the truth predicate in Field's artificial language). It might seem that we want to say the same thing about (ii), but there is a catch. Since the sentence on the blackboard is indeterminate (i.e., not determinately true and not determinately not true), this reading implies that Carl's sentence

---

[53] All of this is spelled out in Field (2008a); I also discussed it in Chapters Two and Three.
[54] Field (2008a: 349).

is indeterminate as well. Moreover, Field claims that it is inappropriate to say of an indeterminate sentence that it is true or that it is not true, and this seems like a good norm to use when assessing assertions of sentences that could turn out to be indeterminate.[55] Thus, if 'true' in Carl's sentence expresses the concept of truth, then his assertion is illegitimate. That move would seriously impact the expressive role of the truth predicate. Instead, we can read 'true' in Carl's sentence as meaning *determinately True*. If we do this, then his assertion is legitimate. Therefore, either Carl's sentence is indeterminate (and so his assertion is unwarranted), or 'true' in his sentence means *DTrue*.

The same problem arises for Lenny's sentence but at a level higher. That is, since the target of Lenny's attribution is not determinately determinately true and not determinately determinately not true, treating 'true' in his sentence as expressing either truth or determinate truth results in his sentence being indeterminate. Again, that would render his assertion infelicitous and this choice would negatively impact the expressive power accorded to the truth predicate. Instead, we should read 'true' in Lenny's sentence as meaning *determinately determinately True*. On this reading, his assertion is fine. Therefore, either Lenny's sentence is indeterminate (and so his assertion is unwarranted), or 'true' in his sentence means D^2True.

As I see it, Field takes some of the work traditionally given to the truth predicate and outsources it to his determinateness operators. The problem for Field's view comes when we think about how to interpret an ordinary speaker of English who has never heard of Field's solution or even the liar paradox itself; we are forced to make a hard choice as to how we should interpret some of that person's uses of 'true' to express disagreement. We have to conclude that either 'true' is not univocal and invariant or that some sentences containing the truth predicate that are legitimately used to express disagreement are indeterminate. It seems to me that either Field's theory is seriously revisionist (not just about our logic, but about our use of 'true'), which even in the best-case scenario

[55] See Field (2008a: 350-353). See Maudlin (2004) for an alternative view.

would leave us without a theory of truth *as we use it*, or he is forced to treat truth predicates as ambiguous or context-dependent. Each of these alternatives has major costs associated with it; the first makes it seem as though Field's points about how we use truth predicates are just cherry-picked to support his theory. However, in light of the criticism of ambiguity and context-dependence approaches to the aletheic paradoxes given in the last two sections, we can see that the second option is equally bad. It should be obvious that the features determining which meaning 'true' should have in these three cases are often unavailable to the participants in the conversations in question. Thus, the claim that 'true' is context-dependent in this way violates the Gricean Condition, in just the same way we saw in the last subsection; again, going with ambiguity over context-dependence would not help. I find it remarkable and surprising that even a theory of truth designed around truth's expressive role could encounter this kind of trouble when it is applied to a natural language.

The lesson here is that facts about truth's expressive role, facts about empirical paradoxicality, and the Gricean Condition make for an extremely powerful combination. Even theories of truth that do not seem to require interpreting truth predicates of natural language as ambiguous or context-dependent might inadvertently have these consequences, and the considerations in this chapter bring that out.

7.5 Impact

The results of this chapter have the most impact on philosophical approaches to the paradoxes. Grammaticality, Meaningfulness, Assertibility, Ambiguity, and Context-Dependence approaches to the aletheic paradoxes all come in for serious problems when one considers the combination of truth's expressive role, the Gricean Condition, and empirical paradoxicality. I am inclined to think that this problem is decisive, but of course, one has to evaluate it in light of the problems faced by

other views. In addition, Field's version of the indeterminacy approach has serious hidden costs in light of what is essentially the same problem. Finally, considerations arising from empirical paradoxicality force disquotationalists and minimalists to adopt a non-classical logical approach to the paradoxes—either the paraconsistent approach or the paracomplete one. That is a serious hidden cost to the two most prominent versions of deflationism.

Chapter 8

Aletheic Vengeance

The first two chapters of Part II covered the expressive role of truth and empirical paradoxicality. The third key idea in Part II is that approaches to the aletheic paradoxes tend to give rise to revenge paradoxes. This topic has already come up a few times, but I have yet to explain it in detail or discuss its significance.

8.1 Revenge Paradoxes

Revenge paradoxes have been known for decades, but they used to be called 'strengthened liars'.[1] Sometime in the late 1990s the term 'revenge paradox' caught hold and now seems to be the accepted locution. Just as there are many kinds of liar paradoxes, there are many kinds of revenge paradoxes. A revenge paradox is structurally similar to a liar paradox and the sentence causing the revenge paradox often contains not only a truth predicate or falsity predicate, but also some additional linguistic expression that features in an approach to the liar paradox. For example, the following sentence is a liar sentence:

(1) (1) is false.

Assume for a moment that we accept the inner theory of Kripke's strong Kleene minimal fixed point. According to this approach, (1) is not in the extension of 'true' and not in the anti-extension of 'true'. We could introduce the term 'gappy' into the language in question, and characterize (1) by saying that it is gappy. However, by doing so, we also accept that the language has the following sentence:

[1] See van Fraassen (1968).

(2) (2) is either false or gappy.

This sentence gives rise to a revenge paradox for the approach under consideration. Recall that the liar paradox is generated by principles of logic and the following aletheic principles:

(T-In) If p, then $\langle \mathsf{p} \rangle$ is true

(T-Out) If $\langle \mathsf{p} \rangle$ is true, then p.

(Sub) If $\langle \mathsf{p} \rangle = \langle \mathsf{q} \rangle$, then $\langle \mathsf{p} \rangle$ is true $\leftrightarrow$ $\langle \mathsf{q} \rangle$ is true.

Now, let us compare the liar reasoning with the reasoning concerning our revenge paradox:

	Liar Reasoning	*Revenge Reasoning*	
1.	Assume (1) is true	Assume (2) is true	
2.	'(1) is false' is true	'(2) is either false or gappy' is true	[Sub]
3.	(1) is false	(2) is either false or gappy	[T-Out]
4.	If (1) is true, then (1) is false	If (2) is true, then (2) is either false or gappy	[$\rightarrow$ Intro]
5.	Assume (1) is false	Assume (2) is either false or gappy	
6.	'(1) is false' is true	'(2) is either false or gappy' is true	[T-In]
7.	(1) is true	(2) is true	[Sub]
8.	If (1) is false, then (1) is true	If (2) is either false or gappy, then (2) is true	[$\rightarrow$ Intro]
9.	(1) is true iff (1) is false	(2) is true iff (2) is either false or gappy	[$\leftrightarrow$ Intro]
10.	(1) is true and (1) is false	(2) is true and (2) is either false or gappy	[CTF Conseq][2]

Using the exact same reasoning from the liar paradox, we get a contradiction by reasoning about (2).[3] The approach in question blocks the liar reasoning from step 9 to step 10. However, the same move does not block the reasoning from step 9 to step 10 in the revenge case. It is consistent to say that (1) is gappy, but from the claim that (2) is gappy, it follows that (2) is true since it says of itself that it is either false or gappy.

Notice that the revenge paradox generated by (2) is paradoxical only if we accept the above approach to the liar paradox. That is an important feature of revenge paradoxes: whether a sentence generates a revenge paradox is relative to a particular approach to the liar. (2) is a revenge paradox

[2] This stands for 'Classical Truth Functional Consequence'.

[3] Note that the contradiction is not a consequence of (2)—the contradiction in this derivation has no undischarged assumptions, just as in the case of the liar.

for the inner theory of Kripke's strong Kleene minimal fixed point. It is not a revenge paradox for other approaches.

Other examples of sentences that generate revenge paradoxes are:

(3) (3) is either false or unstable [revision]

(4) (4) is not true in any context [context dependence]

(5) (5) is either false or indeterminate [paracomplete]

(6) (6) is just false [paraconsistent]

(7) (7) is Xnot true[4] [paracomplete]

(8) (8) is Bnot true[5] [paraconsistent]

Revision theories say that (1) is unstable since its truth value never stabilizes in a revision sequence. The truth value of (3) never stabilizes in a revision sequence either, but if (3) is unstable then (3) is true since it says of itself that it is either false or unstable. So (3) is a revenge liar for approaches that appeal to revision sequences.

Context-dependence approaches say that (1) is true in some contexts and false in others. That approach blocks the liar reasoning since it stipulates that the context shifts in the midst of the liar reasoning. However, (4) poses a serious problem for context dependence views because the claim that it is true in one context seems to imply that it is not true in any context, so the contextualist seems to have a problem.

Paracomplete approaches typically say that (1) is indeterminate and they reject the principles of classical logic involved in the liar reasoning. However, if the paracomplete approach calls (5) indeterminate, then it implies that (5) is true since it says of itself that it is indeterminate. Since sentences are not both indeterminate and true, (5) poses a problem for these approaches.

---

[4] See below for how to understand 'Xnot'.
[5] See below for how to understand 'Bnot'.

Paraconsistent views say that (1) is both true and false and they offer a non-classical logic on which some contradictions can be true (though they hold that not all contradictions are true). Paraconsistentists hold that (1) says of itself that it is false, and it is false, so it is true as well; (1) is both true and false. However, there is a problem with saying that (6) is both true and false since (6) says of itself that it is false only. The paraconsistent view on (1) does not work for (6) since the claim that (6) is both true and false should be incompatible with what (6) says of itself—i.e., that it is only false. So the standard paraconsistent treatment of the liar seems to get the wrong answer for (6).

Sentences (7) and (8) contain non-standard negations. The 'Xnot' in (7) expresses exclusion negation, and 'Bnot' in (8) expresses Boolean negation. In multi-valued logics like paracomplete logic, exclusion negation takes indeterminacy to truth. So a theory that implies that (7) is indeterminate also implies that (7) is true. Thus, the revenge paradox generated by (7) is a variant of the revenge paradox generated by (5). In paraconsistent logics, Boolean negation takes gluts to truths, so a theory that implies that (8) is glutty also implies that (8) is just true. Hence, the revenge paradox generated by (8) is a variant of the revenge paradox generated by (6).

These examples barely scratch the surface of the revenge paradox literature, which is largely scattered and disorganized. I have focused on common revenge paradoxes for the most prominent approaches to the aletheic paradoxes. Notice that the distinction between philosophical approaches and logical approaches I took pains to draw in Chapters Two and Three gets blurred when it comes to revenge paradoxes. Some revenge paradoxes feature terms specific to philosophical approaches (e.g., 'true in a context'), while others involve terms from logical approaches (e.g., 'unstable').

8.2 Negation, Denial, and Rejection

Negation plays a big role in the aletheic paradoxes and in many revenge paradoxes. Two of the most promising logical approaches, paracomplete and paraconsistent, require non-classical logics, which impose strict conditions on negation. Both these approaches require non-standard views on the nature of denial and rejection. Denial is typically thought of as a kind of speech act and the standard explanation is that to deny p is to assert the negation of p. However, according to paracomplete theorists, one ought not assert indeterminate sentences and if p is indeterminate, then p's negation is as well. So, one needs some other way of indicating that one does not endorse p. Paracomplete theorists typically say that denial is an independent speech act that is not the assertion of negation. Instead, one can deny p and deny p's negation when p is indeterminate. Just as denial is an independent speech act, rejection is thought to be an independent attitude. Instead of thinking of rejection of p as the acceptance of p's negation, paracomplete theorists take it that one can reject p and reject p's negation; indeed, this is the correct attitude to take if p is paradoxical.[6]

Paraconsistent approaches need independent accounts of denial and rejection as well since they hold that some sentences are both true and false. For a paraconsistent theorist, one may assert p and p's negation since paraconsistent theorists take some contradictions to be true. Thus, one cannot explain denial in terms of the assertion of negation in a paraconsistent context. The same goes for rejection; one may accept p and p's negation, so the paraconsistent theorist needs an account of rejection as well.[7]

How does one indicate that one is denying a sentence instead of asserting its negation? In English, the word 'not' is usually enlisted for this role. Linguists mark this distinction with the terms 'logical negation' and 'metalinguistic negation'.[8] Logical negation is the familiar truth-functional

---

[6] See Maudlin (2004) and Field (2008a).
[7] See Priest (2006a, 2006b) and Beall (2009).
[8] See Horn (2001) for discussion.

operator, whereas metalingusitic negation is used to indicate that one does not agree with some aspect of a sentence, proposition, or utterance. For example, consider the following exchange:

Rod: Jessica stopped going to church.

Todd: Jessica did not stop going to church; she never went in the first place.

Here the 'not' expresses metalinguistic negation rather than logical negation. Reading 'not' as logical negation gives the first clause of Todd's sentence the wrong truth conditions since it would mean that Jessica is still going to church, which contradicts the second clause. Instead, Todd is indicating that he does not accept the presupposition of Rod's utterance, namely, that Jessica at one time went to church.[9]

One kind of revenge worry is that one could use metalinguistic negation in a sentence like:

(9) (9) is not true.

If so, then (9) would mean that (9) is being denied by whoever is uttering (9). If, say, a paracomplete theory implies that (9) ought to be denied, then one might think that the paracomplete theory has this very sentence as one of its consequences.[10]

There are many complex issues in this area, but I will not be delving into them since these kinds of revenge worries play no role in how I think about the problems associated with truth and the aletheic paradoxes. However, I will say that if 'not' in (9) expresses metalinguistic negation, then it does not contribute to the truth-conditional content of (9), at least on most views of metalinguistic negation. Instead, it expresses the speaker's intention to perform a certain kind of speech act (compare to the 'not' in Todd's claim above). If that is right, then there simply is no problematic truth-conditional content to be found among the consequences of the paracomplete theory. On this

---

[9] See Horn (2001) for an in-depth discussion of this distinction; see also Horn (1985), Atlas (1989), Chapman (1996), Carston (1996, 1998, 1999), and Burton-Roberts (1999) for discussion.

[10] Most philosophers and logicians working on the paradoxes ignore this kind of worry; however see Richard (2008) and Schroeder (2010).

reading, (9) is not the kind of thing that can be a consequence of a theory. The upshot is that this kind of revenge worry should not be taken seriously since it rests on a confusion between a logical operator, which does contribute to a sentence's truth-conditional content and a force marker, which does not.

8.3 Recipes for Revenge

Jc Beall's introduction to the most important volume on revenge paradoxes, *The Revenge of the Liar* (OUP 2007), contains a helpful discussion of the nature and source of revenge paradoxes and the role they typically play in disputes about the aletheic paradoxes.[11] According to Beall, the revenge phenomenon is a problem with classifying liar sentences. If we use Beall's term 'bugger' to describe their semantic status, then there are two potential problems. First, if 'p is bugger' entails 'p is not true', then the theory according to which liar sentences are bugger is inconsistent. Second, we can avoid inconsistency, but at the price of having some sentences that are both bugger and true. However, whatever status liar sentences have, almost everyone agrees that they are defective and should be rejected (and denied), so it seems that being bugger and being true are incompatible in some sense.

Beall points out (and he is certainly not alone here) that most theorists whose approaches generate revenge paradoxes stipulate that they apply only to languages that are expressively impoverished—they cannot contain the terms that feature in revenge liars. Since the descriptive project or the prescriptive project are primary goals for most of those proposing approaches to the aletheic paradoxes, they have to somehow explain how to use an approach that works only for expressively impoverished languages on a seemingly expressively rich language like English. This is

[11] Beall (2007b)

not a new predicament; indeed, the orthodox approach takes a theory that applies only to languages that do not contain their own truth predicates at all and specifies how to use it on natural languages (by claiming that their truth predicates are seriously ambiguous).

How bad is this problem? Beall has a view on this issue as well. For background, recall that when one gives a formal theory of truth as part of an approach to the paradoxes, one specifies an artificial language, L, that contains its own truth predicate 'true-in-L'. The theorist then shows that 'true-in-L' obeys various principles of the formal theory of truth, and the theorist can use L to show that the formal theory of truth is relatively consistent (often using classical logic and set theory in a metalanguage M). Finally, the theorist claims that natural languages are like L in relevant respects, so the theory of 'true-in-L' is intended to apply to truth predicates of natural languages.

Revenge objections state that L is lacking in some vocabulary that is present in natural language and this vocabulary gives rise to revenge phenomena. Thus, the theory of truth achieves its goals only because L lacks certain expressive resources. Beall lays out three distinct revenge recipes:

1. Find some semantic notion X that is *used in M to classify sentences of L*.
   Show in M that X is not expressible in L unless L is inconsistent or trivial.
   Conclude that L is explanatorily inadequate since it does not explain how natural language, which contains X, is consistent.

2. Find some semantic notion X that is *expressible in M*.
   Show in M that X is not expressible in L unless L is inconsistent or trivial.
   Conclude that L is explanatorily inadequate since it does not explain how natural language, which contains X, is consistent.

3. Find some semantic notion X that is *expressible in natural language*.
   Argue that X is not expressible in L unless L is inconsistent or trivial.
   Conclude that L is explanatorily inadequate since it does not explain how natural language, which contains X, is consistent.

The italics indicate the contrasts between the three recipes. In the first case, the concept X is used by the theory of truth to classify paradoxical sentences, whereas in the second case, the concept X is just expressible in the language of the theory—it need not be explicitly used by the theory. In the third case, the concept X is expressible in natural language and need not even be expressible in the

language of the theory. In each case, the problem is that the theory in question does not apply to natural languages, so it does not really solve the aletheic paradoxes.

Beall points out that if L is a non-classical language and X is a classical, model-dependent notion, then recipes 1 and 2 are not very convincing since the theorist is simply using a classical metalanguage to construct a non-classical artificial language in an effort to argue that natural languages are non-classical. Beall calls this "Too easy revenge". It should not count against a theory of truth that a classically constructed notion or model-dependent notion is not expressible in a non-classical model language (without inconsistency).

Beall also notes that for any revenge strategy to be effective, the revenger needs to argue that the notion in question (notion X in the recipes) is not inconsistent. For, if it is, then it should not count against a theory of truth that it cannot accommodate it. I return to this issue in Part III after introducing the idea of an inconsistent concept.

8.4 Two Problems: Inconsistency and Self-Refutation

I have been discussing revenge paradoxes as if there is a single form they all have, but that is misleading. In my view, there are two distinct kinds of revenge paradoxes and they need to be kept separate.[12]

To illustrate, let us consider an example. Let T be a theory of truth that implies that truth expressions are partially-defined predicates (i.e., a paracomplete approach). Assume as well that T validates the primary aletheic principles and the other rules involved in the derivation of the liar paradox. Thus, '(1) is true if and only if (1) is false' follows from T. However, no contradiction follows from '(1) is true if and only if (1) is false' because we are working in a three-valued scheme.

[12] See Greenough (2001: fn9) and Beall (2007b) for similar distinctions.

Indeed, T implies that (1) is a gap. Hence, (1) is not paradoxical for T. We can say that (1) is *pathological* for T (i.e., (1) has traditionally been involved in a liar paradox, but it poses no problem for T).

So far so good for T; however recall the sentence:

(2) (2) is either false or a gap.

As explained above, T has trouble with (2). That is our first example.

Let us consider how T might be altered to accommodate sentences like (2). One way to do so is to alter the logic we use so that the theory still validates the truth rules and still implies that (2) is a gap, but now the theory implies that '(2) is true if and only if (2) is false or a gap' is a gap as well. Let us call this theory T′. Now we cannot derive a contradiction from '(2) is true if and only if (2) is either false or a gap'. However, this sentence poses another problem for T′. Namely, T′ implies that (2) is true if and only if (2) is either false or a gap; hence, T′ has '(2) is true if and only if (2) is either false or a gap' as a consequence. However, T′ implies that '(2) is true if and only if (2) is either false or a gap' is a gap; that is, T′ implies that '(2) is true if and only if (2) is either false or a gap' is neither true nor false. Therefore, T′ implies that one of its consequences is not true. Consequently, T′ is self-refuting—it implies that it is not true.[13, 14] Call this the *self-refutation problem*.

Let us consider a way of altering the theory so that it is not self-refuting. We need a way of characterizing (2) and '(2) is true if and only if (2) is either false or a gap' that does not result in the theory having a consequence that it labels untrue. One way to do this is to accept that (1) is a truth-value gap, but stipulate that the truth-value gaphood predicate itself is partially defined (i.e., the gaphood predicate has gaps—*gaphood gaps*). Let T″ be such a theory. T″ implies that (1) is a truth-value gap. T″ also implies that (2) is true if and only if (2) is either false or a truth-value gap.

---

[13] For an example of a theory like T′, see Maudlin (2004).

[14] In the last two sentences of this paragraph, I am using 'not' to express exclusion negation.

However, T″ can be constructed so that it implies that '(2) is true if and only if (2) is false or a truth-value gap' is true; the reason is that T″ does not imply that (2) is either true, false, or a truth-value gap. Indeed T″ implies that (2) is a gaphood gap. Of course, one can construct a new problematic sentence for T″:

(3) (3) is either false, a truth-value gap, or a gaphood gap.

However, T″ can follow the same strategy to handle (3) by positing a hierarchy of gaphood predicates, each of which is partially defined. On this account (1) is a truth-value gap, (2) is a gaphood gap, (3) is a gaphood-gap-hood gap, etc. In this way, T″ avoids labeling any of its consequences untrue.[15]

The problem with T″ is that if it applies to a language that contains a completely defined truth-value gaphood predicate, then T″ is inconsistent because it implies that a sentence of this language like (2) (i.e., a sentence that attributes either falsity or truth-value gaphood to itself—where the truth-value gaphood predicate is completely defined) is true if and only if it is either false or a truth-value gap. Call this the *inconsistency problem*. The progression from T to T′ to T″ illustrates the fact that there is something like an oscillation between the two kinds of revenge paradoxes—attempts to avoid one tend to bring on the other.[16]

It is my view that one must distinguish between these two types of revenge paradoxes in order to understand our current predicament regarding truth. In short, there are two broad trends when it comes to theories of truth designed to handle sentences like (1). Some theories can handle sentences like (1), but they still have inconsistent consequences for other sentences (e.g., (2)). That is, they do not provide a way of solving all other paradoxes associated with truth that are structurally identical to the liar. This is the inconsistency problem. Other theories can handle sentences like (1),

---

[15] For an example of a theory like T″, see Field (2008a).
[16] I borrow the term 'oscillation' from McDowell (1994).

but they imply that they have the same status (i.e., untrue) as (1). Because few, if any, theories of truth that are designed to handle sentences like (1) imply that sentences like (1) are true, a theory of truth that implies that it has the same status as a liar sentence implies that it is untrue. This is the self-refutation problem.[17]

The inconsistency problem arises when a theory of truth handles some versions of the liar paradox, but not all of them. There are many different versions of the liar; some versions involve concepts that are often used to classify sentences that figure in other versions. This should not come as a surprise given the prominence of views on which sentences that figure in liar paradoxes are defective in a way that renders them neither true nor false. Once one has a term for the third status, one has a new version of the liar paradox. The most common response is to restrict the theory so that it does not apply to such sentences.

On the other hand, the self-refutation problem arises in connection with the consequences of a theory of truth. The liar paradox is unlike other paradoxes (e.g., Russell's paradox, Grelling's paradox, etc.) in that it concerns truth, which applies to things that can participate in inferential relations (e.g., sentences, propositions, etc.). In other cases, the paradoxical items (e.g., sets, predicates, etc.) are not the type of thing that can be the consequence of a theory. However, for truth, the paradoxical items are sentences, which can be consequences of a theory. A theory of truth that is designed to deal with the liar paradox has to classify many paradoxical sentences like (1). It turns out that for many theories of truth, no matter what they say about such sentences, some of these sentences are going to be consequences of the theory.

---

[17] Both the Curry paradox and the Yablo paradox depend on the primary aletheic principles as well (i.e., they require all three rules for their construction), and approaches to each one generate revenge paradoxes in the same way that approaches to the liar paradox generate revenge paradoxes; thus, one can use structurally analogous arguments to the ones in this chapter to argue for conclusions pertaining to the Curry and the Yablo that are analogous to the conclusions I draw pertaining to the liar. It is my view that all three paradoxes (i.e., the liar, the Curry, and the Yablo) are manifestations of the defectiveness of our concept of truth. The approach to truth that I offer in Part III solves all three without generating revenge paradoxes of any kind.

8.5 A Diagnosis

It is difficult to come up with a single explanation for why revenge paradoxes occur that both does justice to the distinction between inconsistency and self-refutation problems *and* works for all the different kinds of approaches to the aletheic paradoxes. In this section, I offer one.

Any combined approach to the aletheic paradoxes will have a philosophical component and a logical component. In Chapter Three, I classified the logical approaches (following Hartry Field) by the aletheic principles accepted and the logic that is compatible with the approach. Of the two aletheic principles that seem to be constitutive of truth, the most uncontroversial and most widely accepted is (T-Out): $\langle \mathsf{p} \rangle$ is true $\rightarrow \mathsf{p}$. Consider again a standard liar sentence:

(1) (1) is not true.

Notice that the instance of (T-Out) for (1) is '(1) is true → (1) is not true. This instance of (T-Out) is equivalent to (1) itself.[18] That is, given the definition of (1), *an instance of (T-Out) is equivalent to a liar sentence*.

Whatever status a theory of truth that includes (T-Out) assigns to (1), it assigns to '(1) is true → (1) is false'. Moreover, every approach to the liar paradox treats (1) as problematic in some way that renders it untrue—either it is false, or a truth-value gap, or indeterminate, or unstable, or whatever. Thus, theories of truth face a fundamental difficulty:

(i) deny (T-Out),

(ii) weaken the logic to the point that it breaks the equivalence between the instance of (T-Out) and (1), or

(iii) accept (T-Out) and accept that (T-Out) is not true.

---

[18] '(1) is not true' is classically equivalent to '(1) is not true or (1) is not true', which is classically equivalent (given the definition of (1)) to 'if (1) is true then (1) is not true'.

(i) is not plausible since almost everyone takes (T-Out) to be constitutive of the concept of truth. In addition, denying (T-Out) is incompatible with treating 'true' as a device of endorsement and generalization, which are two of its most important functions (I discussed this in Chapter Six). (ii) causes problems when applied to natural languages since languages like English have linguistic resources that force classical reasoning, like exclusion negation. When a theory that denies the equivalence is applied to sentences with these linguistic resources, the theory delivers inconsistent results. This is the *inconsistency problem*. (iii) requires accepting that one's theory of truth is not true. This is the *self-refutation problem*.

In sum, there is immense pressure for a theory of truth to accept (T-Out), but any theory that does so faces either the inconsistency problem or the self-refutation problem because instances of (T-Out) are classically equivalent to paradoxical sentences. That, as far as I can tell, is the source of the revenge paradox phenomenon.

8.6 Consequences

Having described and diagnosed revenge paradoxes, I now turn to the consequences they have for a philosophical understanding of truth. Special emphasis is given to relations between revenge paradoxes, empirical paradoxicality, and truth's expressive role, culminating in the "importation" problem.

8.6.1 The Revenge Argument

The first consequence of the revenge phenomenon is a trilemma for any theory of truth that accepts (T-In) and (T-Out); that is, any theory of truth that accepts the primary aletheic principles faces a choice from among three unattractive options.

Assume that T is a theory of truth and T implies that the truth rules are valid for a class of sentences that includes liars. Assume also that T implies that truth predicates are univocal, invariant, non-circular, etc; in short, truth predicates do not have any "hidden" semantic features that render the reasoning in the liar paradox invalid. We use sentence (1) (i.e., '(1) is false') again as our example of a liar sentence. Let T assign a status Δ to sentence (1), where Δ is incompatible with truth—that is if a sentence is Δ, then it is not true. Now consider the following sentence:

(10) (10) is either Δ or false.

No matter what T says about (10), it has the following as a consequence:

(11) (10) is true iff (10) is either Δ or false.

The argument is identical to the one presented in section 8.1 with 'Δ' replacing 'gappy'. There are four options for how T classifies (11): as true, as false, as Δ, or as having some other status **Ω**. If T implies that (11) is true, then T is inconsistent since (11) is a contradiction. If T implies that (11) is false, then T is self-refuting since (11) is a consequence of T. If T implies that (11) is Δ, then T is self-refuting since (11) is a consequence of T and Δ is incompatible with truth. If T implies that (11) is **Ω**, then simply run the same argument with '**Ω**' in place of 'Δ' in (11). The only other option is that T is restricted so that it does not apply to languages that contain 'Δ'. Therefore, any theory of truth that validates the truth rules is either inconsistent (or trivial), or self-refuting, or restricted so that it does not apply to certain languages.

Inconsistency (triviality), self-refutation, and restriction are all unwelcome features of a theory of truth. Almost all theorists who face up to this trilemma choose restriction; of course, many do not explicitly say that their theories are restricted. Rather, they claim that the expressions that cause problems for their theories are either meaningless or incoherent. I discuss this maneuver below.

8.6.2 Expressive Power and Revenge

As Beall makes clear, there is a close link between the revenge phenomenon and the expressive powers of various languages. In particular, many debates between defenders of various theories of truth and critics who appeal to revenge paradoxes turn on the expressive powers of natural languages. This is a very complex issue and one that deserves its own treatment in the next chapter. However, I do want to comment here on the unfortunate practice of defending a theory from a revenge paradox objection by claiming that the linguistic expression featuring in the revenge paradox is meaningless or unintelligible.

Some philosophers do try to avoid restricting their theories of truth by claiming that the resources that give rise to the revenge paradoxes are meaningless or unintelligible.[19] I call this the *unintelligibility maneuver*. My view is that it is unacceptable to assume that these linguistic expressions are meaningless. As I have said, for a language that contains truth value gaps, one can define two sentential operators that behave like classical negation: choice negation and exclusion negation. A theory like McGee's or Field's or Maudlin's has trouble with sentences that express exclusion negation for two reasons. First, these theories are fixed-point theories and so apply only to languages that do not contain non-monotonic sentential operators.[20] However, exclusion negation is non-monotonic. Thus, they do not even return results for languages that express exclusion

---

[19] See Parsons (1984), Priest (1990), and Tappenden (1999), who claim that there is no such thing as exclusion negation; see also Maudlin (2004), who claims that there are no non-monotonic sentential operators whatsoever.

[20] In a three-valued scheme, a sentential operator is *monotonic* if and only if for a sentence containing that sentential operator, changing a component of that sentence from a gap to a truth-value (i.e., from a gap to true or from a gap to false) never results in changing the sentence from one truth-value to the other or from a truth-value to a gap (i.e., from true to false, from false to true, from true to a gap, or from false to a gap). Intuitively, one can "fill in" the gaps in the components without changing the truth-value of the compound. See Gupta and Martin (1984) who show that by using a weak Kleene scheme, one can arrive at fixed points even though one's language contains certain non-monotonic operators. However, exclusion negation is not among them.

negation. Second, one can easily extrapolate to determine the results they would return if they were capable of returning results. The sentence:

(5) (5) is Xnot true

poses a problem. It means something like *(5) has a status other than that of being true*. One can use (5) to generate a revenge paradox for most theories of truth that imply that (5) is Xnot true. Likewise, one can generate revenge paradoxes using completely defined gaphood predicates (as I did in the revenge argument), paradoxicality predicates, groundedness predicates, certain conditionals, quantification over hierarchies of predicates, and so on. In order to pursue the strategy advocated in the objection, one would have to claim that all these linguistic items are meaningless.

If there is an established practice of using a linguistic expression, then that linguistic expression is meaningful.[21] For each of the linguistic expressions that are labeled unintelligible by these theorists, there is an established practice of using them. Moreover, these linguistic expressions belong to some natural languages, including English. Thus, given the Gricean Condition presented in Chapter Six (i.e., competent speakers of natural languages can determine in normal conversational circumstances what contents words and sentences have), the unintelligibility maneuver is not legitimate.

In addition, the "outlaw" linguistic expressions serve an important explanatory role. If we decided that they are all meaningless and gave them up, then we would rob natural languages of important expressive resources. For example, if an object, A, is in neither the extension nor the anti-extension of a predicate, ϕ, then we need a way of expressing this fact. One way of doing so is to say that A is Xnot ϕ and A is Xnot $\ulcorner \sim \phi \urcorner$. Another is to say that A is a ϕ-value gap. If the

---

[21] Even linguistic expressions like 'tonk' are meaningful; they just express inconsistent concepts; see Prior (1960) for a discussion of 'tonk'. Of course, Wittgenstein (1923) is infamous for claiming that many seemingly meaningful words are "nonsense" (*unsinnig*); I do not have the space to discuss the relation between his views on language and the claim on which this footnote comments.

theorists in question are right that the "outlaw" linguistic expressions are meaningless, then we have no way of expressing these facts.

Finally, simply claiming that the linguistic resources in question are meaningless is not enough to avoid the revenge argument. One would have to provide an independent argument for this claim (e.g., something other than, "that's the only way I can think of to avoid the liar paradox"). No such argument has been forthcoming and it seems it would be impossible to present one whose premises were more plausible than the claim that these items are meaningful.[22]

Analytic philosophy has a long history of claiming that certain linguistic expressions that figure in established linguistic practices are meaningless. It is high time for us to realize that we need no longer resort to this kind of move; we can provide an approach to the liar paradox without it.

8.6.3 Expressive Role and Revenge

One might respond to the revenge argument in the previous subsection by denying (T-In) or (T-Out); indeed, the classical gap theories of truth deny (T-In). However, any theory that rejects these principles also rejects the claim that truth plays an expressive role. In Chapter Six, I explored truth's expressive role and argued (following Field and others) that the primary aletheic principles are essential to this role. We use 'true' as a device of endorsement and as a device of generalization. Since we use 'true' in this way, it is a simple fact about our linguistic practice that we use 'true' as if (T-In) and (T-Out) are true. Indeed, these principles are so entrenched that any acceptable theory of truth should treat them as constitutive of the concept of truth.

Once one admits that our use of 'true' presupposes (T-In) and (T-Out), one is hard pressed to deny that they are true. Any theory that does so seems like it would be a revisionary theory—a

[22] See Eklund (2008a) for another criticism of the unintelligibility maneuver that focuses on exclusion negation.

theory that specifies how we ought to use 'true'—rather than a descriptive theory. Moreover, since it should be plain to everyone that we use 'true' as a device of endorsement and as a device of generalization, anyone who endorses a theory of truth that rejects either (T-In) or (T-Out) is obligated to explain why they fail even though competent users of 'true' act as if they are true.

Since there is tremendous pressure to accept that (T-In) and (T-Out) are constitutive of our concept of truth on the basis of our linguistic practice, there is an equally strong push for a theory of truth to accept that these two principles are true. That, of course, is the major assumption of the revenge argument, so it is really the recognition of truth's expressive role that seems to force one into the trilemma that is the conclusion of the revenge argument.

I take this combination of expressive role and revenge to be one of the most pressing problems for any theory of truth. As will become clear in Part III, an inconsistency view allows one to accept that (T-In) and (T-Out) are constitutive of truth without having to accept any of the three options in the trilemma.

8.6.4 Empirical Revenge

Although it should be obvious, it bears mentioning that revenge paradoxes have empirical variants as well. Recall the discussion in Chapter Seven of empirical paradoxicality—some sentences are paradoxical by virtue of contingent empirical facts (e.g., 'Every sentence in section one of Chapter Seven that begins with an 'E' is false'). Everything from that discussion of empirical paradoxicality carries over to revenge paradoxes. For example, (i) there are empirical revenge paradoxes, (ii) there is no syntactic, semantic, or pragmatic test for whether a sentence gives rise to a revenge paradoxes, and (iii) all the approaches criticized in Chapter Seven for being unable to deal with empirical paradoxes (e.g., ambiguity, context-dependence, etc.) are just as implausible when used as strategies

for avoiding revenge paradoxes. In the last section of this chapter, 8.6.8, these points will be used in the construction of a serious "importation" against theories that give rise to revenge paradoxes.

8.6.5 Paracompleteness and Indeterminateness

Hartry Field's paracomplete approach to the aletheic paradoxes has already received considerable attention in this work. In this section, I explore what is probably its most central feature: the account of determinateness and how it addresses revenge paradoxes.

Again, the key innovation in Field's theory is that he defines a determinateness operator (hereafter I use '$\mathcal{D}$'), which is defined in terms of his new conditional. With the aid of this determinateness operator, the languages Field considers are able to characterize paradoxical sentences by saying that they are not determinately true. Of course, since the determinateness operator is part of the language, *it* features in revenge paradoxes (e.g., (5) = '(5) is false or indeterminate') and it is incorrect to label these as not determinately true. However, the determinateness operator iterates to characterize them. For example, (5) is not *determinately* determinately true. We can adopt the convention of using superscripts (e.g., (5) is not $\mathcal{D}^2$true). Indeed, the determinateness operator can be iterated infinitely many times (for details see the discussion in Chapter Three). The keys to this construction are that the determinateness operator is not idempotent (i.e., $\mathcal{D}\mathcal{D}p \neq \mathcal{D}p$), and the law of excluded middle fails for each determinate truth predicate in the hierarchy (i.e., '$D^{\sigma}p$ or $\sim D^{\sigma}p$' fails to be valid no matter what σ is).

In the last chapter, we saw that this hierarchy of determinate truth predicates actually does important work—work that one would think the truth predicate should be doing. In this section, we see that it does not do enough work. In particular, it does not act as exclusion negation. Instead of a truth-table definition, we can define exclusion negation by the following rules instead:

(LEM) $\vdash p \vee \neg p$
(EFQ) $p, \neg p \vdash \bot$

However, it is easy to show that exclusion negation is not definable in the languages Field considers (and it cannot be added without trivializing the language).

So what? One might wonder why this matters. There are several reasons. First, we seem to have something like exclusion negation in English. Here is a quotation from Kripke:

> Liar sentences are *not true* in the object language, in the sense that the inductive process never makes them true, but we are precluded from saying this in the object language by our interpretation of negation and the truth predicate. If we think of the minimal fixed point, say under the Kleene valuation, as giving a model of natural language, then the sense in which we can say, in natural language, that a Liar sentence is not true must be thought of as associated with some later stage in the development of natural language, one in which speakers reflect on the generation process leading to the minimal fixed point.[23]

The fourth word of the quotation is 'not', but it does not make sense to think that it is choice negation since that reading would make Kripke's own sentence indeterminate. Instead, it makes the most sense to say that it expresses exclusion negation.

Field is well aware that this is a pressing worry and he claims that exclusion negation is "unintelligible." He is careful to point out that he thinks that exclusion negation is meaningful; it is just an incoherent concept.[24] Field defends this response in several ways. First, there is no way to define exclusion negation in the languages he considers. Second, he claims that there is no good argument for the legitimacy of exclusion negation—any purported argument is circular (i.e., it uses exclusion negation). Third, he suggests a way of interpreting someone who mistakenly uses exclusion negation.

I take it that the first point is just that nothing in Field's approach to the aletheic paradoxes requires exclusion negation; nor does it require anything that could be used to define exclusion

---

[23] Kripke (1975: 714).
[24] Field (2008a).

negation. In terms of Beall's three recipes for revenge, it seems that Field is denying that either recipe 1 or recipe 2 will be successful. I agree with him on this—the presence of exclusion negation or anything in terms of which it could be defined would make Field's language trivial, and he has proven that it is not trivial.

The second point does not really make sense. Field seems to think that the onus is on his opponent to show that exclusion negation is coherent. However, that is hardly fair since Field gets to assume that truth is coherent, and he uses that assumption to argue that the other expressions are incoherent. In addition, Field claims that any argument for the coherence of exclusion negation begs the question by assuming that exclusion negation is coherent. Of course, we know that we cannot argue for the coherence of our basic logical concepts without begging the question, so Field's reasoning here is not convincing. The important thing to recognize is that if speakers of natural language use these expressions and have good reason to use them (and I think it is very plausible that they do), then Field's solution will carry serious hidden costs.

That brings us to his third point. Field writes:

> I've heard it argued that even if no *good* theory posits a Boolean negation, we haven't solved the paradoxes until we've given an account of how to apply the term 'true' to the sentences of someone who has a *bad* theory according to which the word 'not' obeys all of Boole's assumptions (or at least, to the sentences of someone for whom this bad theory plays such a central role in his linguistic practices that it determines what 'not' means for him.) But I don't think that this raises an interesting challenge. There is bound to be a certain arbitrariness in how we attribute truth-values to sentences that contain concepts we find defective (e.g., 'tonk' (Prior 1960) or 'Boche' (Dummett 1973: 454)). We can't, for instance, regard both 'All Germans are Boche' and 'All Boche are prone to cruelty' as true, unless we share the racist beliefs of those who possess the concept; but a decision as to which (if either) should count as true, in the language of the racist for whom both beliefs enter into his practices in a "meaning-constituting" way, seems both pointless and unlikely to have a principled answer. Similarly for the use of 'not' by someone with a defective theory. Probably the simplest course is to translate his 'not' with our 'not', in which case his claims to the validity of excluded middle will come out false even though they were "meaning-constitutive"; but there's no reason to take a stand on this translation, or to think that there's a determinate fact of the matter as to whether it's correct.[25]

---

[25] Field (2008a: 309-10fn.1).

There is a lot going on in this passage. The mention of 'tonk' and 'Boche' is a reference to previous work on defective concepts. 'tonk' is A. N. Prior's term, which has the introduction rule of disjunction and the elimination rule of conjunction. Adding it to most languages renders them trivial.[26] 'Boche' is an old French derogatory term for Germans, which has the connotation that Germans are barbaric. Michael Dummett popularized it in a discussion of inferential role semantics. It seems inappropriate to use 'Boche' since it seems that any use is an implicit endorsement of the claim that all Germans are barbaric (the same point holds for other pejoratives).[27] When Field writes about 'not' obeying all Boole's assumptions, he is talking about exclusion negation (at least in the context of a paracomplete approach). He describes those who use exclusion negation as having a bad theory of negation. By this, he means that someone who uses 'not' to express exclusion negation has false beliefs about negation. Finally, Field accepts Quine's criticism of analyticity and his argument for the indeterminacy of translation, which are controversial and immensely influential theses in the philosophy of language. The former attempts to undermine the idea that any sentences are true by virtue of their meanings alone (i.e., meaning-constituting), while the latter purports to show that meaning and translation are largely indeterminate.[28] Putting all this together: someone who uses 'not' to express exclusion negation has a bad theory of negation and we should attribute truth values to his or her sentences by simply translating them into Field's paracomplete language since there is no fact of the matter as to what they "really" mean. The effect is that Field advocates treating everyone who uses 'not' as meaning exactly what Field means when he uses 'not'.

Why is this a problem? Most people do not have theories of negation—they utter sentences that contain 'not' in certain circumstances and they interpret others who utter sentences that contain

---

[26] See Prior (1960), Belnap (1961), Bonnay and Simmonauer (2005), Cook (2005), and Wansing (2006) for discussion.
[27] See Dummett (1973: 454), Brandom (1994: 125-132; 2009: 124), Williamson (2003, 2006), Whiting (2008), and Hom (2008).
[28] See Quine (1951, 1960).

'not' in certain circumstances. The question for a philosopher of language or a linguist who works on semantics or pragmatics is: what is the best way to make sense of how English users behave? If it turns out that a theory on which 'not' expresses exclusion negation on some occasions provides the best explanation of how English users use 'not', then Field's approach has a major cost—namely, it cannot be applied to English.

Is there any reason to think that people use 'not' in this way? Yes. The intuitive evidence is the following. If I say that p is not true, and I mean to be saying that p is something other than true (which includes falsity, indeterminacy, or whatever), then I take myself to be saying something true, no matter how indeterminate p might be. But Field cannot accommodate this intuition. If Field is correct, then no matter how much I try, if p is indeterminate enough, my sentence will be indeterminate as well. To be a bit more careful about this point: Field will translate my claim that p is not true as 'p is not D^{σ}true' for some σ, but no matter how large σ is, if p's level of indeterminacy is higher, 'p is not D^{σ}true' is indeterminate. Moreover, according to Field, it is incoherent for me to say or believe that p is not true, where this claim or belief is true no matter what level of indeterminacy p has.[29]

In addition, linguists claim that 'not' in English (at least sometimes) is properly interpreted as exclusion negation, and linguists use exclusion negation in their theories. Here are two examples. Jay Atlas in *Philosophy without Ambiguity* (1989) argues that 'not' has a general sense and on particular occasions of use it can express either choice negation or exclusion negation. There is linguistic evidence that 'not' is univocal and invariant because it fails ambiguity tests and context-dependence tests; thus, it is neither ambiguous nor context-dependent. Nevertheless, on many occasions, it makes the most sense to interpret English speakers as meaning exclusion negation when they use

[29] Field (2008a)

'not'.[30] A second example is that Laurence Horn in *A Natural History of Negation* (2001) surveys views on negation from Aristotle to present, the evidence for choice negation readings of 'not' vs. exclusion negation readings of 'not', and how these readings interact with other linguistic phenomena (presupposition, conversational implicature, scope, etc.). He too argues that 'not' is not ambiguous or context dependent. Rather, exclusion negation provides the semantics for natural language descriptive (non-metalinguistic) negation or predicate denial (in Aristotle's sense), and what seems like choice negation is an artifact of pragmatic tendencies like that of reading topical/definite subjects as taking wide scope with respect to ordinary predicate denial.[31]

On the other hand, there is no evidence that English contains a transfinite hierarchy of determinate truth predicates. No scientists studying English have ever found any data to support such a view. The only support Field can offer is "well, that's just a consequence of the best way I can think of to solve the liar paradox." Perhaps this kind of armchair justification would be compelling if its consequences were obscure enough to have avoided any scientific inquiry, but that is not the case—linguists have plenty of data that suggest we often use 'not' to express exclusion negation. Again, blanket dismissal of these results is on par with being a Creationist or a flat-Earther. Scientific results trump armchair speculation every time.

8.6.6 Paraconsistency and Just True

In the previous section, we saw that the paracomplete approach works only for those languages that do not express the concept of being other than true. In this one, we find that the paraconsistent approach has the opposite problem: it works only for languages that do not express the concept of being only true (or just true).

---

[30] Atlas (1989: ch. 3).
[31] Horn (2001).

Recall, the paraconsistent approach claims that paradoxical sentences are both true and false. Thus, it takes truth to be compatible with falsity. That seems problematic since many English speakers take truth and falsity to be incompatible. Imagine the following conversation:

Martin: I've been taking a linguistics class, and I now think that discourse representation theory (DRT) is true.

Ralph: Oh, I agree with you—DRT is true. You know, though, it also happens to be false.

Martin: Uh, no. I don't think it's false. I just said it's true.

Ralph: Right. It is true. And it's false.

Martin: No. I don't think it's false. It's just true.

Ralph: Oh! You mean that it is true and it is not false. I agree with you—it is true and not false. However, it also happens to be false as well.[32]

Martin: Look, I don't think it is false at all. It is just true. It's not true and false, nor is it true and not false and also false. It's just true.

Ralph: I agree with everything you've just said. But just so you know, it's also false.

It is Martin's sense of 'just true' that causes problems for the paraconsistentist. This problem has been around for a long time as a worry about how paraconsistentists deal with disagreement—normally, if I say that something is false, that indicates that I disagree with it. The paraconsistentist denies this—thinking that something is false is compatible with accepting it (as the toy conversation above makes clear). So, one wonders how the paraconsistentist would express disagreement; one suggestion is that he or she might use 'just false'.

Jc Beall has a detailed discussion of this issue in his recent defense of the paraconsistent approach to the aletheic paradoxes. Beall accepts that a paraconsistent language cannot contain the kind of negation that would be required to define 'just true' as 'true and not false' in the sense in which Martin is using it above. However, Beall claims that when someone like Martin uses 'just

[32] Note that paraconsistent approaches take some sentences to be true and not true and false and not false, so this response is actually faithful to the position.

true', there are two things going on. First, 'just true' has the same meaning as 'true', so Martin means exactly the same thing as Ralph. Second, a use of 'just true' has as a conversational implicature that the speaker *rejects* the negation of the proposition in question.[33] Recall that both paraconsistentist and paracomplete approaches need a distinction between the speech acts of *assertion* and *denial* (where denial is not assertion of negation) and a distinction between the attitudes of *acceptance* and *rejection* (where rejection is not acceptance of negation). The paraconsistentist claims that sentences can be both true and false, but no one may both assert and deny the same proposition, and no one can accept and reject the same proposition. Thus, the proposal from Beall is that 'just true' is to be pragmatically distinguished from 'true'—the former, but not the latter, is used to convey the speaker's attitude of rejection.

There are several problems with this proposal. First, we do use 'false' to express disagreement or rejection in English. Indeed, even a cursory familiarity with the history of philosophy shows quite clearly that by far the most popular way of criticizing a theory is to argue that it is false. Beall offers no linguistic evidence that 'true', 'false', 'just true', and 'just false' function in the way he suggests, and there are mountains of data showing that they do not.

Even if we disregard all the empirical facts, there is still a serious problem with Beall's proposal—it does not work for embedded uses of 'just true'. For example, imagine Martin asserts 'if DRT is true, then semantic minimalism is false'. He can tell Ralph that he is not considering the possibility that DRT is both true and false, and he is not considering the possibility that semantic minimalism is both true and false. However, Beall might suggest that he reformulate this sentence as 'if DRT is just true, then semantic minimalism is just false'. Notice that this would not do, even on Beall's view. The problem is that Martin is not expressing rejection of the negation of DRT, and he is not expressing rejection of semantic minimalism. He is asserting a conditional, and there is no

[33] Beall (2009: ch. 3).

indication of whether he accepts or rejects either of these views. There is simply no way to use the kind of proposal Beall offers for embedded uses of 'just true'.[34]

8.6.7 Self-Refutation

I have focused quite a bit on only one of the revenge problems without really evaluating or pressing the other. Perhaps that is because the vast majority of theorists opt for theories that are threatened by inconsistency problems and then attempt to block thee problems or overcome them. Self-refutation is much more difficult to block and so seems like a more decisive problem. Still, at least one theorist, Tim Mauldin, has taken the heroic stance of endorsing a classical gap approach to the aletheic paradoxes and accepting that it implies that it is untrue. That is, Maudlin accepts (T-Out), rejects (T-In) and accepts classical logic. Since the liar sentence is equivalent (modulo its own definition) to its instance of (T-Out), his approach implies that not all instances of (T-Out) are true. Thus, Maudlin's theory is self-refuting in the sense that it implies that it is not true.

More specifically, Maudlin's approach allows three exclusive statuses: true, false, and ungrounded. By 'ungrounded' he means just what Kripke means; namely, that the sentence is not determined to be either true or false by virtue of the truth or falsity of non-aletheic sentences (i.e., those not containing occurrences of the truth predicate). All paradoxical sentences are ungrounded, but there are many non-paradoxical sentences that are ungrounded as well. Some of the principles of the theory turn out to be non-paradoxical, but ungrounded. Thus, the theory implies that some of its principles (e.g., (T-Out)) are ungrounded. In the terminology of Chapter Three, he accepts the outer theory of Kripke's strong Kleene recursive construction.

---

[34] See also Parsons (1990) and Shapiro (2004) for discussion.

One might wonder: how can Maudlin coherently assert his own theory when it implies that it is untrue? His answer is: truth is not a necessary condition for assertibility. He stipulates that it is permissible to assert some ungrounded sentences, including those that compose his theory of truth. However, all false sentences are not assertible and all true sentences are assertible. He is clear on this point:

> If one accepts the theory and the standard of permissibility, then one is permitted to assert the theory and also to assert that the theory is not true. This would only be self-contradictory if one also claimed that only true sentences should be asserted, but this is something we deny.[35]

Thus, once one rejects the received view that truth is a necessary condition for assertibility, one can both assert Maudlin's theory and assert that it is not true.

However, this move brings with it two additional problems. The first is that once one adds the predicate 'permissible' to the language to which the theory applies, one can generate new paradoxes with sentences like:

(12) (12) is not permissible.

If (12) is permissible, then (12) is false and permissible, but if (12) is not permissible, then (12) is true and impermissible. Either way one is forced to accept that either some permissible sentences are false or some true sentences are impermissible, which is incompatible with Maudlin's permissibility standard. Note that this is a different kind of revenge paradox—an inconsistency problem. In his attempt to lessen the blow of one kind of revenge paradox (i.e., the self-refutation problem), Maudlin stumbles into the other kind of revenge paradox (i.e., the inconsistency problem). This is not a coincidence—as I have mentioned, there is an oscillation between these revenge paradoxes so that when a theorist turns his attention to fighting one, the other strikes. Then when he turns his attention to the other, the first returns.

---

[35] Maudlin (2004: 178).

In response to this permissibility version of the inconsistency problem, Maudlin bites the bullet and accepts that every set of rules for permissibility is inconsistent (i.e., in the parlance of this book, that permissibility is an inconsistent concept). "Indeed, the practical advice one is likely to offer with respect to rules of permissibility is this: simply try to avoid conversational contexts which lead into problematic areas (e.g., the discussion of sentences like [(12)].[36] Maudlin seems to think that this is "a problem we must learn to live with."[37] But notice the evidence he marshals to convince us that permissibility is an inconsistent concept—that it gives rise to paradoxes structurally identical to the liar. His book could have been much shorter if he had just made these claims about truth from the start and left it at that.

The second problem with revising the received view of permissible assertion is that it is incompatible with truth's expressive role. Although Maudlin fails to consider this problem, it follows easily from his unorthodox views on permissibility. Maudlin writes:

> If a sentence is ungrounded, then it is not appropriate to assert that the sentence is true or that the sentence is false. The claim that an ungrounded sentence is either true or false, such as [the liar is true or the liar is false], is, we shall say, impermissible.

Imagine a situation in which Ned wants to assert some sentence, say, the semantic rule for the biconditional (i.e., a biconditional is true iff both components have the same truth value), but he cannot remember exactly how it is formulated. All he needs to do is use 'true' as a device of endorsement—he should assert 'the biconditional rule is true'. Of course, Ned's utterance is impermissible since the biconditional rule is ungrounded and according to Maudlin, it is impermissible to assert that ungrounded sentences are true. Thus, Maudlin's view gives us the wrong predictions with respect to truth's expressive role. Therefore, it is not a good descriptive theory of our aletheic practice.

---

[36] Maudlin (2004: 175).
[37] Maudlin (2004: 177).

The problem has nothing to do with Maudlin's claim that any system of principles for permissibility is inconsistent. No matter what one says about how to explain permissibility, Ned's utterance is permissible. It is a datum to be explained.

Notice also that this problem is far more vulgar than the one Kripke finds with the orthodox approach (described in Chapter Seven). In Kripke's objection, the speaker does not know the level of the sentence she wants to endorse, but if she did, then she would be able to use the truth predicate as a device of endorsement. At least the orthodox approach gets the latter situations right. For Maudlin, no matter how much the speaker knows about the ungrounded target, it is still impermissible to assert that it is true. And remember, Kripke's objection to the orthodox approach convinced an entire generation of theorists to seek new philosophical and logical approaches to the aletheic paradoxes; it is the meter bar of objections. Maudlin's failure to take any of this seriously should serve as a red flag: avoid tangling with the self-refutation problem.

8.6.8 Importing Revenge

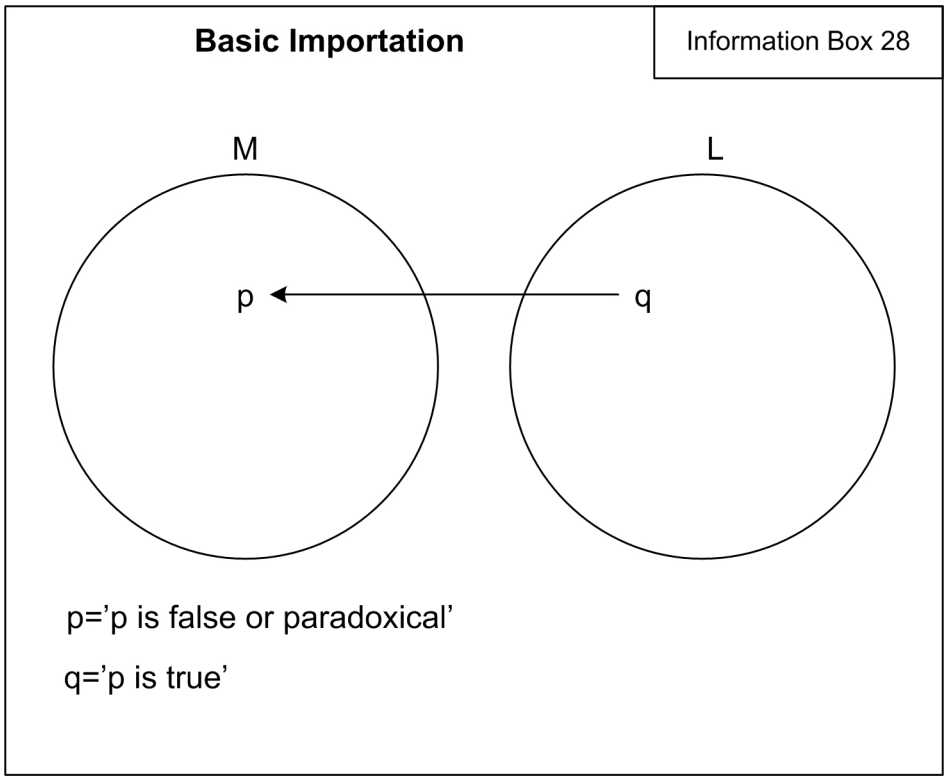
So far in this chapter, there have been discussions of the kinds of revenge paradoxes that affect a wide range of approaches to the aletheic paradoxes, indications of how revenge paradoxes interact with truth's expressive role, and with empirical paradoxicality, and a look at the revenge paradoxes for the most popular non-classical logical approaches. In this final section, I bring together the points made about truth's expressive role in Chapter Six, empirical paradoxicality in Chapter Seven, and revenge paradoxes in this chapter into a powerful kind of argument I call an *importation* argument.

We have already seen an elementary form of importation problem in our discussion of language-specific truth predicates, but let me present it in that guise as an introduction to the topic. Let T be a theory of truth that accepts both (T-In) and (T-Out) and assume that 'pathological' is the status T

assigns to liar sentences (I am taking it for granted that if a sentence is pathological, then it is not true, and that excluded middle holds for 'pathological'). Let ρ be a sentence that gives rise to a revenge paradox for T; for example ρ might be 'ρ is either false or pathological'. The standard response to this kind of problem is to say that T is restricted so that it does not apply to languages containing anything like ρ. Let ρ be a sentence of language M. Finally, assume that L is a language that does not contain ρ or any translation of it, L contains an unrestricted truth predicate (i.e., one that is not language-specific), and L contains a singular term that refers to ρ (we might as well use 'ρ'). So, L does not seem to contain anything that would give rise to a revenge paradox for T since it does not contain ρ itself. However, L does contain the sentence 'ρ is true'; call this ν. If ρ were in the scope of T, then T would imply that ρ is both true and either false or pathological; that, again, is the conclusion of the argument with which this chapter began, and it is the reason why T is restricted to avoid ρ. Nevertheless, ν *is* in the scope of T, and an argument with the same form shows that T implies that ν is both true and either false or pathological. The arguments are set out in Information Box 28. The upshot is that if ρ generates a revenge paradox for T, then ν generates a revenge paradox for T. Thus, to avoid revenge paradoxes, T should be restricted so that it does not apply to ν either. Moreover, it must be restricted so that it does not apply to sentences that attribute truth to ν, and those that attribute truth to those that attribute truth to ν, and so on. Call this an *importation problem*.[38]

---

[38] Importation problems can also take the form of liar pairs. If L contains an unrestricted truth predicate but no pathologicality predicate, and L′ contains a falsity predicate and a paradoxicality predicate, but no truth predicate, then L contains a sentence β='α is true' where α= 'β is false or pathological' and β is a sentence of L′. T seems to apply unproblematically to both L and L′, but that is not the case; although neither language, by itself, has the resources to construct a revenge paradox, together they have a revenge liar pair (α and β).



One way to avoid this problem is to formulate a theory of language-specific truth (e.g., 'true-in-English'), and to claim that truth predicates of natural language can be explained in terms of language-specific truth predicates. If L contains 'true-in-L' instead of 'true', then ν would be 'ρ is true-in-L', which is false since ρ is not a sentence of L. Thus, for approaches to the aletheic paradoxes that give rise to revenge paradoxes, language-specific truth predicates play an absolutely essential role.

Matti Eklund presents a different kind of importation argument in his recent discussion of revenge paradoxes. Eklund argues that even if a language L cannot *directly* formulate revenge liars, it might *indirectly* express the concepts sufficient to formulate revenge liars. Here is his definition of indirect expressibility:

> A property ϕ is *indirectly expressible* in a language L iff there is a predicate F of L such that for some context c, an utterance of F by a speaker of L is such that for all x, 'F(x)' and 'x has ϕ' have the same truth-value.

For example, L might have the predicate 'falls under Frink's favorite predicate', where Frink's favorite predicate belongs to some other language sufficient to formulate revenge liars. (Note that 'falls under' is a satisfaction predicate.) Imagine that Frink's favorite predicate is 'pathological' as used by theory T. Then L would have a sentence μ, which is 'μ is either false or falls under Frink's favorite predicate'. It should be obvious that roughly the same argument as above shows that T implies that μ is true iff it is either false or pathological.[39] So, indirect expressibility poses serious problems for approaches to the aletheic paradoxes that give rise to revenge paradoxes.

It seems to me that the proponent of an approach that faces this kind of importation problem should follow the same strategy as above. That is, she should stipulate that only language-specific satisfaction predicates are legitimate. For example, if L contains only 'falls-under-in-L' instead of 'falls under', then Eklund's importation argument is thwarted. Presumably, one would want to stick with the same policy for all semantic predicates of L.

There are two further points I want to make about importation arguments. First, as I argued in Chapter Six, natural language truth predicates cannot be explained in terms of language-specific truth predicates. That means the importation arguments pose serious problems for anyone who restricts a theory of truth to avoid revenge paradoxes.

Second, one can run an importation argument even on language-specific truth predicates. The argument is more complex and involves a detour through contingently paradoxical sentences. For example, assume we have a theory of truth T that accepts (T-In) and (T-Out), and classifies paradoxical sentences as pathological (so far, this is just as above). In this example, we have three languages, L, M, and N. L is the language to which T applies. L, M, and N each have their own respective L-S truth predicates. In addition, L has 'true-in-M', and M has 'true-in-N'. L does not

[39] Eklund (2007).

have 'pathological'. Instead, only N has 'pathological', and as such, only N contains an explicit revenge liar for T. The importation argument pertains to the following five sentences:

(i) Sentence t is in N, and t= 't is either false-in-N or pathological'

(ii) Sentence s is in M, and s is the translation of t, and s is the only sentence on the blackboard

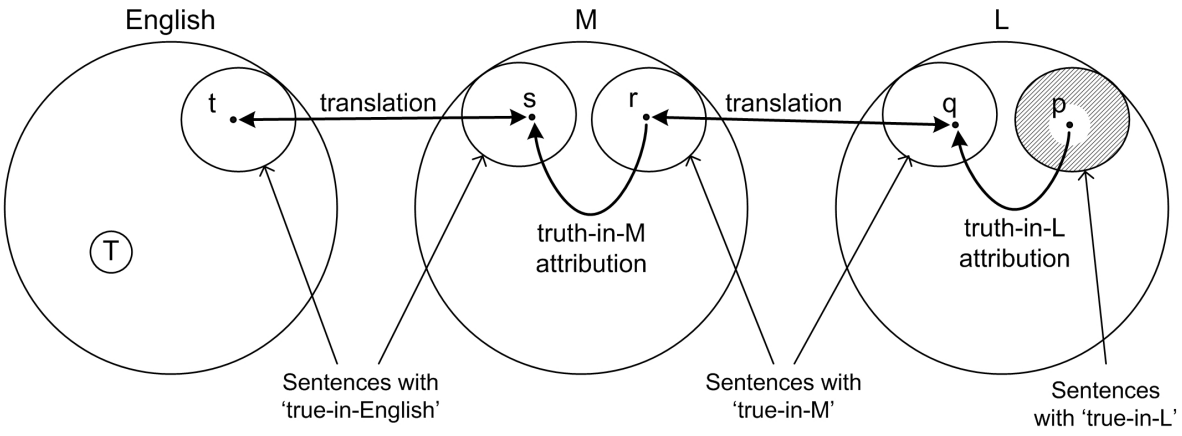
(iii) Sentence r is in M and r= 'the sentence on the blackboard is true-in-M'

(iv) Sentence q is in L and q is the translation of r

(v) Sentence p is in L and p= 'q is true-in-L'

So, q attributes truth-in-L to another sentence, q, of L. q is in L, but it is the translation of r, a sentence of M. r attributes truth-in-M to another sentence of M, s, which happens to be the lone sentence on the blackboard. s is in M, but s is not a revenge paradox because it says of itself that it is either false-in-N or pathological. However, it is not in N, it is a member of M. Finally, t is in N and is a revenge liar for T. This situation is depicted in Information Box 29.

Language-Specific Importation Information Box 28



t is 't is either false-in-English or paradoxical-in-English'
s is the translation of t into M, and s is on the blackboard
r is 'the sentence on the blackboard is true-in-M'
q is the translation of r into L
p is 'q is true-in-L'
T is the semantic theory for truth-in-L
▨ is the scope of T

The following is the argument that p is an imported revenge paradox for T, even though L contains only an LS truth predicate and no pathologicality predicate.

1. If t is true-in-N, then r is true-in-M. (Argument: If t is true-in-N, then s is true-in-M because s is a translation of t. Sentence r ('the sentence on the blackboard is true-in-M') says that s is true-in-M. Sentence s is true-in-M. Thus, r is true-in-M.)

2. If r is true-in-M, then p is true-in-L. (Argument: If r is true-in-M, then q is true-in-L because q is a translation of r. Sentence p ('q is true-in-L') says that q is true-in-L. Thus, p is true-in-L.)

3. If t is false-in-N or pathological, then r is false-in-M or pathological. (Argument: If t is false-in-N or pathological, then s is false-in-M or pathological because s is a translation of t. Sentence r says that s is true-in-M. Sentence s is false-in-M or pathological. Thus, r is false-in-M.)

4. If r is false-in-M or pathological, then p is false-in-L or pathological. (Argument: If r is false-in-M or pathological, then q is false-in-L or pathological because q is a translation of r. Sentence p says that q is true-in-L. Thus, p is false-in-L or pathological.)

5. Therefore, *if* t is true-in-N iff t is false-in-N or pathological, *then* p is true-in-L iff p is false-in-L or pathological.

Despite the fact that L has no truth expression for the language to which t belongs and has no sentence that is a translation of t, we have still managed to construct a sentence of L that is a revenge liar for T if t is a revenge liar for T. Thus, if p is in the scope of T, then T implies that p is both true-in-L and not true-in-L.

Notice that this importation argument brings together the expressive role of truth (i.e., q and r are being used to endorse whatever is on the blackboard), empirical paradoxicality (i.e., s is empirically paradoxical—if it were not the sole sentence on the blackboard, it would not be paradoxical), and revenge paradoxes (i.e., t is a standard revenge paradox for T). It seems to me that these importation arguments show that the standard move in response to revenge paradoxes—restricting one's theory of truth—is useless. This goes a long way toward showing that approaches to the aletheic paradoxes that face revenge paradoxes are unacceptable, at least when it comes to natural languages.

8.7 Impact

The impact of the considerations in this chapter is hard to overestimate. Since virtually every philosophical and logical approach to the paradoxes generates revenge paradoxes, the considerations here affect all of these. Moreover, given what I said about the connection between truth's expressive role and revenge paradoxes, it seems as if any theory of truth that explains truth's expressive role faces revenge paradoxes. Finally, given the importation problem, the standard move in the face of revenge paradoxes, restricting one's theory, is a non-starter. Thus, although the considerations in this chapter affect particular views on truth more than others, one should start to get the feeling that the points raised here pose a general problem for anyone formulating or defending a theory of truth.

Chapter 9

No Metalanguage Required

We saw in the last chapter that revenge paradoxes confront virtually every approach to the aletheic paradoxes. By far the most popular response is restriction: when confronted with a revenge-paradoxical sentence, a theorist will usually say that his or her theory does not apply to languages that contain sentences like that. Of course, many theorists justify the restriction by claiming that some of the linguistic resources occurring in the revenge-paradoxical sentence are defective or even meaningless. In this chapter I present a problem for approaches to the aletheic paradoxes that are restricted in this way.

I am not the first to find this kind of restriction troublesome (I discuss others in the first section). However, making the case against approaches that are restricted has been notoriously difficult. The problem is that there is an extremely complex constellation of issues surrounding revenge paradoxes, the expressive features of natural languages, and the adequacy of approaches to the aletheic paradoxes. In order to get clear on these topics, we need a new conceptual framework to serve as a background against which discussions of these issues can take place. The centerpiece of that framework is the concept of internalizability (to be defined below). Once that framework is in place, I use it to argue that approaches that are restricted to avoid revenge paradoxes are unacceptable.

The biggest problem I have with the current literature is that its terminology allows for a rough characterization of *properties of* theories of truth (e.g., 'requires a distinction between object-language and metalanguage') and it allows for a rough characterization of *properties of* languages (e.g., 'semantically self-sufficient'), but it makes it difficult to explain how theories of truth and languages are related. In other words, properties of theories and properties of languages are privileged. One

can try to formulate relations between theories and languages from them, but it is unfortunate that this structure makes it exceedingly difficult to argue convincingly for the claim that theories of a certain type (i.e., those that are restricted so as not to apply to the languages in which they are formulated) do not work well for languages of a certain type (i.e., natural languages).

The terminology I advocate begins with *relations between* theories and languages. With these in hand, one can then define properties of languages and properties of theories. There are (at least) two advantages to the terminology I advocate. First, it is much more precise than the old vocabulary. That is, it allows one to make distinctions between properties of theories, properties of languages, and relations between theories and languages that are unavailable in the old terminology. For example, with it, one can formulate several conditions on theories of truth that are similar to an intuition voiced by William Reinhardt and Vann McGee. Second, it simplifies arguments for the claim that theories of a certain type (i.e., ones that are restricted in certain way) do not work well for languages of a certain type (i.e., natural languages).

9.1 Reinhardt and McGee on Theories and Languages

Several theorists working on the liar paradox have voiced the opinion that if an approach to the liar must be restricted to avoid revenge paradoxes, then it is unacceptable. For example, William Reinhardt expresses this sentiment in the following passage:

> Let us suppose, as I believe is intuitively correct, that one of the primary features of [truth] is that it is one notion: in particular it does not split into some hierarchy of notions. … Let us explain that the truth predicate of our formal language (call the language L) is intended to be taken in the sense of our preexisting informal notion of truth. … Unless we are prepared to entertain splitting the notion of truth, we are forced to admit that the metalanguage is included in the object language. If the formal language is to provide an adequate explication of the informal language that we use, it must contain its own metalanguage. I take it that this is in fact a desideratum for success in formulating a theory of truth.[1]

[1] Reinhardt (1986: 227-228).

Likewise, Vann McGee proposes the "integrity of language" requirement, which states, "[i]t must be possible to give the semantics of our language within the language itself."[2] McGee says of his requirement that it "is intended to hold open the possibility that the methods we develop can be applied to natural languages. If in developing the theory of truth for a language, we required the services of an essentially richer metalanguage, that possibility would be closed off. … [It] makes it reasonable to hope that our methods can be used to get a semantics of a natural language."[3] Both McGee and Reinhardt suggest that if a theory of truth does not apply to languages in which it is formulated, then the theory cannot be applied successfully to natural languages.[4] I call this the *Reinhardt-McGee intuition*.

When one adopts the terminology I suggest, one can formulate the Reinhardt-McGee intuition precisely and provide an argument for it. Moreover, it turns out that the Reinhardt-McGee intuition is too weak—there are theories of truth that do not require a distinction between object-language and metalanguage but still face revenge paradoxes.

9.2 Definitions

In order to expose and avoid certain difficulties, I first present and reject three suggestions for capturing something like the Reinhardt-McGee intuition. Roughly, their idea is that a theory of truth that applies to a natural language is acceptable if and only if it applies to everything in that language to which it is supposed to apply and it does not use anything to which it is supposed to apply but to which it cannot apply. I assume that a condition on the terminology I introduce is that

---

[2] McGee (1991: 159).
[3] McGee (1991: 159).
[4] For similar sentiments, see Kearns (1970), Priest (1979, 1984b, 1999, 2006a, 2006b), Simmons (1993), and Martin (1997). See also Field (2003a, 2003b, 2005b), Eklund (2008a), and L. Shapiro (2010).

it should work for many different types of theories, not just theories of truth. I also want the terminology to work for many different types of languages, not just natural languages.

The first suggestion is that theories should be expressible in the languages to which they apply. However, this requirement is too strong. The Reinhardt-McGee intuition is that theories of truth that *cannot* be expressed in the languages to which they apply are unacceptable. I do not want to say that a theory that *happens to be* inexpressible in a language to which it applies is unacceptable.

A second suggestion is that a theory should be expressible in an extension of a language to which it applies. This condition is too weak because for any theory and any language to which it applies, there exists an extension of that language in which the theory can be expressed, so long as the theory does not apply to the extended language. The condition I am trying to capture is that a theory should apply to the very languages in which it can be expressed.

As a final suggestion one might say that for a given theory and a language to which it applies, the theory should both be expressible in an extension of that language and apply to that extended language. The problem with this suggestion is that a theory might apply to some sentences of a language but not others. For example, one might propose a theory of truth that applies to some sentences of English that contain 'true', but not to all of them. Such a theory would apply to English, but it would not apply to all the relevant sentences of English.

Another problem with all these suggestions is that they do not adequately distinguish between kinds of theories. Take theories of truth as an example. Theories of one type include claims about the nature of truth, while theories of another type attribute meanings to sentences that express the concept of truth. I reserve the term *theories of truth* for the former and call the latter *semantic theories for truth*. An adequate idiom for discussing theories of truth that have been restricted to avoid revenge paradoxes must be sensitive to this distinction. In presenting the concept of internalizability and its

relatives, I privilege semantic theories for truth. Then I define analogous concepts for theories of truth.

The best suggestion is that a semantic theory that applies to a language is acceptable only if it is expressible in an extension of that language and applies to everything in that extension to which it is supposed to apply. A semantic theory of this type is *internalizable* for that language. The following is a more elaborate and precise definition of internalizability:

> A semantic theory T that purports to specify the meanings of sentences that express a concept X is *internalizable for a language L* if and only if there exists an extension of L such that all the sentences that compose T can be translated into sentences that belong to the extension of L and T specifies the meanings of all the sentences of the extension of L that express X.

That is rather long-winded and contains many expressions whose meanings are unclear. In the interest of clarity, I discuss four aspects of this definition: semantic theory, language, expression, and application. Semantic theories and languages are objects (in a loose sense that includes abstract entities), while expression and application are relations between a semantic theory and a language.

9.2.1 Semantic Theories

I begin with 'semantic theory'. First, a *theory* is a set of declarative sentences all of which belong to a single language.[5] The term 'semantic theory' is tricky to define because it has been used in so many ways. I follow Dummett in distinguishing between meaning-theories and the theory of meaning. For Dummett, *the theory of meaning* is the branch of philosophy that deals with the nature of meaning,

[5] This account of theoryhood is not without its problems. First, we usually think of theories as things that can be expressed in different languages. We find it natural to say that two physics textbooks, one written in English, the other written in French, both contain Newton's theory of mechanics. However, on my account, they contain two different theories. I attempt to defuse this problem by speaking of a theory and its translations into other languages. Furthermore, although philosophers (and many other people) use the term 'theory' quite often, it is rather difficult to say which sentences constitute a particular informal theory. I have no doubt that I would have trouble specifying the sentences that constitute Lewis's theory of natural laws or Davidson's paratactic theory of indirect discourse. Nevertheless, I stick with the idealization. In addition, I do not require that a theory be closed under logical consequence.

while a *meaning-theory* is a particular theory that specifies the meanings of the words or sentences of a particular language or languages. I use the term 'a theory of meaning' in a Dummettian spirit to designate a theory that specifies the nature of meaning.[6] Theories of meaning provide necessary and sufficient conditions on meaning-theories. According to my usage, a semantic theory is a type of meaning-theory. In particular, a *semantic theory* is a theory that specifies the meanings of certain sentences that belong to some particular language or languages.[7] I use the locution 'semantic theory for X', where X is a placeholder for the name of a concept (e.g., a semantic theory for moral obligation, a semantic theory for truth). A semantic theory for X specifies the meanings of the sentences of certain languages that express the concept X (e.g., a semantic theory for truth specifies the meanings of sentences that express the concept of truth).[8] I also use the locution 'theory of X'; a *theory of X* is a theory that makes claims about the nature of X.[9]

9.2.2 Languages

---

[6] Dummett (1991: 20-22). See Peacocke (1981) for this use of 'a theory of meaning'.
[7] Dummett also uses the term 'semantic theory' but his account differs from mine. For Dummett, a semantic theory must specify the truth-value of each sentence in a given language (see Dummett 1991: 25, 33, 35). King (1994: 57) and Soames (2002c: 97) define 'semantic theory' as I do.
[8] I do not discuss the nature of concepts because I prefer to accommodate a range of views on this issue.
[9] Although I do not make much of it, the distinction between a semantic theory for X and an X definition is important. An *X definition* provides the extension, the intension, or the sense of a word that expresses the concept X. (The *extension* of a predicate is the set of things of which the predicate is true; the *intension* of a predicate determines its extension across possible worlds, and the *sense* of a predicate is something like its cognitive significance.) There is a certain amount of overlap between an X definition and a theory of X, but the distinction between them is important. For example, a theory of planethood makes claims about the nature of planets—what it is for something to be a planet. A planethood definition might specify the extension of 'planet'—which things are planets. A semantic theory for planethood specifies the meanings of the sentences that contain 'planet' and its synonyms. It is also important to keep in mind the distinction between a semantic theory for X and a semantic theory for things that are X. For example, a semantic theory for vagueness specifies the meanings of sentences that contain 'vague' and its synonyms, whereas a semantic theory for things that are vague specifies the meanings of sentences that contain vague terms. A semantic theory for quantification specifies the meanings of sentences that contain 'quantifier' and its synonyms, whereas a semantic theory for things that display quantification specifies the meanings of sentences that contain quantifiers.

A *language* is a function from sets of sentences (syntactic strings) to a set of sentential meanings.[10] A *sublanguage* L_0 of a language L_1 is a language whose set of sentences is a subset of the set of sentences of L_1, whose set of sentential meanings is a subset of the set of sentential meanings of L_1, and whose function from sentences to meanings agrees with that of L_1. A language L_1 is an *extension* of a language L_0 if and only if L_0 is a sublanguage of L_1. Although this definition of language leaves much to be desired, I use it because it simplifies discussions of language, and because it is the one that is assumed by most of those who propose semantic theories for truth.[11] An attempt to construct a more plausible account of language would take me too far afield.

The above definition of 'language' focuses on sentences. What about words? The most familiar languages have syntax according to which words are combined into sentences. Moreover, the sentential meaning of a sentence is often taken to be determined by its structure and the meanings of its parts. Thus, if one has a finite list of words and their meanings, rules for combining words into sentences, and rules for determining the meaning of a sentence from its structure and the meanings of its words, then one has a language (as defined above).

I follow most people who study languages by appealing to the distinction between types and tokens.[12] A *token* of a word or sentence is a physical entity (e.g., ink marks on a page, sound waves,

---

[10] My discussion of sentential meanings is intended to be compatible with a wide range of views on the nature of sentential meanings (sets of possible worlds, structured propositions, inferential roles, etc.).

[11] See Lewis (1969), Soames (1984), Stalnaker (1987), and Davidson (1992) for examples of this account of language. Perhaps the most difficult issue facing proponents of this account of language is specifying the relation between languages and the humans who use them: the *actual language relation*. A specification of the actual language relation explains what it is about the mental, physical, and social activities of a group of humans that makes them users of a particular language. I say nothing about what the actual language relation is or how one determines which language a group of people use. See Lewis (1969, 1975) and Schiffer (1993) for more on this issue; see Hawthorne (1990) and Field (1994a) for criticism. Another problem with this account of language is that any change in the syntactic or semantic features of the expressions used by a person or group of people results in a change in the language they use. Consequently, the language one uses changes almost continuously. There are some alternative accounts of language in the literature. Mental definitions of language usually focus on the brain states of the humans that have linguistic capacities (e.g., Chomsky 1995), while those who favor pragmatic definitions of language often concentrate on the dispositions, regularities, or rules associated with the members of a linguistic practice (e.g., Sellars 1954 and Lewis 1975).

[12] See Kaplan (1973, 1990), Hugly and Sayward (1981), Simons (1982), Wetzel (1993, 2008), Horwich (1998: 98-103), Cappelen (1999), Dummett (1999), Szabó (1999), and Truncellito (2000) for discussion of the distinction between types

pulses of light, etc.), while a *type* is an abstract entity. There might be many different tokens of the same type. There might be many different tokens of the same type. For example, the previous two sentences are two tokens of the same sentence type. All the languages I consider have an infinite set of sentence types because they have sentential operators (e.g., 'and') and allow unlimited iterations of some term functions (e.g., 'the father of x').

One more point about languages—I do not assume that languages have their logics "built in". It is common to see phrases like 'let L be a first order classical language'. The presupposition is that it is essential to L that its sentences obey the classical consequence relation. However, it seems to me that the best way to understand a logic is as a theory of the consequence relation, not as some part or aspect of a language. It makes sense to consider a single language or single theory from the perspective of classical logic, from the perspective of intuitionist logic, or from the perspective of some other logic.

One might protest that that the best way we have of making sense of logical connectives is in terms of their inferential roles; thus, the meanings of a language's logical connectives determine its logic. I agree that it makes sense to say that, for example, modus ponens is constitutive of the conditional. However, it does not follow that modus ponens is valid. There are many terms that have constitutive principles that are incompatible with one another or with some features of the environment in which they are used. 'tonk', 'Boche', 'mass', and 'up above' are familiar from philosophical discussions (I discuss these at length in Part III).[13] Furthermore, there is a growing tradition that takes this view on concepts that give rise to philosophical problems and paradoxes

---

and tokens. I assume that sentence tokens are the primary bearers of truth and I prefer to treat sentence tokens as pairs of possible objects and contexts, so long as contexts are individuated finely enough. However, there are some tricky issues here having to do with the fact that the syntactic and semantic features of a truth bearer are not always sufficient to determine whether it is paradoxical and the fact that paradoxicality affects truth value (on some views).

[13] See Prior (1960), Dummett (1973), Field (1973), and Gupta (1999), respectively.

including the liar paradox.[14] On these views, a concept's constitutive principles can turn out to be false. Thus, one can accept that inference rules and logical principles are constitutive of logical terms without accepting that a language has its logic "built in".

9.2.3 Expressibility

The next topic is the expressibility relation. A theory T that belongs to one language L_0 is *expressible* in another language L_1 if and only if for every sentence q that composes T, there exists a sentence p of L_1 such that p is a translation of q.[15] This definition of 'expressible' relies on a notion of translation from one language into another. I assume that a sentence of one language is a translation of a sentence that belongs to another if and only if they have the same or relevantly similar meanings (or contents).[16] Although I say very little about meaning and what makes two meanings relevantly similar, I assume that two sentences with the same or relevantly similar meanings have the same truth conditions. For the most part, I ignore issues related to the indeterminacy of translation, the indeterminacy of interpretation, the inscrutability of reference, and their implications for defining suitable notions of meaning and translation.[17] I do not want to give the impression that I think these issues are not worth discussing. Quite the contrary; there is so much to say about them that I could

---

[14] See Chihara (1979), Mates (1981), Tappenden (1993), Yablo (1993a, 1993b), Eklund (2002a, 2002b), Schiffer (2003), Patterson (2006), and Scharp (2007, 2008).

[15] I use the word 'express' in two different ways. Words or sentences express concepts, while languages express sentences or theories. I define the latter in terms of translation and content. I say little about the former.

[16] For the most part, I ignore the distinction between meaning and content, but the standard way of drawing it is that a context dependent expression has the same meaning in every context, but its content differs from context to context. The distinction for sentential meaning and sentential content is analogous.

[17] See Quine (1960) and Davidson (1973, 1979). It is my view that there is room for accepting both Quine's thesis on the indeterminacy of translation given the behavioristic evidence he allows and Davidson's thesis on the indeterminacy of interpretation given the evidence base he allows, while at the same time accepting a perfectly legitimate notion of translation that does not commit one to the distinction between analytic and synthetic truths. However, I do not argue for this claim here. See Leitgeb (2001a, 2001b) for a formal notion of translation.

not possibly give these issues the space they deserve and still discuss everything that is required to start coming to terms with the problems bequeathed to us by our concept of truth.[18]

9.2.4 Theory Application

The fourth aspect of the definition of internalizability on which I comment is application. Before defining it, I want to mention that, because of the prevalence of revenge paradoxes, it is common to treat semantic theories for truth as if they do not apply to languages that cause revenge paradoxes. However, I treat semantic theories as if they apply by default to every language, and I treat restrictions as explicit parts of a semantic theory—I often use the locution 'version of a semantic theory' to distinguish between semantic theories that differ only in the way they are restricted.[19] For example, in the case of first-order classical languages that do not contain their own truth predicates, Tarski's truth definition has come to be used as the standard semantic theory for truth. However, one version of this semantic theory applies to first-order classical languages that *do* contain their own truth predicates. Tarski showed that if this semantic theory applies to certain languages that contain their own truth predicates, one can derive a contradiction. That is, the semantic theory implies that some sentences of these languages are both true and not true (the sentences for which this occurs are like liar sentences). According to my convention, the version of Tarski's theory that applies only to languages that do not contain their own truth predicates is one semantic theory and the version that applies to languages that do contain their own truth predicates is another. The former is

---

[18] Given that I do not define 'sentential meaning', one can think of my definitions of 'language', 'expression', and 'application' as definition schemata—as the forms of definitions. It does not matter for my purposes how one explains sentential meanings (e.g., in terms of sets of possible worlds, structured propositions, inferential roles, causal relations, nomic relations, dispositions, etc.). One might worry that translation depends on one's choice of semantic theory. A defender of this view can still accept my arguments by relativizing translation to a semantic theory. Of course, all the definitions that depend on translation are then relativized to a semantic theory as well, but that does not affect the cogency of the arguments.

[19] My approach is considerably more liberal than Davidson's; for Davidson's views on what it is for a certain theory to apply to a given language, see Davidson (1967, 1973).

consistent and the latter is not. This convention of treating the restrictions as explicit additions to the theory is intended to avoid equivocations.

A *restriction* for a semantic theory T is a claim that T does not provide the meanings for the sentences of certain languages or that T does not provide the meanings of certain sentences of certain languages. An *unrestricted semantic theory* is one that has no restrictions, and a *restricted semantic theory* is the conjunction of an unrestricted semantic theory and its restrictions. A semantic theory *applies to a language L* if and only if the semantic theory does not contain a restriction specifying that it does not provide the meanings for sentences of L. A semantic theory *applies to a sentence of L* if and only if it is not restricted from doing so. I say that the *scope* of a semantic theory T is the set of sentences to which T applies.[20] I call the sentences of a language that express a certain concept X the *X-sentences* of the language, and a language that contains an X-sentence I call an *X-language*. I assume that a semantic theory for X is restricted by default to X-languages and to the X-sentences of X-languages.

I employ a deductive account of semantic theory application. Assume that T is a semantic theory for X and that S is the set of all the sentences in T's scope. For each member of S, an assignment follows from the union of the set of sentences that constitute T and a set of additional claims.[21] An *assignment* is a specification of the meaning of a sentence in the scope of the semantic theory in question.[22] The assignments of a semantic theory need not have the form: ⟨p⟩ means that

---

[20] The assumption that this collection is a set plays no role in my presentation or arguments other than ease of exposition.

[21] The set of additional claims might involve syntactic, semantic, or pragmatic information about the sentences in S (e.g., that a certain sentence is declarative, that a certain name names a particular object, or that a certain sentence token has been used to make an assertion).

[22] To determine the assignments of a given semantic theory, one requires a theory of logical consequence for a set of sentences, which specifies the sentences that follow from each subset of that set. I assume that each language will require its own theory of logical consequence (that is not to say that a notion of logical consequence is essential to the identity of each language). When comparing sentences from different languages, I invoke the notion of translation. However, I often ignore this complication and write as if a sentence of one language is a logical consequence of a set of sentences that belong to another. A semantic theory T together with a set of auxiliary claims (e.g., claims about the

q; instead, most semantic theories assign truth-values to sentences under certain conditions. Theorists skeptical of determinate meanings (e.g., Field) should realize that what I am calling 'an assignment of meaning' might be as weak as specifying the logical role of a sentence.

9.2.5 Internalizability

There is a sense in which a semantic theory for X that applies to a language L *should* provide an assignment for every X-sentence of L. It will be helpful to have a term for semantic theories that satisfy this demand.

> A semantic theory T for X is *descriptively complete for L* if and only if T provides an assignment for every X-sentence of L.

The notion of descriptive completeness plays an important role in the definition of internalizability and in the arguments of sections three and four.

Now that I have discussed the components of the definition of 'internalizable' and I have introduced some new terminology, I can provide a more economical definition of it and related terms.

> A semantic theory T for X is *internal for L* if and only if T is expressible in L and T is descriptively complete for L.
>
> A semantic theory T for X is *internalizable for L* if and only if there exists an extension L′ of L such that T is internal for L′.

9.3 The Importance of Descriptive Completeness

---

syntactic structure of the sentences in the scope of T) entail, on the theory of logical consequence, an assignment for each member of the scope of T. For example, if T is a semantic theory for truth and T applies to a language L, then T provides the meanings for the sentences of L that are members of the scope of T. An assignment for each member of L that is in the scope of T follows from T.

One can always extend a given language to express a particular semantic theory; if T is a semantic theory for X that applies to an X-language L, then L can be extended to a language L′ in which T can be expressed. It might even be the case that T is not internalizable for L, but T is descriptively complete for L (e.g., if T is not descriptively complete for L′ because it is restricted from applying to some of the X-sentences of L′). Thus, the notion of descriptive completeness does much of the work in the definition of internalizability. Internalizability for L requires that T is both expressible in L′ and descriptively complete for L′.

9.4 Descriptive Correctness

The definition of internalizability does not have any implications for the correctness or truth of a semantic theory. If a semantic theory T for X is descriptively complete for an X-language L, but is not internalizable for L, then one can construct a new theory T′ that is internalizable for L. One simply picks an extension L′ of L in which T can be expressed, and one stipulates that T′ agrees with T on the sentences of L and T′ assigns the meaning of 'Roquefort is yummy' to every sentence of L′ that is not a sentence of L. T′ is internalizable for L because there exists an extension L′ of L such that T′ is expressible in L′ and T′ is descriptively complete for L′. Of course, T′ is certainly false because for every X-sentence p of L′ that does not belong to L, T′ implies that both p and ⌈∼ p⌉ are synonymous. The lesson is that internalizability is relatively easy to achieve if one is willing to sacrifice correctness. The difficult task, when it comes to semantic theories for truth, is to give a theory that is both correct and internalizable for some language (by 'correct', I mean that it is correct not only for the truth-sentences of L, but for all the sentences in its scope); even more difficult is the task of constructing a semantic theory for truth that is correct and internalizable for a natural language.

What is it for a semantic theory to be correct? The answer is sure to be something like: the semantic theory accurately describes the sentences in its scope. I introduced the notion of descriptive completeness above to capture what it means for a semantic theory to describe what it is supposed to describe. Now I define descriptive correctness to capture the 'accurately' in 'accurately describes'.

> A semantic theory T for X is *descriptively correct for a language L* if and only if T is consistent and for every sentence s of L in T's scope, there exists a meaning m such that s has m and T assigns m to s.

The consistency clause forbids theories that have correct assignments but contradictory consequences.[23] I assume that a consistent theory has no contradictions as consequences (in first-order classical logic).[24]

9.5 Internalizability Across Languages

It is possible that a semantic theory for X is internalizable for some X-language and not internalizable for a different X-language. For example, let T be a semantic theory for X that is restricted so that no sentences that express the concept Y are within its scope and none of the sentences of T expresses Y. T might be internalizable for some X-language that does not express Y, but it will not be internalizable for those X-languages that express Y and have a completely defined two-place sentential operator (e.g., 'and')—such languages will have X-sentences that express Y.

9.6 Semantic Teamwork

---

[23] I am ignoring dialetheism for the purposes of exposition; however, it is easy to alter the definition to accommodate it by substituting 'non-trivial' for 'consistent'.

[24] I am assuming that self-refutation can be explained in terms of logical inconsistency. Thus, the consistency clause is intended to rule out both semantic theories that face inconsistency problems and those that face self-refutation problems. If it turns out that self-refutation cannot be explained in terms of logical inconsistency, then a non-self-refutation clause will need to be added to the definition of descriptive correctness.

It should be obvious that, in order to provide assignments to X-sentences, a semantic theory for X will have to work in conjunction with other semantic theories. For example, if a semantic theory for truth is to provide an assignment to the English sentence 'Herb's belief that grass is green is true', it will have to incorporate some assumptions about names (e.g., 'Herb'), propositional attitude terms (e.g., 'belief that'), natural kind terms (e.g., 'grass'), color terms (e.g., 'green'), and copulas (e.g., 'is'). I do not expect a semantic theory for X to provide assignments for all the X-sentences of an X-language *by itself*. A semantic theory for X should work together with semantic theories for the other phenomena displayed by the language (concepts, quantifiers, names, demonstratives, sentential operators, etc.). That is, I assume that it is a *team* of semantic theories that provides assignments to the sentences of a given language. Given this assumption, it is reasonable to expect a semantic theory for X to be able to provide assignments for all the X-sentences of an X-language when it can work together with other semantic theories that apply to that X-language. For example, if a semantic theory for necessity were unable to provide assignments for sentences that both express necessity and contain pronouns, even when working together with a satisfactory semantic theory for pronoun expressions, then that semantic theory for necessity would be inadequate. A semantic theory for X should be able to work together with other semantic theories to provide an assignment for any X-sentence whatsoever.

9.7 Theories of Truth vs. Semantic Theories for Truth

I mentioned at the beginning of section 9.2 that I would formulate my definitions in terms of semantic theories for X; one can then formulate analogous definitions for theories of X. I want to briefly indicate how to carry out the second step. Let me begin by discussing the relation between a theory of X and a semantic theory for X; I will use truth as my example.

A theory of truth is a theory that specifies some aspect of the nature of truth, while a semantic theory for truth is a theory that assigns meanings to some collection of sentences that express the concept of truth. Revenge paradoxes confront both theories of truth and semantic theories for truth. For example, a theory of truth that validates both the truth rules and classical logic implies that the liar is both true and false. Indeed, the revenge paradoxes pose a greater threat to theories of truth than they do to semantic theories for truth. The reason is that one can construct a semantic theory for truth that is not based on a truth conditional-theory of meaning. Such a semantic theory might assign conceptual roles, assertibility conditions, or nomological roles to sentences; consequently, a theory of this sort might not imply that the sentences within its scope have any particular truth status. However, a theory of truth will imply that the sentences within its scope have certain truth statuses. Of course, the theory of truth alone might not have these consequences, but when combined with a set of auxiliary claims, it will. For example, a theory of truth might imply that a sentence is true if and only if there is a fact to which it corresponds. Alone, that theory of truth has no implications for the truth status of 'dogs are mammals', but when combined with the auxiliary claim that it is a fact that dogs are mammals and 'dogs are mammals' corresponds to this fact, it implies that 'dogs are mammals' is true. One can formulate this point by saying that theories of truth imply truth definitions—that is, some of the consequences of a theory of truth (when combined with auxiliary hypotheses) are assignments of truth statuses to the sentences within its scope. Thus, although a semantic theory for truth might be able to avoid the revenge paradoxes by assigning non-truth-conditional meanings, a theory of truth cannot.

There are two important "based on" relations pertaining to semantic theories. A semantic theory for X is based on both a theory of meaning and a theory of X. The theory of meaning on which a given semantic theory is based determines what sort of meanings it assigns to the sentences in its scope (e.g., a semantic theory for truth that is based on an inferential role theory of meaning

assigns inferential roles to the sentences within its scope, a semantic theory for truth that is based on a truth-conditional theory of meaning assigns truth conditions to the sentences within its scope, etc.) The theory of X on which a semantic theory for X is based lays down necessary and sufficient conditions on the semantic theory for X (e.g., a semantic theory for truth that is based on a contextual theory of truth assigns meanings to the sentences within its scope that conform to the dictates of the theory of truth—e.g., these meanings might be functions from contexts to sets of possible worlds). If a theory of truth implies that all declarative sentences are either true or false, then the semantic theory for truth that is based on this theory of truth must assign meanings to the sentences within its scope that are consistent with this principle. Obviously, the theory of meaning and the theory of X on which a semantic theory for X is based must be consistent, at least with respect to the sentences within the scope of the semantic theory. In sum:

> A semantic theory T for X is *based on* a theory of meaning $\mathfrak{M}$ if and only if the meanings T assigns to the sentences in its scope are consistent with $\mathfrak{M}$'s claims about meanings
>
> A semantic theory T for X is *based on* a theory $\mathfrak{T}$ of X if and only if T is consistent with $\mathfrak{T}$'s claims about X.

One important consequence of the fact that a semantic theory for truth is based on a theory of truth is that if the theory of truth in question is restricted to avoid revenge paradoxes, then any semantic theory for truth that is based on this theory of truth will inherit these restrictions. Of course, one can construct a semantic theory T for truth that is based on a theory $\mathfrak{T}$ of truth such that T is restricted in ways that $\mathfrak{T}$ is not.

There are several ways to define descriptive completeness and descriptive correctness for theories of X. I begin by defining relations between a theory of X and a language (just as I did for semantic theories for X).

> A theory $\mathfrak{T}$ of X is *descriptively correct for L* if and only if $\mathfrak{T}$ is consistent and there exists a semantic theory T for X such that T is based on $\mathfrak{T}$ and T is descriptively correct for L.

> A theory $\mathfrak{T}$ of X is *descriptively complete for L* if and only if there exists a semantic theory T for X such that T is based on $\mathfrak{T}$ and T is descriptively complete for L.
>
> A theory $\mathfrak{T}$ of X is *internalizable for L* if and only if there exists a semantic theory T for X and there exists an extension L′ of L such that T is based on $\mathfrak{T}$, T is expressible in L′, $\mathfrak{T}$ is expressible in L′, and T is descriptively complete for L′.

Notice that these definitions appeal to the analogous terms that express relations between semantic theories for X and languages. For the most part, I focus on semantic theories for truth instead of theories of truth. I present the above definitions to illustrate how to extend the conceptual apparatus I introduce to theories of truth.

9.8 Consequences

I want to emphasize that this chapter is intended to *start* a discussion of these issues, not *finish* one. I do not take any of the formulations or arguments in here to be definitive. Rather, the notion of internalizability and related notions (descriptive completeness and descriptive correctness) are intended to serve as a framework in which discussions of these incredibly complex issues can take place. I hope that, at a minimum, I will have demonstrated the power of this framework.

9.8.1 Object-Language and Metalanguage

The terms 'object-language' and 'metalanguage' are familiar from Tarski's pioneering work on truth. For Tarski, the object language is the language for which one is defining a truth predicate and the metalanguage is the one in which the definition is formulated. In Tarski's case, the truth predicate marks the difference—the object language cannot contain a predicate that is true of all and only the true sentences of the object language, but the metalanguage does contain such a predicate.[25]

[25] Tarski (1933).

Since the work of Kripke and others in the mid 1970s, theorists have been studying languages that contain their own truth predicates. The terms 'object-language' and 'metalanguage' have come to mean something like: *language to which a theory of truth is intended to apply* and *language in which a theory of truth is formulated*. These theorists distinguish between object-language and metalanguage to avoid revenge paradoxes. That is, if the theory were to apply to a language capable of formulating it, then it would be inconsistent. So, for these theorists, the distinction between object-language and metalanguage marks the languages to which the theory applies and those to which it is restricted from applying.

9.8.1.1 Properties of Theories

If a semantic theory for truth requires an expressively richer metalanguage, then it is not internalizable for any language. Assume otherwise; then there is a language L such that T is expressible in L and descriptively complete for L, but then T does not require an expressively richer metalanguage.

On the other hand, a semantic theory for truth that does not require an expressively richer metalanguage is internalizable for some language, but it might not be internalizable for every language or even for natural languages. As we saw in section 9.5, it is possible that a semantic theory T is internalizable for a language L but not internalizable for another language L′. If T is expressible in L, and T is restricted so that it does not apply to L′, then T does not require an expressively richer metalanguage, but it is not internalizable for every language. If L′ is a natural language, then it does not apply to a natural language. As we will see in a bit, Hartry Field's approach to the liar has this feature.

These points suggest that we can use the notion of internalizability to define *properties* of semantic theories (as opposed to *relations* between semantic theories and languages, which was the

subject of the previous section). Here are some examples. We can define three descriptive completeness properties for semantic theories:

> A semantic theory T for X is *weakly descriptively complete* if and only if for some X-language L, T is descriptively complete for L.
>
> A semantic theory T for X is *naturally descriptively complete* if and only if for every natural language L, T is descriptively complete for L.
>
> A semantic theory T for X is *strongly descriptively complete* if and only if for every X-language L, T is descriptively complete for L.

Likewise, we could define three analogous descriptive correctness properties for semantic theories, but they would be of limited interest. Rather we are almost always interested in semantic theories that are descriptively correct for every sentence in their scope. I reserve the term 'descriptively correct' as a predicate of semantic theories for this purpose:

> A semantic theory T for X is *descriptively correct* if and only if for all X-sentences s, if s is in the scope of T and s belongs to the X-language L, then T is descriptively correct for L.

Finally, we have three internalizability properties for semantic theories:

> A semantic theory T for X is *weakly internalizable* if and only if there exists an X-language L such that T is internalizable for L.
>
> A semantic theory T for X is *naturally internalizable* if and only for every natural language L, T is internalizable for L.
>
> A semantic theory T for X is *strongly internalizable* if and only if for every X-language L, T is internalizable for L.

For each group of properties, strong implies natural and natural implies weak.

To recap, if a semantic theory for truth requires an expressively richer metalanguage, then it is not weakly internalizable. A semantic theory for truth that does not require an expressively richer metalanguage is weakly internalizable, but it might not be naturally internalizable or strongly

internalizable. The internalizability properties presented above do a much better job of describing semantic theories than the traditional object-language / metalanguage distinction.[26]

9.8.1.2 The Reinhardt-McGee Intuition

In section 9.1, I also discussed the Reinhardt-McGee intuition—semantic theories for truth that require expressively richer metalanguages are unacceptable for use on natural languages. We can formulate the Reinhardt-McGee intuition in the new terminology: an acceptable semantic theory for truth that applies to a natural language should be internalizable for some language.[27] We can also formulate other requirements that are similar in spirit:

> (STRONG) A semantic theory for X should be internalizable for every X-language (i.e., should be strongly internalizable).
>
> (WEAK) A semantic theory for X should be internalizable for some X-language (i.e., should be weakly internalizable).

Any semantic theory that satisfies (STRONG) satisfies (WEAK) as well. However, these requirements make no mention of natural languages. The following are three internalizability requirements that are specific to natural languages:

> (STRONG_{N}) A semantic theory for X that applies to a natural language should be internalizable for every X-language.
>
> ($\text{MODERATE}_{\text{N}}$) A semantic theory for X that applies to a natural language should be internalizable for that language.
>
> (WEAK_{N}) A semantic theory for X that applies to a natural language should be internalizable for some X-language.

---

[26] It is relatively straightforward to formulate similar properties for theories of truth (instead of properties of semantic theories for truth).

[27] Although in the passage I quoted above, McGee seems to claim that for a semantic theory to apply successfully to a natural language, it must be internal for that language, in later writings he makes it clear that he is interested in internalizability instead of internality; see McGee (1997: 405-406).

A semantic theory that satisfies (STRONG_N) satisfies (MODERATE_N) and one that satisfies (MODERATE_N) satisfies (WEAK_N), but the converses of these claims are false.

The Reinhardt-McGee intuition is that a semantic theory that applies to a natural language should apply to some language in which it is formulated. In my terminology, they suggest (WEAK_N). Above, I claimed that this condition is too weak. Now I can provide an argument for this claim. Let T be a descriptively correct semantic theory for truth that is internalizable for some language L. Thus, there exists an extension L′ of L such that T is expressible in L′ and T is descriptively complete for L′. Assume that there is some concept Y that is not expressed by any of the sentences that constitute T; however, T is restricted so that any sentence that expresses Y is outside its scope. Finally assume that a natural language N is a Y-language (i.e., some sentences of N express Y). Given that N is a truth-language (i.e., it contains a truth predicate) and it contains a completely defined two-place sentential operator (e.g., 'and'), there are sentences of N that express both truth and Y. T might apply to N, but because no Y-sentences are in the scope of T, T is not descriptively complete for N; that is, no sentence of N that expresses both truth and Y is in the scope of T. Therefore, it is possible that a descriptively correct semantic theory for truth satisfies (WEAK_N), but is not descriptively complete for a natural language.

9.8.2 On Being Revenge Immune

I have already mentioned Hartry Field's approach to the aletheic paradoxes that relies on a paracomplete logic, in which the law of excluded middle and several other classically valid principles fail.[28] One very attractive feature of the approach is that its metalanguage is a fragment of its object language. That is, the theory of truth can be formulated in a language to which it applies. Field seems to think that this

[28] Field (2003a, 2003b, 2005b, 2008a).

feature of his theory renders it "revenge-immune".[29] This sentiment is certainly in keeping with the Reinhardt-McGee intuition. That is, if only theories of truth that require an object language / metalanguage distinction are those that face revenge paradoxes, then a theory of truth whose metalanguage is a fragment of its object-language does not face revenge paradoxes.

Field's approach to the aletheic paradoxes does not require an expressively richer metalanguage. However, Field's semantic theory for truth provides an excellent example of a theory that is weakly internalizable, but not naturally internalizable (and thus not strongly internalizable). Field shows that, for a certain artificial language L, his semantic theory for truth is descriptively correct for L, descriptively complete for L, and expressible in L. Thus, his semantic theory is internalizable for L (indeed, it is internal for L). However, Field's semantic theory is restricted so that it does not apply to sentences that contain certain expressions (e.g., exclusion negation). The reason for the restriction is that if such sentences were in the scope of the theory, it would be inconsistent; in other words, his theory is restricted so that it does not face a revenge paradox. Moreover, the concepts that feature in revenge paradoxes for Field's theory are expressible in English.[30] If that is right, then Field's semantic theory for truth is not internalizable for English even though it does not require an expressively richer metalanguage.

One might wonder, is there a property of a semantic theory for X that renders it revenge-immune? It seems to me that being descriptively correct and descriptively complete for every X-language is just such a feature. If a semantic theory for X has this property, then we know that it faces no revenge paradoxes whatsoever.

---

[29] Field does not make this claim explicitly in the text, but the title of Field (2003) is "A Revenge-Immune Solution to the Liar Paradox," and the solution he offers has the feature in question; moreover, he discusses how this feature prevents his theory from giving rise to revenge paradoxes in one of the final sections of the paper; see also Field (2008a: chs. 21-23).

[30] This claim is contentious; see Field (2008a: ch. 23).

Of course, it is difficult to argue that a semantic theory for X has this feature since it involves every language that expresses concept X. However, we can show that there is an important necessary condition for being strongly descriptively complete and descriptively correct: strong internalizability. The argument I present depends on the claim that a semantic theory for X that is not internalizable for an X-language is not descriptively complete for every X-language. In other words, if a semantic theory is strongly descriptively complete, then it is strongly internalizable.

I rely on the following two assumptions about languages:

(i) *There are no concepts that are language-specific* (if there exist two languages, L_1 and L_2, such that L_1 can express a concept that L_2 cannot, then L_2 can be extended to a new language L_3 that can express the concept in question), and

(ii) *Any two languages are quasi-intertranslatable* (for any two languages L_1 and L_2, L_1 can be extended to a language L_3 and L_2 can be extended to a language L_4 such that L_3 and L_4 are intertranslatable).

Theorem: If T is a semantic theory for X and T is not internalizable for an X-language L, then T is not strongly descriptively complete.

Proof: Because T is not internalizable for L, either: (i) T is not descriptively complete for L or (ii) T is not expressible in L. On option (i), T is not strongly descriptively complete (which is my conclusion); thus, if T is *not* descriptively complete for L, then I am done. Consider option (ii), and assume that T *is* descriptively complete for L. Given that T is a theory, there is some language L′ in which T is expressible. L and L′ are quasi-intertranslatable because all languages are quasi-intertranslatable. Hence, L can be extended to a language L″ and L′ can be extended to a language L‴ such that L″ and L‴ are intertranslatable. If T is expressible in L′, T is expressible in L″; if T is expressible in L‴, then T is expressible in L″. Hence, T is expressible in L″. If T is expressible in L″, then T is not descriptively complete for L″ (because T is not internalizable for L). Hence, T is not descriptively complete for L″. Therefore, T is not strongly descriptively complete. ■

Here is an intuitive summary of the argument. T is not internalizable for L, but T has to be formulated in some language or other; if we consider the result of adding to L whatever expressive resources it takes to express T, then T will be descriptively incomplete for that extended language. For example, assume that T is a semantic theory for truth that treats sentences like the liar as truth-value gaps. Assume that if a revenge liar (e.g., 'this sentence is either false or a truth value gap') is in the scope of T, then T is inconsistent (e.g., it implies are that the revenge liar is both true and not true). In order to keep T consistent, one restricts T so that it does not apply to sentences that contain gaphood predicates. Although T might be descriptively complete for a language that does not contain a gaphood predicate, T will not be descriptively complete for languages that have both a truth predicate and a gaphood predicate because such a language will contain a sentence in which both a truth predicate and a gaphood predicate occur; such a sentence will not be in the scope of T. Hence, T does not provide assignments for all the truth-sentences of this language. Therefore, T is not descriptively complete for this language. Consequently, T is not strongly descriptively complete.

Why should one care about strong descriptive completeness? The most obvious requirement for a semantic theory for X is that it should explain the concept X. Philosophical explanation is a notoriously slippery concept but it is fairly straightforward in the case of semantic theories: a semantic theory T for X explains the concept X if and only if T assigns the right meaning to every sentence containing a word that expresses X. In my terminology, that means that a semantic theory for X should be both strongly descriptively complete and strongly descriptively correct. I have shown that only strongly internalizable semantic theories can satisfy this demand; indeed, I have shown that only strongly internalizable semantic theories can satisfy the strong descriptive completeness condition, regardless of whether they are descriptively correct. A semantic theory for X that is not internalizable for some X-language will not be able to specify the meanings of all the

sentences that express X. Hence, a semantic theory for X that is not internalizable for an X-language fails to explain X.

From these materials, we can construct an easy argument for (STRONG)—i.e., that a semantic theory for X should be internalizable for every X-language:

(a) If T is descriptively complete for every X-language, then T is internalizable for every X-language.

(b) Being descriptively complete for every X-language is a necessary condition on any adequate semantic theory for X.

∴ (c) Being internalizable for every X-language is a necessary condition on any adequate semantic theory for X.

I have argued for (a) and (b) above. The result is an easy argument for a condition on semantic theories for truth (and thus on theories of truth) that is far stronger than the Reinhardt-McGee intuition.

One might protest: we have no idea whether a semantic theory that is both strongly descriptively complete and strongly descriptively correct is even *possible*. Consider the case of truth. We are talking about a theory that works for *every* language that has a truth predicate. There is no good reason to require semantic theories to live up to these expectations. It hardly makes sense to criticize a semantic theory for failing to be applicable to a handful of sentences in a handful of languages.

This sort of "burden of proof" objection is common in the literature on truth; I address it here in some detail and refer back to it when it comes up later.[31] The obvious and intuitive view is that truth is a concept and there is something all sentences that express this concept have in common. That is the view we take on other areas of inquiry. A semantic theory for X that is consistent and provides correct assignments for all the X-sentences in all the X-languages is the ideal for a semantic

---

[31] See Gupta and Belnap (1993: 257) and Gupta (1997) for similar comments on semantic self-sufficiency.

theory for X. Such a semantic theory is strongly descriptively complete and strongly descriptively correct. Any semantic theory that fails to live up to this ideal is inadequate. Semantic theories that are not strongly internalizable are not strongly descriptively complete. Thus, they are inadequate.

A proponent of a semantic theory T for X that is not internalizable can respond to this fact in several different ways. She could claim either that T is actually a semantic theory for *X in some contexts* or that it is a semantic theory for a restricted version of X. I have no problem with either of these responses, and there is an important place for such theories. However, they leave us without a semantic theory for X and so without an explanation of X.

Instead, a theorist might argue either that the obvious and intuitive view is wrong—that there really is nothing in common that calls for explanation or that no matter how hard we try, the concept in question will remain inexplicable.[32] Either one of these responses requires substantial argumentation. However, a theorist who points out that we do not know whether it is possible to achieve the ideal of descriptive completeness and descriptive correctness has not adequately responded to the problem. *Of course* we do not know if it can be done. If it turns out that we cannot accomplish it then there remains something we cannot explain. If a theory that is not strongly internalizable is the best we can do, then the best we can do is inadequate given our understanding of what needs to be done. I admit that I cannot *prove* that a given semantic theory is strongly descriptively complete and descriptively correct because I do not have access to all the languages to which it applies. However, I can prove that some theories are *not* strongly descriptively complete and descriptively correct. Thus, even though we do not know whether it is possible to provide a descriptively complete and descriptively correct semantic theory for X, we do know both that semantic theories that are not strongly internalizable are not strongly descriptively complete, and

---

[32] See Williams (1996) for the claim that the intuitive view of knowledge is wrong—there really is nothing in common that calls for explanation. See McGinn (2000) for the claim that the intuitive view of consciousness is wrong—no matter how hard we try, it will remain inexplicable.

that semantic theories that are not strongly descriptively complete are inadequate. In conclusion, unless we already have good reason to believe that a strongly descriptively complete and descriptively correct semantic theory for X is impossible, we have good reason to be unhappy with a semantic theory for X that is not internalizable for some X-language.[33]

I argued earlier that most, perhaps all, of the familiar theories of truth and semantic theories for truth have to be restricted to avoid revenge paradoxes. One might claim that if that is the case, then the fact that a semantic theory for truth fails to satisfy (STRONG) does not really count against it. Because every theory has the same problem, it is illegitimate (or at least unhelpful or uninteresting) to criticize a theory for having this problem.

It makes no difference whether every theory of truth has a certain problem—it is still a problem. If every theory of truth faces the same problem, then that is the problem on which we should concentrate when trying to explain truth. If (STRONG) is a condition on acceptable semantic theories for truth and it turns out that no known semantic theory for truth satisfies it, then we are left without an acceptable semantic theory for truth. Of course, a problem of this sort cannot be used to justify one theory over the others. Moreover, the unacceptable theories can still prove to be helpful—one might be able to learn something true and important from studying them.

A diagnosis: it is as if the objector is saying that the problem is so severe and so devastating that one would rather ignore it than expend the effort to address it. That is, this sort of objection is part of an attempt to justify philosophical laziness, ineptitude, or cowardice. Either one is too lazy or too inept to address the problem, or one would rather cling to a discredited theory that gives the illusion of understanding rather than have the courage to face the fact that we clearly do not understand the phenomenon in question.

---

[33] A different burden of proof objection is that I have not shown that it is possible to construct internalizable semantic theories. My reply should be obvious.

It is my view that if it were the case that no known theory avoids the criticism I present here, then this objection would be impotent. However, there are theories of truth that do not face revenge paradoxes, and there are semantic theories for truth that are strongly internalizable (at least there does not seem to be any reason to think they are not). It should not come as a surprise that I advocate such a theory; it is the topic of Part III.

A different sort of objection comes from the semantic teamwork point. Let us assume that when we use a semantic theory for X to derive an assignment for a sentence that expresses both concept X and concept Y, we use both a semantic theory for X and a semantic theory for Y to do it. In order to demonstrate that a semantic theory for X is descriptively complete for an X-language L, we must have semantic theories for all the other linguistic items that occur in L. Hence, whether a semantic theory for X is strongly descriptively complete depends on the existence of semantic theories for all the other linguistic items that appear in X-languages. Achieving strong descriptive completeness is a team effort. Hence, a semantic theory for X can fail to be strongly descriptively complete if it turns out to be impossible to provide a semantic theory for some other concept that appears in an X-language. Why should a semantic theory be held accountable for that?

Perhaps it will turn out that there are inexplicable linguistic items. If there are, then no semantic theory will be strongly descriptively complete. I agree that we should not fault a semantic theory for X for failing to be strongly descriptively complete just because some X-language contains an inexplicable linguistic phenomenon that is relatively unrelated to X. If such a phenomenon occurs in an X-language, then a descriptively incomplete semantic theory for X will be the best we can hope for. That is, if we have good reason to believe that some concept Y is inexplicable, then we would have good reason to think that a semantic theory for X need not be strongly descriptively complete. However, in the case of semantic theories that are not strongly internalizable, the theory itself is responsible for its inadequacy, not some unrelated inexplicable concept.

Moreover, the fact that descriptive completeness is a team effort cuts both ways. If one accepts a semantic theory for X that is not strongly internalizable, then one accepts that it is impossible to provide a strongly descriptively complete semantic theory for *any* other linguistic item that occurs in an X-language.[34] Assume that T is a semantic theory for X that is not internalizable for L. There exists an X-language for which T is not descriptively complete. Hence, there exists an X-sentence of this language for which T provides no assignment. Consider a semantic theory T′ for Y, where Y is expressed by this X-language. There exists a sentence of this X-language that is both an X-sentence and a Y-sentence and is outside the scope of T′.[35] Because T cannot provide an assignment for this sentence, T′ cannot either. Hence, T′ is not descriptively complete for this language. Therefore, T′ is not descriptively complete. That is a serious problem. It means that if we accept a semantic theory for truth that is not strongly internalizable, then we give up on descriptive completeness for semantic theories for knowledge, for semantic theories for necessity, for semantic theories for moral obligation, for *every* other semantic theory one can imagine. It means that if we accept *even one* semantic theory that is not strongly internalizable, we effectively give up trying to explain any of our concepts. The best we would be able to do is explain restricted versions of them. I, for one, am not willing to damn humanity to eternal ignorance.

A final objection comes from the role of idealizations in semantics. Even when we focus on natural languages, we often study artificial languages and ignore certain aspects of natural languages (e.g., indexicals, demonstratives, intensional expressions, pronouns, indefinite descriptions, vagueness, ambiguity, empty names, interrogatives, imperatives, etc.). If we try to explain everything all at once, then it becomes difficult to make any progress. The criticism of semantic theories that are not strongly internalizable is like saying that a semantic theory for X is inadequate because one

---

[34] Semantic theories that are internalizable for some languages would still be possible though.
[35] I assume that the language has some completely defined two-place sentential operator (e.g., 'and').

can make up some new term that the theory was not designed to handle. That hardly seems fair. Thus, if we accept the internalizability requirement (STRONG), we would no longer be able to make idealizations in semantics.

I am not arguing that we should give up idealizations in semantics. In the current state of the discipline, idealizations are important and helpful. The internalizability requirement (STRONG) does not imply that idealizations in semantics are illegitimate. When presenting a semantic theory, one might want to ignore certain linguistic phenomena. However, the assumption is that the semantic theory should be compatible with an account of those phenomena. That is, the idealization can be dropped at a later time. However, one can make an idealization without *restricting* one's semantic theory. For example, Bob presents a theory of necessity and a semantic theory for necessity. He presents both theories by explaining how they apply to a certain artificial language that does not have any context-dependent expressions. Does that mean that his theories are restricted so that they do not apply to languages that have context-dependent expressions? No. Of course, his semantic theory for necessity will have to work with a semantic theory for context-dependent expressions to provide assignments to sentences that contain both 'necessary' and a demonstrative (e.g., 'that sentence expresses a necessary proposition'). Because semantic theories work together as a team to provide assignments to sentences, there is no need to restrict a semantic theory unless one has good reason to think either that it will provide the wrong assignments to certain sentences or that it will be rendered inconsistent or self-refuting if certain sentences are within its scope. Therefore, one restricts a semantic theory only if it is going to run into trouble otherwise.

If one thinks of the restrictions placed on non-internalizable semantic theories as idealizations, then they are idealizations that cannot be dropped. Consider Kripke's semantic theory for truth, which employs a gaphood predicate. A restricted version of this theory is not applicable to sentences that contain gaphood predicates. One might think of this as an idealization. However, it

constitutes a permanent idealization. It cannot be used in combination with a semantic theory for gaphood without being inconsistent.

Furthermore, when one makes an idealization for a semantic theory for X by excluding some linguistic phenomenon, one must be in a position to say that the phenomenon in question is relatively unrelated to X. For example, we feel justified when giving a semantics for truth in excluding color predicates from the ideal languages we consider because the two are relatively independent. However, according to non-internalizable semantic theories for X, the very concepts that get excluded are intimately related to X. For example, a semantic theory for truth that uses gaps and is not strongly internalizable posits an important relation between truth predicates and gaphood predicates. Yet it excludes gaphood predicates from the languages it considers. And, because it is not strongly internalizable, it is impossible to drop this idealization.

9.8.3 Semantic Self-Sufficiency

It should come as no surprise that arguments for the Reinhardt-McGee intuition often turn on the expressive properties of natural languages. Since the metalanguage for a theory of truth is almost always a natural language, it does not seem that a theory that requires an object-language/metalanguage distinction could apply to a natural language.[36]

Anil Gupta, in an effort to defend the revision theory of truth (which requires a distinction between object-language and metalanguage), formulates the following version of this argument:

(a) It is possible to provide a semantic description of a natural language L.
(b) A semantic description of L is expressible in L (i.e., L is semantically self-sufficient).
(c) The revision theory must be formulated in a language that is expressively richer than the one it describes.

---

[36] For discussion see Priest (1987), McGee (1991, 1997), Simmons (1993), Martin (1997), Gupta (1997), and Eklund (2008).

(d) The revision theory is not suitable for L (from (a), (b) and (c)).

∴ (e) The revision theory fails to explain truth in L (from (d)).

Gupta then attacks this argument by claiming that we have no reason to believe that natural languages are semantically self-sufficient (i.e., attacking (b)).[37]

I agree with Gupta that we have no reason to believe (b). Still, it is not obvious how a theory of truth that requires an expressively richer metalanguage successfully applies to a natural language. Gupta suggests that some of the vocabulary of the theory is not present in natural language—it is technical vocabulary made up just for the theory in question. Thus, a theory of this kind can apply to a natural language as it was prior to the formulation of the new vocabulary.[38] It is hard to see how a defender of the Reinhardt-McGee intuition would counter this move.

Gupta goes on to distinguish between two projects: the truth project and the semantic self-sufficiency project. The *truth project* is to "give the semantics of the predicate 'true in L' of L," whereas the goal of the *semantic self-sufficiency project* is "the construction of a language L that can express its own semantic theory."[39] Gupta then argues that the truth project is independent of the semantic self-sufficiency project because the former is still viable even if the latter is impossible.

In the framework I offer, it becomes clear that there is not just one notion of semantic self-sufficiency. Moreover, there are compelling arguments that the truth project does depend on the semantic self-sufficiency project (in certain senses of 'semantic self-sufficiency'). Using the notion of internalizability, we can define several notions of self-sufficiency (I omit 'semantic' from here on for ease of reading).

> A language L is *weakly self-sufficient* if and only if for every concept X expressible in L, there is a semantic theory T for X s.t. T is expressible in L, descriptively complete for L, and descriptively correct for L.

---

[37] Gupta (1997: 437). Note that Gupta formulates (b) with a 'must be' in place of the 'is', which is a mistake. According to his definition, a language L is semantically self-sufficient if and only if L can express its own semantic theory.
[38] Gupta (1997: 439); a similar move is made in Kripke (1975: 714n34).
[39] Gupta (1997: 438).

> A language L is *strongly self-sufficient* if and only if for every concept X expressible in L, there is a semantic theory T for X s.t. T is expressible in L, T is descriptively complete for every X-language and T is descriptively correct for every X-language.

These are the two most obvious notions of self-sufficiency, but there are others. In particular, we can talk about whether a language is self-sufficient with respect to a particular concept:

> A language L is weakly-X-self-sufficient if and only if L is an X-language and there is a semantic theory T for X s.t. T is expressible in L, T is descriptively complete for L, and T is descriptively correct for L.

> A language L is strongly-X-self-sufficient if and only if L is an X-language and there is a semantic theory T for X s.t. T is expressible in L, T is descriptively complete for every X-language, and T is descriptively correct for every X-language.

These properties of languages are interesting, but there is another sense of self-sufficiency that Gupta and McGee discuss. Consider the following passage from McGee:

> It is quite likely, in fact, that to describe the semantics of natural language will require new concepts and new vocabulary. What I do want to express is the earnest expectation that the conceptual tools we would need to describe, sketchily but coherently, the semantics of natural language do not lie irretrievably beyond our reach. I want to express the hope that, someday in the future, say by the year 2050, linguists and philosophers will develop at least a coherent outline of a semantic theory of natural language that comprehends, among other things, English as it is spoken in the year 2050.[40]

Given the definition of 'language' in use throughout this chapter, we can think of English (or any other natural language) as a sequence of languages. The notion of self-sufficiency in the passage from McGee is not any of the ones above—it is not that English in its current state is self-sufficient; rather, he hopes that some extension of English in its current state is self-sufficient. Of course, given the different kinds of self-sufficiency above, we can interpret this hope in several different ways:

> A language L is *potentially weakly self-sufficient* if and only if there is an extension L′ of L s.t. for every concept X expressible in L, there is a semantic theory T for X s.t. T is expressible in L′, T is descriptively complete for L′ and descriptively correct for L′.

---

[40] McGee (1997: 406); for Gupta's discussion see Gupta (1997: 439-440).

Or, more generally:

> L is *potentially* (*weakly* / *strongly*) *self-sufficient* if and only if there is an extension of L that is (weakly / strongly) self-sufficient.

Using the internalizability framework, we have captured a variety of notions.

One big question is: do any of these formulations allow us to argue that natural languages are self-sufficient? Indeed, they do. Let T be a semantic theory for X. We saw in the last subsection that if T is descriptively complete for all X-languages, then T is internalizable for all X-languages (i.e., T is strongly internalizable). Notice now that if T is strongly internalizable and descriptively correct, then for every X-language L, L is potentially strongly X-self-sufficient (that is, L has an extension L′ such that T is expressible in L′, T is strongly descriptively complete and T is descriptively correct). If T is strongly internalizable, then T is internalizable for every X-language. If T is internalizable for every X-language and descriptively correct, then every X-language has an extension in which T is internal. That extension is strongly X-self-sufficient.

Gupta claims that the self-sufficiency project and the truth project are independent. We can now see that if a semantic theory for truth is descriptively correct and descriptively complete for every truth-language, then every truth-language is potentially strongly truth-self-sufficient. That is, if a semantic theory for truth is descriptively correct and descriptively complete for every truth language, then every language that contains a truth predicate has an extension that is strongly self-sufficient with respect to truth. Here we have shown (contra Gupta) that there is an important sense in which the truth project depends on the self-sufficiency project.

We know already that the semantic theory for truth offered by Gupta is not even weakly internalizable. If that is the best we can do, then there is no possibility of a semantic theory for truth that is descriptively correct and strongly descriptively complete. However, if we think that such a semantic theory for truth is possible, then we have good reason to believe that natural languages are

potentially strongly truth-self-sufficient, and that gives us reason to reject Gupta's theory along with any other that shares its inadequacies.

9.8.4. Tarski's Indefinability Theorem

I admitted that (STRONG) is not binding if one shows that there is no possibility of satisfying it. Indeed, we do have results that suggest that a descriptively correct internalizable semantic theory for truth is impossible; namely, Tarski's indefinability theorem and related results. One can use these results to show that it is impossible to construct an internalizable semantic theory for truth. Every semantic theory for truth faces a revenge paradox. Each revenge paradox can be used to prove an indefinability theorem. Each indefinability theorem implies that the semantic theory in question is not internalizable (so long as it is descriptively correct). Therefore, we have good reason to believe that a strongly internalizable semantic theory for truth is impossible.[41]

Let us first review Tarski's indefinability theorem. He proved that if a language L is bivalent (i.e., every sentence of L is either true or false), L is monoaletheic (i.e., no sentence of the language is both true and false), and L has the capacity to describe its own syntax, then L does not contain a predicate that is true of all and only the true sentences of L. That is, truth-in-L is indefinable in L. Given that a semantic theory for truth-in-L contains sentences that express the concept of truth-in-L, no such language can express a semantic theory for truth-in-L. Tarski proves his theorem by reductio; he shows that if L does contain its own truth-in-L predicate and satisfies the conditions of the theorem, then it contains a paradoxical sentence (i.e., a sentence for which one can derive that it is both true-in-L and not true-in-L).[42]

---

[41] See Herzberger (1970, 1981), Parsons (1983), and McCarthy (1985), which contain remarks that suggest that these philosophers would be sympathetic to this argument. See Field (2003a, 2003b, 2005b) for discussion.
[42] Tarski (1933). See also McGee (1985, 1991), Gupta and Belnap (1993), Simmons (1993), Halbach (1996, 1997), Soames (1999), Ketland (2000), Field (2003a, 2003b, 2008a: chs. 1 and 2), and Maudlin (2004).

One can prove similar results using revenge paradoxes. For example, one can prove that a language that is not bivalent but satisfies the other conditions of Tarski's theorem does not contain a predicate with an extension that is the set of true sentences of the language and an anti-extension that is the set of untrue sentences of the language. One can prove this theorem using a revenge liar for gap approaches (e.g., sentence (2)). Another example comes from Gupta and Belnap's semantic theory for truth (i.e., the revision theory). The revision theory employs the notion of categoricality; it implies that pathological sentences are uncategorical. Accordingly, one can construct a revenge paradox for the revision theory using the sentence:

(6) (6) is either false or uncategorical.

On the revision theory, (6) is both categorical and uncategorical. Gupta and Belnap use this result to prove the indefinability of categoricality in languages that have the capacity to construct this revenge liar. One consequence is that no language that satisfies these conditions can express the revision theory. Thus, the revision theory is not internalizable so long as it is consistent.[43]

There are a number of problems with the argument given in the objection. First, the indefinability results are proven using sentences that figure in only one type of revenge paradox: inconsistency problems. However, not all semantic theories face inconsistency problems. For example, a semantic theory based on an error theory of truth (i.e., all sentences with truth predicates are false) does not face an inconsistency problem (of course, it faces a horrible self-refutation problem, but that is quite different—one cannot prove indefinability results with self-refutation problems). One might argue that any *plausible* semantic theory for truth faces an inconsistency problem or that any semantic theory for truth that does not face a self-refutation problem faces an inconsistency problem, but those are not the arguments under consideration. Moreover, in my view,

[43] Gupta and Belnap (1993: 229-230).

there are plausible semantic theories for truth that do not face any revenge paradoxes. All of them are in the inconsistency tradition. That is, they all imply that truth is an inconsistent concept.[44] It is my view that revenge paradoxes result from treating what is essentially an inconsistent concept as if it were consistent (more on this in Part III). My point is that some semantic theories for truth do not face the type of revenge paradox used to prove indefinability results and hence, the above reasoning is unsound.

Second, the indefinability results are not as strong as they appear. Let us take a look at them in more detail. Each one has the following form: if L is a language with properties ϕ_1, ϕ_2, etc., and a semantic concept τ is definable in L, then L contains a sentence σ such that σ both has and does not have some property ψ. The proof concludes by rejecting the claim that τ is definable in L.

Each indefinability result has a number of hidden premises. First, they assume that the semantic concept in question is consistent. If it is an inconsistent concept, then it should not come as a surprise that it both applies and fails to apply to certain items. Moreover, if the concept in question is inconsistent, then it is not obvious that reductios are valid forms of reasoning for sentences that express this concept. Second, each result uses certain claims about the semantic concept in question to derive the contradiction. For example, Tarski's indefinability result relies on convention T (i.e., that for each sentence $\langle p \rangle$ of the language, one can show that $\langle p \rangle$ is true if and only if p). A second example is that Gupta and Belnap's indefinability result for categoricality relies on the claim that the truth predicate in question obeys the revision theory for truth. Thus, each indefinability result depends on a certain theory. Moreover, each depends on the claim that the theory in question applies to the language in question. Finally, each result depends on the claim that the theory in

---

[44] For examples of theories of truth in the inconsistency tradition, see Chihara (1973, 1979, 1984), Yablo (1985, 1993a, 1993b), Priest (1987), Eklund (2002a, 2002b), Patterson (2006), and Scharp (2007, 2008).

question is consistent. Obviously, if the revision theory for truth is inconsistent, then we should expect that one could show that it implies that sentence (6) is both categorical and uncategorical.

Once the hidden premises of the indefinability results are made explicit, it is obvious that they pose no threat to the internalizability requirements. Each result has the form: if such and such theory is consistent, correct, and applies to such and such languages, then these languages cannot express such and such consistent concept. However, that is considerably weaker than the formulation in the objection, and it does not support the claim that no descriptively correct semantic theory for truth can be strongly internalizable. Therefore, there is no good argument from the indefinability results to the claim that a descriptively correct strongly internalizable semantic theory for truth is impossible.

9.9 Impact

This chapter continues the theme presented in the previous one; as such its impact is similar. Since virtually every philosophical and logical approach to the aletheic paradoxes generates revenge paradoxes and these render the semantic theories for truth associated with these approaches non-internalizable for natural languages, the results in this chapter suggest that all these approaches are unacceptable.

Chapter 10

What is the Problem?

10.1 The Challenge for Unified Theories of Truth

Part II has focused on four key issues, which are relatively straightforward, but there are myriad complicated ways they relate to one another and to the literature on truth presented in Part I. In this section, I attempt to summarize the systematic challenge that, together, they pose for those of us interested in a unified theory of truth (i.e., a theory of the nature of truth together with a philosophical approach and a logical approach to the aletheic paradoxes).

10.1.1 Against Analysis

The vast majority of work on the nature of truth is engaged in trying to find a conceptual analysis of truth. This would be something like a definition of truth in terms that are more primitive or more fundamental or better understood or less controversial. Many have criticized this kind of project before, mostly on the grounds that there is nothing more primitive, fundamental, better understood, or less controversial.[1] I too see no hope for such a project, and I want to say a bit about my reasons.

All the purported analyses agree that the primary aletheic principles,

(T-In) If p, then ⟨p⟩ is true, and

(T-Out) If ⟨p⟩ is true, then p,

hold. However, when we look at logical approaches to the aletheic paradoxes that validate (T-In) and (T-Out), they all require a rejection of classical logic. Now, I am not one to think that classical

[1] Davidson (1996) summarizes his reasons, which seem like good ones to me.

logic is sacrosanct. The debates about classical logic vs. intutionistic logic vs. relevance logic are intricate and subtle; perhaps there is good reason to give up classical logic for one of these alternatives.[2] Nevertheless, this change will not help with the aletheic paradoxes—the two major kinds of logic that permit (T-In) and (T-Out) are paracomplete logics and paraconsistent logics. There is no independent motivation for these besides considerations having to do with (T-In) and (T-Out). Moreover, each of these leaves us without a conditional that allows anything like ordinary reasoning. In fact, both of them use conditionals that are cooked up to avoid the paradoxes. It seems to me that many of the inference rules common to classical logic, intuitionistic logic, and relevance logic (R) are constitutive of the logical connectives of natural language, including the conditional, negation, conjunction, and disjunction. As I have stressed again and again, the move to paracomplete or paraconsistent logic requires giving some of these up. Thus, if one accepts (T-In) and (T-Out), then one has to give up principles that seem to be constitutive of our logical concepts (even if one accepts contradictions as in paraconsistent logic). Since there is no independent reason to think that there is anything wrong with these principles, and, giving them up affects much more than just our use of the truth predicate, it seems to me that, if given the choice, one should avoid any theory that has this consequence. In other words, the principles we take to be constitutive of truth are incompatible with the principles we take to be constitutive of our basic logical vocabulary (again, given that our languages are not trivial and that they have the capacity for self-reference).

Let us consider this problem in a bit more detail. Nearly everyone takes the inference rule modus ponens,

(MP) $A, A \rightarrow B \vdash B,$

[2] See Anderson and Belnap (1975), Anderson, Belnap, and Dunn (1992), Dummett (1975, 1991, 2000), Tennant (1997), Priest (2001), and Mares (2004). See also the debate about whether there is a single correct logic (logical monism) or a plurality of correct logics (logical pluralism): Beall and Restall (2000, 2001, 2006), Wyatt (2004), Allo (2007), Moretti and Ciprotti (2009), and Cook (2010).

to be constitutive of the conditional.[3] In fact, those who reject modus ponens (e.g., McGee) are now used as stock examples of someone who rejects a constitutive principle and still possesses the concept in question.[4] In addition, the meta-rule, conditional proof,

(CP) If $A \vdash B$, then $\vdash A \rightarrow B$,

is also constitutive of the conditional. It just says that if B is derivable from A, then the conditional 'A→B' is derivable. So it seems that any acceptable approach to the aletheic paradoxes is compatible with a logic that includes both (MP) and (CP).[5]

Here is the problem. Non-classical logical approaches accept both (T-In) and (T-Out) and revise classical logic to avoid triviality (the two types are paracomplete and paraconsistent—see Chapter Three for details). The trick to weakening the background logic is to have a conditional that does not contract because Curry's paradox shows us that there is no way to have both truth principles and a conditional that obeys contraction.[6] Contraction is the odd looking axiom:

(Contraction) $\vdash (A \rightarrow (A \rightarrow B)) \rightarrow (A \rightarrow B)$.

It turns out that one can also formulate a Curry paradox using both truth principles and an axiom that is sometimes called material modus ponens:

(MMP) $\vdash A \wedge (A \rightarrow B) \rightarrow B$.

---

[3] However, see McGee (1985) for a potential counterexample.
[4] See Williamson (2006).
[5] (CP) and (MP) are the standard introduction and elimination rules (respectively) for the conditional. There is a long history of taking introduction and elimination rules for logical connectives to be constitutive; see Peregrin (2008) for an overview.
[6] In what follows, I am assume the standard structural rules and identity (i.e., $\vdash A \rightarrow A$); Tennant's system rejects the structureal rule of transitivity to avoid this problem, but rejecting transitivity seems to be even more counterintuitive (see Chapter Three).

However, in a system with conditional proof, modus ponens (MP) and material modus ponens (MMP) are equivalent—any such system includes either both or neither. Therefore, the following are incompatible, where L is any logic:

(i) L accepts modus ponens.

(ii) L accepts conditional proof.

(iii) The theory consisting of (T-In) and (T-Out) is non-trivial in L.

The upshot is that (T-In) and (T-Out) are incompatible with the intuitive theory of the conditional (i.e., the theory that that conditionals obey modus ponens and conditional proof) no matter what the logic. Non-classical solutions to the aletheic paradoxes are forced to say either that there is no such thing as an intuitive conditional or that it is inconsistent. Either way, the non-classical logics used to solve the aletheic paradoxes are incompatible with what is perhaps the most basic and important element of deductive reasoning—an intuitive conditional.

Unless one is going to treat something as an inconsistent concept, it does not seem as though there is any way out of this mess. Since the aletheic paradoxes (including the problems canvassed pertaining to empirical paradoxes and revenge paradoxes) give us good reason to think that truth is an inconsistent concept, and we have no independent reason to suspect that our logical concepts are inconsistent, it seems best to avoid giving anything like an analysis of truth as part of a unified theory. No analysis of truth is going to do justice both to the primary aletheic principles and to the principles constitutive of our logical concepts. Any analysis of truth will have to, at the very least, respect the constitutive principles. I have made a case for thinking that (T-In) and (T-Out) are constitutive, based on the fact that, without them, 'true' could not play its stereotypical expressive role. An analysis that does not entail the primary aletheic principles is inadequate, while one that does is either inconsistent or incompatible with the constitutive principles of logical connectives.

There is a deeper point here. Again, classical logic is not sacrosanct—perhaps it is wrong. However, the case for adopting a non-classical logic should be made by considering … logic! It should not be made by trying to accommodate the constitutive principles of some other concept. I think this point can be made against any attempt to alter our logic in the face of paradoxes. Sure, we could do so, but this move is a "language-wide" change, which means that arguments having nothing to do with the concept in question (truth in our case) that were previously considered valid, will now turn out to be invalid. For example, imagine a conversation between Martin and Ralph.

Martin: You know, there are infinitely many prime numbers.

Ralph: Why do you believe that?

Martin: Suppose that there are only finitely many primes. Let P be one more than the product of all the primes. Since P is greater than all the primes, it cannot be prime, so let q be a prime dividing P. However, q cannot be any of the primes whose product is P-1; otherwise q would divide the difference between P and the product of the primes, which is 1. But that is impossible. Therefore, there are infinitely many primes.

Ralph: That argument is invalid.

Martin: What!? Why?

Ralph: Because of the liar paradox.

Martin: What are you talking about? Prime numbers have nothing to do with the liar paradox.

Ralph: Reductio arguments are invalid because otherwise the set of all T-sentences for expressively rich languages would be inconsistent. Your argument is missing a premise—either there are infinitely many primes are there are not.

Martin: This argument has been taken to be valid for thousands of years by the greatest minds in history—it is the paradigm of a valid argument!

Ralph: Sorry, but my solution to the liar paradox implies that it is invalid.

This sort of "collateral damage" is often overlooked when thinking about approaches to a paradox. My attitude is that altering our standards of reasoning is always open to us, but the reasons for doing so should be that there is some independent reason to think that there is something wrong with

them (e.g., the paradoxes of implication—for relevance logic—or global anti-realism—for intuitionism). In the case of the aletheic paradoxes, it makes much more sense to think that the problem is not with our accepted ways of reasoning, but rather with the concept of truth. How do we know that we should blame truth instead of blaming logic? Easy—if we try to blame the wrong culprit, we get very clear indications—in the form of revenge paradoxes. Of course, we can keep on adding epicycles to our approach by blaming more and more culprits to avoid the revenge paradoxes as well, but it should be clear to any unbiased observer that this is a desperate move.

That leaves us without much in the way of a view on the nature of truth. I suggested in Chapter One that Davidson's view on the nature of truth is underappreciated and can be thought of as a measurement theory for truth. Moreover, in Chapter Four, I used measurement theory as a way of explaining the relation between the philosophical approaches and the logical approaches to the aletheic paradoxes. It makes sense that these two treatments might go well together, and indeed they do. Of course, that alone does not constitute a unified theory of truth. One needs a formal theory to be interpreted by the measurement theory, and if we want that formal theory to be a theory of truth, then we will run into the same problem as above—namely, if the formal theory is acceptable, then it will have (T-In) and (T-Out) among its axioms, but then it is either inconsistent or incompatible with our accepted ways of reasoning. So switching from conceptual analysis to measurement theory does not seem to offer any help. In Part III of the book, I argue that this pessimism is unwarranted, for a measurement-theoretic account gives us a kind of flexibility to model defective concepts like truth that we do not get with conceptual analysis.

10.1.2 Evaluating Approaches

So far, I have been discussing views on the nature of truth, but now I want to turn to approaches to the aletheic paradoxes. Again, given the considerations in Part II, the outlook is bleak. Although I pointed out problems for a wide range of approaches, both philosophical and logical, I have not addressed every approach. The goal was *not* to discredit all available approaches in order to clear the way for my own; rather, I intended to demonstrate the significance of the four key issues by relating them to some of the literature described in Part I. Although I think that the theory of truth I offer is the best one currently available, this fact is not the primary justification for it.

One of the biggest worries facing approaches to the aletheic paradoxes is accounting for truth's expressive role, which requires (T-In) and (T-Out). The only logical approaches that include these principles are paracomplete and paraconsistent ones, which require non-classical logics and face serious revenge paradoxes. Again, it is important to note that a theory of any predicate that plays truth's expressive role will require a non-classical logic. It is also notable that truth's expressive role requires (on the basis of the arguments given in Chapter Six) that truth predicates be general instead of 'language-specific'. However, every attempt to deal with revenge paradoxes begins by focusing only on language-specific truth predicates. Because of the importation arguments in Chapter Eight, there is no way to avoid revenge paradoxes if one offers a theory of a general truth predicate. So it seems that there is no way for any theory of truth to accommodate truth's expressive role—no theory of a general truth predicate can validate (T-In) and (T-Out) upon pain of inconsistency (or triviality for paraconsistent views). That result *seems* to show that there is no adequate logical approach to the paradoxes whatsoever.

On the philosophical side, the combination of truth's expressive role, the empirical paradoxes, and the Gricean condition poses a major problem for any view that posits some hidden semantic

feature of the truth predicate (e.g., ambiguity or context-dependence). In light of these considerations, none of these approaches is acceptable.

10.1.3 Paradox and Persons

In the opening paragraph of the book, I said that the aletheic paradoxes pose a serious threat to us. That probably sounded like hyperbole at the time, but now I would like to justify it.

> Consider what David Lewis calls our "general theory of persons" in the following passages:
>
> Imagine that we have undertaken the task of coming to know Karl as a person. We would like to know what he believes, what he desires, what he means, and anything else about him that can be explained in terms of these things. We see a two-fold interpretation: of Karl's language, and of Karl himself. And we want to know his beliefs and desires in two different ways. We want to know their content as Karl could express it in his own language, and also as we could express it in our language. Imagine also that we must start from scratch. At the outset we know nothing about Karl's beliefs, desires, and meanings. Whatever we may know about persons in general, our knowledge of Karl in particular is limited to our knowledge of him as a physical system.
>
> Both **Ao** and **Ak** are to be specifications of Karl's propositional attitudes—in particular, of Karl's system of beliefs and desires. **Ao** specifies Karl's beliefs and desires as expressed in our language; **Ak** specifies them as expressed in Karl's language; until we find out what the sentences of Karl's language mean, the two sorts of information are different.
>
> **M**, the third component of our desired interpretation of Karl, is to be a specification, in our language, of the meanings of expressions of Karl's language.
>
> What are the constraints by which the problem of radical interpretation is to be solved? Roughly speaking, they are the fundamental principles of our general theory of persons. They tell us how beliefs and desires and meanings are normally related to one another, to behavioral output, and to sensory input. The general theory of persons serves as a schema for particular theories of particular persons. A particular theory of Karl, for instance, may be constructed by ascribing particular beliefs, desires, and meanings to him. That is, by filling in **Ao**, **Ak**, and **M**.[7]

Lewis goes on to suggest a way of filling in **Ao**, **Ak**, and **M** given only physical facts about Karl. The main point of this passage, for my purposes, is that specifying the meanings of a person's words is

[7] Lewis (1974: 108-111.

an integral part of characterizing that person *as a person*, rather than as a merely physical system. Moreover, specifying the meanings of a person's words goes hand in hand with specifying their propositional attitudes.

If Karl is like any person you have ever met, then Karl possesses the concept of truth and Karl's language contains a truth predicate. One's specification of the meaning of Karl's truth predicate and one's specification of the content of Karl's propositional attitudes involving truth ought to respect the fact that truth obeys (T-In) and (T-Out). That is, for any proposition p, if Karl accepts p (as specified by either **Ao** or **Ak**), then Karl accepts that p is true; likewise, if Karl accepts that p is true, then Karl accepts p. Moreover, Karl's language and thought have the capacity to represent the sentences of Karl's language and the propositions he entertains, and Karl possesses basic logical concepts like negation, conjunction, disjunction, and the conditional (and his language has the associated expressions).

Here is the problem. There is no way to fill in **Ao**, **Ak**, and **M** for Karl. As we have seen, it is impossible to specify the meaning of Karl's truth predicate or the content of Karl's concept of truth so that (T-In) and (T-Out) are true while specifying the contents of Karl's logical terms and concepts accurately (e.g., so that the conditional obeys modus ponens and conditional proof) and respecting the fact that Karl's language contains liar sentences and there are liar propositions that he might entertain.[8] The fact that Karl possesses the concept of truth and his language has a truth predicate seems to render it impossible to treat Karl as a person. Far from being harmless puzzles, the aletheic paradoxes threaten the very idea that we are people, at least if people conform to the general theory of persons Lewis articulates. These considerations lend urgency to our task of constructing an acceptable unified theory of truth.

---

[8] Notice that this problem does *not* presuppose that we are specifying the meanings of Karl's sentences by giving their truth conditions. It is a problem for everyone, no matter what one's prefereed theory of meaning.

10.2 Conditions of Adequacy

On the basis of the discussion so far in Part II, we can formulate conditions for an acceptable unified theory of truth (T):

(1) T implies that (T-In) and (T-Out) are constitutive of truth.

(2) T is compatible with classical logic.

(3) T is a theory of a general truth predicate.

(4) T implies that truth predicates are univocal and invariant.

(5) T does not give rise to revenge paradoxes (of either kind).

(6) T is internalizable.

(7) T enjoys generic theoretical virtues (e.g., consistency, simplicity, modesty, power, and depth).

A few comments on these conditions are in order.

(1) stems from truth's expressive role and the intuition that anyone who asserts p together with 'p is not true' or '~p' together with 'p is true' seems to be misusing the word 'true'. Yes, there are theoretical reasons for the former (e.g., p is not in the business of stating facts), but the point here is that liar sentences do not seem to be in this category, so anyone who asserts either of these combinations where p is a sentence in the same category as liar sentences seems to misunderstand the word 'true'. Notice that (1) does not require that T implies that (T-In) and (T-Out) are *true*; rather, T must imply that they are constitutive. Of course, constitutive principles are almost always taken to be true (indeed, true by definition). I shall have much to say about constitutive principles in the next chapter.

(2) is based on the considerations throughout Chapters Eight and Nine; in particular, it seems that we use logical devices (e.g., exclusion negation) in natural language. Non-classical approaches

to the aletheic paradoxes and the unified theories of truth that incorporate them are incapable of applying to languages with these features. Again, there might be good reason to give up classical logic, but any such reason will have to be based on the norms of correct reasoning, not on trying to solve paradoxes that pertain to a specific concept.

(3) stipulates that truth predicates are not language-specific. As argued in Chapter Six, there is no way to explain natural language truth predicates in terms of language-specific truth predicates. The only alternative is to admit that 'true' applies equally to sentences of English and sentences of other languages.

(4) is justified by the considerations in Chapter Seven on empirical paradoxes; since (i) truth plays an expressive role in our linguistic practice, (ii) the Gricean condition governs our communicative practices, and (iii) there are empirical paradoxes, it does not make sense to think that truth predicates (or the sentences in which they occur) are ambiguous or context-dependent in a way that would obviate the aletheic paradoxes. This issue will come up again in Chapter Fourteen, where I suggest that 'true' is assessment-sensitive in virtue of expressing an inconsistent concept.

(5) should be obvious at this point; the revenge paradoxes studied in Chapters Eight and Nine are as debilitating as they are ubiquitous. Theories of truth that give rise to revenge paradoxes are non-starters—they do not apply to natural languages, they require language-specific truth predicates, and they do not solve the paradoxes.

(6) is justified by the internalizability argument that was the focus of Chapter Nine. If a theory of truth is not internalizable for natural languages, then it is unacceptable. This point, too, should be non-negotiable.

(7) has nothing to do with truth in particular, but rather with philosophical theories in general. They should be *consistent* (inconsistent theories are unacceptable, and switching from classical to

paraconsistent logic does not make them any more so). They should, other things being equal, be *simple*; this is a widely accepted condition on theories. They should, other things being equal, be *modest* in the sense that accepting them should not require giving up other, independent views (this is one major problem with non-classical theories of truth—they go hand in hand with rejecting accepted ways of reasoning even when those have nothing to do with truth). They should be *powerful*, which means that they explain a wide range of issues associated with truth, not just 'true' as it is used in certain specialized circumstances. Finally, they should be *deep*, in the sense that they give us some insight into a diverse set of phenomena that had previously been not as well understood.

In Part III, I offer a unified theory of truth that satisfies these seven conditions. Its central claim is that truth is an inconsistent concept. While there might be some other unified theory that is preferable to the one I offer here, it is clear that it would be in the same category (i.e., it would treat truth as an inconsistent concept). There is simply no other way to satisfy the conditions.

Chapter Eleven is an introduction to inconsistent concepts. It explains what they are, why there are any, and gives several examples. It also contains an extended discussion of analyticity and constitutive principles, which feature prominently in the account of inconsistent concepts.

Chapter Twelve contains several arguments for the claim that truth is an inconsistent concept. It also argues that inconsistent concepts should be replaced for many explanatory purposes, which is what I advocate in the case of truth.

Chapter Thirteen introduces the two concepts I propose as replacements for truth—ascending truth and descending truth. I justify this choice of replacements and discuss some hurdles for them. This chapter ends with a formal theory of ascending and descending truth along with a new kind of

possible-worlds semantics to accompany the theory. An Appendix contains a proof that that the theory is sound with respect to the semantics.

Chapter Fourteen begins with a discussion of the various proposals for a theory of inconsistent concepts, focusing on logic, semantics, and pragmatics. One essential point of the book is that a theory of truth as an inconsistent concept should not employ truth in any kind of explanatory role; instead, it should appeal to the replacement concepts. This chapter contains a detailed suggestion for just such a theory—it treats truth predicates as assessment-sensitive and uses relativized versions of ascending truth and descending truth.

Chapter Fifteen, the final chapter, brings together the theory of ascending truth and descending truth from Chapter Eleven and the theory of truth in Chapter Twelve, and it contains discussions of several issues related to these proposals. Topics include: other paradoxes, truth's connection to other concepts, and several potential objections.

Part III

The Proposal

> The greatest enemy of any one of our truths may be the rest of our truths.
> —William James, *Pragmatism*, p. 38

Chapter 11

Inconsistent Concepts

The central claim of this book is that truth is an inconsistent concept; however, the term 'inconsistent concept' is not found in common usage or in contemporary analytic philosophy discussions, so the first step is to provide an adequate explanation. This chapter introduces the idea, provides several examples, and considers the relation between someone who possesses an inconsistent concept and that concept's constitutive principles.

11.1 Concepts

Before discussing inconsistent concepts, I should say a bit about concepts in general.[1] There are three main views on the nature of concepts:

(i) *Mental representations*: concepts are mental particulars that are the constituents of beliefs and other propositional attitudes. As such, concepts are internal symbols with representational properties.[2]

(ii) *Abstract entities*: concepts are abstract (i.e., non-spatio-temporal) entities that are the constituents of propositions (i.e., Fregean senses).[3]

(iii) *Abilities*: concepts are cognitive abilities or capacities—e.g., the ability to draw certain inferences, classify objects based on perceptions, or react to stimuli in various ways.[4]

The debate about the nature of concepts is rough terrain, and I do not intend to take a stand on this issue.[5] Rather, I do not think that anything I say about inconsistent concepts commits me to one of these views on the ontological nature of concepts. A related issue on which I shall commit myself is

---

[1] My presentation is based on Margolis and Laurence (1999).
[2] Advocates include Fodor (1975, 1987, 1998, 2004), Carruthers (1996, 2000)
[3] Advocates include Peacocke (1992) and Zalta (2001).
[4] Advocates include Evans (1982), Dummett (1993), Brandom (1994), and Millikan (2000).
[5] For background, see the papers in Margolis and Lawrence (1999); for the contemporary debate, see Fodor (1998), Prinz (2002), Murphy (2004), Machery (2009), and Carey (2009).

concept possession; i.e., a view on what it is to possess a concept, but that needs to wait until we get some examples of inconsistent concepts.

11.2 Inconsistent Concepts

A concept is *inconsistent* if and only if its constitutive principles are incompatible. For example, consider the following definition:

(1a) 'rable' applies to x if x is a table.

(1b) 'rable' disapplies to x if x is a red thing.[6]

These rules are *constitutive* for rable in the sense that they determine (in part) the meaning of 'rable'. There are several ways of explaining the relationship between agents and constitutive principles, but a *prima facie* plausible explanation is that anyone who possesses a certain concept accepts that concept's constitutive principles. According to this view, if someone uses 'rable' but does not believe (1a) and (1b), then that person's word 'rable' does not mean rable.[7] However, for reasons I discuss below, a more subtle account of the relation is required.

The definition of 'inconsistent concept' might cause some confusion since the constitutive principles for 'rable' are not *logically* inconsistent. The problem with 'rable' is instead that its constitutive principles have false consequences (e.g., there are no red tables). We could stipulate that an inconsistent concept has constitutive principles that are incompatible with the empirical facts, or we could say that an inconsistent concept has some false constitutive principles.[8] I do not see much difference between these amendments.

---

[6] I use 'disapplies' as an antonym for 'applies'.

[7] One might notice that I am not distinguishing between the meaning of an expression and the concept it expresses; although there is a place for this distinction, it does not affect any of the points I make in this chapter.

[8] Notice that this formulation implies that familiar examples like 'Boche' and 'tonk' express inconsistent concepts (e.g., the constitutive principles for 'Boche' imply that all Germans are cruel and prone to barbarism). See Dummett (1973) on 'Boche' and Prior (1960) on 'tonk'.

A person who employs the concept rable might believe and assert that a red shirt is not a rable and that a brown table is a rable. However, such a person will run into trouble when confronted with a red table because the constitutive principles for 'rable' imply that it both applies and disapplies to red tables. For example, let R be a red table. R is a table; hence, R is a rable. R is red; hence, it is not the case that R is a rable. Thus, R is a rable and it is not case that R is a rable. We have arrived at a contradiction via intuitively plausible steps from intuitively plausible assumptions. Consider another example. Assume for *reductio* that some red tables exist. Let R a red table. The reasoning above shows that R is a rable and R is not a rable. Contradiction. Therefore, no red tables exist. We have proven an obviously false sentence via intuitively plausible steps from intuitively plausible assumptions. If one accepts some basic logical principles and treats 'rable' as univocal and invariant, then it will be difficult to avoid these unacceptable conclusions. Since most people do not believe that any contradictions are true (even ones involving odd concepts like rable) and they believe in the existence of red tables, it seems that adding rable to one's conceptual repertoire corrupts it in a certain way.

I can imagine a reader who has been protesting: there is no such thing as an inconsistent concept! The attempted stipulation above failed to define any term at all because the definition is illegitimate. Therefore, 'rable' does not mean anything, and no conceptual harm has been done.[9]

Hartry Field considers an actual case of conceptual revolution and an excellent example of an inconsistent concept: mass as it occurs in Newtonian mechanics.[10] In Newtonian mechanics, physical objects have a single physical quantity: mass. According to this theory, mass obeys two laws (which are considered equally fundamental): (i) mass = momentum / velocity, and (ii) the mass of an object is the same in all reference frames. We can think of these as constitutive principles for mass.

---

[9] See Wright (1975) and Patterson (2007b).
[10] Field (1973).

In relativistic mechanics, physical objects have two different "kinds" of mass: proper mass and relativistic mass. An object's *proper mass* is its total energy divided by the square of the speed of light, while an object's *relativistic mass* is its non-kinetic energy divided by the square of the speed of light. Although relativistic mass = momentum / velocity, the relativistic mass of an object is not the same in all reference frames. On the other hand, proper mass ≠ momentum / velocity, but the proper mass of an object is the same in all reference frames. Thus, relativistic mass obeys one of the principles for mass and proper mass obeys the other. Since we live in a relativistic universe (i.e., one where momentum over velocity is not the same in all reference frames), mass is an inconsistent concept. That is, before the 20th century, we used a concept whose constitutive principles are inconsistent with the way the world is (i.e., they imply that momentum/velocity is the same in all reference frames).[11]

Although the objection in question (i.e., that there are no inconsistent concepts) might seem convincing for 'rable', it is not plausible to claim that 'mass' is simply meaningless. It has an established use, sentences containing it participate in inferential relations, people use these sentences to express propositional attitudes, etc. To say that such an expression is meaningless severs the concept of meaning from most of the things for which we use it. In addition, if the objection were correct, then when we discovered that the constitutive principles for mass are incompatible, we would have also discovered that our word 'mass' is meaningless. However, it does not seem that an entire community of people can be wrong about whether a word is meaningful. It does not even seem possible to discover that a word one has a history of using is meaningless. Finally, by the Gricean Condition defended in Chapter Six, this move is ruled out. Perhaps that is why there are no examples of this sort of thing actually happening.

---

[11] For more information on this example, see Jammer (2000) and Petkov (2009).

The objection might seem plausible at first because it also seems plausible that if a concept is inconsistent, then anyone who possesses the concept knows that it is inconsistent. However, the 'mass' example should dispel this impression. The rules for the employment of a concept often incorporate features of the environment in which it is used; if the employers of a concept are ignorant or mistaken about some features of their environment, then the concept in question can be inconsistent without their knowledge. No amount of "reflection on their concepts" will inform them that their concept is inconsistent; they have to go out into the world and discover empirical facts to discover the conceptual inconsistency.

Another worry is voiced by John Earman and Arthur Fine, who argue that mass and proper mass are identical, and that we simply had the false belief that mass is momentum/velocity.[12] Fine gives evidence that Einstein himself held this opinion, while Earman provides an argument for it. The basis of his argument is that when Newtonian mechanics and special relativity are written in four-dimensional intrinsic form, mass and proper mass satisfy three fundamental principles that have the same form, whereas relativistic mass does not. Since this formulation of the two theories is so important, we should think that mass just is proper mass and not relativistic mass. Moreover, in contemporary physics, proper mass is considered to be much more fundamental than relativistic mass because the latter is relative to a reference frame while the former is not. Thus, Earman agrees that 'mass' is meaningful, but argues that it expresses a consistent concept. The general objection would be that whenever it seems like the constitutive principles for a concept are incompatible, it must be that some of those principles are not really constitutive.

It seems to me that these are not good reasons to think that there are no inconsistent concepts. Consider a person living in 1850 who denies that mass is momentum / velocity. No one at the time

---

[12] Earman and Fine (1977). They focus on Field's claim that the reference of 'mass' is indeterminate and argue that the reference of 'mass' is the property of proper mass. I am interested in how one might use the points they make to object to the existence of inconsistent concepts.

would say that that person's term 'mass' expressed *mass*. Everyone treated the claim that mass is momentum / velocity as a constitutive principle for mass. It is hard to find a better example of what people at the time would have called an analytic claim. So if Earman is correct, then we could all be wrong about the constitutive principles for our concepts. However, unless we are right about some of these principles, it hardly makes sense to say that we are using the concept in question. For example, if I do not think that grass is green or that it is a plant or that it is growing in my back yard or any of the other familiar claims about grass, then it does not make sense to interpret my word 'grass' as expressing *grass*. Moreover, Earman's point can be handled in stride. It is not surprising that mass satisfies some constitutive principle satisfied by proper mass but not by relativistic mass. We should expect that once we have the conceptual resources to distinguish between the two notions of mass (and all the concomitant mathematical and physical innovations), we can formulate principles like the ones to which Earman appeals. It also is not surprising that proper mass seems like a more important or fundamental notion than relativistic mass. Once an inconsistent concept has been replaced, one of its replacements might come to be more useful than the other. Indeed, that might happen in the case of truth; but it would not cast doubt on the claim that truth is inconsistent.

I want to make several points about inconsistent concepts. First, it is essential to distinguish between inconsistent concepts and unsatisfiable concepts. An *unsatisfiable concept* is one that is consistent but which does not apply to anything. An unsatisfiable concept places incompatible demands on the objects for which it is defined, while an inconsistent concept places incompatible demands its employers. For example,

(2) x is a *squircle* if and only if x is a square and x is a circle.

Squircle is an unsatisfiable concept, but it is not inconsistent. Someone who possesses squircle has no problem employing it. It should be disapplied to everything.[13]

Second, attempting to place the definition of an inconsistent concept in the standard form results in a consistent concept that is either conjunctive or disjunctive. Notice the difference in definitions (1) and (2). (2) prescribes both the application conditions and the disapplication conditions for squircle at once, while (1) has two separate clauses for rable. When considering a definition like (2), it is common to assume that if something is not both a square and a circle, then it is not a squircle. This assumption fits well with consistent concepts because their application conditions and disapplication conditions are disjoint. However, the application conditions and disapplication conditions for inconsistent concepts overlap. That makes it impossible to introduce them with definitions that are in the form of (2). Consider another definition:

(3) x is a *non-red-table* if and only if x is a table and x is not red.

There is a big difference between non-red-table and rable. Non-red-table is consistent and applies to things that are both tables and not red; it disapplies to everything else.

Third, inconsistent concepts characteristically give rise to paradoxes, as evidenced by the arguments above where we derived a contradiction using a red table on one hand, and showed that no red tables exist on the other. It is obvious that something has gone wrong in these arguments, but what? I take it as a condition on any account of inconsistent concepts that it must explain the fallacy in them. It should not be surprising that arguments like this one feature prominently in criticisms of theories that posit inconsistent concepts.

---

[13] I mention the distinction between inconsistent and unsatisfiable concepts because it is a common mistake to assume that inconsistent concepts are merely unsatisfiable. Even some theorists in the inconsistency tradition still make this basic mistake; for example, see Patterson (2010: 16). See Stenius (1972), Chihara (1979), and Yablo (1993b) for discussions of the distinction and the mistake.

The next point is that there is an affinity between inconsistent concepts and partial concepts. A *partial concept* is one that has a limited range of applicability. Some concepts are partial by definition. Here is Scott Soames' example of a partial concept:

(4a) 'smidget' applies to x if x is greater than four feet tall;

(4b) 'smidget' disapplies to x if x is less than two feet tall.[14]

Smidget is a partial concept because it is undefined for entities that are between two and four feet tall. I want to introduce several terms that are helpful in discussing partial concepts and inconsistent concepts. When discussing any partial concept, I assume that there is a set of all the objects that exist; I call it the *domain*. This assumption brings with it several obvious and difficult set-theoretic problems that I will not go into; they do not matter for my purposes. I say that the *range of applicability* of a concept is the subset of the domain to which it either applies or disapplies. The *range of inapplicability* is the complement of the range of applicability. I say that a concept is *inapplicable to an object* if that object falls within its range of inapplicability. 'smidget's range of applicability is the set of objects that are either greater than four feet tall or less than two feet tall. 'rable's range of applicability is the set of objects that are either tables or non-red things. I call the set of things to which a concept applies its *application set* and the set of things to which a concept disapplies is its *disapplication set*. The application sets of consistent concepts are their extensions and the disapplication sets of consistent concepts are their anti-extensions. A concept's *overdetermined set* is the intersection of its application set and its disapplication set. One must be especially careful dealing with negation and partial concepts. 'α is not a smidget' can mean *smidget disapplies to* α or it can mean *smidget either disapplies to* α *or it is inapplicable to* α. The former reads 'not' as choice negation and the latter reads it as exclusion negation.

---

[14] Soames (1999). See Glanzberg (2003) for criticism.

Up to this point I have discussed only inconsistent concepts whose application sets and disapplication sets are not disjoint. However, if a concept's range of applicability and its range of inapplicability are not disjoint, then it is inconsistent as well. For example:

(5a) 'mammamonkey' applies to x if x is a mammal;

(5b) 'mammamonkey' disapplies to x if x is an animal and x is not a mammal;

(5c) 'mammamonkey' is inapplicable to x if x is either a monkey or x is not an animal.

Although the application set and disapplication set for mammamonkey are disjoint, it is an inconsistent concept because its range of applicability and range of inapplicability overlap. A concept can exhibit both types of inconsistency as well. I mark this distinction by saying that an *application-inconsistent* concept (e.g., rable) is one whose application set and disapplication set are not disjoint; a *range-inconsistent* concept (e.g., mammamonkey) is one whose range of applicability and range of inapplicability are not disjoint. I focus primarily on application-inconsistent concepts in the remainder of this chapter, but most of my comments and results hold for range-inconsistent ones as well.

It is possible to define an inconsistent concept that poses no practical difficulty in any physically possible situation. Consider the following definition:

(7a) 'uranicube' applies to x if x is a cube whose volume is at least one cubic mile;

(7b) 'uranicube' disapplies to x if x is composed entirely of uranium.

As I have defined it, uranicube is both partial and inconsistent. Its range of applicability is the union of the set of cubes whose volumes are greater than one cubic mile and the set of things composed entirely of uranium. I assume that, according to the laws of nature, it is physically impossible for a cube of pure uranium whose volume is at least one cubic mile to exist (this is a stock example from philosophy of science discussions). Thus, an employer of uranicube will not run into any practical difficulty while applying it to objects of the actual world. Although uranicube is an inconsistent

concept and an employer of it faces a normative difficulty, he will never have to decide whether to apply it or disapply it to an object in its overdetermination set.[15, 16]

I want to emphasize that in most cases, the inconsistency arises by virtue of the environment in which it is used. The following example illustrates this point and is based on a discussion of Anil Gupta's.[17] Consider a community of people who speak a language that is similar to English except that in their language, the rules for using the expression 'x is up above y' (where 'x' and 'y' are replaced by singular terms) are different. I call the members of this community *Higherians*. Two equally important features of the Higherian's 'up above' talk are that they can perceptually distinguish situations in which one object is up above another (these situations are similar to the ones in which an English speaker would say that one object is up above another), and that they can determine when the ray connecting two objects is parallel to a particular ray that is designated as "Standard Up" (where Standard Up is orthogonal to a tangent plane for the surface of the object on which the Higherians live). 'Up above' applies to an ordered pair <A, B> if either (i) both A and B are constituents of one of the perceptually distinguishable situations (call this the *perceptual criterion*), or (ii) the ray connecting A and B is parallel to Standard Up and A is farther from the surface than B (call this the *conceptual criterion*). 'Up above' disapplies to an ordered pair <A, B> if either (i) A and B are not in the proper perceptually distinguishable relation to one another, or (ii) it is not the case that both the ray connecting A and B is parallel to Standard Up and A is further from the surface than B.

Assume that 'up above' is defined only for perceptible objects and only for objects within the national borders of the Higherian's country. When a Higherian can perceive two objects at the same time then that person can perceive whether they are in the right perceptually distinguishable relation

---

[15] Depending on one's views on counterfactuals and laws of nature, an employer of uranicube might run into trouble by using it in certain subjunctive conditionals or by formulating natural laws with it.
[16] One can construct a concept that is inconsistent by virtue of the natural laws of the world in which it is used (e.g., the pre-relativistic concept of simultaneity). I suggest the term 'nomically inconsistent' for such concepts.
[17] Gupta (1999).

to one another. In addition, every Higherian can determine the ray that connects any two perceivable objects and can determine whether any two rays are parallel. Thus, if a Higherian can perceive object A and he can perceive object B (not necessarily simultaneously), then he can determine whether the ray that connects them is parallel to Standard Up. Assume that the Higherians do not know that their concept is inconsistent because when they can perceive two objects at the same time, they employ the perceptual criterion and when they cannot, they employ the conceptual criterion. Assume also that whether one object is up above another does not depend on any of the Higherians taking them to be in this relation and that the notion of warrant is not relative to anyone's epistemic situation. Finally, assume that there is no difference between the Higherian's idiolects and their common language, that there is no conversational implicature associated with statements containing 'up above', and that the conventions governing 'up above' are common knowledge (i.e., there is no division of linguistic labor for this expression).

If the Higherians live on the surface of a spherical planet, and their nation consists of more than just a single point, then 'up above' is inconsistent. If A and B are two objects that are located some distance from where Standard Up intersects the surface of their sphere and are in the right perceptually distinguishable relation then 'up above' both applies to <A, B> and disapplies to <A, B> because they are in the right perceptually distinguishable relation, but the ray connecting them is not parallel to Standard Up. However, if the Higherians' country is confined to one flat surface of a rectangular solid, then 'up above' is consistent because it is defined only within their national borders. Hence, up above is an empirically inconsistent concept in the case where the Higherians live on the surface of a sphere.

The rules for the employment of a concept often incorporate features of the environment in which it is used; if the employers of a concept are ignorant or mistaken about some features of their environment, then the concept in question can be inconsistent without their knowledge. Again, no

amount of "reflection on their concepts" will inform them that their concept is inconsistent; they have to go out into the world and learn empirical facts to discover the conceptual inconsistency. Consider the history of human inquiry—we (humans) discover false empirical beliefs alarmingly often. Given the degree of our ignorance and error, there is a good chance that many, perhaps most, of our concepts are empirically inconsistent. That sobering thought should lend urgency to the task of constructing an adequate theory of inconsistent concepts and a descriptively complete and descriptively correct semantic theory for inconsistent concepts.

11.3 Policies for Handling Inconsistent Concepts

What should a person do if she discovers that she employs an inconsistent concept? Here I want to discuss three potential answers to this question. To have a concrete example, assume that Troy is a person who has discovered that one of his dearly beloved concepts, concept X, is inconsistent. I do not address the difficult issue of how one discovers such a thing.

11.3.1. The Reinterpretation Policy

Suspicions of conceptual inconsistency are invariably accompanied by efforts to reinterpret the conceptual employment in question. The reinterpretation policy makes this reaction the official strategy for dealing with inconsistent concepts. In my example, one thing Troy could do is reinterpret his past actions and beliefs so that either he never employed X or X is not inconsistent. The first option leaves X alone and posits a consistent concept, Y, as the one Troy was using all along. The second option reinterprets X. The sort of reinterpretation I have in mind here is similar to the maneuvers found in Kripke's rule-following argument,[18] Quine's argument for indeterminacy

[18] Kripke (1982) contains an example where someone reinterprets 'plus' as *quus*.

of translation,[19] and Goodman's new riddle of induction.[20] I do not doubt that such a reinterpretation is possible, but I do question the legitimacy of the reinterpretation policy as a way of dealing with the discovery of an inconsistent concept. It seems to me that it would not be hard to construct situations of inconsistent concept employment that would force very strange and uncharitable reinterpretations. It is far better to have an alternative option ready to hand that can be used in the event of such a discovery. I want to emphasize that I have no argument to show that the reinterpretation strategy is impossible or that people do not do use it. On the contrary, it is the most common response. My qualm is with having it as a general strategy for dealing with inconsistent concepts.

11.3.2 The Containment Policy

According to the containment policy, we should identify the overdetermined items for concept X and treat them in a way so as to render them benign. That means we should determine which objects are in X's overdetermination set and avoid applying or disapplying X to them. We should refrain from asserting sentences associated with these employments of X and avoid having propositional attitudes associated with them as well. (For example, if R is a red table, then we should assert neither 'R is a rable' nor 'R is not a rable' and we should believe neither that R is a rable nor that R is not a rable). A number of prominent philosophers have advocated one form or another of the containment policy for dealing with the aletheic paradoxes.[21] It is also a common view among non-philosophers who are presented with paradoxes that arise in connection with inconsistent concepts.

---

[19] Quine (1960) contains an example where someone reinterprets 'rabbit' as *undetached-rabbit-part* or *rabbit stage*.
[20] Goodman (1955) contains an example where someone reinterprets 'green' as *grue*.
[21] See Popper (1954), Katsoff (1953), van Bentham (1978), Chihara (1979, 1984), Yablo (1985, 1989).

I have several reservations about the containment policy. First, it can turn out to be difficult or impossible to avoid either uttering paradoxical sentences or entertaining attitudes toward the propositions they express (if such propositions exist). Consider the case of truth (a good candidate for an inconsistent concept). As we saw in Chapter Seven, the knowledge required to determine whether a particular sentence containing 'true' is paradoxical goes far beyond that which any normal speaker has in everyday situations, and in some cases it is beyond anyone's knowledge. Of course, sometimes we can figure out whether a sentence is paradoxical, but in general, determining whether any given sentence is paradoxical is incredibly difficult because it can depend on the semantic properties of sentences to which we no longer have access. Likewise, in many cases it is far too demanding to restrict people from applying or disapplying an inconsistent concept to overdetermined items.

My biggest concern about the containment policy is that does not get to the root of the problem. The problem is that an inconsistent concept is defective. If the concept is useful, then it should be replaced; otherwise it will cause serious problems for those who use it. Without replacements, it is difficult or impossible to identify the situations in which it can be safely used. However, a proponent of the containment policy tries to deal with the problems that result from the continued employment of the inconsistent concept. It treats the symptoms instead of the disease.

I think that there is a limited place for the containment policy in an effective paradoxicality response program. The containment policy is an important first step in the eventual replacement of an inconsistent concept. While we (humans) are deciding what changes to make to our conceptual repertoire, the containment policy is the best one for the interim. Nevertheless, we must actually go on to alter our concepts so as to remove the inconsistency.

11.3.3. The Replacement Policy

I advocate the replacement policy for inconsistent concepts. According to it, we should determine the best way to replace an inconsistent concept with consistent ones, and do so. That means we should stop employing an inconsistent concept and begin employing a different concept or group of concepts. One difficult issue with pursuing the replacement policy is the choice of replacement(s). I say very little on how to go about choosing a replacement for an inconsistent concept and I am not sure that it is possible to provide a strategy that will result in the best replacement each time. It seems to me that judging the best replacement involves the weighing of factors that are not easily quantifiable, as in considerations of simplicity, economy, and charity.

11.4 Possessors and Principles

Here is a worry I hear often:

> There are no inconsistent concepts. Any attempt to introduce a term that behaves according to incompatible rules fails to introduce a meaningful term at all. Thus, it is impossible that a term obeys incompatible rules of employment. One reason for thinking this is that interpretation requires one to use the logic one endorses when interpreting another. Thus, it is inappropriate to ever attribute an inconsistent concept to someone, since the interpreter would have to attribute something that defies the logic she endorses.[22] Moreover, even if one could introduce a term that obeys incompatible rules, it would be overdetermined for every item, so it would be unemployable.[23]

First, the claim that we interpret others as if they endorse our logical standards is simply false. If it were true then there would be no distinction between criticizing someone for failing to follow an inference rule she endorses and criticizing someone for endorsing the wrong inference rule. It is obvious that there is such a distinction and it plays an important role in philosophical discussions. Second, charity can cut both ways. One might simply introduce an inconsistent concept, begin using it, and describe it as inconsistent (I did this with the concept rable). It seems to me that it

---

[22] One can find a similar objection in Stebbins (1992).
[23] See Gupta and Belnap (1993: 13-15) for this objection; see also Chihara (1984) for discussion.

would be quite difficult to go on interpreting someone who does this as if they had misunderstood their own stipulative definition and their claims about it. Indeed, one might give an account of all the relevant factors in charitable interpretation and present two situations, one in which the weighted sum of all the factors is higher than that of the second, while in the first one attributes an inconsistent concept, but in the second one does not. The point here is that attributing an inconsistent concept is sometimes the most charitable thing to do. No matter what constraints one imposes on charitable interpretation (except of course, a conceptual consistency constraint), there will be situations in which it is more charitable to attribute an inconsistent concept. I agree that a major problem for a theory of inconsistent concepts is showing that a concept can be both inconsistent and employable (i.e., not overdetermined for every item). I take up this task in Chapter Fourteen.

One problem raised by inconsistent concepts is how they could be possessed. By far the most popular theory of concept possession is *concept pragmatism*, which Jerry Fodor characterizes in the following way:

> *The* characteristic doctrine of 20th Century philosophy of mind/language … was that *concept possession is some sort of dispositional, epistemic condition.* Maybe it's some sort of "knowing that"; or maybe it's some sort of "knowing how"; or maybe it's a bit of both. In any case, "knowing", "believing" and the like must come into the story somewhere, and what you have to know in order to have a concept ipso facto constitutes the concept's content.[24]

The central claim of concept pragmatism is that if an agent S possesses a concept C, then S knows something, or knows how to do something or believes something and this feature of S constitutes S's possession of C. Let whatever epistemic or cognitive capacities S must have in order to possess C be C's *possession conditions*.

As Fodor mentioned in the passage above, one very popular view on possession conditions is that they involve belief. Let us explore this idea. Assume that, for any concept C there is some

[24] Fodor (2004); note that Fodor rejects concept pragmatism.

proposition that p such that believing that p is a necessary condition for possessing C. This account is clearly inadequate since there are many concept/principle pairs that fail this condition. As discussed above, the concept of mass (as defined in Newtonian mechanics) has the following constitutive principles:

(3a) an object's mass = its momentum/its velocity

(3b) an object's mass is the same in all reference frames

Of course, it follows from these two principles that momentum/velocity is the same in all reference frames. But we all know that this is not correct—special and general relativity imply that momentum/velocity is relative to a reference frame. So, despite the fact that I possess the concept of mass, I do not accept or believe both (3a) and (3b). Paul Boghossian summarizes the point in the following passages:

> The concept itself should not be designed in such a way that, only those who believe a certain creed are allowed to possess it.
>
> You don't ever want the *possession conditions* for a concept to foreclose on the possible falsity of some particular set of claims about the world, if you can possibly avoid it. You want the possessor of the concept to be able coherently to ask whether there is anything that falls under it, and you want people to be able to disagree about whether there is.[25]

There are several options for dealing with this problem.

Boghossian suggests that we pursue an idea proposed by Frank Ramsey of thinking of constitutive principles as conditionalized—following this suggestion we arrive at these constitutive principles for mass:

(4a) if objects have mass, then an object's mass = its momentum/its velocity

(4b) if objects have mass, then an object's mass is the same in all reference frames

---

[25] Boghossian (2003a: 245) and Boghossian (2003a: 246); see also Williamson (2003, 2006) for a similar point.

If we say that (4a) and (4b) are the constitutive principles for mass, then one can believe them (and thereby possess the concept of mass on some views) without believing their consequents. So a person who thinks that mass is defective can still believe these two conditionals (since, presumably, the person rejects their antecedents and their consequents).[26] However, I still use 'mass'. What are the principles according to which I should use it? Well, presumably, they are (4a) and (4b) (plus possibly others). But I do not really use it according to them since I deny their antecedents. I use it, in certain circumstances, according to the consequents of these principles. So the constitutive principles of mass do not really say anything about how I use it. It seems to me that the conditionalization approach does not work well for useful yet defective concepts like mass.

Matti Eklund suggests that being disposed to believe that p is a necessary condition for possessing C.[27] There are several problems with this view. Let us consider a person, Otto, who has come to possess the concept of rable and realizes that it is an inconsistent concept. Assume also that Otto does not like the idea of accepting contradictions because he thinks that rational agents should avoid doing such things if at all possible. What can Otto do? If Eklund is right, then Otto needs to get rid of his newly acquired disposition to accept the constitutive principles for 'rable'. Let us assume that he does rid himself of the offending dispositions. Now that he is no longer disposed to accept its constitutive principles, in what sense does he still possess the concept? If Eklund is right that *being disposed to accept* is the relation between concept possessors and constitutive principles, then Otto no longer possesses the concept once he has eliminated those dispositions. Perhaps Eklund would amend his view so that Otto still possesses the concept because he *used to have* the dispositions. If that is correct, then it does not seem like anyone could ever lose possession of a concept.

---

[26] Boghossian (2003a); see also Ramsey (1929) and Lewis (1970a).
[27] Eklund (2002a, 2007).

Another option is to invoke a subpersonal attitude. One can say that we cognize the principles—we feel primitively compelled to accept them even if we do not in fact accept them.[28] The analogy is with visual illusions—it still seems that the lines in the Müller-Lyer diagram are different lengths even though one does not believe it. The problem with this suggestion is that the principles governing mass do not seem true—there is no sense in which I am primitively compelled to accept them. Rather, I think that they are approximately true in certain circumstances. That is a big difference, and it suggests that the subpersonal view is inadequate.

I suggest that instead of using cognitive relations like belief to explain the relation between concept possessors and constitutive principles, we should consider epistemic relations. Of course, knowledge is too strong since it implies belief. However, the notion of *entitlement*, which is introduced by Tyler Burge and taken up by Boghossian and Crispin Wright, is perfect for the job.[29, 30] Burge claims that justification and entitlement are kinds of warrant, and he offers the following characterization:

> The distinction between justification and entitlement is this: Although both have positive force in rationally supporting a propositional attitude or cognitive practice, and in constituting an epistemic right to it, entitlements are epistemic rights or warrants that need not be understood by or even accessible to the subject.[31]

Entitlement is defeasable—if an agent finds some reason to doubt the proposition in question, then the warrant is lost. Entitlement is also non-evidential; as Wright puts it: "there is a distinction between being rationally entitled to proceed on certain suppositions and the having of evidence that

---

[28] See Patterson (2007a).

[29] See Burge (1993), Wright (2004a, 2004b), and Boghossian (1996, 2003b). There are subtle differences between the way these theorists use 'entitlement' (Wright discusses three different varieties—the "cognitive project" variety seems most relevant to my discussion), but they do not matter for my purposes.

[30] Two other theorists come close to suggesting entitlement as the relation between possessors and principles: see Eklund (2005a:50), which discusses default acceptability, and Ray (2002: 166-167), which invokes subtle conceptual warrant.

[31] Burge (1993: 458).

those suppositions are actually true."[32] Being entitled to some proposition does not require having evidence for it.[33]

Someone who possesses a certain concept is *entitled* to the constitutive principles of that concept. That is, the person is warranted in believing the constitutive principles provided he or she has no reason to doubt them. However, one can be entitled to a principle without believing it, and entitlement is defeasible. Thus, if a person has evidence to the contrary, then he or she is not warranted in believing the principle. In most cases, concept possessors will not only be entitled to the constitutive principles in question, they will also accept them, since they will not have any reason to doubt them. Only in cases where a person has evidence that his or her concept is inconsistent would the person reject its constitutive principles. If one knows that a concept is inconsistent, one will reject one or more of the concept's constitutive principles. Instead of accepting the concept's constitutive principles, a person in this situation will probably accept similar principles that permit exceptions. For example, one might accept that *in non-relativistic situations*, the mass of an object is the same in all reference frames.[34]

Since entitlements are defeasible, we need to make a choice about how to use 'entitlement'. Let us say that a person is entitled to p but discovers some reason to doubt that p. Is the person still entitled but no longer warranted in believing that p, or is the person no longer entitled to believe that p? It seems that the latter sounds better. If so, then we need a term that means 'would be entitled provided one had no countervailing evidence'. Call this *quasi-entitled*. Now, we can say a

---

[32] Wright (2004b: 167).

[33] To accept entitlement as the relation between concept possessors and constitutive principles does not commit one to accepting the doctrines Burge, Boghossian, or Wright defend. Indeed, it seems to me that accepting entitlement as the possessor-principle link runs counter to some of these projects.

[34] Thinking of the relation between an agent and constitutive principles as one of entitlement represents a change from the account given in Scharp (2008) in which I distinguish between concept possession and concept employment, and I argued that someone who *employs* a concept is *committed* to its constitutive principles.

subject S possesses concept C if and only if S is quasi-entitled to the constitutive principles for C.[35] Even if speakers disagree on the constitutive principles for a given concept, this definition will work as long as they think there is a fact of the matter as to what the constitutive principles are. If they do not think there is a fact of the matter, then the definition will need to be relativized to speakers, hearers, or linguistic communities. I will not worry about this complication.

This view of constitutive principles is somewhat different from the received view. First, constitutive principles need not be true. That is a welcome result since inconsistent concepts have constitutive principles that could not all be true. Moreover constitutive principles in my sense are not cut out to explain analyticity, apriority, or necessity since these notions are all factive (if coherent). Second, one need not believe a concept's constitutive principles to possess that concept. A person who possesses a concept and has no reason to think it is defective will almost certainly believe its constitutive principles, but another person who possesses the concept in question and suspects that it is defective will probably not believe all its constitutive principles and might even believe the negation of one or more of them. Third, constitutive principles on this reading still serve as a guide to interpretation. If P is a constitutive principle for concept C then P contains a word W that typically expresses C. If a person denies P then that is good evidence that that person's word W does not express C. However, it is not conclusive evidence since the person might think that the concept in question is defective. Fourth, whether a principle is constitutive for a concept is a status the principle can have or lack and it is a status that reasonable people can disagree about.

We can understand what it is to treat a principle as constitutive for a certain concept by the role this status has in interpretation. Hearers take everyone to be committed by default to the constitutive principles of concepts being deployed (unless they know otherwise). When a speaker denies a principle that the hearer takes to be constitutive of a concept that the hearer takes the

[35] Thanks to Michael Miller on this point.

speaker to be using, this is an interpretive "red flag". After all, the hearer will take the speaker's denial of what the hearer takes to be a constitutive principle as evidence that the hearer is misinterpreting the speaker. So, when a speaker denies one of these, the hearer has to either take the principle off the conversational record or change the interpretation of the word in question. Either way, when a speaker denies a constitutive principle, the hearer is not just engaging in business as usual—adding something to the record. It can be difficult to determine which of these two options is the right one to pursue—usually the speaker has to know that the principle is typically taken to be constitutive in order for the hearer to take it off the record—denying a principle that is taken to be constitutive is good evidence that the person's word does not express the concept in question. The hearer will often want to know that the speaker has good reason either to treat the principle in question as not constitutive or to deny it even though it is constitutive. Taking the concept to be inconsistent is good reason to deny it even though it is constitutive.

Knowingly using words in a way that violates what are taken to be constitutive principles by most of the members of one's linguistic community without indicating that one is using them this way is simply failing to cooperate. So, the hearer has the "nuclear" option of treating the speaker as uncooperative and refusing to engage in communication any further. However, as long as a hearer is confident that the speaker is cooperating, the hearer will try to determine whether his interpretation of the speaker's word is correct. Notice that if a speaker does not understand that most members of his linguistic community treat some principle as constitutive for a given concept, then that is evidence that the speaker is not competent with that word.

A hearer will consider: (i) the speaker's recognition that the principle is taken to be constitutive for the concept in question by other members of the community, (ii) the speaker's reasons for either saying that the principle is not constitutive for that concept or saying that the concept is legitimate, (iii) the speaker's recognition that the concept is taken to be legitimate by other members of the

community, and (iv) the speaker's recognition that the word in question is taken to express the concept in question by other members of the community. If any of these fail, chances are that the hearer will conclude that the speaker is not competent with the word in question or the concept in question or both.

Since constitutive principles need not be true, one of the major problems associated with them, i.e., that they saddle one with a pernicious distinction between analytic and synthetic, is avoided. However, a genuine problem remains: what is the source of the entitlement that is the link between possessors and principles? It seems to me that these entitlements stem from the epistemic nature of the concept's possession conditions. Recall that the received view on possession conditions is that they involve an agent's knowledge, beliefs, or abilities. Consider, for example, Christopher Peacocke's influential discussion of possession conditions as they apply to the concept square:

> For a thinker to possess the concept square (C):
> (S1) he must be willing to believe the thought Cm_1 where m_1 is a perceptual demonstrative, when he is taking his experience at face value, the object of the demonstrative m_1 is presented in an apparently square region of his environment, and he experiences that region as having equal sides and as symmetrical about the bisectors of its sides …
> (S2) for an object thought about under some other mode of presentation m_2, he must be willing to accept the content Cm_2 when and only when he accepts that the object presented by m_2 has the same shape as perceptual experiences of the kind (S1) represents objects as having.[36]

This example is merely meant to illustrate what possession conditions might be like; although Peacocke's account seems right in this case, nothing hinges on this. The main point is that an agent comes to possess of a concept by acquiring certain practical abilities when it comes to thinking, judging, and perceiving.

For what it is worth, I prefer an interpretive approach to possession conditions, which is based on Donald Davidson's hypothetical radical interpreter (described in Chapter One).[37] On such a view, an agent S possesses a concept C iff a unified theory of S's beliefs and desires and the

---

[36] Peacocke (1992: 108).
[37] Davidson (1973). See Peacocke (1992: ch. 1) for discussion.

meanings of the sentences in S's language entails that some of S's beliefs or desires have C as a constituent. Of course, the justification for this particular unified theory of S comes in the form of S's interaction with items in the world shared by S and the radical interpreter.

I want to emphasize that nothing turns on accepting the interpretive view of possession conditions. Instead, the important point is that, acquiring a concept takes effort on the part of the agent. No matter whether one accepts Peacocke's theory or Davidson's or some other theory of possession conditions, as long as a necessary condition of possession conditions is some kind of cognitive achievement on the part of the agent in question, this is enough to ground the agent's entitlement to the constitutive principles of the concepts thereby possessed. The agent's process in acquiring the abilities that leads up to the agent's possession of some concept institute the agent's entitlement to that concept's constitutive principles. That is the heart of a concept-pragmatist theory of concept possession, and it remains intact even when one admits that some concepts are inconsistent.

To summarize: we can understand what constitutive principles are by understanding the status of being a constitutive principle, and we can understand this status by understanding its role in our practice of interpretation. I find scorekeeping pragmatics especially useful in this regard (as evidenced by the above discussion). Moreover, the relation between a concept possessor and the possessed concept's constitutive principles is entitlement. Finally, the possession conditions for the concept in question involve some kind of cognitive achievement or ability, which underwrites the entitlement to the concept's constitutive principles.

Chapter 12

Reasons for Replacement

In the last chapter, I introduced inconsistent concepts. In this one, I argue that truth is an inconsistent concept, and that it should be replaced. In the following two, I present a theory of the replacements for truth, and a theory of truth itself.

There are four main arguments in this chapter for the claim that truth is an inconsistent concept. The first and second arguments turn on untoward consequences of treating truth as a consistent concept. They focus on reasoning and meaning, respectively. The third is an inference to the best explanation—i.e., that truth is inconsistent is the best explanation for the aletheic paradoxes and the revenge paradoxes. The fourth follows the general strategy of arguing that the best theory of truth implies that truth is an inconsistent concept. The reason for this strategy is simple. The most intuitive argument for the claim that truth is an inconsistent concept is that (T-In) and (T-Out) are constitutive of truth and the liar paradox shows them to be inconsistent (given some basic logical principles and the availability of syntax). That is the kind of argument used to show that 'rable', 'mass', and 'up above' are inconsistent. Of course, every approach to the liar paradox except those in the inconsistency category is designed to avoid this very argument. Thus, my fourth argument is intended to bring out the problems with attempts to avoid the intuitive argument for inconsistency.

12.1 The Expressive Argument

In Chapter Six, we saw that truth serves an expressive role as a device of generalization. I argued there that if truth is a device of generalization, then truth obeys the aletheic intersubstitutability principle (at least). That is, if we can use truth to formulate generalizations like 'a rational agent

ought to believe a proposition only if that proposition is true', then it must be that substituting p and 'p is true' in extensional contexts preserves truth value. That argument presupposes that truth is a consistent concept. Moreover, we also saw in Chapter Six that if truth obeys the aletheic intersubstitutability principle then only a non-classical approach (either paracomplete or paraconsistent) is acceptable. We just saw (in Chapter Ten) that non-classical approaches have to reject either conditional proof (since no structural logic with conditional proof is compatible with the aletheic intersubstitutability principle[1]), or transitivity of logical consequence (since the only non-trivial logics that allow both the intersubstitutability principle and conditional proof are substructural). Putting all this together, we get that if truth is a consistent concept and serves as a device of generalization, then either conditional proof or transitivity of consequence is unacceptable. Remember, everyone thinks truth serves as a device of generalization, so denying that is not a viable option. Thus, a fundamental choice for anyone proposing a theory of the nature of truth, a philosophical approach to the aletheic paradoxes, or a logical approach to the aletheic paradoxes is: accept that truth is an inconsistent concept or give up either conditional proof or the transitivity of logical consequence.

The following is a summary of the first argument for treating truth as an inconsistent concept:

(i) If truth is a consistent concept and truth is a device of generalization, then the aletheic intersubstitutivity principle holds.

(ii) Truth is a device of generalization.

(iii) If the aletheic intersubstitutivity principle holds, then either conditional proof or the transitivity of logical consequence must be rejected.

(iv) Conditional proof and the transitivity of logical consequence should be accepted.

∴ (v) Truth an inconsistent concept.

---

[1] Again, assuming that the conditional obeys identity (i.e., $\vdash A \rightarrow A$).

Obviously, this argument is only as strong as premises (ii) and (iv), with (iv) being the one that is most likely to be given up. The real point here is that treating truth as a consistent concept has massive costs for our reasoning practices. As we will see in the next two chapters, if we accept the unified theory of truth I offer, then we need not give up any logical principles whatsoever since my account is compatible with classical logic.

12.2 The Meaning Argument

Back in Chapter Five I argued that the aletheic paradoxes pose a serious threat to the truth-conditional theory of meaning. The problem is that anyone who accepts a truth-conditional theory of meaning is committed to giving truth conditions for all meaningful sentences of a language, but virtually all approaches to the aletheic paradoxes are restricted to avoid revenge paradoxes. When a truth-conditional theory of meaning is applied to a language that contains paradoxical sentences, it has to be paired with an approach to the aletheic paradoxes, otherwise, it would be straightforwardly inconsistent. However, owing to the restrictions on approaches to the aletheic paradoxes, there will be meaningful sentences that cannot be given truth conditions. The upshot is that since any view on which truth is a consistent concept is bound to be restricted to avoid revenge paradoxes, accepting that truth is a consistent concept is incompatible with accepting a truth-conditional theory of meaning. The problem affects most of formal semantics (dynamic semantics[2] and game-theoretic semantics[3] aside) insofar as it purports to explain meaningful discourse in general.

However, if one treats truth as an inconsistent concept, then one can save truth-conditional semantics. The details are given in Chapters Fourteen and Fifteen, but the basic idea is that the

---

[2] See van Eijck and Visser (2010) for an overview.
[3] See Hodges (2009) for an overview.

unified theory of truth I offer faces no revenge paradoxes of any kind and so does not need to be restricted in any way. It can be paired with a truth-conditional semantics, and the combination applies to any meaningful sentence whatsoever.

The following is a summary of the second argument for treating truth as an inconsistent concept:

(i) If truth is a consistent concept, then there are meaningful sentences that cannot be treated by truth-conditional semantics.

(ii) If there are meaningful sentences that cannot be treated by truth-conditional semantics, then truth-conditional semantics is unacceptable.

(iii) Truth-conditional semantics is acceptable.

∴ (iv) Truth is an inconsistent concept.

Again, the force of the argument rests on an assumption that is seemingly independent of truth—premise (iii). The acceptability of truth-conditional semantics comes from linguistics, where it is firmly entrenched and has many explanatory and predictive successes, and the modest attitude toward the relation between philosophy and the sciences defended in Chapter Six as part of the discussion of the Gricean Condition.

Hold on! If truth is an inconsistent concept, then how can it be used legitimately in a truth-conditional theory of meaning? The central theme of this whole book is that truth ought to be replaced for certain purposes, and it sure seems like giving truth-conditions for paradoxical sentences has to be one of those purposes. So it does not seem like one can accept that truth is an inconsistent concept and accept truth-conditional semantics. Moreover, if that is right, then it seems as if I have violated my own modest attitude in using philosophical considerations to reject an established tenet of the sciences.

Let us revisit a passage from Hofweber on the modest attitude: "To have the modest attitude is not to have science worship. One can have the modest attitude and be critical of various sciences."[4] As I said in Chapter Six, linguists' assumption that propositions are sets of possible worlds can be overruled by philosophers. However, in the case of propositions, these are really just empirical considerations—the claim that propositions are sets of possible worlds gets the wrong results. Moreover, semantic theories in linguistics deliver the same results when we replace that assumption with the view that propositions determine sets of possible worlds. The same goes for the point about truth-conditional theories of meaning. I am *not* saying that truth-conditional theories of meaning are unacceptable because they are incompatible with my favorite theory of truth (that would be like Hartry Field's view—he thinks that truth-conditional theories of meaning are unacceptable, but he only thinks this because of his philosophical commitments, specifically, his disquotationalism).[5]

The truth-conditional theory of meaning has tremendous explanatory power, but explanatory power does not trump empirical inadequacy, at least when there is an alternative on the table (consider Newtonian mechanics and the procession of the perihelion of Mercury). Fine, but in order for this line of argument to work, the replacement theory would have to have as much explanatory power as truth-conditional semantics, right? Right. So the meaning argument only works for prescriptive theories that have this feature. And mine does. Ascending and Descending Semantics (presented in Chapter Fifteen) reduces to truth-conditional semantics when the distinction between ascending and descending truth is negligible. Just as general relativity reduces to Newtonian mechanics when the distinction between relativistic mass and proper mass is negligible. So it has as much explanatory power as truth-conditional semantics. Therefore, far from violating

---

[4] Hofweber (2009: 263).
[5] See Field (1994a).

my modest attitude toward the sciences, my inconsistency theory of truth turns on the legitimacy of truth-conditional semantics—it is worth saving. However, it cannot be saved if truth is a consistent concept. Obviously it cannot be saved in its current form if truth is an inconsistent concept either. But it can be preserved in the new theory in the way that Newtonian mechanics is preserved in relativistic mechanics—that is, if one accepts the prescriptive theory in the next chapter.

12.3 The Abductive Argument

The third argument is that if one decides to treat truth as an inconsistent concept, then one has available a satisfying explanation of the current situation in truth studies. That is, one can explain why other theories of truth face revenge paradoxes, both inconsistency problems and self-refutation problems. No other theory of truth has managed to do this.[6]

The explanation for why theories of truth that imply truth is a consistent concept face revenge paradoxes or self-refutation problems is straightforward. Our concept of truth is inconsistent in the sense that its constitutive principles are incompatible. That is, there are objects that these rules classify as both true and not true. (In the last chapter, I called the set of such objects the *overdetermination set* for truth.) All the paradoxical sentences considered so far are members of the overdetermination set for truth.[7] Any theory of truth that implies that truth is a consistent concept and that includes these principles is inconsistent and can be rendered consistent only by restricting it. If a theory of truth implies that some of the sentences in the overdetermined set for truth are gaps, then the theory's fate depends on which of these sentences it classifies as gaps. Recall that many of the members of the overdetermination set for truth are truth attributions, and no matter

---

[6] See Glanzberg (2005), Cook (2007), and Field (2007) for the only alternative explanations of which I am aware; I consider them in a reply to an objection below.

[7] It seems to me that truth-tellers (e.g., sentence τ, 'τ is true', is a truth-teller) are in the *under*determination set for truth, but these sentences are not paradoxical and none of my claims or arguments hang on this opinion.

what truth status (e.g., true, false, gappy, etc.) one assigns them, they are consequences of the assignment. No matter whether one's theory of truth classifies these paradoxical sentences as true, false, or gappy, some of these paradoxical sentences are consequences of the theory. Thus, if a theory of truth implies that all the sentences in the overdetermined set for truth are gaps, then the theory implies that some of its consequences are gaps. On the other hand, if a theory of truth does not classify some of these sentences as gaps, then the truth rules imply that they are both true and not true. On the first option, the theory is self-refuting, while on the second, it faces an inconsistency problem. Therefore, both types of revenge paradoxes can be explained if we assume that truth is an inconsistent concept.

In Chapter Eight, I argued that theories of truth that validate the primary aletheic principles face revenge paradoxes. If we admit that truth is an inconsistent concept, then we can explain why this occurs. Therefore, by accepting that truth is an inconsistent concept, we arrive at a deeper explanation for why theories of truth that validate the truth rules fail are unacceptable.

The following is a summary of the third argument for treating truth as an inconsistent concept:

(i) If we assume that truth is an inconsistent concept, then we can explain the presence of the liar paradox and the presence of revenge paradoxes.

(ii) The inconsistency explanation of the liar paradox and the revenge paradoxes is better than any of the others.

∴ (iii) Probably, truth is an inconsistent concept.

Only by admitting that truth is an inconsistent concept can we satisfactorily explain the most significant feature of our long battle with the liar paradox.

There are at least two well-developed explanations of the revenge paradox phenomenon, one from Hartry Field and one from Michael Glanzberg. Why is the explanation I offer superior to the ones they offer?

I begin with Glanzberg, who offers a context-dependence approach to the liar. However, instead of claiming that truth predicates are explicitly context dependent, Glanzberg argues that sentences that contain truth predicates display an implicit context dependence that is due to the presence of quantification. Glanzberg offers a theory of background domains of propositions for the quantifiers involved, which includes an infinite hierarchy of domains and no "biggest" domain.[8]

If one accepts Glanzberg's theory, then one has to admit that there is no unrestricted quantification. Indeed, one has to accept that we can express the notion of truth-in-a-context and we can even quantify over contexts to a limited degree, but we cannot express an unrestricted notion of truth. "One way or another, hierarchical theories all require that speakers cannot in any one instance express the entirety of a unified concept of truth."[9] He argues that the sort of fragmentation we see in our concept of truth is familiar to us (i.e., it occurs in the concept of mathematical proof as well) and that it occurs because truth fails to be closed under reflection.

Glanzberg's defense of this feature is based on the idea that any characterization of truth permits one to reflect on the truth of the characterization, and this reflection both shows that the initial characterization is inadequate and points the way toward a stronger one. This process of reflection is unending; hence the infinite hierarchy of contexts.[10] The motivation for this view comes from what has been called the strong liar reasoning. Consider the following sentence:

(1) (1) is not true.

---

[8] Glanzberg (2001, 2004, 2005).

[9] Glanzberg (2004: 289).

[10] This view about the relation between reflection and revenge seems to stem from some of Kripke's remarks: "Such semantical notions as 'grounded,' 'paradoxical,' etc. belong to the metalanguage. This situation seems to me to be intuitively acceptable; in contrast to the notion of truth, none of these notions is to be found in natural language in its pristine purity, before philosophers reflect on its semantics (in particular, the semantic paradoxes). If we give up the goal of a universal language, models of the type presented in this paper are plausible as models of natural language at a stage before we reflect on the generation process associated with the concept of truth, the stage which continues in the daily life of nonphilosophical speakers" (Kripke 1975: 714).

The indeterminacy approach to the liar implies that (1) is indeterminate. We know that if a sentence is indeterminate, then it is not true. Thus, the indeterminacy approach implies that (1) is not true. Hence, the indeterminacy approach implies that '(1) is not true' is true; therefore, it implies that (1) is true.[11] It is by reflection on the way the approach classifies (1) that drives us to conclude that (1) is true after all. The claim, 'if the indeterminacy approach implies that p, then p is true', is similar to what is called a *reflection principle*. It states something about a formal theory that cannot be captured by the formal theory on pain of contradiction.[12]

Glanzberg argues that one can begin with a basic formal theory of truth, formulate a reflection principle for that theory, which illustrates the theory's inadequacy, and arrive at a new formalization of the theory that effectively incorporates the reflection principle. We can continue this process to arrive at a transfinite hierarchy of formal theories of truth, which is analogous to the hierarchy of contexts for truth attributions. He claims that truth is a *Kreiselian concept* in this sense: any formal theory of truth points the way to a stronger formal theory, and the process of theory construction is unending.[13]

Glanzberg's point is that what seem to be revenge paradoxes are really just the effects of the Kreiselian aspect of truth. A theory of truth should not be expected to treat as true the claim that its consequences are true. Nor should a theory of truth be found lacking if the result of conjoining a reflection principle to it results in an inconsistent theory. These phenomena are just consequences of the fact that truth is a Kreiselian concept.

There are several places at which I disagree with Glanzberg's analysis. The first is that I do not find the strong liar reasoning compelling. Because the strong liar reasoning involves a move from 'p is indeterminate' to 'p is not true', (1) should be read as:

---

[11] See Burge (1979a) for discussion of the strong liar reasoning.
[12] See Feferman (1991) for an overview of reflection principles.
[13] Glanzberg (2005).

(1′) is Xnot weak true.[14]

We already know that indeterminacy approaches have troubling handling sentences like this, but the trouble has nothing to do with reflection on how the theory of truth in question classifies (1′). In fact, most indeterminacy approaches to the liar are based on fixed-point constructions and so have no consequences for sentences like (1′) at all. Thus, there is no reason to think that one derives a contradiction only by assuming that the theory implies that (1′) is indeterminate. Therefore, the view that reflection on the dictates of a theory of truth has any special role to play in reasoning about the status of paradoxical sentences seems to be a mistake.

A second issue is that there seems to be no reason to think that the concepts described by each of the formal theories Glanzberg identifies have anything in common. In particular, I see no reason to think that they are all "more or less" theories of truth. However, Glanzberg claims that each of the formal theories provides a rough characterization of the unified concept of truth. The problem with this view is that on Glanzberg's own account, it is impossible to express the unrestricted notion of truth that each of these theories is supposed to describe. Thus, we have this concept of truth, but we can never actually use it. That sounds fairly counterintuitive. Moreover, if Glanzberg is right, then it is impossible to arrive at a theory of truth that correctly and completely describes our concept of truth. The best we can achieve is stronger and stronger theories that are always lacking.

On the other hand, Field presents a sophisticated paracomplete logical approach together with an inconsistency philosophical approach that is designed to handle not just liar sentences but certain revenge sentences as well. In Chapter Eight, I argued that Field's approach to the liar paradox faces revenge paradoxes.

---

[14] Again, 'Xnot' expresses Boolean negation

How does Field avoid that argument? He rejects the law of excluded middle for all of the indeterminateness predicates he defines. That is, he rejects '(1) is indeterminate$_0$ or (1) is not indeterminate$_0$'. Thus, all of his indeterminacy predicates are analogous to truth (on his view) in that they are partially defined. Of course, as one goes further and further up the hierarchy of indeterminacy predicates, one gets closer and closer to a completely defined indeterminateness predicate, but Field argues convincingly that one never gets there. That is, it is impossible to define a completely defined indeterminacy predicate in the language he constructs. Thus, he never has to deal with a real revenge paradox, like:

(2) (2) is either false or indeterminate*,

where 'indeterminate*' is completely defined (i.e., every sentence is either indeterminate* or not indeterminate*). Field's theory clearly cannot handle sentences like (2). What does he say about them?

Field argues (convincingly in my view) that one need not use any such linguistic expression to formulate his theory (his semantic theory uses a completely defined notion of semantic value, but it is relative to a model and so cannot be used to construct a revenge paradox). In addition, he argues against the claim that the artificial language he constructs avoids revenge paradoxes only because it has expressive limitations. Instead, he claims that sentences like (2) are unintelligible. Indeed, Field argues, any linguistic expression that is not in his artificial language and seems to give rise to a revenge paradox (e.g., 'indeterminate*') is unintelligible; however, by 'unintelligible' he does not mean *meaningless*:

> I don't want to deny that we have these notions; but not every notion we have is ultimately intelligible when examined closely. A large part of the response to the counterintuitiveness qualm will be an argument, in Part Four, that the notion of "the" hierarchy of iterations of D has a kind of inherent vagueness that casts doubt on there being a well-behaved notion of "Dα-true for every α"; and without that there is no reason to suppose that there is a well-behaved notion of "determinately

> true in every reasonable sense of that term". The apparent clarity of such notions is an illusion.[15]

Field then argues that there is no way to extrapolate from the hierarchy of determinateness predicates to define a well-behaved (i.e., intelligible) notion of hyper-determinateness.[16]

I am willing to admit that if we have only the resources provided by Field's artificial language, then we will be unable to define a well-behaved notion of hyper-determinateness. However, it seems to me that this point does little to quell the revenge worries. The problem is *not*: how can we use the resources Field gives us to generate a paradox his theory cannot handle? The problem is: we have a notion of determinateness that obeys excluded middle (i.e., what Field calls hyper-determinateness) and one cannot express this notion in Field's artificial language. Thus, Field avoids revenge only by an expressive limitation on his language.

I assume that it is obvious how Field would respond. He would probably claim that his artificial language can express any *intelligible* notion we have. Furthermore, he might continue, the problem of revenge paradoxes that I keep pressing is a problem that arises only when truth, which is intelligible, is combined with other resources (e.g., exclusion negation, other non-monotonic sentential operators, hyper-determinateness operators, etc.), which are not intelligible. Thus, it is not that truth is responsible for these revenge paradoxes; rather, truth has been keeping company with some other notions, which are responsible for the trouble.

There are several issues to consider when deciding which linguistic items should be blamed for the paradox. One issue involves the sort of explanation we get. The objector suggests that we should blame exclusion negation, and blame all the other non-monotonic sentential operators, and blame the conditional, and blame completely defined gaphood predicates, and blame idempotent

---

[15] Field (2007: §11).
[16] Field also discusses what he calls "model-theoretic revenge," but it is distinct from the sort of worry that I have pressed in this book; see Field (2007: §9).

determinacy operators, and blame quantification over partially defined gaphood predicates, and blame paradoxicality predicates, and blame groundedness predicates, and blame truth expressions that are not language-specific, … the list goes on.

I suggest that we should blame truth. That is it. Thus, my explanation is much simpler. It is also much more plausible. We can construct artificial languages that contain the outlaw linguistic expressions and they are perfectly well-behaved as long as they do not contain truth predicates (or related semantic terms). Of course, we can also construct artificial languages with truth predicates that are perfectly well behaved as long as they do not contain the outlaw linguistic expressions. However, the difference is that there are many different ways to construct revenge paradoxes; one involves *truth* and exclusion negation, one involves *truth* and another non-monotonic sentential operator, one involves *truth* and the conditional, one involves *truth* and an idempotent determinacy operator, etc. Exclusion negation is not involved in each case, nor are any of the other outlaw linguistic expressions. However, *truth* is involved every time. Truth is the only suspect that has no alibi—it is present at every crime scene; none of the others is. It does not take a Holmes, or a Spade, or a Columbo to identify the perpetrator; even a Wiggum could get this one right.

However, there is another, even more compelling reason to prefer my approach. As I argued in Chapter Eight, as long as one has a truth predicate that obeys the truth rules (even a language-specific one) and some minimal resources (e.g., common descriptions), one can "import" a revenge paradox into the language by way of inter-linguistic truth attributions. Field avoids this problem only because he does not consider other languages at all. Thus, even if we follow Field's advice (i.e., blame the outlaw linguistic resources for the liar and revenge paradoxes and revise our language so that it does not contain any outlaw linguistic resources), then we still have to restrict the theory so that it does not apply to certain sentences of the revised language. Thus, Field's strategy does not really secure a language and a theory such that the theory both applies to and is expressible in the

language. Hence, the "blame everything but truth" strategy does not work. Not only is my strategy simpler and more plausible, it is the only one that works.

I have argued that there are two kinds of revenge paradoxes: self-refutation problems and inconsistency problems. Glanzberg addresses self-refutation problems, which confront theories of truth that imply that they are Xnot true. He argues that this kind of revenge paradox has its source in the fact that truth is a Kreiselian concept (i.e., it is not closed under reflection). However, Glanzberg does not explain or even address inconsistency problems, and there are good reasons to doubt his explanation of the self-refutation problem. On the other hand, Field addresses inconsistency problems and argues that these sorts of revenge paradoxes arise when what are ultimately unintelligible—read that as inconsistent or not well-defined—concepts (e.g., hyper-determinateness) are combined with truth. However, Field does not explain or even address self-refutation problems, and there are good reasons to doubt his explanation of the inconsistency problem. In contrast to both Glanzberg and Field, I offer an explanation of both types of revenge paradoxes, and my explanation of each type is superior to the one offered by Glanzberg and by Field, respectively.

12.4 The Revenge Argument

The argument in this section depends on the revenge argument presented in Chapter Eight. Here is a summary of that argument. Let T be a theory of truth that validates the primary aletheic principles and implies that liar sentences have status Δ, where a sentence is Δ only if it is Xnot true. Let '(ρ)' be our revenge liar:

(ρ) (ρ) is either false or Δ.

There are three options for T:

(i) T implies that (ρ), '(ρ) is either false or Δ', is true.

(ii) T implies that (ρ) is false.

(iii) T implies that (ρ) is Δ.

On any of these options, T implies '(ρ) is true if and only if (ρ) is false or Δ'. If T classifies this sentence as true, then T is inconsistent. If T classifies it as false, then T is self-refuting. If T classifies it as Δ, then T is self-refuting. Therefore, T is either inconsistent or self-refuting, unless T is restricted so as to avoid applying to languages that contain sentences like (ρ). Consequently, any theory of truth that validates the primary aletheic principles (and implies that liar sentences have status Δ, where a sentence is Δ only if it is Xnot true) is either inconsistent, self-refuting, or restricted.

Given that any theory of truth that is inconsistent (trivial), self-refuting, or restricted is unacceptable, any theory of truth that implies both that the primary aletheic principles are true and that liar sentences have status Δ is unacceptable. That is the conclusion of the revenge argument. I have argued repeatedly that we should accept that the primary aletheic principles are constitutive of truth. Thus, on my view: (i) if a theory of truth is acceptable, then it implies that the truth rules are constitutive for truth, and (ii) if a theory of truth validates the truth rules, then it is unacceptable. There seems to be very little wiggle room here. However, there is an additional assumption that connects the two conditionals: if a theory of truth implies that the truth rules are constitutive of truth, then it implies that they are true. I reject this claim. One of the key aspects of the account of inconsistent concepts I endorse is that one or more of the constitutive rules governing the concept are invalid. Thus, I accept that the truth rules are constitutive of truth, but it is not the case that they are all true; the theory of truth presented in Chapter Fourteen has this result as a consequence. This result allows anyone with a suitable account of inconsistent concepts to avoid the revenge argument.

The following is a summary of the revenge argument for treating truth as an inconsistent concept:

(i) If a theory of truth is acceptable, then it implies that truth obeys the primary aletheic principles.

(ii) If a theory of truth implies that truth is a consistent concept and it implies that truth obeys the primary aletheic principles, then it implies that the primary aletheic principles are valid.

(iii) If a theory of truth implies that the primary aletheic principles are valid, then it is inconsistent, self-refuting, or restricted.

(iv) If a theory of truth is inconsistent, self-refuting or restricted, then it is unacceptable.

∴ (v) If a theory of truth is acceptable, then it does not imply that truth is a consistent concept.

Only an inconsistency theory of truth, on which principles (like the primary aletheic principles) that are constitutive for a concept need not be valid, has a chance of being an acceptable theory of truth. That is the conclusion of the revenge argument for an inconsistency approach.

Since the standard ploy, by those who think truth is a consistent concept, to avoid this kind of argument is to deny (iv) and restrict their theory of truth. The next three subsections argue that this kind of restriction is unacceptable.

12.4.1 Restriction and Importation

Chapter Eight covered revenge paradoxes and had a discussion of what I call *importation problems*. The idea is that even if one restricts one's theory of truth to languages that do not have the resources for explicit revenge liars, seemingly innocent languages can harbor revenge paradoxes because they have the ability to refer indirectly to languages with explicit revenge liars. That is, using the resources of the seemingly innocent language, one can import a revenge paradox from the restricted language.

The easiest way to see this is that the seemingly innocent language will have a truth predicate. If it also has a singular term for the revenge liar of a restricted language, then it will also have a revenge liar, even though it does not contain all the linguistic expressions needed for the explicit revenge liar. The upshot is that restricting a theory of truth so that it does not apply to languages capable of formulating revenge liars does not work. There is no good way to ensure that a language in the scope of one's theory cannot import a revenge paradox, since seemingly innocuous singular terms (e.g., 'the first sentence of the most popular book in Springfield's library') could refer to an offending revenge liar.

As I emphasized in Chapter Eight, those who restrict their theories of truth also stipulate that their theories apply only to language-specific truth predicates, which might seem to be immune to importation problems. There are two serious difficulties here—the first is that, because of truth's expressive role and the Gricean condition, one cannot explain a natural language truth predicate in terms of language-specific truth predicates; I made that argument in Chapter Six. The second is that even if one could get by with a theory of language-specific truth predicates, importation problems arise again, just in a more subtle and complicated way. Thus, there really is no escape from importation problems.

12.4.2 Restriction and Internalizability

Chapter Nine is dedicated to internalizability—the relation between a theory and a language when that theory applies to and is expressible in an extension of that language. There I argued that an acceptable semantic theory for truth ought to be internalizable for any natural language. Moreover, if one restricts one's theory of truth to avoid the revenge argument, then one's semantic theory is not internalizable for natural languages.

12.4.3 Restriction and Inconsistent Concepts

Those who restrict the scope of their theories in the face of revenge paradoxes offer one of two justifications: (i) the linguistic expressions that occur in revenge liars are meaningless, so the theory is not really restricted, or (ii) the linguistic expressions that occur in revenge liars express inconsistent concepts, so the theory is only restricted from applying to defective concepts, which hardly seems objectionable. I have talked about these two justifications already, but the point I want to make in this subsection is new. Option (i) is totally unacceptable, given the Gricean condition on communication—anyone who offers this kind of justification simply does not understand how languages are used. Option (ii) is much less problematic, so I concentrate on it.

What is wrong with option (ii)? Maybe nothing, provided that its proponents could make the case that these concepts are indeed inconsistent. However, their arguments all assume that truth is consistent. So, at best, we can conclude that either truth is inconsistent or these concepts that feature in revenge paradoxes are inconsistent. Notice what has happened here. Both sides in the debate about how best to approach the aletheic paradoxes agree that inconsistent concepts are at fault. The inconsistency theorists think that the inconsistency of truth explains why the liar paradox occurs, while the consistency theorists think that the inconsistency of these other concepts explains why revenge paradoxes occur. So, either way, an adequate approach to the aletheic paradoxes is going to have to say something about inconsistent concepts. It must incorporate a theory of inconsistent concepts. Thus, no matter which approach to the aletheic paradoxes one prefers, an essential component of it will be a theory of inconsistent concepts. No matter what one thinks is the source of the aletheic paradoxes, inconsistent concepts are somehow to blame for our inability to solve them.

But there is an asymmetry between the roles played by inconsistent concepts consistency views and in inconsistency views. Those who offer an inconsistency approach to the aletheic paradoxes

claim that truth is an inconsistent concept; that is it. Those who offer consistency approaches to the aletheic paradoxes claim that any concept that can be used to formulate a revenge paradox is inconsistent. That results in saying that lots of concepts are inconsistent. For example, a paracomplete theorist has to say that exclusion negation, determinateness, paradoxicality, groundedness, classical negation, the classical conditional, and the intuitionistic conditional are inconsistent. Moreover, if one looks at the arguments here, the consistency theorist has nothing new to offer—the reason for thinking that these concepts are inconsistent is that one can derive a contradiction from their constitutive principles; of course, that is, taking it for granted that truth is consistent. Notice, however, that it is much simpler to treat truth as an inconsistent concept—it avoids having to find inconsistent concepts all over a wide terrain. Moreover, it is more modest to say that truth is an inconsistent concept since all the revenge-prone concepts only give rise to paradoxes in conjunction with truth! Truth is always there. It seems almost like willful ignorance to blame all these other concepts for the faults of truth.

In summary, every party to the debate has to say that inconsistent concepts are at fault for some of the paradoxes associated with truth. It results in a much better theory to say that truth itself is an inconsistent concept than to say that any concept that figures in a revenge paradox is inconsistent. That point alone should justify the inconsistency approach over the consistency approaches. However, when one considers the futility of restricting one's theory (given importation) and the severe consequences of doing so for the rest of our theorizing (given internalizability), the case for an inconsistency view is strengthened considerably.

12.5 Inconsistency Views

The feature of the inconsistency view that I propose that sets it apart from the others is that I think truth should be replaced because it is inconsistent. We need new concepts to do the work that truth

was supposed to do. In order to get a feel for just how different this view is from the rest, let us take a moment to consider a couple of these views. Then, in section six, I argue that inconsistency theorists should endorse the replacement strategy. Finally, in section seven, I argue that a unified theory of truth that implements the replacement strategy should not appeal to the inconsistent concept of truth—rather, it should appeal to the replacements.

12.5.1 Dialetheism

Dialetheism is most often treated as a theory of inconsistent concepts.[17] According to dialetheists, the argument in the liar paradox is sound—that is, it is valid and all its premises are true, which makes its conclusion true. The conclusion is a contradiction, namely that the liar sentence is both true and not true. Usually the dialetheist argues that since the premises of the argument are constitutive of the concepts involved, and so they must be true; however, the argument is obviously valid, so the conclusion must be true as well. Since the conclusion is a contradiction, some contradictions must be true. Of course, the dialetheist does not think that everything is true, so accepting dialetheism goes hand in hand with adopting a paraconsistent logic (i.e., a logic on which it is not the case that everything following from a contradiction).

I have said several times that I reject dialetheism, and since it is probably the view that pops into most philosophers heads when hearing the term 'inconsistent concept', a word or two is in order about why I reject it. First, as should be clear from the revenge argument in the previous section and the discussion of paraconsistent approaches in Chapter Eight, dialetheism faces revenge paradoxes and has to be restricted so as to avoid them. Of course, as I argued, the restrictions do not really work, as evidenced by importation arguments; and they are intellectually crippling, as the

[17] See Priest (2006a, 2006b) and Beall (2009) for discussion.

internalizability arguments demonstrated. So, as a theory of truth, dialetheism fails just as badly as Field's theory or any other one that generates revenge paradoxes.

However, since it is a kind of inconsistency theory, it might be helpful to see exactly which parts of dialetheism I reject. I agree with dialetheists that truth has inconsistent constitutive principles—that is a point on which all inconsistency theorists (as I have defined them) can agree. However, the view that constitutive principles are true is highly suspect, and that is the major point of contention. As I emphasized in Chapter Eleven, there are no principles that are true by virtue of their meaning—that is, there are no analytic principles. Nevertheless, there are constitutive principles, for these play an important role in our practice of interpretation. The meanings of our words and the contents of our concepts incorporate some elements of the world—they take a stand on the way the world is. Constitutive principles often turn out to be true, but when they do it is because the world is as they take it to be. Simply stipulating that a certain word has a certain meaning is enough to establish that it does indeed have that meaning, but it does not ensure that the constitutive principles in question are true. If the concept expressed by that word is consistent, then its constitutive principles are true; whether it is consistent depends on what the world is like. Notice that the dialetheist is more willing to accept contradictions than accept that constitutive principles might not be true; that seems to me like a serious flaw in reasoning.

The other major problem I want to mention is that the dialetheist's hands are tied when it comes to responding to revenge paradoxes. Paraconsistent approaches to the aletheic paradoxes face standard revenge paradoxes (e.g., pertaining to 'just true' as discussed in Chapter Eight). Paracomplete approaches face revenge paradoxes as well; however, the paracomplete theorist can say that the concepts involved in the revenge paradoxes are inconsistent. At that point, the debate between someone like me and someone like Field turns on which concepts are inconsistent; I think I win that argument, but that is not the point. Rather, the paraconsistent theorist cannot follow this

strategy—claiming that the concepts in question are inconsistent does not justify eliminating them from the scope of the paraconsistent approach. After all, the central claim of paraconsistent dialetheism is that truth is an inconsistent concept. Instead, the dialetheist has to make the radically implausible move of saying that the offending concepts do not exist and that the words that feature in revenge paradoxes for paraconsistent approaches are simply meaningless. As I have stressed, this move flies in the face of the Gricean Condition, which is presupposed by virtually all contemporary work in pragmatics. To reject it without any empirical evidence simply because it conflicts with one's favorite theory of truth is as preposterous as rejecting contemporary evolutionary theory without any empirical evidence simply because it conflicts with one's favorite theology.

12.5.2 Patterson

Douglas Patterson is an inconsistency theorist who starts with the idea that an agent who understands a language bears some more or less cognitive relation (knowing, believing, etc.) to a semantic theory for that language. If the language contains a word that expresses an inconsistent concept, then the semantic theory for that language is inconsistent.[18] Thus, linguistic competence is being cognitively related to a semantic theory, and competence with an inconsistent concept is being cognitively related to an inconsistent semantic theory. Patterson argues that because the semantic theory for an inconsistent language is inconsistent, the expressions of that language are meaningless. He claims that even though the sentences of an inconsistent language are meaningless, communication is still possible as long as the participants bear the same cognitive relation to an

[18] He does not actually discuss inconsistent concepts; instead he focuses on inconsistent languages. However, I assume that a language is inconsistent if and only if it contains a word that expresses an inconsistent concept.

inconsistent semantic theory.[19] Furthermore, we can translate from an inconsistent language into a consistent one if the need arises (where translation is preservation of perceived meaning).

I can imagine someone thinking: the traditional approaches to the liar paradox have to be better than saying that all the sentences of English are meaningless! *I agree.* I would much rather accept a traditional approach than say that English is meaningless. I cannot imagine an argument for this claim whose premises I trust more than the negation of the conclusion. Of course, Patterson downplays the radically implausible consequences of his approach by arguing that it does not matter that all natural languages are meaningless. All that matters is that we take them to be meaningful. As long as we treat a language as meaningful, we can get along without any problems. Moreover, most of us never notice that our language is meaningless, because we do not bother to follow out the consequences of our beliefs.

These moves do not make much difference to the overall plausibility of his view because if they were correct, then being meaningful would not be a very important feature of a language. However, most philosophers think that being meaningful is such an important characteristic of languages that it does not even make sense to say that a natural language is meaningless. In fact, on every theory of language of which I am aware, being meaningful is the defining feature of a language, since languages are individuated in part by the meanings of the words they contain. Thus, if Patterson is right, then English does not even count as a language—it is just a bunch of grunts and marks.

In Chapter Six, I argued on the basis of Gricean considerations that we cannot discover that a word with an established usage is meaningless. Now try to imagine reading in the newspaper that scientists have discovered that the entire French language is meaningless. I, for one, cannot do it (unless I have mistakenly picked up a copy of *The Onion* or some other satirical paper). The reason is

---

[19] See Patterson (2006, 2007a, 2007b, 2009).

that it is inconceivable for us to discover that an entire natural language is meaningless, much less *every* natural language.

Even if we ignore all the empirical evidence, there is an additional problem. Patterson thinks that our language is inconsistent *because* it contains a truth predicate (although there might be other troublesome words as well). If the problem has to do with truth, why should it spill over into meaning? In other words, why do they think that a suitable approach to the liar should have consequences for the meanings of sentences that have nothing to do with truth? It seems to me that Patterson adopts his view because of a commitment to truth-conditional semantics. He reasons that no truth conditional semantic theory for a natural language can respect the principles that everyone takes the truth predicate to have unless it is inconsistent; thus, if natural languages are meaningful, then they have inconsistent semantics. Patterson then performs a *modus tollens*. The problem, of course, is that he seems to be more confident that meaning should be explained in terms of truth conditions than he is that English is meaningful. That is, he is so sure that meaning should be explained in terms of truth conditions, that it has convinced him that there is no such thing as meaning (at least as it is commonly understood); he retains the *explanans* at the expense of the *explanandum*. That is hardly a promising explanatory story; it is more like throwing out the baby *instead of* the bathwater.

12.5.3 Ludwig

Kirk Ludwig's inconsistency approach (part of which appears in work with Amil Badici) has two parts: (i) a characterization of the truth predicate of a natural language and (ii) a meaning theory for

languages that contain their own truth predicates.[20] The first part consists of the claim that truth predicates of natural languages do not express the concept of truth; indeed, they do not express any concept whatsoever. Natural language speakers are under the impression that their truth predicates do express the concept of truth, but they are mistaken. That is not to say, however, that there is no concept of truth. Indeed, Ludwig and Badici claim that there is a concept of truth, but it cannot be expressed by a predicate of a language to which it applies.[21] They set out to show how an attempt to introduce the concept of truth into a language to which it is intended to apply runs into difficulty.

They model the difficulty by considering a language L and a metalanguage M. M contains the predicate 'true-in-L', translations of all the sentences of L, and the means to give structural descriptions of the sentences of L. With these resources, M contains a true T-sentence for each sentence of L. Consider now an attempt to extend L by adding a predicate, '$T(x)$' to it, which is supposed to express the concept of truth-in-L; let the extended language be L^+. They then appeal to what is essentially Tarski's theorem on the indefinability of truth to show that '$T(x)$' does not express the concept of truth-in-L upon pain of contradiction. Their conclusion: '$T(x)$' "expresses no concept, and, hence, fails to have a meaning."[22]

Badici and Ludwig consider the objection that the fact that a contradiction follows from the T-sentence for a liar sentence of L shows that the concept expressed by the predicate 'true-in-L' of M is an inconsistent concept. Their reply:

> We can show that this thought is incorrect. Suppose one adds two truth-predicates, '$T_1(x)$' and '$T_2(x)$' to an extension of L, L^{2+}, with the intention that they express the concept of truth, and suppose that $\lambda 1$ and $\lambda 2$ refer respectively to '$\sim T_1(\lambda_1)$' and '$\sim T_2(\lambda_2)$', and λ_3 and λ_4 refer respectively to '$\sim T_2(\lambda_1)$' and '$\sim T_1(\lambda_2)$'. The two predicates should have the same

---

[20] Ludwig (2002) and Ludwig and Badici (2007). The first part of the approach appears in the latter paper so, in the text, I attribute it to the two authors; the second part appers in the former paper in detail and is mentioned in the latter.
[21] Ludwig and Badici (2007: 623). This should seem like an odd claim; I think the best way to understand it is that they think that there are only language-specific concepts of truth, but no language can express its own language-specific concept of truth.
[22] Ludwig and Badici (2007: 628).

> meaning, because they are intended to capture the same conceptual content, and on the view in question the two corresponding T-schemas, (T1) and (T2), determine the same meaning.
>
> (T1) T_1(s) iff p
>
> (T2) T_2(s) iff p
>
> Nevertheless, they do not. For λ_1 and λ_2, given (T1) and (T2), lead directly to contradictions and so are pathological, while the result of replacing each truth predicate in these sentences by the other, λ_3 and λ_4, do not, but rather seem to say the right thing about the pathological sentences λ_1 and λ_2. T_1 and T_2 do not have the same meaning then, and it follows that one or the other fails to express the concept of truth with respect to the language. Since there is perfect symmetry between them, the proper conclusion is that neither does.[23]

Their argument depends on the claim that λ_1 (i.e., '$\sim T_1\lambda_1$') is pathological, while λ_3 (i.e., '$\sim T_2(\lambda_1)$' "seems like the right thing to say" (along with the analgous claims about λ_2 and λ_4). However, the fact is that λ_3 and λ_4 are just as pathological as λ_1 and λ_2. Here is the argument for λ_3:

1.	$\sim T_2(\lambda_1)$	(assume)
2.	$\sim T_2(\text{`}\sim T_1\lambda_1\text{'})$	(definition of λ_1)
3.	$\sim\sim T_1(\lambda_1)$	(by T2)
4.	$T_1(\lambda_1)$	(by double negation elimination)
5.	$T_1(\text{`}\sim T_1\lambda_1\text{'})$	(by definition of λ_1)
6.	$\sim T_1(\lambda_1)$	(by T1)
7.	$\bot$	

1.	$T_2(\lambda_1)$	(assume)
2.	$T_2(\text{`}\sim T_1\lambda_1\text{'})$	(definition of λ_1)
3.	$\sim T_1(\lambda_1)$	(by T2)
4.	$T_1(\text{`}\sim T_1\lambda_1\text{'})$	(by T1)
6.	$T_1(\lambda_1)$	(by definition of λ_1)
7.	$\bot$	

The argument for λ_4 is analogous. I conclude that Badici and Ludwig's attempt to show that truth predicates do not express the concept of truth fails. So much for the first part of their project.

The second part of this approach consists of Ludwig's suggestion for using an inconsistent theory of truth as a meaning theory for a language. Recall that this came up in Chapter Five, where I showed that that his suggestion fails badly—it implies that there are infinitely many correct but

---

[23] Ludwig and Badici (2007: 629-630).

incompatible specifications for the meanings of liar sentences. That result dispatches the second part of this inconsistency approach.

12.5.4 Eklund

Matti Eklund focuses on the constitutive principles for the concepts expressed by the words of an inconsistent language. For Eklund, a language is inconsistent if and only if its constitutive principles (i.e., the constitutive principles for the concepts expressed by the words the language contains) are inconsistent. Instead of following Patterson, Eklund explains linguistic competence in terms of dispositions to accept the language's constitutive principles. Eklund rejects dialetheism (i.e., the view that some contradictions are true), but he accepts that some of a language's constitutive principles might be false. As for the semantic theory for an inconsistent language, Eklund argues that the words and sentences of an inconsistent language have semantic values that come as close as possible to making the language's constitutive principles true. He admits that some constitutive principles might be more important than others, so that would have to be taken into account when deciding on a semantic theory. Also, there will probably be multiple equally good candidate semantic theories for an inconsistent language; thus, the semantic values of the words and sentences of an inconsistent language are indeterminate to that extent.[24]

Although Eklund's view is far superior to that of Patterson, it still has several problems. First, it has an unacceptable account of the relation between concept possession and constitutive principles (discussed in Chapter Eleven). Furthermore, Eklund's theory is intended to provide an approach to the liar paradox. His view is that truth is an inconsistent concept (or that any language with a truth predicate is an inconsistent language). But his theory appeals to the very concept it deems

[24] Eklund (2002, 2005, 2008a).

inconsistent. That is, his approach to the liar paradox gives truth, which he takes to be an inconsistent concept, a crucial explanatory role. However, the fact that truth is an inconsistent concept should cast doubt on its ability to perform such a role. Consider again the analogy to mass: once we discovered that mass is inconsistent, we stopped using it for serious theorizing. Why is the case of truth any different?

Consider this problem in a little more detail. Let L be an inconsistent language and ML be the language in which Eklund's theory is formulated (L and ML might be the same language since, unlike traditional approaches to the liar, Eklund's theory does not require an expressively richer metalanguage). Both L and ML contain truth predicates, but it is the truth predicate of ML that is used by Eklund's theory (call it τ). We know that since τ is a truth predicate, it has certain constitutive principles. We also know that because τ expresses an inconsistent concept, not all of its constitutive principles are true (or valid). Eklund does not tell us which of truth's constitutive principles fail, but we know that some do. The question is: how can τ function properly in Eklund's theory, which is supposed to provide a semantics for L, if some of its constitutive principles fail? Mono-aletheism (i.e., no sentence is both true and false) is an essential principle for a non-dialetheic semantics, and I do not see how a semantic theory could assign the right truth conditions to the sentences of L unless the truth predicate it employs obeys the ascending and descending truth rules (i.e., $\langle p \rangle$ follows from $\langle\langle p \rangle$ is true$\rangle$, and $\langle\langle p \rangle$ is true$\rangle$ follows from $\langle p \rangle$). However, these are the very principles that give rise to the liar paradox. Thus, some of them have to fail; otherwise, Eklund's theory would be inconsistent. In sum, Eklund's theory casts truth in a crucial explanatory role, and it implies that some of truth's constitutive principles are untrue. It seems, however, that if some of truth's constitutive principles fail, then it is unsuited to play this explanatory role (I see this problem as an analog of a revenge paradox for Eklund's theory). At the very least, Eklund owes us an

explanation of how truth can function properly in his semantic theory even though it is an inconsistent concept (and consequently, some of its constitutive principles fail).

12.6 The Replacement Argument

This section is dedicated to clarifying what I take to be the most important aspect of my particular approach to the aletheic paradoxes: an inconsistency approach should be part of a larger account of conceptual change, and in particular conceptual change with respect to truth. The following is a rough account of the stages of conceptual change:

1. *Pre-revolution*: people possess and use concept X and theory T in which X serves an explanatory role (e.g., mass and Newtonian mechanics).

2. *Early revolution*: people discover that X is an inconsistent concept; they have some idea of which situations cause problems for those who use X; because of these problems, doubt is cast on the explanatory force of X and the acceptability of T as fundamental theory; however, without an alternative, people still use T and X.

3. *Late revolution*: new concepts (say Y_1, …, Y_n) are proposed and a new theory (say U) is proposed in which the Y_is serve an explanatory role (e.g., relativistic mass and proper mass in relativistic mechanics); U reduces to T in familiar cases, and the Y_is agree with X on familiar cases; U is used to determine the cases in which it is acceptable to use T; at this point the conceptual repertoire and language have been extended.

4. *Post-revolution*: U has replaced T as the accepted fundamental theory, and the Y_is have replaced X as the accepted fundamental concepts; people might or might not still use T (and thus X) in certain cases (e.g., phlogiston theory has been totally superseded, but Newtonian mechanics is still indispensable for everyday situations).[25]

The fundamental problem with all the other inconsistency theorists is that although they attempt to give an account of our concepts and language at stage 1, and some of them (e.g., Patterson, Ludwig, and Eklund) consider stage 2, they completely ignore stages 3 and 4. An inconsistency approach to the liar that does justice to stages 3 and 4 would propose replacement concepts for truth, and replacement theories for the theories we currently have that appeal to truth. Truth is a very popular

[25] For more on conceptual revolutions, see Kuhn (1962), Thagard (1992) and Anderson, Barker, and Chen (2006).

concept—it is used in theories of meaning, knowledge, assertion, belief, validity, objectivity, rationality, etc.; thus, replacing it is a big job. Obviously, one wants replacements for truth that can be used to construct new theories to replace the old ones (that is a task of Chapter Fifteen).

In an attempt to understand our language at stages 1 and 2, some inconsistency theorists (e.g., Eklund) have presented traditional semantic theories for inconsistent concepts/languages. Given the role of truth in understanding language, their actions are understandable. However, once we remind ourselves of the stages of conceptual change, we can see that they have jumped the gun—their semantic theories appeal to the concept of truth! Before we can explain our stage 1 and 2 language, we need to find replacements for our concept of truth. Then we can use the replacements to formulate a new semantic theory. Once we have that, we can use it to explain the languages we speak at each of the stages.

Notice that the case of mass is much less complicated than the case of truth because truth is a linguistic concept. We do not use mass to try to explain discourse involving 'mass'. However, we do use truth to explain discourse involving 'true'. If we had used mass in this way, then once we reached stage 2, we would have been tempted to use it to explain our stage 1 and 2 language. Thus, when discussing inconsistency approaches to the liar, it is essential that one maintain one's bearings by keeping the stages of conceptual change firmly in mind.

It is a necessary condition on an acceptable account of inconsistent concepts (and thus, on an inconsistency approach to the liar) that it do justice to all the stages of conceptual change. If we accept a theory that appeals to truth (e.g., Eklund's theory), then we will not be able to progress to stages 3 and 4 without giving up the theory—we will be stuck in stage 2. Of course, as a provisional account of stages 1 and 2, it is fine to use the concept of truth, provided one keeps in mind that one is using an inconsistent concept to describe discourse involving that inconsistent concept and that the provisional theory should be superseded by a more fundamental one once we have acceptable

replacement concepts (that is the way I think of Eklund's theory). Thus, I am suggesting that the other inconsistency theorists suffer from a lack of vision—they do not see the larger enterprise in which they are engaged. The moral is that if one endorses an inconsistency approach to the liar paradox, one should be in the business of replacing truth.

Another reason for replacing truth has more to do with inconsistent concepts in general. Consider the case of mass. Once we discovered that momentum/velocity varies with reference frames, we discovered that the concept of mass is inconsistent.[26] At that point, we knew that using it in certain situations would lead us astray—it would deliver incorrect predictions or even outright contradictions. Nevertheless, we also knew that in many situations, it is perfectly legitimate to use it, just as we had been for hundreds of years. What marks the difference between the two kinds of situations? Only once we had relativistic mechanics, with its two concepts of mass, could we answer this question. The answer is, of course, that when the difference between relativistic mass and proper mass is negligible given one's interests in a given situation, one may use mass; otherwise, one should use the replacement concepts. This example points up a general lesson: a useful inconsistent concept should be replaced since it is only by using the replacements that one can determine in which situations it may legitimately be used. Truth is incontestably useful. Many exploit its *explanatory* roles; but everyone should agree that its *expressive* role is useful. Thus, any inconsistency approach to the aletheic paradoxes should offer replacements for truth.

12.7 Two Theories

The discussion so far illustrates a deep divide between my approach and those of other contemporary inconsistency theorists: I take a *dynamic* attitude toward the liar paradox, while theirs is

[26] Obviously, this is a gross oversimplification of the empirical and theoretical situation that led us to reject Newtonian mechanics and accept special relativity; however, the additional technical details would distract from the philosophical point without any offsetting benefit. See Jammer (2000), Petkov (2009), and the papers in Capria (2005).

static. That is, my approach focuses on what we can do and what we should do about the liar paradox. Their approaches are only about how to describe one aspect of the current mess we are in—they focus on how to understand languages that contain words that express inconsistent concepts. They care about where we are; I care about both where we are and where we want to be. Of course, it is important to understand our language as it is; without such an understanding, we would not know what problem needs to be fixed. In fact, we inconsistency theorists all agree that the biggest mistake made by those who propose traditional approaches to the liar paradox is that they misdiagnose the problem. *We* think that truth is an inconsistent concept (or that a language containing a truth predicate is an inconsistent language), while *they* think that everyone taken in by the reasoning involved in the liar paradox is making some mistake.

We think that competence with the concepts involved in the paradox predisposes those who employ them to accept all the assumptions and inferences involved in the liar reasoning. Although we (inconsistency theorists) all agree on this matter, we disagree about how to characterize inconsistent concepts and languages. From my point of view, the biggest mistake made by the other inconsistency theorists is that they do not consider what can be done to change our language and our conceptual repertoire in order to eliminate the liar paradox and its vengeful brethren. Only by understanding the process by which we change our concepts and our language can we really understand both our inconsistent language and what we should do to fix it.

Accordingly, my approach to the liar paradox has two parts: (i) a descriptive theory, which explains our inconsistent language and our inconsistent concept of truth, and (ii) a prescriptive theory, which explains how we should change our language and which introduces new concepts to take the place of our inconsistent concept of truth. It is essential that *the descriptive theory depends on the prescriptive theory*. That is, the theory that explains our inconsistent concept of truth does not appeal to our inconsistent concept of truth. Instead, the descriptive theory appeals to the replacement

concepts introduced by the prescriptive theory. Otherwise, one could not accept the explanation of our inconsistent concept of truth without giving our inconsistent concept of truth a crucial explanatory role.

12.8 The Parable of Mindy

One might have the following worry about my replacement strategy: we need to understand our current linguistic practice *before* we can figure out how to fix it; otherwise, we have no reason to think that the fix will be successful; that is, we need to know where we are *before* we figure out where we should go. Thus, a descriptive theory should not be based on the prescriptive theory.

My reply is that we need a good enough understanding of our current linguistic practice to figure out what to do. However, we have good reason to think that although many of our tools for understanding our linguistic practice rely on truth (e.g., truth-conditional semantics), truth is a defective concept. Thus, we simultaneously think that truth is a key to understanding our language and that it is defective. That puts us in the position of being able to understand *well enough* what our current practice is like. We understand what is wrong well enough to see that, using our current concepts, we cannot successfully describe what is going on. Moreover, we understand what is wrong well enough to place some conditions on potential conceptual revolutions.

I would like to use an analogy from Hasok Chang's recent work on the development of the concept of temperature to illustrate my point (I have changed the story a bit).[27] Imagine that a very nearsighted person, Mindy, has been using a monocle, but now finds that it does not work very well; in particular, there is a major distortion in her field of view. She takes off the monocle and looks at it but cannot see any defects owing to her inability to see much at all without it. She gets an idea and

[27] Chang (2004).

puts on the monocle and looks in a mirror—she can see well enough with the monocle to notice a very large and deep scratch across the lens. Of course, the scratch prevents her from seeing the reflection of the monocle in the mirror perfectly. Nevertheless, she sees it well enough to diagnose the problem—the scratch on the lens—and well enough to figure out what needs to be done—replace the lens. So, it would be impossible for Mindy to see her monocle perfectly given that the tool she is using—that very monocle—is defective. Nevertheless, it works well enough for her to diagnose the problem and arrive at a course of action to fix it. Let us say Mindy calls her optometrist and there is no replacement lens in stock, so she decides to buy a pair of glasses instead. With her new pair of glasses, she can see the scratch on the monocle far better than she could while using the monocle and a mirror—after all, she is not looking through a defective lens anymore. So her monocle was able (with some help from the mirror) to give her enough information about her predicament to allow her to determine what the problem was even though she did not have a perfect understanding of it since the scratch on the lens distorted her view of the scratch on the lens. It also gave her enough information to figure out what to do about it, even though it did not give her perfect understanding of it since the scratch on the lens distorted her view of the scratch on the lens.

The monocle represents our concept of truth, the scratch on the lens is truth's conceptual defectiveness, and the new pair of glasses represents the replacements for truth. Even though we cannot use truth to get a perfect understanding of our linguistic practice and our truth predicate, our tools (which involve the notion of truth) give us enough information to figure out that truth is a defective concept and they give us enough information to figure out what to do about it.

We get a better understanding of our linguistic practice by using the replacement concepts rather than our concept of truth, just as Mindy gets a better view of her monocle using the glasses. That is not to say that the monocle did no work—it allowed her to figure out what was wrong even if it did not give her a perfect view of what was wrong. The same point holds in the case of truth—using it,

we have figured out that it is defective; using it, we have figured out what its replacements should be like. It is a mistake to think that defective tools, whether they are physical or conceptual, are good for nothing.

Chapter 13

The Prescriptive Theory

This chapter and the next one present the two major theories of this book. This chapter outlines the replacement concepts: ascending truth and descending truth. The next presents a theory of truth. This chapter is prescriptive in the sense that it offers a suggestion for changing our conceptual scheme. The next is descriptive in the sense that it purports to describe our concept of truth. One of the central tenets of the unified theory of truth I offer is that the descriptive theory of truth should depend on the prescriptive theory of truth—the replacements, not truth, should serve in explanatory roles for the descriptive theory.

13.1 The Replacements: Ascending Truth and Descending Truth

I am hardly the first philosopher to suggest replacements for truth. I think Tarski can be read as suggesting a sanitized replacement.[1] More recently, Vann McGee suggested that we replace truth with two concepts—a vague concept of truth and a concept of definite truth.[2] Indeed, since most approaches to the aletheic paradoxes require giving up something that seems integral to the everyday concept of truth, most of them can be read as offering a replacement concept.[3]

We can determine first that replacing truth with a single concept is not a promising strategy. Here is why—a single concept that does the job we require of truth would need to obey both (T-In) and (T-Out); otherwise it would not perform the expressive roles discussed in Chapter Six. However, a concept that obeys both (T-In) and (T-Out) is *inconsistent* in classical logic, and the move

---

[1] Tarski (1933).
[2] McGee (1991).
[3] For recent discussion see Field (2008a), Richard (2008), and Ebbs (2009).

to a non-classical logic gives rise to revenge paradoxes. Field (2008) does a nice job of cataloguing the problems faced by theories of truth that deny (T-In) or (T-Out). It is for this reason that I regard the single-concept replacement strategy is a non-starter.

Vann McGee is the only other person I know of who has suggested a pair of replacements for truth. However, one of his pair is a truth predicate that violates (T-In) and (T-Out)—his is a classical symmetric theory—and it is subject to revenge paradoxes. For these reasons, it should be clear that it does not satisfy the conditions of adequacy set out in Chapter Eight.

Instead of trying to pack both (T-In) and (T-Out) in a single concept and introduce a new concept to clean up after the resulting mess, we can introduce two concepts, where one obeys one of the primary aletheic principles and the other obeys the other; that is, one obeys (T-In) and the other obeys (T-Out). A reason for thinking that (T-In) and (T-Out) should be split up in the replacement concepts is that in the three major aletheic paradoxes (the liar, Curry, and Yablo), each paradoxical argument uses *both* of these principles. Thus, it makes the most sense to separate these principles so that they serve distinct concepts. Recall that this strategy worked well with the concept of mass. Discovering that momentum/velocity is not the same in all reference frames led us to posit one replacement for mass that obeys one of the constitutive principles for mass and another replacement for mass that obeys the other constitutive principle.

I suggest we do the same thing with truth—the two constitutive principles of truth that give rise to the liar paradox are:

(T-In) If p, then $\langle \mathsf{p} \rangle$ is true.

(T-Out) If $\langle \mathsf{p} \rangle$ is true, then p.

We should replace truth with two concepts; one that obeys (T-In) but not (T-Out), and another that obeys (T-Out) but not (T-In). Inspired by Quine's comment that (T-In) encapsulates truth's

function of semantic ascent, I call the concept that obeys (T-In) *ascending truth*.[4] To the other I give the name *descending truth*. We will consider which other principles these concepts obey below.

13.2 Aletheic Principles

Let us begin by considering the principles commonly assumed to hold of truth. We need to make some hard decisions about which of these govern ascending truth and which govern descending truth. It turns out that this is the most difficult part of offering a theory of the replacement concepts. The following is a list of aletheic principles:

Disquotational Principles		*Truth-functional Principles*	
(T-Out)	$T(\langle p\rangle) \rightarrow p$	(~-Imb)	$\sim T(\langle p\rangle) \rightarrow T(\langle \sim p\rangle)$[5]
(T-In)	$p \rightarrow T(\langle p\rangle)$	(~-Exc)	$T(\langle \sim p\rangle) \rightarrow \sim T(\langle p\rangle)$[6]
(T-Elim)	$T(\langle p\rangle) \vdash p$	($\wedge$-Imb)	$T(\langle p\rangle) \wedge T(\langle q\rangle) \rightarrow T(\langle p \wedge q\rangle)$
(T-Intro)	$p \vdash T(\langle p\rangle)$	($\wedge$-Exc)	$T(\langle p \wedge q\rangle) \rightarrow T(\langle p\rangle) \wedge T(\langle q\rangle)$
(~T-Elim)	$\sim T(\langle p\rangle) \vdash \sim p$	($\vee$-Imb)	$T(\langle p\rangle) \vee T(\langle q\rangle) \rightarrow T(\langle p \vee q\rangle)$
(~T-Intro)	$\sim p \vdash \sim T(\langle p\rangle)$	($\vee$-Exc)	$T(\langle p \vee q\rangle) \rightarrow T(\langle p\rangle) \vee T(\langle q\rangle)$
(Cat)	$\vdash p \rightarrow \vdash T(\langle p\rangle)$	($\rightarrow$-Imb)	$(T(\langle p\rangle) \rightarrow T(\langle q\rangle)) \rightarrow T(\langle p \rightarrow q\rangle)$
(Co-Cat)	$\vdash T(\langle p\rangle) \rightarrow \vdash p$	($\rightarrow$-Exc)	$T(\langle p \rightarrow q\rangle) \rightarrow (T(\langle p\rangle) \rightarrow T(\langle q\rangle))$

Quantificational Principles *Misc. Principles*

[4] Quine (1960: 271).
[5] 'Imb' is short for 'imbibe'.
[6] 'Exc' is short for 'excrete'.

(∀-Imb) $(\forall x)T(\langle\phi(x)\rangle) \rightarrow T(\langle(\forall x)\phi(x)\rangle)$

(∀-Exc) $T(\langle(\forall x)\phi(x)\rangle) \rightarrow (\forall x)T(\langle\phi(x)\rangle)$

(∃-Imb) $(\exists x)T(\langle\phi(x)\rangle) \rightarrow T(\langle(\exists x)\phi(x)\rangle)$

(∃-Exc) $T(\langle(\exists x)\phi(x)\rangle) \rightarrow (\exists x)T(\langle\phi(x)\rangle)$

(Taut) $T(\langle\top\rangle)$

(Contra) $\sim T(\langle\bot\rangle)$

(T-Del) $T(\langle T(\langle p\rangle)\rangle) \rightarrow T(\langle p\rangle)$

(T-Rep) $T(\langle p\rangle) \rightarrow T(\langle T(\langle p\rangle)\rangle)$

(TT) $T(\langle T(\langle p\rangle \rightarrow p\rangle)$

Implication Principles

(MPC) $(p_1 \wedge \ldots \wedge p_n \rightarrow q) \rightarrow (T(\langle p_1\rangle) \wedge \ldots \wedge T(\langle p_n\rangle) \rightarrow T(\langle q\rangle))$

(SPC) $(p \rightarrow q) \rightarrow (T(\langle p\rangle) \rightarrow T(\langle q\rangle))$

(Sub-in) $p \leftrightarrow q \rightarrow T(\langle p\rangle) \leftrightarrow T(\langle q\rangle)$

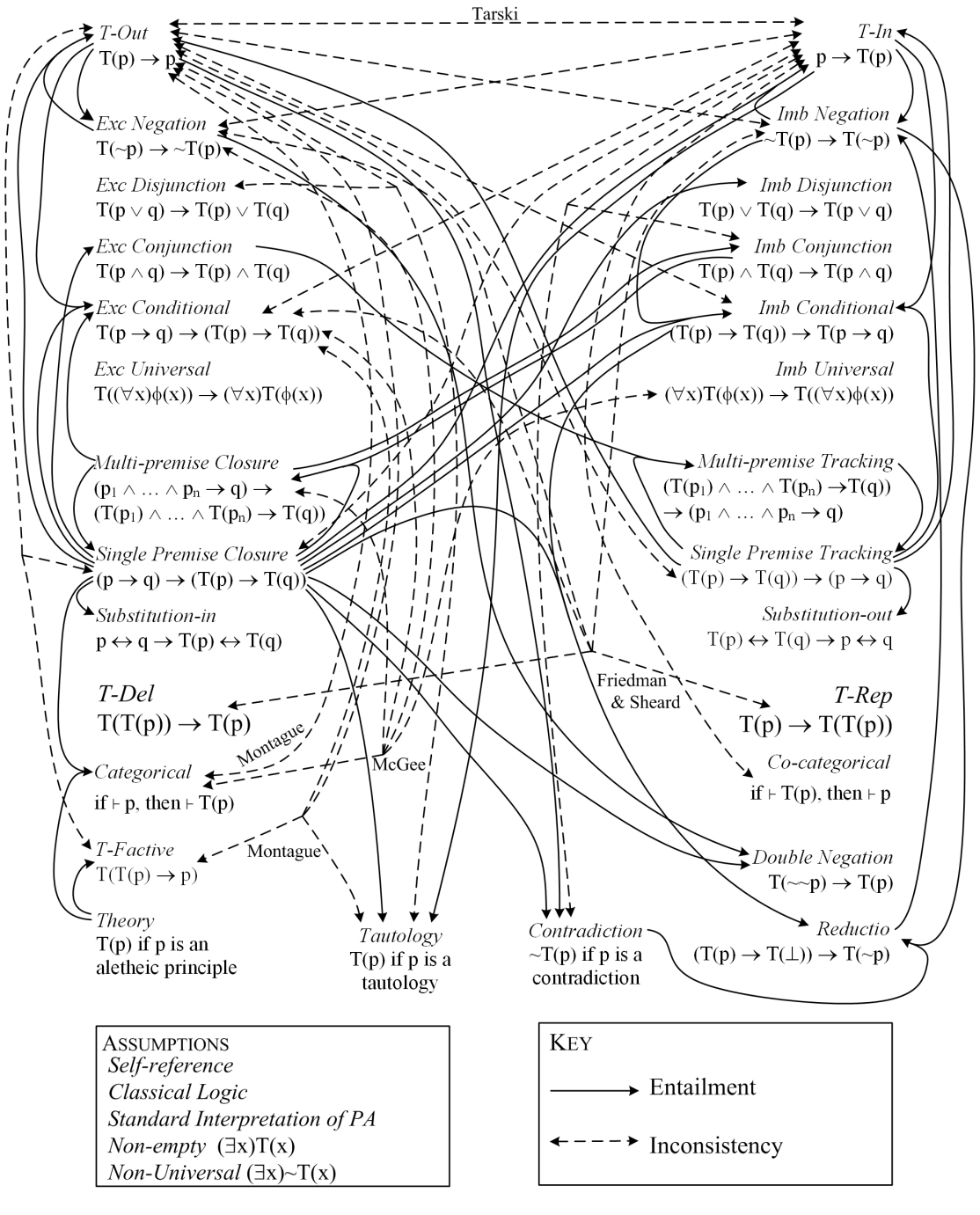
(MPT) $(T(\langle p_1\rangle) \wedge \ldots \wedge T(\langle p_n\rangle) \rightarrow T(\langle q\rangle)) \rightarrow (p_1 \wedge \ldots \wedge p_n \rightarrow q)$

(SPT) $(T(\langle p\rangle) \rightarrow T(\langle q\rangle)) \rightarrow (p \rightarrow q)$

(Sub-out) $T(\langle p\rangle) \leftrightarrow T(\langle q\rangle) \rightarrow (p \rightarrow q)$

There is a very complex constellation of inconsistency results and entailments among these principles. See Information Box 30 for some of them (note that the diagram is not exhaustive—one should not infer that all the entailments and incompatibilities are depicted).

Entailments and Inconsistencies Among Aletheic Principles Information Box 30



The goal is to have ascending truth and descending truth be as close as possible to truth without engendering paradoxes of any kind or requiring a weakening of logic. Moreover, on sentences that

do not pose any kind of threat (e.g., sentences not containing semantic predicates), ascending and descending truth should obey all these principles. It turns out that there are several serious obstacles to satisfying these demands, and some tough choices have to be made.

13.3 Montague's Theorem

One might think that once one has replaced truth with ascending and descending truth, our problems with paradoxes are over. Not so. In fact there are other hurdles to clear, and the most difficult is a theorem Montague proved in 1963 that has had much more impact on the philosophical discussion of necessity than the discussion of truth. Montague proved that a theory of some predicate H(x) with the following features is inconsistent:

(i) All instances of 'H(⟨p⟩) → p' are theorems.

(ii) All instances of 'H(⟨H(⟨p⟩) → p⟩)' are theorems.

(iii) All instances of 'H(⟨p⟩)' where ⟨p⟩ is a logical axiom are theorems.

(iv) All instances of 'H(⟨p→q⟩) → (H(⟨p⟩)→H(⟨q⟩)) are theorems.

(v) Q (i.e., Robinson arithmetic) is a subtheory.[7]

Condition (v) is present to ensure that the language in which the theory is expressed has the ability to refer to its own sentences. The other four conditions are highly desirable for *descending* truth. On this reading, note that (i) is just (T-Out), (ii) says that all instances of (T-Out) are descending true, (iii) says that all tautologies are descending true, and (iv) says that descending truth is closed under modus ponens (i.e., if a conditional is descending true and its antecedent is descending true, then its

[7] Montague (1963).

consequent is descending true). Montague's theorem shows that if descending truth is a consistent concept, then it does not obey all four of these principles. Since I am taking (T-Out) to be constitutive of descending truth, my options are to deny (ii), deny (iii), or deny (iv). Denying (ii) results in a theory of descending truth that is not descending true, which is a version of a revenge paradox (self-refutation problem). That leaves (iii) or (iv). We have already seen (in Chapter Five) that no logical approach to the paradoxes is compatible with the claim that valid arguments are necessarily truth preserving. The same considerations sink any attempt to define validity in terms of descending truth. So, any theory of descending truth will have to admit that descending truth is not closed under some deducibilities. However, it is open to say that all logical truths are descending true. Thus, it makes the most sense to reject (iv) and accept (iii). As such, I stipulate that all classical tautologies are descending true; it follows by Montague's theorem that descending truth is not closed under modus ponens.

We know several things about ascending and descending truth already. First, we are using classical logic and we are not restricting the expressive resources of the languages we consider. So we know that we will have to deal with sentences like the following:

(α) $\sim A(\alpha)$

(δ) $\sim D(\delta)$

From (δ) and the fact that D(x) obeys (T-Out) we can prove $\sim D(\delta)$ and $\sim D(\text{`}\sim\delta\text{'})$; from ($\alpha$) and the fact that A(x) obeys (T-In), we can prove $A(\alpha)$ and $A(\text{`}\sim\alpha\text{'})$. So we know that there are some sentences such that they and their negations are ascending true and there are some sentences such that they and their negations are not descending true.

One rather tricky issue is the relation between ascending truth and descending truth (if any). Consider the relation between a sentence p, 'p is ascending true' and 'p is descending true'. p

follows from 'p is descending true' but not vice versa; hence, 'p is descending true' is stronger than p. On the other hand, 'p is ascending true' follows from p, but not vice versa; hence p is stronger than 'p is ascending true'. Given these claims, 'it is not the case that p is descending true' is weaker than ⌜~p⌝ and 'it is not the case that p is ascending true' is stronger than ⌜~p⌝. So, what is the relation between 'p is ascending true' and 'it is not the case that ⌜~p⌝ is descending true'? Further, what is the relation between 'p is descending true' and 'it is not the case that ⌜~p⌝ is ascending true'? I stipulate that, in both cases, they are equivalent.[8] That is, $A(\langle \mathsf{p} \rangle) \leftrightarrow \sim D(\langle \sim \mathsf{p} \rangle)$ and $D(\langle \mathsf{p} \rangle) \leftrightarrow \sim A(\langle \sim \mathsf{p} \rangle)$. Thus, ascending truth and descending truth are *dual* predicates.[9] They have the same relation that obtains between possibility and necessity, between permission and obligation, between consistency and provability, etc. We will see below how this assumption plays out in formal treatments of ascending and descending truth.

Above, I chose to have tautologies be descending true over having descending truth be closed under modus ponens. So we know that tautologies are descending true. Likewise, by duality contradictions are not ascending true.

It is pretty straightforward to add principles for negation to each concept. We can say that descending truth obeys (~-Exc) and ascending truth obeys (~-Imb), but not vice versa. Given what has been said already, we know that ascending truth does not obey (∧-Exc) and that descending truth does not obey (∧-Imb). However, we can have descending truth obey (∧-Exc) and (∨-Imb) and ascending truth obey (∧-Imb) and (∨-Exc).

Since we do not want to block the resources for generating self-reference in any way, it makes sense to require that the theories that ensure the kinds of expressive resources needed to construct

---

[8] There are alternative ways of defining descending truth and ascending truth, but I do not consider them in this work.
[9] Many thanks to Dana Scott who impressed upon me the importance of duality in the theory of ascending and descending truth.

potentially paradoxical sentences are descending true. That is, the axioms of a theory of syntax are descending true. Also, since the most important way to achieve self-reference in mathematical theories is via arithmetization (i.e., Gödel numbering), we need at least the axioms of PA to be descending true.

Finally, given the choice made above in light of Montague's theorem, we want all the axioms of the theory of ascending and descending truth to be descending true.

13.4 Safety

We have already said that descending truth obeys the principle '$\mathrm{D}(\langle \mathsf{p} \rangle) \rightarrow \mathsf{p}$' and ascending truth obeys the principle '$\mathsf{p} \rightarrow \mathrm{A}(\langle \mathsf{p} \rangle)$'. However, neither can obey the inverse principle upon pain of contradiction. However, the potential problems raised by the inverse principles arise only for a few sentences like (α) and (δ). It does no harm (and a lot of good) to let descending truth obey a restricted version of (T-In) and let ascending truth obey a restricted version of (T-Out). We can formulate them in the following way:

(M3) $\mathrm{S}(\langle \mathsf{p} \rangle) \wedge \mathsf{p} \rightarrow \mathrm{D}(\langle \mathsf{p} \rangle)$

(M4) $\mathrm{A}(\langle \mathsf{p} \rangle) \wedge \mathrm{S}(\langle \mathsf{p} \rangle) \rightarrow \mathsf{p}$

where 'S(x)' is a predicate that stands for 'safety'. Intuitively, a safe sentence is one for which both directions of the principles for ascending truth and for descending truth hold. Unsafe sentences are those for which they do not.

Using the defining principles for safety and for ascending truth and descending truth we can derive the following principle of safety:

(M2) $S(\langle \mathbf{p} \rangle) \leftrightarrow D(\langle \mathbf{p} \rangle) \vee \sim A(\langle \mathbf{p} \rangle)$

That is, a safe sentence is either descending true or not ascending true. Conversely, an unsafe sentence is both ascending true and not descending true. A consequence of this result is a clearer picture of the relation between descending truth and ascending truth. We know that any sentence that is descending true is ascending true. From this it also follows that any sentence that is not ascending true is not descending true. Moreover, some sentences are ascending true and not descending true and no sentence is both descending true and not ascending true.

Given the guiding analogy between the concept of truth and the concept of mass, and the fact that I want to be able to explain when one can use 'true' without running into problems, it makes sense to have several additional constraints on safety. It is acceptable to use 'mass' if and only if one is dealing with a situation in which the difference between relativistic mass and proper mass is negligible. Likewise, it is acceptable to use 'true' if and only if one is dealing with a situation in which the difference between ascending truth and descending truth is negligible. These are exactly the situations in which one is dealing with safe sentences. Thus, one should expect that if we restrict our attention to them, then ascending truth and descending truth obey all the principles we take truth to obey (i.e., all the aletheic principles from section 13.2).

13.5 A Formal Theory: ADT

Given what has been said above, we can summarize the principles that any theory of ascending truth and descending truth should include. I do not require such a theory to be axiomatizable, so the following is not meant to be *the* theory of ascending truth and descending truth; rather any theory of ascending truth and descending truth should have the following as a subtheory. I call the following theory ADT:

D1 $D(\langle\phi\rangle) \rightarrow \phi$
D2 $D(\langle\sim\phi\rangle) \rightarrow \sim D(\langle\phi\rangle)$
D3 $D(\langle\phi\wedge\psi\rangle) \rightarrow D(\langle\phi\rangle) \wedge D(\langle\psi\rangle)$
D4 $D(\langle\phi\rangle) \vee D(\langle\psi\rangle) \rightarrow D(\langle\phi\vee\psi\rangle)$
D5 $D(\langle\phi\rangle)$ if ϕ is a logical truth (i.e., a tautology of first order predicate calculus)
D6 $D(\langle\phi\rangle)$ if ϕ is a theorem of PA
D7 $D(\langle\phi\rangle)$ if ϕ is an axiom of ADT (i.e., if ϕ is an instance of D1-D6, A1-A6, M1-M4, or E1-E3)

A1 $\phi \rightarrow A(\langle\phi\rangle)$
A2 $\sim A(\langle\phi\rangle) \rightarrow A(\langle\sim\phi\rangle)$
A3 $A(\langle\phi\rangle) \vee A(\langle\psi\rangle) \rightarrow A(\langle\phi\vee\psi\rangle)$
A4 $A(\langle\phi\wedge\psi\rangle) \rightarrow A(\langle\phi\rangle) \wedge A(\langle\psi\rangle)$
A5 $\sim A(\langle\phi\rangle)$ if ϕ is a logical falsity (i.e., a contradiction of first order predicate calculus)
A6 $\sim A(\langle\phi\rangle)$ if ϕ is the negation of an axiom of PA[10]

M1 $D(\langle\phi\rangle) \leftrightarrow \sim A(\langle\sim\phi\rangle)$
M2 $S(\langle\phi\rangle) \leftrightarrow (D(\langle\phi\rangle) \vee \sim A(\langle\phi\rangle))$
M3 $\phi \wedge S(\langle\phi\rangle) \rightarrow D(\langle\phi\rangle)$
M4 $A(\langle\phi\rangle) \wedge S(\langle\phi\rangle) \rightarrow \phi$

E1 If $\sigma = \tau$ and ψ results from replacing some occurrences of σ with τ in ϕ, then $D(\phi)\leftrightarrow D(\psi)$.
E2 If $\sigma = \tau$ and ψ results from replacing some occurrences of σ with τ in ϕ, then $A(\phi)\leftrightarrow A(\psi)$.
E3 If $\sigma = \tau$ and ψ results from replacing some occurrences of σ with τ in ϕ, then $S(\phi)\leftrightarrow S(\psi)$.

One question that naturally arises is: is this theory consistent or are there new paradoxes hiding in here? Given Gödel's second incompleteness theorem, all we can hope for is a proof of relative consistency (i.e., if some uncontroversial theory is consistent, then ADT is consistent), but given the extreme difficulty with saying anything at all consistent about the liar paradox, a relative consistency proof seems in order, which is given in the appendix to this chapter and is based on the semantics presented in the next section.[11]

[10] Axioms D5, D6, A5, and A6 hold for sentences that contain 'ascending true' and 'descending true'.
[11] My appreciation goes again to Dana Scott for his help in formulating and defending ADT.

13.6 Semantics for ADT

Because of M1, D(x) and A(x) are *dual* predicates—they have the same relationship that obtains between possibility and necessity, between obligation and permission, between provability and consistency, etc. For this reason, it makes sense to model their behavior using modal logic. This insight is the basis for the semantics given in this section.

Throughout this discussion it is essential to keep in mind the distinction between the theory ADT and its semantics. The theory is the set of theorems, where a theorem is either an axiom or a sentence that is deducible from the axioms by some combination of rules and classical logic. The list above contains the axioms. The semantics is a mathematical structure that we can use to prove certain things about the theory. In every case, the semantics uses a structure that is definable in set theory. A crucial part of the semantics is the definition of a valid sentence. Once we have the definition of a theorem (for the theory) and a definition of a validity (for the semantics), we can prove results about how they relate. For example, one might want to prove that the theory is sound with respect to the semantics, which requires showing that every theorem of the theory is a validity of the semantics. Or, more generally, if a formula ϕ is provable from a set of formulas Γ, then the argument from Γ to ϕ is valid. Once one has a soundness result, one can be sure that anything that is provable from the axioms will be valid in the semantics. In particular, one has a relative consistency proof for the theory in question. On the other hand, one might want to prove that the theory is complete with respect to the semantics, which requires showing that every validity of the semantics is a theorem of the theory (more generally, if ϕ is a consequence of Γ, then ϕ is provable from Γ).

13.6.1 Normal Modal Logic and Relational Semantics

A *normal* modal logic has the following form. Let L be a classical sentential language with the usual connectives and a 1-place operator '$\Box$' on sentences (i.e., if ϕ is a sentence then $\Box\phi$ is a sentence). All normal modal logics have axioms and rules. The axioms include the K axiom, which is:

(K) $\Box(\phi\rightarrow\psi) \rightarrow (\Box\phi\rightarrow\Box\psi)$.

Many normal modal logics include other axioms as well. All normal modal logics include the Necessitation rule:

(Nec) If $\vdash \phi$, then $\vdash \Box\phi$.

In a normal modal logic, we can define the '$\Diamond$' operator in terms of the '$\Box$' operator in the following way:

(Duality) $\Diamond\phi \leftrightarrow \sim\Box\sim\phi$[12]

Every normal modal logic has Duality as a theorem. Notice that Duality is similar to M1 of ADT and this similarity is the inspiration for appealing to modal logic to model ADT.

Let us turn to the semantics for normal modal logics. Let W be a set of worlds and let R be a relation on W (called the *accessibility relation*). Together, W and R are called a *relational frame*, $\mathfrak{F}$=<W, R>. I call any semantics based on a frame of this kind a *relational semantics*. A valuation function, V, assigns to each sentential variable of L a truth value at each world in W. Together, F and V are called a *relational model*, $\mathfrak{M}$=<$\mathfrak{F}$, V> for L. Each world in W is classical in that a classical scheme determines the value of each truth-functionally compound sentence. That gives us the following clauses for defining truth at a world in a model (i.e., <$\mathfrak{M}$, w>$\models\phi$):

(ϕ) <$\mathfrak{M}$, w>$\models\phi$ if and only if w$\in$V(ϕ) for ϕ atomic

($\sim$) <$\mathfrak{M}$, w>$\models\sim\phi$ if and only if it is not the case that <$\mathfrak{M}$, w>$\models\phi$

[12] This principle fails in intuitionistic logic, but I assume classical logic in what follows.

($\wedge$) $\langle\mathfrak{M}, w\rangle\models\phi\wedge\psi$ if and only if $\langle\mathfrak{M}, w\rangle\models\phi$ and $\langle\mathfrak{M}, w\rangle\models\psi$

($\vee$) $\langle\mathfrak{M}, w\rangle\models\phi\vee\psi$ if and only if $\langle\mathfrak{M}, w\rangle\models\phi$ or $\langle\mathfrak{M}, w\rangle\models\psi$

($\rightarrow$) $\langle\mathfrak{M}, w\rangle\models\phi\rightarrow\psi$ if and only if if $\langle\mathfrak{M}, w\rangle\models\phi$, then $\langle\mathfrak{M}, w\rangle\models\psi$

($\leftrightarrow$) $\langle\mathfrak{M}, w\rangle\models\phi\leftrightarrow\psi$ if and only if $\langle\mathfrak{M}, w\rangle\models\phi$ iff $\langle\mathfrak{M}, w\rangle\models\psi$

The clause for sentences of the form $\Box\phi$ is:

($\Box$) $\langle\mathfrak{M}, w\rangle\models\Box\phi$ if and only if $\forall u\in W$ if Rwu, then $\langle\mathfrak{M}, u\rangle\models\phi$

(i.e., $\Box\phi$ is true at w if and only if ϕ is true at all worlds accessible from w). These clauses constitute an inductive definition of truth at a world (i.e., $\langle\mathfrak{M}, w\rangle\models\phi$) whatever the complexity of ϕ. A sentence ϕ is *valid in a model* $\mathfrak{M}$ (i.e., $\mathfrak{M}\models\phi$) if and only if $\forall w\in W$ $\langle\mathfrak{M}, w\rangle\models\phi$. A sentence ϕ is *valid on a frame* $\mathfrak{F}$ (i.e., $\mathfrak{F}\models\phi$) if and only if for all $\mathfrak{M}$ based on $\mathfrak{F}$ $\forall w\in W$, $\langle\mathfrak{M}, w\rangle\models\phi$.

If our language contains a second modal operator, $\Diamond$, then we include a separate clause, which would read:

($\Diamond$) $\langle\mathfrak{M}, w\rangle\models\Diamond\phi$ if and only if $\exists u\in W$ s.t. Rwu and $\langle\mathfrak{M}, u\rangle\models\phi$.

Notice that given these clauses for $\Box$ and $\Diamond$, Duality is valid in any relational semantics.

By imposing constraints on the accessibility relation in a relational semantics, one can control which sentences are valid on the resulting frame. For example, a *reflexive* frame is any frame where the accessibility relation is reflexive (i.e., $\forall w, Rww$). Any sentence of the form $\Box\phi\rightarrow\phi$ is valid on

the class of reflexive frames. Any model that is to serve as a semantics for a theory that includes $\Box\phi\rightarrow\phi$ will be based on a reflexive frame.[13]

13.6.2 Problems with Using Relational Semantics for ADT

There are several problems with using normal modal logic and its semantics for ADT.

Problem #1: every normal modal logic includes the K axiom and every relational semantics validates the K axiom. However, Richard Montague proved that the predicate version of the K axiom (i.e., $\forall\phi\forall\psi(\Box(\langle\phi\rightarrow\psi\rangle)\rightarrow(\Box(\langle\phi\rangle)\rightarrow\Box(\langle\psi\rangle))$ is inconsistent with the combination of D1, D5, D6, and D7. Thus, no semantics that validates the K axiom can serve as a semantics for ADT.[14]

Problem #2: every normal modal logic includes the Necessitation rule and every relational semantics validates the Necessitation rule. However, Montague proved that the predicate version of the Necessitation rule (i.e., if ϕ is a theorem, then $\Box(\langle\phi\rangle)$ is a theorem) is inconsistent with the combination of D1 and D6. Thus, no semantics that validates the Necessitation rule can serve as a semantics for ADT.[15]

Problem #3: normal modal logic is a *sentential* modal logic (i.e., it deals with whole sentences and operators on whole sentences), but ADT is a theory of predicates, so I need a *first-order* modal logic that deals with parts of sentences, like predicates. First order *normal* modal logic is well understood. We expand our language to include individual constants, individual variables, n-place predicates, and quantifiers. We expand our semantics with a domain, and treat individual constants, variables, n-place predicates, and quantifiers in the usual way. The major issue here is how to set up the domain of quantification—should every world have the same domain or should the domain differ from world to world? The former is called *constant-domain semantics* and the latter is called *variable-domain semantics*. However, given problems 1 and 2, even first-order normal modal logic is inadequate for ADT.

Problem #4: normal modal logics and relational semantics are designed for *operators* (an operator takes a sentence as input and has a sentence as output), but according to ADT, ascending truth and descending truth are *predicates* (a predicate takes a singular term as input and has a sentence as output). One major difference between operators (e.g., 'it is true that') and predicates (e.g., 'is true') is their expressive power. For example, one cannot use an operator to construct a self-referential sentence, but it is possible with a predicate; in addition, operators apply only to sentences of the same language, whereas predicates can apply to anything that can be referred to in the same language. Finally, quantification into a predicate is well-understood and uncontroversial, while quantification into an operator is contentious and complicated.

---

[13] See Chellas (1980), Fitting and Mendelsohn (1998), Portner (2009), and Garson (2009) for background on modal logic.
[14] Montague (1963).
[15] Montague (1963).

These are *major* problems. I deal with them in this order.

13.6.3 Classical Modal Logic and Neighborhood Semantics

Problem #1 and problem #2 can be remedied by using a more general semantics for modal logics—neighborhood semantics. Again, L be a sentential language with the usual connectives and operators. Let W be a set of worlds, but let N be a function from W to the set of sets of subsets of W (N is called the *neighborhood function*); so N assigns a set of subsets of W to each world in W. Together, W and N are called a *neighborhood frame*, $\mathfrak{F}$=<W, N>. I call any semantics based on a frame of this kind a *neighborhood semantics*. Just as in relational semantics, a valuation function, V, assigns to each sentential variable of L a truth value at each world in W. Together, $\mathfrak{F}$ and V are called a *neighborhood model*, $\mathfrak{M}$=<$\mathfrak{F}$, V> for L. Each world in W is classical in the sense that a classical scheme determines the value of truth-functionally compound sentences. Thus, we can keep all the previous clauses for defining truth at a world in a model except the (□) clause, which becomes:

(□) <$\mathfrak{M}$, w>⊨□ϕ if and only if ∃X∈N(w) s.t. ∀u∈W (<$\mathfrak{M}$, u>⊨ϕ ↔ u∈X).

(i.e., □ϕ is true at w iff a neighborhood of w contains all the worlds at which ϕ is true). One can think of the neighborhoods assigned to a world, w, as a list of the sets of worlds that are assigned to sentences that are necessary at w. That is, if we think of the set of worlds in which a sentence ϕ is true as the proposition expressed by ϕ (symbolized as P(ϕ)), then we can rephrase the (□) clause as:

(□) <$\mathfrak{M}$, w>⊨□ϕ if and only if P(ϕ)∈N(w)

In neighborhood semantics, the clause for ◇ becomes:

(◇) <$\mathfrak{M}$, w>⊨◇ϕ if and only if ~(∃X∈N(w) s.t. ∀u∈W (<$\mathfrak{M}$, u>⊨~ϕ ↔ u∈X)).

or:

($\Diamond$) $\langle\mathfrak{M}, w\rangle\models\Diamond\phi$ if and only if $P(\sim\phi)\notin N(w)$.

Validity is defined just as it was in relational semantics.

It is relatively easy to show that there are neighborhood frames on which the K axiom is invalid. For example, let $\mathfrak{M}_k$ be the neighborhood model with $W_k=\{w, x, y, z\}$, $N_k(w)=\{\{w, x\}, \{w, y, z\}\}$, $N_k(x)=\{\{x\}\}$, $N_k(y)=\{\{y\}\}$, $N_k(z)=\{\{z\}\}$, $V_k(P)=\{w, x\}$, and $V_k(Q)=\{w, y\}$. Then we have: $\langle\mathfrak{M}_k, w\rangle\models\Box P$, $\langle\mathfrak{M}_k, w\rangle\models\Box(P\rightarrow Q)$, but $\langle\mathfrak{M}_k, w\rangle\not\models\Box Q$. So the K axiom is invalid in $\mathfrak{M}_k$. This solves problem #1. One can impose conditions on the neighborhood function to ensure that the K axiom is valid, but since we want it to be invalid, we will not consider these.

Likewise, is relatively easy to show that there are neighborhood frames on which the Necessitation rule is invalid. For example, let $\mathfrak{M}_n$ be the neighborhood model with $W_n=\{w, x\}$, $N_n(w)=\{\{w\}\}$, $N_n(x)=\{\{x\}\}$, $V_n(P)=\{w, x\}$. Then we have: $\mathfrak{M}_n\models P$ but $\mathfrak{M}_n\not\models\Box P$. So the Necessitation rule is invalid in $\mathfrak{M}_n$. This solves problem #2. One can impose conditions on the neighborhood function to ensure that the Necessitation rule is valid, but since we want it to be invalid, we will not consider these.

Even though it is not the case that the K axiom is valid in any neighborhood frame, and it is not the case that the Necessitation rule is valid in any neighborhood frame, there are axioms and rules that are valid in any neighborhood frame. Duality is an axiom that is valid on any neighborhood frame, and the following rule is valid on any neighborhood frame as well:

(E) If $\vdash\phi\leftrightarrow\psi$, then $\vdash\Box\phi\leftrightarrow\Box\psi$.

Call any modal logic that includes the Duality axiom and rule E a *classical* modal logic. Notice that rule E is similar to the Necessitation rule, but it is weaker. So, the move from normal modal logics and their relational semantics to classical modal logics and their neighborhood semantics solves

problem #1 and problem #2. That is, this move avoids the K axiom and the Necessitation rule, both of which would render ADT inconsistent.

13.6.4 Yet Another Problem

Although the move from normal modal logic to classical modal logic solves problem #1 and problem #2, it presents us with a new problem that is similar to problem #2.

> *Problem #5*: We want to replace our operator, $\Box$, with a predicate for descending truth, D(x). To do that, we need to move from a sentential language to a first order language. We know that when we make the move to first order logic and predicates, Gödel's Diagonalization Lemma guarantees that if our language can express Peano Arithmetic or its own theory of syntax (these are pretty minimal expressive constraints), then it will have a sentence δ s.t. $\sim D(\delta)$ is provably equivalent to δ. We know that $\vdash_{ADT} \sim D(\delta)$ [assume $D(\delta)$; if $D(\delta)$, then δ; if δ, then $\sim D(\delta)$; so if $D(\delta)$, then $\sim D(\delta)$; thus, $\sim D(\delta)$]. We also know that $\vdash_{ADT} 0=0$. So we have $\vdash_{ADT} \sim D(\delta) \leftrightarrow 0=0$. By rule (E) we would get $\vdash_{ADT} D(\delta) \leftrightarrow D(\text{`}0=0\text{'})$. We already have $\vdash_{ADT} D(\text{`}0=0\text{'})$; so we would have $\vdash_{ADT} D(\delta)$. $\bot$. Therefore, ADT cannot include rule E.[16] This argument shows that rule E is incompatible with D1 and D5; similar arguments show that rule E is also incompatible with D1 and D6, and that it is incompatible with D1 and D7. These results show that no semantics that validates rule E will work for ADT.

13.6.5 Xeno Semantics

At this point, we have left the well-traveled paths of modal logic and are off on our own. What we need is something more general than neighborhood semantics; as far as I know, there is no such thing. So we will have to break new ground to solve problem #5.

In relational semantics, the extension of '$\Box$' at each world is determined by a binary accessibility relation on the set of worlds; we can think of this as a function that assigns each world a set of worlds (i.e., those accessible from it). Moreover, each sentence is assigned a proposition, which is a

[16] Dana Scott first noticed this problem.

set of worlds (i.e., those in which it is true). In neighborhood semantics, the extension of '$\Box$' at each world is determined by a function that assigns each world a set of sets of worlds; as in relational semantics each sentence is assigned a proposition, which is a set of worlds. In the new semantics, which I call *xeno semantics*, the extension of '$\Box$' at each world is determined by both an accessibility relation *and* a neighborhood function.[17] As before, each sentence is assigned a set of worlds as its proposition. However, although the neighborhood function is unchanged, the key to xeno semantics is that the accessibility relation is relative to each type of sentence in the language. Indeed, we can think of xeno semantics as involving as many binary accessibility relations as there are syntactic types of sentences.

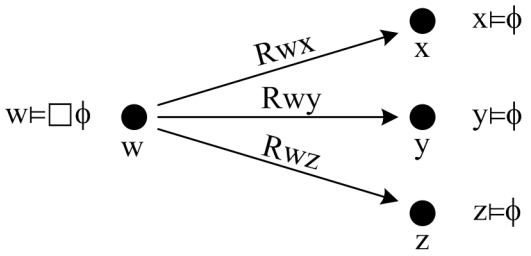
In *relational* semantics '$\Box\phi$' is true at a world w if and only if ϕ is true at all worlds accessible from w. In *neighborhood* semantics, '$\Box\phi$' is true at a world w if and only if the set of worlds in which ϕ is true is a neighborhood of w. In *xeno* semantics, '$\Box\phi$' is true at a world w if and only if the set of worlds in which ϕ is true is a neighborhood of all worlds accessible$_\phi$ from w, where 'accessible$_\phi$' is the accessibility relation assigned to ϕ's syntactic type. So one can think of xeno semantics as a blend of relational semantics and neighborhood semantics with a relativization to syntactic types. In xeno semantics, each sentence is assigned a proposition (a set of worlds) and a relation on the set of worlds. We can think of this as a sentence granting accessibility from one world to others, or we can say that the accessibility relation is relative to each sentence. Moreover, the accessibility relation alone does not determine the extension of $\Box$ at each world; rather, together the accessibility relation and the neighborhood relation determine the extension of $\Box$ *for that particular sentence* at each world. Alternatively, we can think of a proposition as a pair of a subset of W and a relation on W. But

[17] Xeno semantics is named after our dog; thanks to Alison Duncan Kerr for the suggestion.

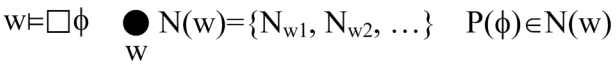
neighborhoods of a world are still just subsets of W. □'s extension at a world is then an operation on propositions, and it is determined by the whole neighborhood function, not just the neighborhoods of that world. Information Box 31 illuminates the three kinds of semantics.

Three Kinds of Possible-Worlds Semantics Information Box 31

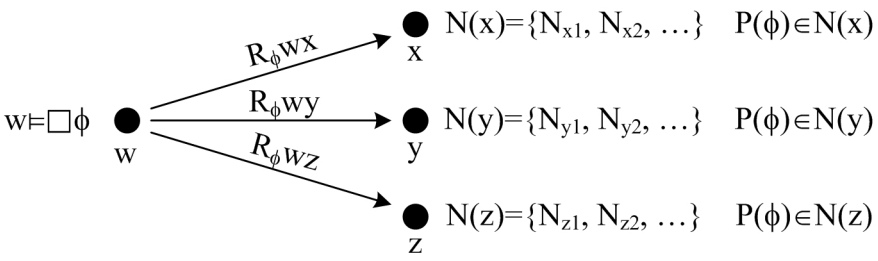
Relational Semantics



Neighborhood Semantics



Xeno Semantics



We need to define the syntactic type of a sentence. Let the formation rules of L be the usual ones (since it has the usual connectives and a single operator). Let any two sentences that have the same syntactic decomposition into components according to the formation rules be of the same syntactic type. So, syntactic types are equivalence classes of sentences. For example, if ϕ and ψ are

distinct sentential variables then $\phi\wedge\psi\rightarrow\psi$ and $\phi\wedge\psi\rightarrow\phi$ have the same syntactic type whereas $\phi\rightarrow\psi$ differs.

Now that we have the basic idea for xeno semantics, I am going to provide a particular xeno semantics for ADT. This will be accomplished in stages. First, I provide a xeno semantics for a classical sentential language with descending truth operator, $\Box$, an ascending truth operator, $\Diamond$ and a safety operator, Σ; then we switch from a sentential language to a first-order language; finally, we consider a xeno semantics for a first order language with a descending truth *predicate*, an ascending truth *predicate*, and a safety *predicate*.

Let $\mathcal{L}$ be a sentential language with the usual connectives and three sentential operators: $\Box$, $\Diamond$, and Σ. Let L be the set of well-formed formulas of $\mathcal{L}$. Let a *xeno frame* $\mathfrak{F}$=<W, R, N> where W is a set of worlds, R is a denumerable set of binary relations on W, and N is a neighborhood function from W to 2^{2^W}.

Let a *xeno model* M=<$\mathfrak{F}$, $\mathfrak{R}$, V> where $\mathfrak{F}$ is a xeno frame, $\mathfrak{R}$ is a function from L to R, and V is a function from the sentential variables of L to 2^W. $\mathfrak{R}$ assigns an accessibility relation to each sentence of L, and V assigns a set of worlds to each sentential variable (R_ϕ is the accessibility relation $\mathfrak{R}$ assigns to ϕ).

We can give an inductive definition of truth at a world (i.e., $\langle\mathfrak{M}, w\rangle\vDash\phi$) in the following way:

(ϕ) $\langle\mathfrak{M}, w\rangle\vDash\phi$ if and only if $w\in V(\phi)$ (for ϕ atomic)

($\sim$) $\langle\mathfrak{M}, w\rangle\vDash\sim\phi$ if and only if it is not the case that $\langle\mathfrak{M}, w\rangle\vDash\phi$

($\wedge$) $\langle\mathfrak{M}, w\rangle\vDash\phi\wedge\psi$ if and only if $\langle\mathfrak{M}, w\rangle\vDash\phi$ and $\langle\mathfrak{M}, w\rangle\vDash\psi$

($\vee$) $\langle\mathfrak{M}, w\rangle\vDash\phi\vee\psi$ if and only if $\langle\mathfrak{M}, w\rangle\vDash\phi$ or $\langle\mathfrak{M}, w\rangle\vDash\psi$

($\rightarrow$) $\langle\mathfrak{M}, w\rangle \vDash \phi \rightarrow \psi$ if and only if if $\langle\mathfrak{M}, w\rangle \vDash \phi$, then $\langle\mathfrak{M}, w\rangle \vDash \psi$

($\leftrightarrow$) $\langle\mathfrak{M}, w\rangle \vDash \phi \leftrightarrow \psi$ if and only if $\langle\mathfrak{M}, w\rangle \vDash \phi$ iff $\langle\mathfrak{M}, w\rangle \vDash \psi$

($\Box$) $\langle\mathfrak{M}, w\rangle \vDash \Box\phi$ if and only if $\forall u \in W\ R_{\phi}wu \rightarrow \exists X \in N(u) \forall v \in W(\langle\mathfrak{M}, v\rangle \vDash \phi \leftrightarrow v \in X)$

or:

($\Box$) $\langle\mathfrak{M}, w\rangle \vDash \Box\phi$ if and only if $\forall u \in W\ R_{\phi}wu \rightarrow P(\phi) \in N(u)$

where $P(\phi)$ is the set of worlds in which ϕ is true.

We can introduce a definition for the dual operator $\Diamond$ in the following way:

($\Diamond$) $\langle\mathfrak{M}, w\rangle \vDash \Diamond\phi$ if and only if $\sim(\forall u \in W\ R_{\sim\phi}wu \rightarrow P(\sim\phi) \in N(u))$

or:

($\Diamond$) $\langle\mathfrak{M}, w\rangle \vDash \Diamond\phi$ if and only if $\exists u \in W\ R_{\sim\phi}wu \wedge P(\sim\phi) \notin N(u))$

It should be obvious from these definitions that $\Box$ and $\Diamond$ are dual operators in any xeno model.

Finally, the clause for the safety operator Σ is:

(Σ) $\langle\mathfrak{M}, w\rangle \vDash \Sigma\phi$ if and only if $\forall u \in W\ (R_{\phi}wu \rightarrow P(\phi) \in N(u)) \vee \exists u \in W\ (R_{\sim\phi}wu \wedge P(\sim\phi) \notin N(u))$

As I mentioned above, we eventually define Σ in terms of $\Box$ and $\Diamond$, and one advantage of working with a modal logic where duality is presupposed is that we can even define $\Box$ or $\Diamond$ in terms of the other with the help of negation.

As with all our semantics, the most important part is to define a notion of validity—in this case the obvious choice is to say that a sentence is *valid in a xeno model* if and only if it is true in all worlds of that model, and a sentence is *valid in a xeno frame* if and only if it is valid in all xeno models based

on that frame. We could alter these definitions so that there a proper subset of W on which validity is defined (the so-called "normal worlds")—we end up employing this option in the appendix.

Notice that if we stipulate that all the accessibility relations of a xeno frame are reflexive (i.e., $\forall\phi\ \forall w{\in}W\ R_\phi ww$) and co-reflexive (i.e., $\forall\phi\ \forall w{\in}W\ \forall u{\in}W\ (R_\phi wu \rightarrow w{=}u)$), then our xeno frame is equivalent to a neighborhood frame. Thus, xeno semantics is a natural generalization of neighborhood semantics. As we will see, not every xeno frame has an equivalent neighborhood frame.

For our purposes, we will need to introduce more structure on xeno frames and xeno models for them to serve as a semantics for ADT. Note first that rule E does not hold in all xeno models. For example, let S and T be logically equivalent sentences of $\mathcal{L}$ with distinct syntactic types, and let $\mathfrak{M}$ be the following xeno model:

$W=\{a, b\}$

$N(a)=\{\{a, b\}\}$

$N(b)=\{\{b\}\}$

$R_S=\{\langle a, a\rangle, \langle b, b\rangle\}$

$R_T=\{\langle a, a\rangle, \langle b, b\rangle, \langle a, b\rangle\}$

$P(S)=\{a, b\}$

$P(T)=\{a, b\}$

It is obvious that $\forall w{\in}W\ \langle\mathfrak{M}, w\rangle{\models}S$ if and only if $\langle\mathfrak{M}, w\rangle{\models}T$. Thus, $S{\leftrightarrow}T$ is valid in $\mathfrak{M}$. Furthermore, $\forall u\ R_S au \rightarrow P(S){\in}N(u)$. Thus, $\langle\mathfrak{M}, a\rangle{\models}\Box S$. However, $\exists u\ R_T au \wedge P(s){\notin}N(u)$. Thus, $\langle\mathfrak{M}, a\rangle{\models}{\sim}\Box T$. Therefore, we have an easy violation of rule E. We have to be sure that the additional conditions we impose on our xeno frames and models preserve this feature.

Call a sentential xeno frame *acceptable* if and only if it has the following features:

(i) $\forall w \in W\ N(w) \neq \varnothing$

(ii) $\forall w \in W\ \forall X \in N(w)\ X \neq \varnothing$

(iii) $\forall w \in W\ \forall X \in N(w)\ w \in X$

(iv) $\forall \phi \in L\ \forall w \in W\ R_\phi ww$ (i.e., R_ϕ is reflexive)

(v) if ϕ and ψ have the same syntactic type then $R_\phi = R_\psi$

Acceptable xeno frames have some nice features from our perspective. For example we can show that $\Box\phi \rightarrow \phi$ is valid on any xeno model based on an acceptable xeno frame. Assume that $\mathfrak{M}$ is such a model. Assume $\langle \mathfrak{M}, w \rangle \vDash \Box\phi$. Thus, $\forall u \in W\ R_\phi wu \rightarrow P(\phi) \in N(u)$. By condition (iv) $R_\phi ww$; hence $P(\phi) \in N(w)$. By condition (iii) $\forall X \in N(w)\ w \in X$; hence, $w \in P(\phi)$. Therefore, $\langle \mathfrak{M}, w \rangle \vDash \phi$.

Another nice feature is that $\Box\sim\phi \rightarrow \sim\Box\phi$ is valid on any xeno model based on an acceptable xeno frame. Assume $\langle \mathfrak{M}, w \rangle \vDash \Box\sim\phi$. Thus, $\forall u \in W\ R_\phi wu \rightarrow P(\sim\phi) \in N(u)$. By condition (iv) $R_\phi ww$; hence $P(\sim\phi) \in N(w)$. By condition (iii) $\forall X \in N(w)\ w \in X$; hence, $w \in P(\sim\phi)$. Therefore, $\langle \mathfrak{M}, w \rangle \vDash \sim\phi$. Since all worlds are classical, it follows that $w \notin P(\phi)$. By condition (iv), $P(\phi) \notin N(w)$, and by condition (iii), $\sim(\forall u \in W\ R_\phi wu \rightarrow P(\phi) \in N(u))$. Therefore, $\langle \mathfrak{M}, w \rangle \vDash \sim\Box\phi$. Notice that these two principles are the operator equivalents of D1 and D2 in ADT.

Acceptable xeno frames cannot do all the work, however. We need to introduce the notion of an acceptable xeno model, but that job is made significantly more complex by the apparatus of

quantifiers, individual constants, and predicates. So, let us first see how xeno semantics works for first order languages.

13.6.6 First Order Modal Logic

So far I have addressed only problem #1 and problem #2. We saw that they are solved by moving from normal modal logic and relational semantics to classical modal logic and neighborhood semantics. However, that move brought problem #5. I have solved problem #5 by moving from classical modal logic and neighborhood semantics to xeno semantics (the kind of modal logic validated by bare xeno semantics does not have a name—we could use 'traditional modal logic' for it). That leaves us with problem #3 (how to deal with first order languages) and problem #4 (how to use possible worlds semantics for predicates, rather than for operators) to solve still. I deal with problem #3 in this subsection and problem #4 in the next.

The next step is to define a xeno semantics for a first-order language, which will involve the whole quantifier apparatus. Luckily, the move to first order modal logic is largely independent of the issues that forced the move to xeno semantics. It involves a change in the language, the theory, and the semantics.

Let $\mathcal{L}$ be a first order classical language with two modal operators, $\Box$ and $\Diamond$. $\mathcal{L}$ has a countable set of individual variables, a countable set of n-place predicate symbols, two quantifiers, and the usual logical operators. Let L be the set of well-formed formulas of $\mathcal{L}$.

As for the theory, let $\phi[y/x]$ denote a sentence just like ϕ except that free variable x is replaced with free variable y at all and only its free occurrences, without y thereby becoming bound at any of those occurrences. Add the following axiom and rule to the theory:

(Inst) $\forall x\phi(x) \rightarrow \phi[y/x]$

(Gen) if $\phi \rightarrow \psi$ is a theorem, then $\phi \rightarrow \forall x\psi$ is a theorem, where x is not free in ϕ.

The theory that results deals with quantifiers in the usual way.[18]

The additions to the semantics require a decision about how to treat the domain—I select a constant domain framework where each world has the same domain (variable domain frameworks are more complex and the additional complexity does not add anything). Add to the xeno frame, $\mathfrak{D}$, a non-empty set, called the *domain*; so a *constant domain xeno frame* is $\mathfrak{F}=<W, N, R, \mathfrak{D}>$. Instead of a valuation function, V, constant domain xeno models will have an interpretation function, I, such that for each n-ary predicate symbol F, we have $I(F, w) \subseteq \mathfrak{D}^n$; so a *constant domain xeno model* is $\mathfrak{M}=<\mathfrak{F}, \mathfrak{R}, I>$. Let a *substitution* be a function from the set of individual variables to the domain. A substitution ν' is said to be an *x-variant* of ν if $\nu(y) = \nu'(y)$ for all variables y except possibly x; this will be denoted by $\nu \approx_x \nu'$. Truth in a model is defined at a world relative to a substitution.

Let $\mathfrak{M} = <W, N, R, \mathfrak{D}, \mathfrak{R}, I>$ be any constant domain xeno model and ν any substitution.

(F) $<\mathfrak{M}, w>\vDash_\nu F(a_1, \ldots, a_n)$ (where a_i is either an individual constant or an individual variable) if and only if $<f(a_1), \ldots, f(a_n)> \in I(F, w)$, where if a_i is a variable x_i, then $f(a_i)=\nu(x_i)$, and if a_i is an individual constant c_i, then $f(a_i)=I(c_i)$ (for each n-place predicate F).

(~) $<\mathfrak{M}, w>\vDash_\nu \sim\phi$ if and only if it is not the case that $<\mathfrak{M}, w>\vDash_\nu \phi$

($\wedge$) $<\mathfrak{M}, w>\vDash_\nu \phi\wedge\psi$ if and only if $<\mathfrak{M}, w>\vDash_\nu \phi$ and $<\mathfrak{M}, w>\vDash_\nu \psi$

($\vee$) $<\mathfrak{M}, w>\vDash_\nu \phi\vee\psi$ if and only if $<\mathfrak{M}, w>\vDash_\nu \phi$ or $<\mathfrak{M}, w>\vDash_\nu \psi$

($\rightarrow$) $<\mathfrak{M}, w>\vDash_\nu \phi\rightarrow\psi$ if and only if if $<\mathfrak{M}, w>\vDash_\nu \phi$, then $<\mathfrak{M}, w>\vDash_\nu \psi$

---

[18] The first order classical modal logic and neighborhood semantics presented here is adopted from Arló Costa and Pacuit (2006).

($\leftrightarrow$) $\langle\mathfrak{M}, w\rangle\models_v\phi\leftrightarrow\psi$ if and only if $\langle\mathfrak{M}, w\rangle\models_v\phi$ iff $\langle\mathfrak{M}, w\rangle\models_v\psi$

($\forall$) $\langle\mathfrak{M}, w\rangle\models_v\forall x\phi(x)$ if and only if for each x-variant v' $\langle\mathfrak{M}, w\rangle\models_{v'}\phi(x)$

($\exists$) $\langle\mathfrak{M}, w\rangle\models_v\exists x\phi(x)$ if and only if there is an x-variant v' s.t. $\langle\mathfrak{M}, w\rangle\models_{v'}\phi(x)$

The clause for sentences of the form $\Box\phi$ or $\Diamond\phi$ are:

($\Box$) $\langle\mathfrak{M}, w\rangle\models_v\Box\phi$ if and only if $\exists X\in N(w)$ s.t. $\forall u\in W$ ($\langle\mathfrak{M}, u\rangle\models_v\phi \leftrightarrow u\in X$).

($\Diamond$) $\langle\mathfrak{M}, w\rangle\models_v\Diamond\phi$ if and only if $\sim(\exists X\in N(w)$ s.t. $\forall u\in W$ ($\langle\mathfrak{M}, u\rangle\models_v\sim\phi \leftrightarrow u\in X$)).

Notice that the clauses for the truth functions and the modal operators do not change except that they are relativized to substitutions. In this example, we do not have individual constants, but they are easy to add once we understand how quantifiers work. This solves problem #3.

13.6.7 Revision Sequences and Modal Logic

To summarize the discussion so far: we have seen how first order classical modal logic with neighborhood semantics allows us to solve some of the problems we encountered with normal modal logics and relational semantics. However, classical modal logics face a problem, which is that logically equivalent sentences have the same modal status. No logic with this feature can work for ADT since we want all instances of $D(\langle\phi\rangle)\rightarrow\phi$ to be descending true, and we want δ (i.e., the sentence such that $\delta\leftrightarrow\sim D(\langle\delta\rangle)$ is a theorem of syntax or arithmetic) to be not descending true. I have presented a new kind of semantics—xeno semantics—and I have shown how to do xeno semantics for sentential languages and first order languages with modal operators. Moreover, I have presented some of the formulas that are valid in all acceptable xeno frames.

However, the biggest problem with this entire project, problem #4, has yet to be addressed. Problem #4 is that in all the modal logics considered so far, '$\Box$', '$\Diamond$', and 'Σ' are operators. The fact that they are operators allows us to give an inductive (recursive) definition of truth at a world in a model (i.e., $\langle\mathfrak{M}, w\rangle\vDash\phi$) based on the complexity of a formula (since $\Box\phi$, $\Diamond\phi$, and $\Sigma\phi$ are more complex formulas than ϕ). However, in ADT, the items to be explained are not operators, but predicates. We could try to use the xeno semantics for first order non-classical modal logic as a semantics for ADT (altering the clauses ($\Box$), ($\Diamond$), and (Σ) appropriately), so that '$\Box$' is a descending truth predicate (i.e., 'D(x)'), '$\Diamond$' is an ascending truth predicate (i.e., 'A(x)'), and 'Σ' is a safety predicate (i.e., 'S(x)'). However, we can no longer define truth at a world in a model in the standard way (since $D(\langle\phi\rangle)$, $A(\langle\phi\rangle)$, and $S(\langle\phi\rangle)$ are atomic). Thus, this strategy does not arrive at a semantics at all, much less a semantics for ADT.

Some work has been done on using modal logic for predicates instead of operators, and one way to do it involves revision sequences. Revision sequences were originally designed to handle circular definitions, in which the definiens occurs as part of the definiendum. They can be adapted to modal logics for predicates by thinking of the definition of truth at a world in a model as a circular definition by virtue of the modal clauses. For example, '$D(\langle\phi\rangle)$' can occur in the definiens for $\langle\mathfrak{M}, w\rangle\vDash_{\sigma}D(\langle\phi\rangle)$, which makes the overall definition circular. We can then use a revision sequence to arrive at particular frames and models.[19]

A revision sequence begins with a particular interpretation of the circularly defined term in question, and then one generates a sequence of interpretations through a revision rule, which is

[19] Stewart Shapiro suggested this strategy to me; see Gupta and Belnap (1993) and Halbach, Leitgeb, and Welch (2003) who use revision sequences to give possible-worlds semantics for predicates.

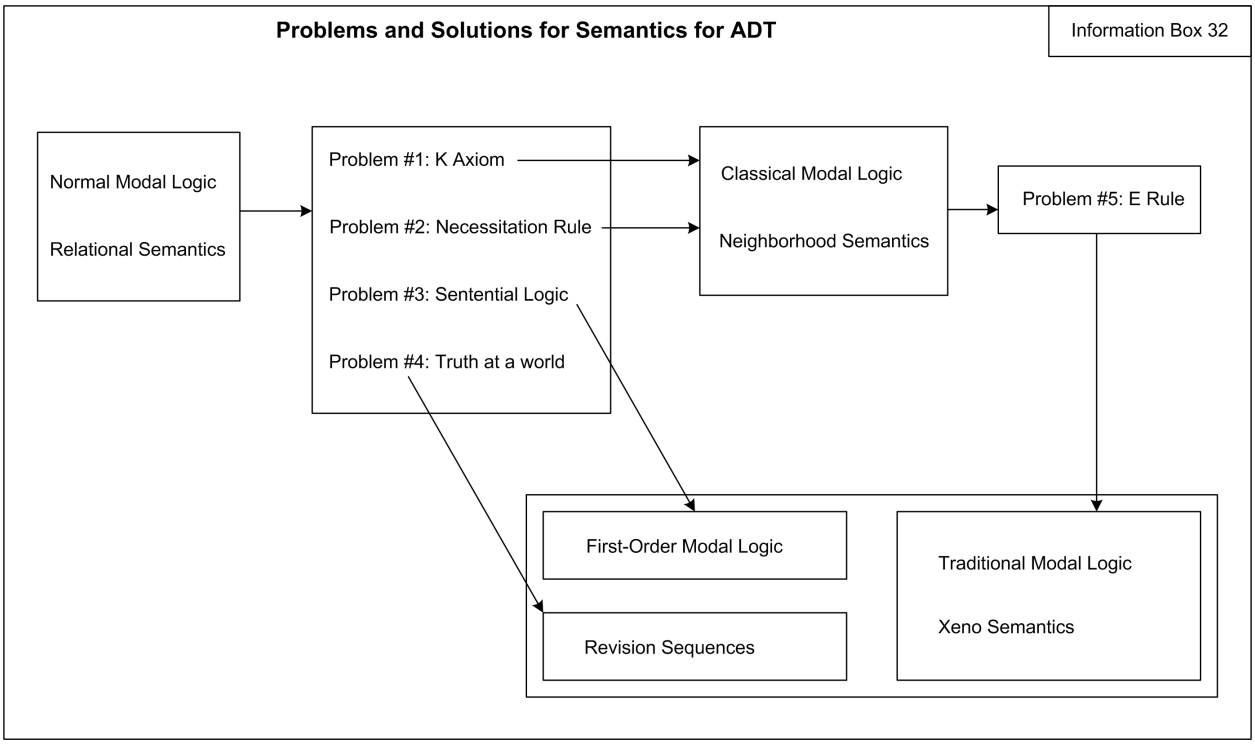
based on the circularly defined term. In our case, we start with a first order language that contains a predicate D(x), which will serve as our descending truth predicate (we worry just about D(x) first, and then see if we can define A(x) and S(x) in terms of it). The revision sequence begins with a model of the language that is similar to the first-order xeno models discussed above, except this model will not satisfy the (D) clause. Instead, we use the (D) clause to generate a new model of the language, but it will not satisfy the (D) clause either; by repeating this process over and over, we generate a sequence of models of the language. The goal is to reach a fixed point—i.e., a point in the sequence where it stops changing. If we can reach such a point, we would then have a legitimate definition of truth at a world in a xeno model for our descending truth *predicate*—a model of the language that satisfies the (D) clause. A construction of this type solves problem #4 (providing a possible-worlds semantics for a descending truth *predicate* instead of for an *operator*).

Note that Gupta and Belnap use revision sequences to formulate a revision theory of truth, but that is not a project I endorse. Instead, I use revision sequences to define truth in a xeno model. Xeno models serve as the semantics for the theory of ascending and descending truth, which I present as replacements for our concept of truth.

One might wonder about the intuitive significance of xeno semantics. How should we interpret the accessibility relations and the neighborhood function? I take no stand on this issue in this work. With respect to this project, they should be thought of as technical devices that allow us to prove things about ADT.

13.6.8 Summary of Problems and Solutions

Information Box 32 contains a diagram depicting our starting point, normal modal logic and relational semantics, the problems it faces, and the solutions to these problems. Notice that three elements of the overall package (at the bottom) are relatively independent of one another.



All that is left to do is show that one can define truth-in-a-model for Xeno semantics by proving that a revision sequence of Xeno models eventually reaches a fixed point. I present this proof in an appendix.

13.7 Features of ADT

Now that we have a rudimentary theory of ascending truth and descending truth and a semantics for it, let us explore some of its features. Remember, I do not claim that ADT is *the* theory of ascending and descending truth—it is *rudimentary* in the sense that it is a subtheory of any adequate theory of ascending and descending truth, but there is no reason to think that it contains all the important principles for ascending and descending truth. Indeed, I have given no reason to think that ADT captures all the interesting truths validated by acceptable xeno models; the theorem in the appendix is effectively a soundness proof, but I have not given a completeness proof (in fact, I think ADT is probably incomplete with respect to acceptable xeno models).

13.7.1 Principles of Ascending and Descending Truth

So, which principles do ascending truth and descending truth obey? To begin with, ascending truth and descending truth are just normal predicates—they are fully compatible with classical logic, so reasoning with them does not require one to give up any of the intuitive canons of reasoning. Since they are compatible with classical logic, they are compatible with any of the weakenings of classical logic, including intuitionistic logic, relevance logic, free logic, paracomplete logic, and paraconsistent logic.[20]

Another point bears repeating—ascending truth and descending truth predicates are not context dependent, or ambiguous, or vague, or typed in any way. There is nothing remarkable about them as predicates. Moreover, they are fully compatible with theories of syntax—there is no reason to think that languages that can refer to and quantify over their own terms and sentences would have any problem with ascending and descending truth. Indeed, they were developed with this point specifically in mind.

As for the particular principles they obey, aside from the axioms of ADT, there are other important principles for ascending and descending truth. Indeed, it would be nice to get analogs of each principle in the list of aletheic principles above—each one is a principle of either ascending truth or descending truth or both. However, that is not what we find. Instead, we do find that some of these principles hold of either ascending truth or of descending truth (or both), but for some of them that have multiple occurrences of the truth predicate, the theorem of ADT involves both ascending truth and descending truth. For example, two of the aletheic principles are single premise closure and single premise tracking:

[20] However, there are elements of classicality built in to ADT. If one wanted a version of ADT that does not collapse, say intuiitonistic logic into classical logic, then one would need different axioms; Neil Tennant has done some work on this in Tennant (MS3).

(SPC) $(p \rightarrow q) \rightarrow (T(\langle p \rangle) \rightarrow T(\langle q \rangle))$

(SPT) $(T(\langle p \rangle) \rightarrow T(\langle q \rangle)) \rightarrow (p \rightarrow q)$

If we substitute an ascending truth predicate for the truth predicate throughout, then we can find easy counterexamples to each of the principles that results; the same goes for descending truth. However, if we use both ascending truth and descending truth, then we can get analogs:

$(A(p) \rightarrow D(q)) \rightarrow (p \rightarrow q)$

$(p \rightarrow q) \rightarrow ((D(p) \rightarrow A(q))$

These relate ascending truth and descending truth to the conditional in an interesting way. The upshot is that expanding our conceptual scheme with two concepts allows us an unexpected freedom in trying to accommodate the principles previously accepted for the inconsistent concept in question. We need not assign each principle to one or the other replacement concept; some principles might be *hybrid* in the sense just described. With this idea on board, we can formulate a condition of adequacy for a theory of replacements for truth: every aletheic principle should be either a principle of one of the replacement concepts or it should be hybrid. That way, every aletheic principle gets represented in the theory of the replacements. Other hybrid principles include:

$D(\langle p \rangle) \wedge D(\langle q \rangle) \rightarrow A(\langle p \wedge q \rangle)$ – a hybrid version of ($\wedge$-Imb)

$D(\langle p \vee q) \rightarrow A(\langle p \rangle) \vee A(\langle q \rangle)$ – a hybrid version of ($\vee$-Exc)[21]

Quantifiers deserve some mention. I have not gone to the trouble in this elementary exposition, but I am confident that quantifier principles could be added to ADT. In particular:

$D(\langle (\forall x)\phi(x) \rangle) \rightarrow (\forall x)D(\langle \phi(x) \rangle)$

[21] It might be instructive to survey all the aletheic principles above, but space considerations prevent it.

$(\forall x)A(\langle\phi(x)\rangle) \rightarrow A(\langle(\forall x)\phi(x)\rangle)$

The first says that if a universal generalization is descending true, then every instance is descending true. The second says that if every instance of a generalization is ascending true, then that generalization is ascending true. In order to add these to ADT, one would have to alter the definition of an acceptable xeno model in the proof in the appendix.

Another matter is that ADT might seem too weak since it is hard to know how to think about ascending and descending truth values for empirical sentences (e.g., one might wonder how descending truth differs from mathematical provability). However, once one takes into consideration the comments about ascending and descending truth being equivalent on empirical sentences, this worry should disappear. That is, one condition for the replacements is that empirical sentences (i.e., those without occurrences of semantic expressions like 'true', 'refers', 'ascending true', 'descending true', etc.) are all safe. That is, they are either descending true or not ascending true. Ascending truth and descending truth differ only on the unsafe sentences, all of which involve semantic notions in some way. Although this condition is not built into ADT, it is a crucial element of how ADT is applied to languages (it could have been built into the theory at the expense of complicating it).

13.7.2 Non-Principles

There are also many notable principles one might expect to find among the theorems of ADT that are absent. For example, the following are *not* principles of ascending and descending truth:

(i) p is descending true → 'p is descending true' is descending true.

(ii) 'p is ascending true' is ascending true' → p is ascending true.

(iii) if p is a theorem of ADT, then p is descending true.

(iv) if p is ascending true, then p is a theorem of ADT.

Probably the most disturbing is that neither ascending truth nor descending truth is preserved under valid arguments. That is, one can have a valid argument with all ascending true sentences but a conclusion that is not ascending true; the same goes for descending truth. How disturbing is this result? Not very. Recall that no logical approach to the aletheic paradoxes is consistent with the claim that valid arguments preserve truth. Thus, as part of an approach to the aletheic paradoxes, one that advocates replacing truth with ascending and descending truth is no worse off than the others. One big issue is how to explain validity; I say a bit about this below. Moreover, it is not the case that valid arguments might lead one seriously astray. At worst, if the premises of a valid argument are descending true, then its conclusion might not be descending true, but it will be ascending true.[22] This kind of thing will only come up in cases of unsafe sentences.

A consequence is that although all the *axioms* of ADT are descending true (by virtue of axiom schema D7), it is not the case that all *theorems* of ADT are descending true. In fact, it is easy to find theorems that are not descending true—I present some of these in the next subsection.

Another, more far-reaching consequence is that logically interdeducible sentences might have different descending truth values or different ascending truth values. This feature can be counterintuitive to those with experience thinking about logical systems because most of us are used to equivalence classes respecting truth values, but with ADT, they do not. This is an essential feature of ADT due to the choice made back in Section 13.3 for dealing with Montague's theorem. One can see the connection between the problem associated with rule E (described in the previous section) and this consequence. Recall that my diagnosis of why revenge paradoxes occur is that some instances of (T-Out) are equivalent to liar sentences. If logically interdeducible sentences have the same descending truth value, then there is no way for all instances of (T-Out) to be descending

[22] Note that we cannot derive this result *in ADT* because of Gödel's Second Incompleteness Theorem.

true while all ascending liar and descending liars are not descending true. This is a fundamental point that any approach to the aletheic paradoxes must grapple with.

Finally, ascending truth and descending truth iterate non-trivially. That is, one cannot infer that 'p is descending true' is descending true from the claim that p is descending true. "p is descending true' is descending true' is stronger than 'p is descending true'. Likewise, "p is ascending true' is ascending true' is weaker than 'p is ascending true'. Although I am not sure, my hunch is that this is an essential feature of any theory that extends ADT.

13.7.3 The Aletheic Paradoxes

Now that the technical details are out of the way, we can worry about just how ascending truth and descending truth avoid the aletheic paradoxes. Recall that the theory of *truth* is presented in the next chapter—here we only deal with *ascending truth* and *descending truth*. Thus, since liar sentences, Curry sentences, and Yablo sentences all contain truth predicates or falsity predicates, a discussion of them is reserved for the next chapter. Here I want to consider sentences like these that contain 'ascending true' or 'descending true'. Consider the following sentences that are the analogs of liar sentences:

(1) (1) is not descending true.

(2) (2) is not ascending true.

It is easy to show that (1) and (2) are both unsafe—i.e., they are ascending true and not descending true. The standard argument in the liar reasoning uses both (T-In) and (T-Out). However since neither ascending truth nor descending truth obey both these rules, the standard argument is invalid.

Assume (1) is descending true.	Assume (2) is ascending true.
'(1) is not descending true' is descending true.	'(2) is not ascending true' is ascending true.
(1) is not descending true.	*(2) is not ascending true.*
Assume (1) is not descending true.	Assume (2) is not ascending true.

'(1) is not descending true' is descending true.	'(2) is not ascending true' is ascending true.
(1) is descending true.	(2) is ascending true.

The steps leading to the italicized sentences are invalid. In the argument on the left, the inference is from '(1) is not descending true' to "(1) is not descending true' is descending true', which is an instance of 'if p then D(p)'; this inference rule is not valid in general for descending truth. In the argument on the right, the inference is from "(2) is not ascending true' is ascending true' to '(2) is not ascending true', which is an instance of 'if A(p) then p'; this inference rule is not valid in general for ascending truth. So neither of these sentences poses a problem for ADT. Moreover, ADT implies that they are unsafe, i.e., they are ascending true and not descending true.

Since Curry paradoxes and Yablo paradoxes follow the same pattern—they depend on applications of both (T-In) and (T-Out)—the results will be the same there. Those sentences are unsafe, and those arguments do not pose a problem for ADT. The question of how ADT fares against revenge paradoxes is dealt with in section 13.9.

13.8 The Nature of Ascending Truth and Descending Truth

One might wonder about the nature of ascending truth and descending truth. From the discussion in Chapter One, it should be clear that most of those who offer views on the nature of truth take themselves to be offering conceptual analyses; of course, not all of them do—deflationists argue that no analysis is possible—but even deflationists often assume that their view is the only alternative to the philosophical analyses offered (e.g., correspondence theories, coherence theories, epistemic theories, and pragmatic theories). I think that it is outdated to expect that a proper philosophical theory of some concept ought to be an analysis. Throughout this book I have mentioned measurement-theoretic methodological naturalism (MTMN) several times as an alternative to conceptual analysis and other kinds of reductive explanations. That will be the kind of theory of

ascending and descending truth I offer since I am not confident that these concepts can be analyzed or reductively explained.

Recall that measurement theory provides a framework for understanding how mathematical structures apply to the empirical world. A measurement system includes three structures: a mathematical structure, a relational structure, and a physical structure. For example, in the case of a measurement system for length, we have the real numbers (a mathematical structure), a formal model of length (a relational structure), and a group of rigid physical objects (a physical structure). To complete the measurement structure, we need connections between the three structures that permit us to assign numbers to the physical objects as their lengths.

It is the relational structure that acts as the go-between. The relational structure contains a formal theory that implicitly defines 'longer than' and 'concatenation', which apply to a set of ideal rods. 'longer than' is a relation and 'concatenation' is a function (e.g., A is the concatenation of B and C). The connection between the physical structure and the relational structure is that laying two physical rods end to end corresponds to concatenation. If A and B are physical rods, then they get matched to a and b, which are two ideal rods. A and B laid end to end gets matched to an ideal rod c, which is the concatenation of a and b. If A extends beyond B, then a is longer than b. Using the formal theory, we can define 'longer than' and 'concatenation' for the whole set of ideal rods. For every physical rod, there is an ideal rod, but there are many more ideal rods, which correspond to anything in the physical structure that has a length (e.g., rods A, B, and C laid end to end, or the amount by which A extends beyond B).

To connect the relational structure to the mathematical structure, we prove that we can represent 'longer than' and 'concatenation' and the ideal rods in the mathematical structure. How? The ideal rods are represented by positive real numbers (#). 'longer than' is represented by the greater than relation (>), 'concatenation' is represented by the addition function (+). If a is longer

than b, then #a > #b; if a = the concatenation of b and c, then #a = #b + #c. To complete the link between the relational structure and the mathematical structure, we prove two theorems: a *representation theorem* that states that the way described above of representing the ideal rods by the positive real numbers works, and a *uniqueness theorem*, which specifies how many ways of representing the rods there are (e.g., if we have one mapping of ideal rods into real numbers, then we can find another one that does just as well by multiplying the results of the first one by 2).

Finally, the positive real numbers are assigned to the physical objects in the following way: if the real number #a represents the ideal rod a, and the ideal rod a is matched to a particular physical object A, then the real number #a is the length of that physical object A. When one measures the length of any physical object, one presupposes that a complex measurement system like the one just described is in place.[23] Notice that the basis for assigning a property, length, to a physical object is the complex web of relationships that that object bears to all the others in the group. It is this complex web that is responsible for the matching that links the physical structure to the relational structure and the representation that links the relational structure to the mathematical structure. By saying that this table is 2 meters long, I am implicitly comparing it to and contrasting it with all the other rigid physical objects in the universe. It is the measurement system for length that allows me to use the positive real numbers to keep track of the properties of rigid physical objects.

According to MTMN, a theory of ascending and descending truth ought to be a measurement theory for ascending truth and descending truth. Note that these two concepts are to be explained together given the connection between them (i.e., p is descending true iff p's negation is not ascending true). The physical structure is a natural language practice—that is, a group of rational entities that use a system of spoken and written symbols for communication. The goal of the

[23] I think 'presupposition' is the right term here since these effects project.

measurement system is to assign ascending truth values and descending truth values to the utterances in the practice.

The relational structure consists of ADT—the formal theory of ascending and descending truth, and a first-order artificial language of the type that is familiar in formal semantics and formal logic—it has an exactly specified syntax (e.g., it is decidable whether a given string is a well-formed formula of the language) and it contains an ascending truth predicate and a descending truth predicate. This artificial language also has the ability to refer to its own syntax so it can construct sentences like (1) and (2) above.

The mathematical structure is the model theory of xeno semantics, which was described in the previous section. It consists of a domain of entities, a set of nodes (or worlds), and mathematical constructions on the set of nodes (e.g., the division between classical and non-classical, an accessibility relation, and a neighborhood function).[24]

The connection between the physical structure (the natural language) and the relational structure (ADT and the artificial language) is the usual one that linguists, logicians, and philosophers of language have studied for decades; moreover, I went over this connection in Chapter Seven as part of a discussion of formal semantics. Sentences uttered in the linguistic practice are matched with sentences of the artificial language in such a way that the sentences of the artificial language are said to give the *logical form* of the sentences of the natural language. For example, the natural language sentence 'all ravens are black' might be matched to the artificial language sentence '$(\forall x)(Rx \rightarrow Bx)$'. This matching procedure is ridiculously complex and there are still many unresolved issues;

[24] All this should sound familiar since it is essentially the structure described in detail in Chapter Seven under the terms 'interpretive system' and 'linguistic practice'. The interpretive system is the combination of relational structure and mathematical structure. That is, using measurement theory, one can give a precise formulation of the relationship between formal semantics and linguistic practices.

nevertheless, it is the basis for all work in formal semantics.[25] Note that since most natural languages do not (yet!) have ascending and descending truth predicates, chances are that no sentences of the natural language will be matched to the sentences of the artificial language that contain an ascending truth predicate or a descending truth predicate.

The connection between the relational structure and the mathematical structure was described in the previous section (although not in these terms). The linguistic expressions are represented by elements of the mathematical structure in the usual way; e.g., singular terms are assigned elements of the domain, many predicates are assigned subsets of the domain, term functions are assigned functions from the domain to the domain, and so on. The focus of the last section was on how to represent the descending truth predicate, the ascending truth predicate, and the safety predicate. These are given an especially complex interpretation in the mathematical structure, which involves the nodes and the various constructions on the nodes. The key to the representation of the relational structure is a definition of truth-in-a-model, which is used to show that a sentence of the artificial language is a theorem of the formal theory (in our case, ADT) only if every model that makes the axioms of ADT true also makes the sentence in question true. Notice that the soundness theorem in the appendix to this chapter is a representation theorem—it says that the relational structure can be represented in the mathematical structure such that any derivation in the relational structure is valid in the mathematical structure. I have not provided a completeness proof, which would be like a uniqueness theorem; so, to that extent, the measurement system for ADT is incomplete and will have to await further work.

The assignment of ascending truth values and descending truth values goes as follows—'descending true' ends up being represented by a subset of the domain, which consists entirely of

---

[25] Recall Predelli's distinction between sentences uttered in the linguistic practice and clause/index pairs that are input for the interpretive system.

sentences of the artificial language. If a sentence of the natural language has one of these sentences as its logical form, then it is descending true. If not, then it is not descending true. Likewise for ascending truth.[26]

I want to make several points about this measurement system for ADT. First, let us say that our natural language does have words for ascending truth and descending truth. We know that we can assign descending truth values and ascending truth values to the sentences of that language without any inconsistencies since we can do that for the artificial language in the relational structure.

Second, one should think of the whole measurement system as an explanatory superstructure that fits over the natural linguistic practice in order to make sense of it. The entire measurement structure should be fit according to some of the guiding principles already discussed. For example, ascending truth and descending truth should be as close as possible to truth—that is, the class of unsafe sentences should be made as small as possible. A result is that any sentence that contains no semantic vocabulary is safe. These principles guide the application of ADT to particular natural languages rather than appearing as specific axioms of ADT.

Third, the entire measurement system for ascending and descending truth should be thought of as implicitly defining these concepts (along with safety). This attitude fits perfectly with the following passage from Davidson:

> The measurement of length, weight, temperature, or time depends (among many other things, of course) on the existence in each case of a two-place relation that is transitive and asymmetric: warmer than, later than, heavier than, and so forth. Let us take the relation *longer than* as our example. The law or postulate of transitivity is this:
>
> (L) $L(x,y)$ and $L(y,z) \rightarrow L(x,z)$
>
> Unless this law (or some sophisticated variant) holds, we cannot easily make sense of the concept of length. There will be no way of assigning numbers to register even so much as ranking in length, let alone the more powerful demands of measurement on a ratio scale. And this remark goes not only for any three items directly involved in an intransitivity: it is

[26] Note that although the ascending truth predicate and the descending truth predicate have normal semantics (i.e., extensions), the xeno semantics gives us a nice way of calculating what is in these sets.

> easy to show (given a few more assumptions essential to measurement of length) that there is no consistent assignment of a ranking to any item unless (L) holds in full generality. Clearly (L) alone cannot exhaust the import of 'longer than'—otherwise it would not differ from 'warmer than' or 'later than'. We must suppose there is some empirical content, however difficult to formulate in the available vocabulary, that distinguishes 'longer than' from the other two-place transitive predicates of measurement and on the basis of which we may assert that one thing is longer than another. Imagine this empirical content to be partly given by the predicate '*O*(*x, y*)'. So we have this 'meaning postulate':
>
> (M) $O(x,y) \rightarrow L(x,y)$
>
> that partly interprets (L). But now (L) and (M) together yield an empirical theory of great strength, for together they entail that there do not exist three objects *a, b*, and *c* such that *O*(*a, b*), *O*(*b, c*), and *O*(*c, a*). Yet what is to prevent this happening if '*O*(*x, y*)' is a predicate we can ever, with confidence, apply? Suppose we *think* we observe an intransitive triad; what do we say? We could count (L) false, but then we would have no application for the concept of length. We could say (M) gives a wrong test for length; but then it is unclear what we thought was the *content* of the idea of one thing being longer than another. Or we could say that the objects under observation are not, as the theory requires, *rigid* objects. It is a mistake to think we are forced to accept some one of these answers. Concepts such as that of length are sustained in equilibrium by a number of conceptual pressures, and theories of fundamental measurement are distorted if we force the decision, among such principles as (L) and (M): analytic or synthetic. It is better to say the whole set of axioms, laws, or postulates for the measurement of length is partly constitutive of the idea of a system of macroscopic, rigid, physical objects.[27]

Just as it makes sense to think of the entire measurement system for length as partly constitutive of the concepts involved, the entire measurement system for ascending truth and descending truth is partly constitutive of the concepts involved.

Fourth, another point that comes from Davidson is the thought that a Tarskian truth definition serves as a meaning theory for a particular language. Davidson also claims that the Tarskian truth definition should be combined with a theory of probability and a theory of utility into what he calls the unified theory, which serves as a theory of meaning for a target subject's sentences and a theory of belief and a theory of desire for the target subject (see Chapter One for details). This suggestion, in effect, unified formal semantics and formal epistemology. This is a big idea and I cannot explain it or justify it in any detail here; nevertheless, let me say that I think the formal theory of ascending

[27] Davidson (1970: 220-221).

truth and descending truth should ideally occupy the slot currently taken by the Tarskian truth definition in a Davidsonian unified theory. The details will have to wait for another occasion.

13.9 Key Issues

I would like to close out this chapter by considering ADT in light of the four key issues discussed in Part II.

13.9.1 Expressive Role

In Chapter Six, I presented truth's expressive role, which consists in its being a device of endorsement and a device of generalization. How well do ascending truth and descending truth perform these jobs? Let us consider devices of endorsement first. We know that p follows from 'p is descending true', so descending truth functions as a device of endorsement. If a person asserts 'the Flanders hypothesis is descending true', then she has thereby endorsed the Flanders hypothesis. On the other hand, ~p does not necessarily follow from 'p is not descending true'; thus, if one asserts 'the Flanders hypothesis is not descending true', one need not thereby have endorsed the negation of the Flanders hypothesis. For rejections, one would want to use ascending truth; if one asserts 'the Flanders hypothesis is not ascending true' then one has committed oneself to the negation of the Flanders hypothesis. Thus, descending truth serves as a device of endorsement, and ascending truth serves as a device of rejection.

Ascending truth and descending truth as devices of generalization are more complex. Which one is appropriate will depend on the position of the component in question. In our example from Chapter Six, we try to generalize over instances of 'rational agents should believe that grass is green only if grass is green' and 'rational agents should believe that snow is white only if snow is white', etc. In this case, we can use descending truth: rational agents should believe something only if it is

descending true. This sentence has all the instances we were trying to generalize over as consequences. However, if we look at a different generalization, 'rational agents should believe something if it is true', the results are different. In this case we would say 'rational agents should believe something if it is ascending true'. This sentence has all the instances we are after as consequences. The difference is in whether the atomic sentence containing 'true' occurs in positive or negative position inside the compound. Thus, ascending truth and descending truth split the work when it comes to devices of generalization.[28]

It is essential to remember that, on the proposal defended here, it is legitimate to continue using 'true' for most purposes. Only where the difference between ascending truth and descending truth is not negligible does one need to use 'descending true' or 'ascending true' instead of 'true'.

13.9.2 Empirical Paradoxicality

The topic of Chapter Seven, empirical paradoxes, is also an important issue for ascending and descending truth because using the examples of empirical paradoxes we can construct empirically unsafe sentences. Of course, these sentences are not paradoxical since one cannot prove contradictory claims about them, but they bring out an important lesson—whether a sentence is unsafe can depend on seemingly unrelated empirical facts. For example,

(3) Every sentence of section 13.9.2 of *Replacing Truth* that begins with an 'E' is not descending true.

(3) is the only sentence of this section that begins with an 'E', so it says of itself that it is not descending true. We can prove that it is unsafe (using this fact about its location as an assumption).

[28] Thanks to Hartry Field for conversations on this issue.

However if it had occurred in some other section of the book, then it would have been not ascending true instead of unsafe.

13.9.3 Revenge

Revenge was the topic of Chapter Eight. How can I be sure that the approach I offer does not fall prey to revenge paradoxes? We know that ADT, the rudimentary theory of ascending and descending truth, is consistent if (ZFC) set theory is consistent. Moreover, ADT implies that its axioms are descending true (and ascending true). In addition, ADT requires no limitation on self-reference. Finally, ADT is compatible with classical logic, so there is no worry about purchasing the consistency at the expense of expressive limitations on the language. It is easy to have a language that: (i) obeys classical logic, (ii) contains its own ascending truth predicate, descending truth predicate, and safety predicate, and (iii) expresses the theory of ascending and descending truth. The language can even have its own truth predicate as well as the theory of that truth predicate (so long as one treats the truth predicate as assessment sensitive in the way described in the next chapter). No other approach to the liar paradox has this package of benefits. If there is anything like a revenge paradox, it would most likely come in the form of some inadequacy of ascending truth and descending truth to do the work required of truth by some other philosophical theory that relies on truth.

To see how ADT avoids revenge, consider traditionally problematic revenge constructions like:

(4) (4) is either not descending true or unsafe.

(5) (5) is either not ascending true or unsafe.

The following arguments would be used to get a revenge paradox.

Assume (4) is descending true.

'(4) is either not descending true or unsafe' is descending true.

(4) is either not descending true or unsafe.

Assume (4) is either not descending true or unsafe.

'(4) is either not descending true or unsafe' is descending true.

(4) is descending true.

Assume (5) is ascending true.

'(5) is either not ascending true or unsafe' is ascending true.

(5) is either not ascending true or unsafe.

Assume (5) is either not ascending true or unsafe.

'(5) is either not ascending true or unsafe' is ascending true.

(5) is ascending true.

Just as in the cases above with (1) and (2), the arguments break down (at the italicized steps) when one tries to use a (T-In) rule with descending truth or a (T-Out) rule with ascending truth. The argument for ascending and descending liars and the argument of the revenge paradoxes break down at exactly the same point. So (4) and (5) are both unsafe. One might worry that this a problem given what (4) and (5) say of themselves. Since ADT implies that (4) is unsafe we can conclude that (4) is ascending true, but we cannot conclude that (4) is descending true. Likewise, ADT implies that (5) is unsafe; so we can conclude that (5) is ascending true, but not that it is descending true. Neither of these results pose any problem.

I encourage the reader to play around with other examples if these are not enough to give an intuitive understanding of how ADT avoids revenge. Of course, the relative consistency proof in the appendix should convince one that there are no revenge paradoxes, but this kind of technical assurance is no substitute for an intuitive grasp of how the theory works.

Recall that there are two kinds of revenge—inconsistency problems and self-refutation problems. Doesn't ADT face a self-refutation problem? After all, ADT implies that some of its

own theorems are not descending true. For example '(1) is not descending true' is a theorem of ADT and ADT implies that (1) is not descending true. So ADT implies that some of its theorems are unsafe.

It is correct that ADT implies that some of its own theorems are not descending true (they are, of course, ascending true), but there are two points to be made here. First, all the axioms of the theory are descending true. Remember that descending truth is not preserved under logical consequence, so a valid argument might have descending true premises and a conclusion that is not descending true (it will still be ascending true though). Recall that by Montague's theorem, we either get this result or we have to claim that logical tautologies are not descending true. Since we have good reason to think that valid arguments do not preserve truth, it makes sense to set up the theory in the way I have so that its axioms are descending true and all logical theorems are descending true. Moreover, we have good reason to think that any other theory would be in the same situation. Finally, all the theorems of ADT are ascending true, and, moreover, once one has a proper understanding of assertibility with respect to ascending and descending truth (outlined in Chapter Fifteen), one sees that they are all assertible as well.

13.9.4 Internalizability

The final key issue is internalizability. Since ADT is compatible with classical logic and it does not give rise to any revenge paradoxes, it does not have to be restricted in any way. Recall that a theory T is internalizable for a language L iff there is an extension L′ of L such that T applies to L′ and T is expressible in L′. Any natural language, say English, that contains an ascending truth predicate and a descending truth predicate and the other vocabulary needed to express ADT can express ADT without problem. Moreover, following the description of the measurement system for ascending and descending truth given in section 13.8 above, ADT applies to such a language (in the sense that

it attributes ascending truth values and descending truth values to its sentences). Since we can add ascending truth and descending truth predicates to any natural language, ADT is internalizable for any natural language. Recall, from the discussion in Chapter Nine, that no extant theory of truth has that feature.

Appendix: A Fixed Point Theorem

Here we construct a revision sequence of xeno models and prove that it reaches a fixed point. Actually, our construction will be a bit more complicated—we first construct one revision sequence, $\mathbf{\Omega}_0$, using *neighborhood* semantics; we can think of this as our characterization sequence. It does not reach a fixed point, but it does classify our sentences in an illuminating way. We then use the results of this characterization sequence to construct the initial *xeno* model for a second revision sequence, $\mathbf{\Omega}_1$. The second revision sequence will eventually reach a fixed point. So we use a sequence of neighborhood models to construct a sequence of xeno models, and we prove that the sequence of xeno models reaches a fixed point. The fixed point for the sequence of xeno models will be a xeno model and it is our intended model for ADT.[29]

13.A.1 The Characterization Sequence $\mathbf{\Omega}_0$

Let $\mathcal{L}^-$ be a first order language with the usual connectives, quantifiers, individual constants, individual variables, and n-place predicates. We want $\mathcal{L}^-$ to have the resources to express its own syntax. The usual way of ensuring this is to stipulate that PA (Peano Arithmetic) is expressible in $\mathcal{L}^-$; however, there are some complications with this method that we explore below. We stipulate that PA is expressible in $\mathcal{L}^-$, but we also make sure that it can directly refer to its own closed formulas by including them in the domain of any model for it. Let $\mathcal{L}$ be the result of adding the predicate D(x)

[29] I am unaware of anything like this construction in the literature. However, it is based on the work of Gupta and Belnap (1993), and Halbach, Leitgeb, and Welch (2003), who show how to use revision sequences to give a *relational* possible worlds semantics for predicates; but they do not consider neighborhood semantics, xeno semantics (obviously), classical modal logics, or traditional modal logics.

to $\mathcal{L}^{-}$. Let L^{-} be the set of well-formed formulas of $\mathcal{L}^{-}$ and let L be the set of well-formed formulas of $\mathcal{L}$.

We consider a neighborhood frame $\mathfrak{F}=\langle W, N, \mathfrak{D}\rangle$, where W is a set of worlds, N is a neighborhood function from W to 2^{2^W}, and $\mathfrak{D}$ is the domain—a non-empty set. Let $\mathfrak{F}$ be an *suitable frame* if and only if:

(i) every neighborhood of every world in W is non-empty,

(ii) every world in W has a neighborhood,

We consider a neighborhood model $\mathfrak{M}=\langle \mathfrak{F}, I\rangle$, where F is a suitable neighborhood frame, and I is an interpretation function.

Let $\mathfrak{M}_0 = \langle W_0, N_0, \mathfrak{D}_0, I_0\rangle$ be a neighborhood model based on a suitable frame, where:

(i) $\mathbb{N}\subset\mathfrak{D}_0$ (i.e., the domain contains the natural numbers),

(ii) $L\subset\mathfrak{D}_0$ (i.e., the domain contains the sentences of L),

(iii) $\forall w\in W_0$, I_0 assigns the arithmetic vocabulary in L to its standard interpretation in D_0,

(iv) $\forall\phi\in L\ \exists\sigma\ \sigma$ is an individual constant of $\mathcal{L}$ and $I_0(\sigma)=\phi$.

(v) $I_0(D(x), w)=\varnothing$ for all $w\in W_0$.

Let ν be a valuation (i.e., an assignment of elements from the domain to each individual variable of L).

(F) $\langle\mathfrak{M}, w\rangle\vDash_\nu F(a_1, \ldots, a_n)$ (where a_i is either an individual constant or an individual variable) if and only if $\langle f(a_1), \ldots, f(a_n)\rangle\in I(F, w)$, where if a_i is a variable x_i, then $f(a_i)=\nu(x_i)$, and if a_i is an individual constant c_i, then $f(a_i)=I(c_i)$ (for each n-place predicate F).

(~) $\langle\mathfrak{M}, w\rangle\vDash_\nu\sim\phi$ if and only if it is not the case that $\langle\mathfrak{M}, w\rangle\vDash_\nu\phi$

(∧) $\langle\mathfrak{M}, w\rangle\vDash_\nu\phi\wedge\psi$ if and only if $\langle\mathfrak{M}, w\rangle\vDash_\nu\phi$ and $\langle\mathfrak{M}, w\rangle\vDash_\nu\psi$

($\vee$) $\langle\mathfrak{M}, w\rangle\models_\nu \phi\vee\psi$ if and only if $\langle\mathfrak{M}, w\rangle\models_\nu\phi$ or $\langle\mathfrak{M}, w\rangle\models_\nu\psi$

($\rightarrow$) $\langle\mathfrak{M}, w\rangle\models_\nu \phi\rightarrow\psi$ if and only if if $\langle\mathfrak{M}, w\rangle\models_\nu\phi$, then $\langle\mathfrak{M}, w\rangle\models_\nu\psi$

($\leftrightarrow$) $\langle\mathfrak{M}, w\rangle\models_\nu \phi\leftrightarrow\psi$ if and only if $\langle\mathfrak{M}, w\rangle\models_\nu\phi$ iff $\langle\mathfrak{M}, w\rangle\models_\nu\psi$

($\forall$) $\langle\mathfrak{M}, w\rangle\models_\nu \forall x\phi(x)$ if and only if for each x-variant ν' $\langle\mathfrak{M}, w\rangle\models_{\nu'}\phi(x)$

($\exists$) $\langle\mathfrak{M}, w\rangle\models_\nu \exists x\phi(x)$ if and only if there is an x-variant ν' s.t. $\langle\mathfrak{M}, w\rangle\models_{\nu'}\phi(x)$

We can say $\langle\mathfrak{M}, w\rangle\models\phi$ if and only if ϕ is a closed formula and for all valuations ν, $\langle\mathfrak{M}, w\rangle\models_\nu\phi$.

Notice that the extension of the descending truth predicate, D(x), is stipulated to be empty in every world in M_0. Accordingly, $\mathfrak{M}_o$ has no clause for D(x).

$\mathfrak{M}_o$ will serve as the initial model for our first revision sequence. Before presenting the revision sequence, a few definitions are in order.

> A ***revision rule*** ρ is an operation on the set of functions from $\{\{D(x)\}\times\mathfrak{D}\}$ to {t, f}. The members of this set of functions are ***hypotheses***. Each hypothesis interprets D(x). We focus on revision sequences Ω whose length, lh(Ω), is a limit ordinal or On, the class of all ordinals. Let $\Omega@\alpha$ be the αth member of Ω. Let $\Omega\,|\,\alpha$ be the restriction of Ω to ordinal α.
>
> If $x\in$ {t, f} and $d\in\mathfrak{D}$, then d is ***stably x in*** Ω if and only if $\exists\beta$ s.t. $\beta<$lh(Ω) and for all ordinals γ, if $\beta\leq\gamma<$ln(Ω) then $[\Omega@\gamma](d) = x$; the least such β is the ***stabilization point*** of d in Ω. Say d is ***stable in*** Ω if and only if for some $x\in$ {t, f}, d is stably x in Ω.
>
> A hypothesis h ***coheres with a sequence*** Ω if and only if for all $d\in\mathfrak{D}$ and all $x\in$ {t, f}, if d is stably x in Ω then h(d)=x.
>
> Ω is a ***revision sequence*** for ρ if and only if for all $\alpha<$lh(Ω): (i) if $\alpha=\beta+1$, then $\Omega@\alpha = \rho(\Omega@\beta)$, and (ii) if α is a limit ordinal then $\Omega@\alpha$ coheres with $\Omega\,|\,\alpha$ (i.e., for all $d\in\mathfrak{D}$ and all $x\in$ {t, f} if d is stably x in $\Omega\,|\,\alpha$, then $\Omega@\alpha(d)=x$).[30]

[30] Gupta and Belnap (1993: ch. 5).

These definitions are based on those in Gupta and Belnap (1994), which is the standard reference for revision sequences.

Our revision rule, which will generate the revision sequence, is based on the clause (D) that we would have wanted in our first order neighborhood semantics. Let $\mathbf{\Omega}_0$ be the revision sequence of length On with initial model $\mathfrak{M}_0$ generated by the following revision rule ρ_0:

(ρ_0-1) If α is not a limit ordinal, then $\forall w \in W$, if $\exists X \in N(w)$, s.t. $\forall x \in X$, $<\mathbf{\Omega}_0@\alpha, x> \vDash \phi$, then $\phi \in I(D, w)$ for $\mathbf{\Omega}_0@\alpha+1$; otherwise, $\phi \notin I(D, w)$ for $\mathbf{\Omega}_0@\alpha+1$.

(ρ_0-2) If α is a limit ordinal and $D(\langle\phi\rangle)$ is stably true in $\mathbf{\Omega}_0 \upharpoonright \alpha$, then $\forall w \in W$, $\phi \in I(D, w)$ for $\mathbf{\Omega}_0@\alpha$.

(ρ_0-3) If α is a limit ordinal and $D(\langle\phi\rangle)$ is stably false in $\mathbf{\Omega}_0 \upharpoonright \alpha$, then $\forall w \in W$, $\phi \notin I(D, w)$ for $\mathbf{\Omega}_0@\alpha$.

(ρ_0-4) If α is a limit ordinal and $D(\langle\phi\rangle)$ is unstable in $\mathbf{\Omega}_0 \upharpoonright \alpha$, then $\forall w \in W$, $\phi \notin I(D, w)$ for $\mathbf{\Omega}_0@\alpha$

The revision sequence based on this rule will have a fixed set of worlds and a fixed neighborhood function on that set. Obviously, the interpretation, I, changes from step to step, but the only difference between steps will be the interpretation of D(x). The interpretation of all other expressions in L does not change. One can think of this as a set of revision sequences, one for the extension of D(x) at each world. Of course, at $\mathfrak{M}_0$, and indeed at each step throughout $\mathbf{\Omega}_0$, every world satisfies the same formulas.

13.A.2 The Sequence $\mathbf{\Omega}_1$

Remember, $\mathbf{\Omega}_0$ is not the sequence we ultimately care about—its role is to help us assign accessibility relations to the sentences of L in a xeno semantics. That is, we use the results of $\mathbf{\Omega}_0$ to construct a new revision sequence of xeno models that *does* eventually reach a fixed point.

Not just any xeno frame and xeno model will do for these purposes. We need to define acceptable xeno frame and acceptable xeno model. There is one additional complication in the construction—we distinguish between traditional worlds (the set $C \subseteq W$) and non-traditional worlds (the set C'); the clause for D(x) is defined only on traditional worlds and validity is defined as truth at all traditional worlds. The extension of D(x) at non-traditional worlds is stipulated below.[31]

We will consider a constant domain xeno frame $\mathfrak{F}=\langle W, C, N, R, \mathfrak{D}\rangle$, where W is a set of worlds, $C \subseteq W$, N is a neighborhood function from W to 2^{2^W}, R is a denumerable set of binary relations on W, and $\mathfrak{D}$ is a non-empty set. Let $\mathfrak{F}$ be an *acceptable constant domain xeno frame* if and only if:

1. $C \subset W$ [non-traditional worlds]
2. $\forall w \in W, N(w) \neq \emptyset$ [all worlds have neighborhoods]
3. $\forall w \in W\ \forall X \in N(w), X \neq \emptyset$ [non-empty neighborhoods]
4. $\forall w \in W\ \forall X \in N(w), w \in X$ [inclusive neighborhoods]
5. $\forall v \in C'\ \forall X \in N(v)\ \forall x \in X, x \in C'$ [non-traditional neighborhoods]
6. $\forall u \in C, C \in N(u)$ [C is a traditional neighborhood]
7. $\forall w \in C, X \in N(w) \vee Y \in N(w) \rightarrow X \cup Y \in N(w)$ [supplemented neighborhoods]
8. If $X \subset C$ then $\forall u \in C, X \notin N(u)$ [no proper subset of C is a traditional neighborhood]

It should be obvious that acceptable constant domain xeno frames exist.

We consider a xeno model $\mathfrak{M}=\langle \mathfrak{F}, \mathfrak{R}, I\rangle$ where $\mathfrak{F}$ is a xeno frame, $\mathfrak{R}$ is an accessibility function ($\mathfrak{R}$ is a function from L to R, so it assigns each sentence ϕ of L a binary relation on W, designated R_ϕ), and I is an interpretation function (I assigns each individual constant a member of the domain at each world and each n-place predicate a set of ordered n-tuples from the domain at each world). We use the following definitions for accessibility relations:

---

[31] There is a sense in which the logic determined by the particular xeno semantics I provide could be called a *non-traditional* modal logic in the spirit of non-normal modal logics and non-classical modal logics. I do not know if it is possible to avoid this aspect of the construction.

R_ϕ is *reflexive* if and only if $\forall w \in W\ R_\phi ww$

R_ϕ is *coreflexive* if and only if $\forall u \in W\ \forall w \in W, R_\phi wu \rightarrow w=u$

R_ϕ is *closed* if and only if $\forall u \in C\ \forall w \in W, R_\phi uw \rightarrow w \in C$

R_ϕ is *open* if and only if $\forall u \in C\ \exists v \in C', R_\phi uv$

Note that if an accessibility relation is coreflexive, then it is closed (but the converse fails). All accessibility relations in the xeno models we consider are reflexive. Intuitively, the accessibility relations assigned to the instances of axioms of ADT are coreflexive, as are those assigned to sentences of $\mathcal{L}^-$.

Let $\mathfrak{M}$ be an *acceptable xeno model* if and only if:

1. $\mathfrak{F}$ is an acceptable constant domain xeno frame
2. $\forall \phi \in L\ R_\phi$ is reflexive
3. If $I(\sigma)=I(\tau) \in L$, and ψ results from replacing occurrences of σ with τ in ϕ, then $R_\phi=R_\psi$.
4. $R_{\phi \wedge \psi}=R_{\psi \wedge \phi}$
5. $R_{\phi \vee \psi}=R_{\psi \vee \phi}$
6. $R_{\sim\sim\phi}=R_\phi$
7. R_ϕ is coreflexive for $\phi \in L^-$.
8. If R_ϕ is coreflexive then $R_{D\langle\phi\rangle}$ is coreflexive
9. $R_{D\langle\phi\rangle \rightarrow \phi}$ is coreflexive
10. $R_{D\langle\sim\phi\rangle \rightarrow \sim D\langle\phi\rangle}$ is coreflexive
11. $R_{D\langle\phi\wedge\psi\rangle \rightarrow D\langle\phi\rangle \wedge D\langle\psi\rangle}$ is coreflexive
12. $R_{D\langle\phi\rangle \vee D\langle\psi\rangle \rightarrow D\langle\phi\vee\psi\rangle}$ is coreflexive

13. R_ϕ is coreflexive for ϕ a first order classical logical truth.

14. R_ϕ is coreflexive for $PA\vdash\phi$

15. If R_ϕ is coreflexive and R_ψ is coreflexive then $R_{\phi\rightarrow\psi}$, $R_{\phi\wedge\psi}$, $R_{\phi\vee\psi}$ are coreflexive.

16. R_ϕ is coreflexive if and only if $R_{\sim\phi}$ is coreflexive.

17. If $R_{\phi\wedge\psi}$ is closed and $C\subseteq P_v(\phi\wedge\psi)$, then R_ϕ is closed and R_ψ is closed.[32]

18. If R_ϕ is closed and R_ψ is closed then $R_{\phi\vee\psi}$ is closed.

19. $L\subset\mathfrak{D}$ (i.e., the domain contains all closed sentences of L)

20. $\mathbb{N}\subset\mathfrak{D}$ (i.e., the domain contains the natural numbers)

21. $\forall w\in W$, I assigns the arithmetic vocabulary in $\mathcal{L}$ to their standard interpretation in $\mathfrak{D}$

22. All individual constants have the same denotation in every world.

23. All predicates (except possibly D) have the same extension in every world.

24. All predicates have the same extension in all non-traditional worlds.

25. $\forall\phi\in L$, $\exists\sigma$ σ is an individual constant of $\mathcal{L}$ and $I(\sigma)=\phi$

26. If R_ϕ is coreflexive then $C\subseteq P_v(\phi)$ or $C\subseteq P_v(\sim\phi)$

Let Δ_M be the set of sentences of $\mathcal{L}$ closed under the following rules:

($\Delta_{\mathfrak{M}}$ -1) $\forall\phi\in L$ if $\phi=D\langle\psi\rangle\rightarrow\psi$ then $\phi\in\Delta_{\mathfrak{M}}$

($\Delta_{\mathfrak{M}}$ -2) $\forall\phi\in L$ if $\phi\in\Delta_{\mathfrak{M}}$ and σ is an individual constant and $I(\sigma)=\phi$, then $D\sigma\in\Delta_{\mathfrak{M}}$

That is, $\Delta_{\mathfrak{M}}$ is the set of instances of axiom schema D1 closed under applications of D.

Let $\mathfrak{M}_1 = \langle W_1, C_1, N_1, R_1, \mathfrak{D}_1, \mathfrak{R}_1, I_1\rangle$ be an acceptable constant domain xeno model, where:

---

[32] For each ϕ in L, $P_v(\phi)=\{w\in W: \langle\mathfrak{M}, w\rangle\vDash_v\phi\}$

(i) $I_1(D(x), w)=\varnothing$ for all $w\in W_1$.

(ii) if ϕ is stably true in Ω_0, then $R_{1\phi}$ is closed in $\mathfrak{M}_1$.

(iii) if ϕ is stably false in Ω_0, then $R_{1\phi}$ is closed in $\mathfrak{M}_1$

(iv) if ϕ is unstable in Ω_0 and $\phi\notin\Delta_{M1}$ then $R_{1\phi}$ is open in $\mathfrak{M}_1$.

There are several issues to be settled before we can be sure that $\mathfrak{M}_1$ exists.

First, we need to show that there are acceptable constant domain xeno models. That is easy—let $\mathfrak{M}$ be a constant domain xeno model based on an acceptable xeno frame such that the natural numbers and the closed formulas of $\mathcal{L}$ are members of its domain, the arithmetic vocabulary of $\mathcal{L}$ receives its standard interpretation in every world, there is a name in $\mathcal{L}$ for every member of the domain, the interpretation is the same at every world, and every relation in R is the identity relation. Then $\mathfrak{M}$ is an acceptable constant domain xeno model.

Second, we need to show that $\mathfrak{R}_1$ is well-defined. Given the interpretation function I_1, we define $\Delta_{\mathfrak{M}_1}$ as above. There are eighteen conditions under the definition of an acceptable xeno model that pertain to $\mathfrak{R}_1$. None of them conflict with the above specification that defines $\mathfrak{M}_1$. For example, condition 4 is: $R_{1\phi\wedge\psi}=R_{1\psi\wedge\phi}$. It is obvious that $\phi\wedge\psi$ is stable in Ω_0 if and only if $\psi\wedge\phi$ is stable in Ω_0; thus the specification of R_1 does not conflict with this condition. The same holds for all the others. One might worry about condition 9, but all instances of D1 are in $\Delta_{\mathfrak{M}_1}$, so that does not pose a problem. Thus, the above specification of the accessibility relations in M_1 does not conflict with the definition of an acceptable xeno model. Therefore, $\mathfrak{R}_1$ is well-defined.

Let ν be a valuation (i.e., n assignment of elements from the domain to each individual variable of $\mathcal{L}$).

(F) $\langle\mathfrak{M}, w\rangle\models_\nu F(a_1, \ldots, a_n)$ (where a_i is either an individual constant or an individual variable) if and only if $\langle f(a_1), \ldots, f(a_n)\rangle\in I(F, w)$, where if a_i is a variable x_i, then $f(a_i)=\nu(x_i)$, and if a_i is an individual constant c_i, then $f(a_i)=I(c_i)$ (for each n-place predicate F).

(~) $\langle\mathfrak{M}, w\rangle\models_\nu\sim\phi$ if and only if it is not the case that $\langle\mathfrak{M}, w\rangle\models_\nu\phi$

(∧) $\langle\mathfrak{M}, w\rangle\models_\nu\phi\wedge\psi$ if and only if $\langle\mathfrak{M}, w\rangle\models_\nu\phi$ and $\langle\mathfrak{M}, w\rangle\models_\nu\psi$

(∨) $\langle\mathfrak{M}, w\rangle\models_\nu\phi\vee\psi$ if and only if $\langle\mathfrak{M}, w\rangle\models_\nu\phi$ or $\langle\mathfrak{M}, w\rangle\models_\nu\psi$

(→) $\langle\mathfrak{M}, w\rangle\models_\nu\phi\rightarrow\psi$ if and only if if $\langle\mathfrak{M}, w\rangle\models_\nu\phi$, then $\langle\mathfrak{M}, w\rangle\models_\nu\psi$

(↔) $\langle\mathfrak{M}, w\rangle\models_\nu\phi\leftrightarrow\psi$ if and only if $\langle\mathfrak{M}, w\rangle\models_\nu\phi$ iff $\langle\mathfrak{M}, w\rangle\models_\nu\psi$

(∀) $\langle\mathfrak{M}, w\rangle\models_\nu\forall x\phi(x)$ if and only if for each x-variant ν' $\langle\mathfrak{M}, w\rangle\models_{\nu'}\phi(x)$

(∃) $\langle\mathfrak{M}, w\rangle\models_\nu\exists x\phi(x)$ if and only if there is an x-variant ν' s.t. $\langle\mathfrak{M}, w\rangle\models_{\nu'}\phi(x)$

We can say $\langle\mathfrak{M}, w\rangle\models\phi$ if and only if ϕ is a closed formula and for all valuations ν, $\langle\mathfrak{M}, w\rangle\models_\nu\phi$.

Notice that the extension of the descending truth predicate, D(x), is empty in every world in $\mathfrak{M}_1$. Accordingly, $\mathfrak{M}_1$ has no clause for D(x). Let Ω_1 be the revision sequence of length On with initial model $\mathfrak{M}_1$ generated by the following revision rule ρ_1:

(ρ_1-1) α is not a limit ordinal: if $\forall u\in C$, $\forall w\in W$, $R_\phi uw\rightarrow P_{\alpha\text{-}1}(\phi)\in N(w)$, then $\forall w\in W$, $\phi\in I(D, w)$ for $\Omega_1@\alpha$; otherwise, $\forall w\in W$, $\phi\notin I(D, w)$ for $\Omega_1@\alpha$.

(ρ_1-2) α is a limit ordinal: if $D\langle\phi\rangle$ is stably true in $\Omega_1\upharpoonright\alpha$, then $\forall w\in C$, $\phi\in I(D, w)$ for $\Omega_1@\alpha+1$.

(ρ_1-3) α is a limit ordinal: if $D\langle\phi\rangle$ is stably false in $\Omega_1\upharpoonright\alpha$, then $\forall w\in C$, $\phi\notin I(D, w)$ for $\Omega_1@\alpha+1$.

(ρ_1-4) α is a limit ordinal: if $D\langle\phi\rangle$ is unstable in $\Omega_1\upharpoonright\alpha$, then $\forall w\in C$, $\phi\notin I(D, w)$ for $\Omega_1@\alpha+1$.

The revision sequence based on this rule will have a fixed set of worlds and a fixed neighborhood function on that set. As before, the interpretation, I, changes from step to step, but the only

difference between steps will be the interpretation of D(x). The interpretation of all other expressions in $\mathcal{L}$ does not change. The assignment of accessibility relations to sentences of $\mathcal{L}$ does not change.

13.A.3 A Fixed Point for $\boldsymbol{\Omega}_1$

Now we prove that $\boldsymbol{\Omega}_1$ reaches a fixed point. Before we do that, we need several more definitions and results pertaining to revision sequences.

A hypothesis h is ***cofinal in a sequence*** $\boldsymbol{\Omega}$ if and only if for all ordinals $\alpha<\ln(\boldsymbol{\Omega})$ there is a β s.t. $\alpha\leq\beta<\ln(\boldsymbol{\Omega})$ and $\boldsymbol{\Omega}@\beta=h$.

Theorem: $\boldsymbol{\Omega}$ is a sequence of length On. Then:

(i) there is a hypothesis $h\in\{t, f\}^{\mathfrak{D}}$ that is cofinal in $\boldsymbol{\Omega}$
(ii) there is an ordinal α s.t. for all $\beta\geq\alpha$, $\boldsymbol{\Omega}@\beta$ is confinal in $\boldsymbol{\Omega}$; the least such ordinal is the ***initial ordinal*** for $\boldsymbol{\Omega}$.
(iii) for all ordinals α there is an ordinal $\beta>\alpha$ satisfying the condition that for all hypotheses h cofinal in $\boldsymbol{\Omega}$ there is an ordinal γ s.t. $\alpha\leq\gamma<\beta$ and $\boldsymbol{\Omega}@\gamma=h$; such an ordinal is a ***completion ordinal*** for $\boldsymbol{\Omega}$ above α.

Theorem: for all $d\in\mathfrak{D}$ and $x\in\{t, f\}$,

(i) if d is stably x in $\boldsymbol{\Omega}$ then the value of d is x in all hypotheses cofinal in $\boldsymbol{\Omega}$
(ii) if $\mathrm{lh}(\boldsymbol{\Omega})=\mathrm{On}$, then the converse of (i) is true.

An ordinal α is a ***reflection ordinal*** for $\boldsymbol{\Omega}$ if and only if α is a limit ordinal $<\mathrm{lh}(\boldsymbol{\Omega})$ s.t.

(i) $\alpha\geq$ the initial ordinal for $\boldsymbol{\Omega}$, and
(ii) for all $d\in\mathfrak{D}$ and $x\in\{t, f\}$, d is stably x in $\boldsymbol{\Omega}|\alpha$ if and only if d is stably x in $\boldsymbol{\Omega}$.

Theorem: Let $\boldsymbol{\Omega}$ be a revision sequence for ρ and $\alpha<\mathrm{lh}(\boldsymbol{\Omega})$. If $\boldsymbol{\Omega}@\alpha$ is a fixed point of ρ then for all β s.t. $\alpha+\beta<\mathrm{lh}(\boldsymbol{\Omega})$ we have $\boldsymbol{\Omega}@\alpha+\beta=\boldsymbol{\Omega}@\alpha$; furthermore, an object $d\in\mathfrak{D}$ is stably x in $\boldsymbol{\Omega}$ if and only if $\boldsymbol{\Omega}@\alpha(d)=x$.

A hypothesis h is ***recurring*** for ρ if and only if h is cofinal in some revision sequence $\boldsymbol{\Omega}$ of length On for ρ.

Theorem: all and only recurring hypotheses are reflexive.[33]

So if α is a reflection ordinal, then $\mathbf{\Omega}|\alpha$ reflects all the stabilities and instabilities in $\mathbf{\Omega}$.

With these definitions and results in hand, we are ready to show that $\mathbf{\Omega}_1$ reaches a fixed point. I use the convention '$P_\alpha(\phi)$' for $\{w \in W : \langle \mathbf{\Omega}_1@\alpha, w\rangle \vDash \phi\}$.

Let ζ be the initial ordinal for $\mathbf{\Omega}_1$, and let ξ be a reflection ordinal for $\mathbf{\Omega}_1$ s.t. $\xi > \zeta$, and let $\mathfrak{M}_2 = \mathbf{\Omega}_1@\xi$. Thus, $\mathfrak{M}_2$ is a reflexive hypothesis for $\mathbf{\Omega}_1$; it follows that ξ is a limit ordinal.

I rely on the following lemmas:

(1) ϕ is stable in $\mathbf{\Omega}_1|\xi$ if and only if ϕ is stable in $\mathbf{\Omega}_1$. [ξ is a reflection ordinal for $\mathbf{\Omega}$, so by definition, ϕ is stable in $\mathbf{\Omega}|\xi$ if and only if ϕ is stable in $\mathbf{\Omega}$.]

(2) For any ordinal α, if $\exists w \in C$ s.t. $\langle \mathbf{\Omega}_1@\alpha, w\rangle \vDash \phi$, then $\forall w \in C$, $\langle \mathbf{\Omega}_1@\alpha, w\rangle \vDash \phi$. [By induction—the extension of D(x) is the same at all classical worlds in $\mathfrak{M}_1$, and if the extension of D(x) is the same at all classical worlds in $\mathbf{\Omega}@\alpha$, then the extension of D(x) is the same at all classical worlds in $\mathbf{\Omega}@\alpha+1$.]

To show that $\mathfrak{M}_2$ is a fixed point for ρ_1, we prove that the extension of D(x) does not change from $\mathfrak{M}_2$ to $\rho_1(\mathfrak{M}_2)$ on traditional worlds.

Let $u \in C$. Let $Q \in L$. Assume that $\langle \mathfrak{M}_2, u\rangle \vDash D\langle Q\rangle$. Assume for reductio that $\langle \rho_1(\mathfrak{M}_2), u\rangle \vDash \sim D\langle Q\rangle$. $D\langle Q\rangle$ is stably true at u in $\mathbf{\Omega}_1|\xi$ [else $\langle \mathfrak{M}_2, u\rangle \vDash \sim D\langle Q\rangle$, since ξ is a limit ordinal]. $D\langle Q\rangle$ is stably true at u in $\mathbf{\Omega}_1$ [by Lemma 1]. $\langle \rho_1(\mathfrak{M}_2), u\rangle \vDash D\langle Q\rangle$. $\perp$. This result shows that the extension of D(x) does not decrease from $\mathfrak{M}_2$ to $\rho_1(\mathfrak{M}_2)$. Now for the other direction.

[33] For proofs, see Gupta and Belnap (1993: ch. 5).

Assume that $\langle\mathfrak{M}_2, u\rangle\models\sim D\langle Q\rangle$. Assume for reductio 1 that $\langle\rho_1(\mathfrak{M}_2), u\rangle\models D\langle Q\rangle$. $\forall w\in C$, $R_Quw \rightarrow P_\xi(Q)\in N(w)$. It follows that R_Quu. Hence, $P_\xi(Q)\in N(u)$. Thus, $\forall X\in N(u)$, $u\in X$. Therefore, $u\in P_\xi(Q)$, and it follows that $\langle\mathfrak{M}_2, u\rangle\models Q$. Either $D\langle Q\rangle$ is stably false at u in $\boldsymbol{\Omega}_1\,|\,\xi$ or $D\langle Q\rangle$ is unstable at u in $\boldsymbol{\Omega}_1\,|\,\xi$ [else $\langle\mathfrak{M}_2, u\rangle\models D\langle Q\rangle$]. Assume for reductio 2 that $D\langle Q\rangle$ is stably false at u in $\boldsymbol{\Omega}_1\,|\,\xi$. It follows that $D\langle Q\rangle$ is stably false at u in $\boldsymbol{\Omega}_1$ [by Lemma 1]. Hence, $\langle\rho_1(\mathfrak{M}_2), w\rangle\models\sim D\langle Q\rangle$. $\perp$ (for reductio 2). Now for the other disjunct. Assume for reductio 3 that $D\langle Q\rangle$ is unstable at u in $\boldsymbol{\Omega}_1\,|\,\xi$. $D\langle Q\rangle$ is unstable at u in $\boldsymbol{\Omega}_1$ [by Lemma 1]. Assume for conditional proof that Q is stable at u in $\boldsymbol{\Omega}_1$. Hence, $\forall w\in C$, Q is stable at w in $\boldsymbol{\Omega}_1$ [by Lemma 2]. Let β be the stabilization point for Q at u in $\boldsymbol{\Omega}_1$. Then, $\forall\gamma>\beta$ $u\in P_\gamma(Q)$ or $\forall\gamma>\beta$ $u\notin P_\gamma(Q)$. Hence, $\forall w\in C$ ($\forall\gamma>\beta$ $w\in P_\gamma(Q)$ or $\forall\gamma>\beta$ $w\notin P_\gamma(Q)$). Thus, $\forall w\in W$ ($\forall\gamma>\beta$ $w\in P_\gamma(Q)$ or $\forall\gamma>\beta$ $w\notin P_\gamma(Q)$). Hence, $\forall\gamma>\beta$ $P_\gamma(Q)\in N(u)$ or $\forall\gamma>\beta$ $P_\gamma(Q)\notin N(u)$. R_Q is either open or closed; if R_Q is open, then $\sim D\langle Q\rangle$ is stably false at u in $\boldsymbol{\Omega}_1$. Thus, R_Q is closed. Either $\forall\gamma>\beta$ $\langle\boldsymbol{\Omega}_1@\gamma, u\rangle\models D\langle Q\rangle$ or $\forall\gamma>\beta$ $\langle\boldsymbol{\Omega}_1@\gamma, u\rangle\models\sim D\langle Q\rangle$. Hence, $D\langle Q\rangle$ is stable at u in $\boldsymbol{\Omega}_1$. By conditional proof, if Q is stable at u in $\boldsymbol{\Omega}_1$ then $D\langle Q\rangle$ is stable at u in $\boldsymbol{\Omega}_1$. So, by contraposition, if $D\langle Q\rangle$ is unstable at u in $\boldsymbol{\Omega}_1$ then Q is unstable at u in $\boldsymbol{\Omega}_1$. Thus, Q is unstable at u in $\boldsymbol{\Omega}_1$. Hence, Q is unstable at u in $\boldsymbol{\Omega}_1\,|\,\xi$. Therefore, $\langle\mathfrak{M}_2, u\rangle\models\sim Q$. We have $\perp$ for reductio 3. And we have $\perp$ for reductio 1. Consequently, we have a fixed point, and $\mathfrak{M}_2$ is the intended model for $\mathcal{L}$.

Since $\mathfrak{M}_2$ is a fixed point for ρ_1, we know that for $u\in C$:

(D) $\langle\mathfrak{M}_2, u\rangle\models D(\langle\phi\rangle)$ if and only if $\forall w\in W\ R_\phi uw \rightarrow P(\phi)\in N(w)$

So we have a constant domain xeno semantics for $\mathcal{L}$, and it satisfies the intended clause for D(x). We have not said anything about the non-traditional worlds (we have not needed to say anything about them), but to finish the interpretation of $\mathcal{L}$, we can say that if for all $w \in C$ $w \vDash \phi$, then for all $v \in C'$ $v \vDash \phi$.

Recall that we have been concentrating on D(x) and ignoring A(x) and S(x). If we can define them in terms of D(x), then that would do the trick. We could use the following definitions:

(M1) $A(\langle\phi\rangle) \leftrightarrow \sim D(\langle\sim\phi\rangle)$

(M2) $S(\langle\phi\rangle) \leftrightarrow D(\langle\phi\rangle) \vee \sim A(\langle\phi\rangle)$

Let $\mathcal{L}^+$ be the result of adding A(x) and S(x) to $\mathcal{L}$ and let L^+ be the set of formulas of $\mathcal{L}^+$. Here is how to interpret the new predicates. Let $\phi \in L^+/L$. Let ψ result from replacing all occurrences of $A\langle\theta\rangle$ in ϕ with $\sim D\langle\sim\theta\rangle$ and replacing all occurrences of $S\langle\theta\rangle$ in ϕ with $D\langle\theta\rangle \vee D\langle\sim\theta\rangle$. Then $R_\phi = R_\psi$ and $\forall w \in W$ $w \vDash \phi$ if and only if $w \vDash \psi$.

To summarize the constant domain xeno semantics for the descending truth predicate, 'D(x)', the ascending truth predicate, 'A(x)', and the safety predicate, 'S(x)':

(F) $\langle\mathfrak{M}, w\rangle \vDash_v F(a_1, \ldots, a_n)$ (where a_i is either an individual constant or an individual variable) if and only if $\langle f(a_1), \ldots, f(a_n)\rangle \in I(F, w)$, where if a_i is a variable x_i, then $f(a_i) = \nu(x_i)$, and if a_i is an individual constant c_i, then $f(a_i) = I(c_i)$ (for each n-place predicate F).

(~) $\langle\mathfrak{M}, w\rangle \vDash_v \sim\phi$ if and only if it is not the case that $\langle\mathfrak{M}, w\rangle \vDash_v \phi$

(∧) $\langle\mathfrak{M}, w\rangle \vDash_v \phi \wedge \psi$ if and only if $\langle\mathfrak{M}, w\rangle \vDash_v \phi$ and $\langle\mathfrak{M}, w\rangle \vDash_v \psi$

(∨) $\langle\mathfrak{M}, w\rangle \vDash_v \phi \vee \psi$ if and only if $\langle\mathfrak{M}, w\rangle \vDash_v \phi$ or $\langle\mathfrak{M}, w\rangle \vDash_v \psi$

(→) $\langle\mathfrak{M}, w\rangle \vDash_v \phi \rightarrow \psi$ if and only if if $\langle\mathfrak{M}, w\rangle \vDash_v \phi$, then $\langle\mathfrak{M}, w\rangle \vDash_v \psi$

($\leftrightarrow$) $\langle\mathfrak{M}, w\rangle\models_\nu \phi\leftrightarrow\psi$ if and only if $\langle\mathfrak{M}, w\rangle\models_\nu\phi$ iff $\langle\mathfrak{M}, w\rangle\models_\nu\psi$

($\forall$) $\langle\mathfrak{M}, w\rangle\models_\nu\forall x\phi(x)$ if and only if for each x-variant ν' $\langle\mathfrak{M}, w\rangle\models_{\nu'}\phi(x)$

($\exists$) $\langle\mathfrak{M}, w\rangle\models_\nu\exists x\phi(x)$ if and only if there is an x-variant ν' s.t. $\langle\mathfrak{M}, w\rangle\models_{\nu'}\phi(x)$

For all $u\in C$:

(D) $\langle\mathfrak{M}, u\rangle\models_\nu D\langle\phi\rangle$ if and only if $\forall w\in W\ R_\phi uw \rightarrow P_\nu(\phi)\in N(w)$

(A) $\langle\mathfrak{M}, u\rangle\models_\nu A\langle\phi\rangle$ if and only if $\exists w\in W\ R_{\sim\phi}uw \wedge P_\nu(\sim\phi)\notin N(w)$

(S) $\langle\mathfrak{M}, u\rangle\models_\nu S\langle\phi\rangle$ if and only if $\forall w\in W\ (R_\phi uw \rightarrow P_\nu(\phi)\in N(w)) \vee \exists u\in W\ (R_{\sim\phi}uw \wedge P_\nu(\sim\phi)\notin N(w))$

For all $v\in C'$:

(D) $\langle\mathfrak{M}, v\rangle\models_\nu D\langle\phi\rangle$ if and only if $\forall u\in C\ \langle\mathfrak{M}, u\rangle\models_\nu D\langle\phi\rangle$

(A) $\langle\mathfrak{M}, v\rangle\models_\nu A\langle\phi\rangle$ if and only if $\forall u\in C\ \langle\mathfrak{M}, u\rangle\models_\nu A\langle\phi\rangle$

(S) $\langle\mathfrak{M}, v\rangle\models_\nu S\langle\phi\rangle$ if and only if $\forall u\in C\ \langle\mathfrak{M}, u\rangle\models_\nu S\langle\phi\rangle$

We can say $\langle\mathfrak{M}, w\rangle\models\phi$ if and only if ϕ is a closed formula and for all valuations ν, $\langle\mathfrak{M}, w\rangle\models_\nu\phi$.

A sentence ϕ is *valid* in a xeno model $\mathfrak{M}$ if and only if $\forall u\in C\ \langle\mathfrak{M}, u\rangle\models\phi$

13.A.4 A Soundness Theorem

The fixed point theorem from the previous section shows that we have a well-defined notion of truth at a world for an acceptable constant domain xeno model. It follows that we have a well-defined notion of validity for constant domain xeno models. Now, all that is left is to show that

ADT is sound w.r.t. acceptable constant domain xeno semantics. It follows from this result that ADT is consistent (relative to our background set theory).

In order to prove soundness, we need to go through each of the axioms of ADT and prove that they are valid in any acceptable xeno model. It is a tedious but trivial exercise to demonstrate this, and I do not give the details here.[34] The result is, if ϕ is an axiom of ADT, then ϕ is valid in any acceptable constant domain xeno model.

ADT is sound with respect to constant domain xeno semantics if and only if for all acceptable constant domain xeno models μ, any set of sentences Γ and any sentence ϕ, if ϕ is provable from Γ, then the argument from Γ to ϕ is valid. Argument validity is defined in the usual way: for all acceptable constant domain xeno models μ if all the members of Γ are true in μ, then ϕ is true in μ. We know that all the classical logical truths are valid and all classical inference rules are valid. So our proof that all axioms of ADT are valid in any acceptable constant domain xeno model completes our soundness proof. ADT is sound with respect to xeno semantics.

---

[34] For example, to show that all instances of axiom schema D1 are valid, assume for an acceptable constant domain xeno model that $u \in C$ $u \vDash D\langle\phi\rangle$. It follows that $\forall w \in W\ R_\phi uw \rightarrow P(\phi) \in N(w)$. By clause 2 of the definition of an acceptable xeno model, R_ϕ is reflexive. Thus, Ruu. Hence, $P(\phi) \in N(u)$. By clause 4 of the definition of an acceptable xeno frame, N is inclusive. Thus, $u \in P(\phi)$. Hence, $u \vDash \phi$. Therefore, $\vDash D\langle\phi\rangle \rightarrow \phi$. The proofs for the other axioms are similar.

Chapter 14

The Descriptive Theory

This is the second of the two chapters that lay out the central theories defended in this book. In the previous chapter, I introduced two concepts, ascending truth and descending truth, to replace the concept of truth. In this chapter I offer a theory that takes truth to be an inconsistent concept. The theory of truth appeals to ascending truth and descending truth, but not to truth itself.

14.1 Theories of Inconsistent Concepts

In order to decide on a descriptive theory for our inconsistent concept of truth, we need to decide on a theory of inconsistent concepts and apply it to truth. Let me begin by saying that I am not sure which theory of inconsistent concepts is the best one. I have some views on which ones I think will not work, but they do not narrow it down enough to a unique theory. In previous publications I endorsed Joseph Camp's theory, but I am no longer confident that it is the best option.[1] I first propose some criteria for a theory and then consider one promising option.

14.1.1 Conditions of Adequacy

I present four conditions that any acceptable theory of inconsistent concepts should meet. They are:

(i) the theory should imply that the concepts in question are genuinely inconsistent (e.g., it should not reinterpret the concepts so that they have some other semantic features),

(ii) the theory should be inferentially charitable (i.e., the theory should not imply that those who employ inconsistent concepts are poor reasoners),

[1] Scharp (2008).

(iii) the theory should admit that inconsistent concepts have intelligible uses even by those who know they are inconsistent, and

(iv) the theory should apply to empirically inconsistent concepts.

I discuss the conditions in order.

The first condition is that a theory of inconsistent concepts should be a theory of *inconsistent concepts*—it should not imply that there are no such things or that what we take to be an inconsistent concept is really some type of consistent concept. I argued in Chapter Eleven that the strategy of reinterpreting a linguistic practice to avoid attributing an inconsistent concept might work in certain cases, but it fails as a general policy for handling inconsistent concepts.

A theory of inconsistent concepts should include at least: (i) an account of the distinction between consistent and inconsistent concepts, (ii) conditions on the logic that should be used to classify arguments that display inconsistent concepts as valid or invalid, (iii) conditions on the semantic theory that applies to sentences that express inconsistent concepts, (iv) conditions on a pragmatic theory that applies to speech acts involving inconsistent concepts, and (v) a policy for handling inconsistent concepts (i.e., a strategy to follow for those who discover that one of their concepts is inconsistent). Some of the theories of inconsistent concepts I discuss below do not include all five parts, but I do not fault them on these grounds. However, if it seems that a particular theory cannot be amended to include one of these parts, then that is a serious problem.

The second condition is that a theory of inconsistent concepts should be charitable. In particular, a theory of inconsistent concepts is unacceptable if it implies that those who employ inconsistent concepts are irrational. There are plenty of types of rationality and I do not discuss them all here. Instead, I focus on inferential rationality. A theory of inconsistent concepts has implications for the inferential rationality of those who employ inconsistent concepts. Given that an account of inconsistent concepts should include a logic for inconsistent concepts, when one adopts a certain theory of inconsistent concepts, one decides how to treat the reasoning practice of people

who employ inconsistent concepts. Thus, when one adopts a theory of inconsistent concepts, one undertakes a commitment to evaluate arguments in which such concepts are expressed according to a certain standard and to treat people who employ such concepts as if they should reason according to that standard.[2]

Although I do not claim to have an exhaustive list, some of the aspects of inferential rationality include being able to determine when arguments are valid or invalid, being able to determine when inductive arguments are strong or weak, being able to weigh evidence for and against a claim, having the capacity and motivation to follow inference rules in one's reasoning, and having the capacity and the motivation to alter one's beliefs effectively in light of conflicting evidence. One can employ an inconsistent concept and still be inferentially rational in all these ways. A theory of inconsistent concepts should respect this fact.

In particular, a theory of inconsistent concepts should imply that a person who employs an inconsistent concept is: (i) capable of following the formal inference rules he accepts, (ii) capable of following the formal inference rules of the logic used to evaluate his arguments, (iii) motivated to follow the formal inference rules of the logic used to evaluate his arguments, (iv) capable of following the material inference rules he accepts (i.e., capable of following his accepted strategies for weighing evidence), (v) capable of following the material inference rules of the semantic theory used to interpret his utterances and beliefs, and (iv) motivated to follow the material inference rules of the semantic theory used to interpret his utterances and beliefs.

Two features of inconsistent concepts make the inferential rationality condition on theories of inconsistent concepts especially urgent.

First, the potential for paradoxical reasoning accompanies the employment of an inconsistent concept. Recall, for example, the arguments from Chapter Eleven. Let *R* be a red table. R is a

[2] See Camp (2002) and Scharp (2005) for discussion.

table; hence, R is a rable. R is red; hence, it is not the case that R is a rable. Thus, R is a rable and it is not case that R is a rable. We have arrived at a contradiction via intuitively plausible steps from intuitively plausible assumptions. Consider the other example. Assume for reductio that some red tables exist. Let R be a red table. The reasoning above shows that R is a rable and R is not a rable. Contradiction. Therefore, no red tables exist. We have proven an obviously false sentence via intuitively plausible steps from intuitively plausible assumptions. If one accepts classical logic and treats 'rable' as univocal and invariant, then one will have a hard time avoiding these unacceptable conclusions. Hence, there is considerable pressure to endorse non-classical logics for evaluating arguments that involve inconsistent concepts.

Second, a person can possess an inconsistent concept without knowing that it is inconsistent. Indeed, a theory of inconsistent concepts will be used primarily to interpret people who are using an inconsistent concept without knowing that it is inconsistent. Given that many employers of inconsistent concepts are ignorant of their inconsistency and that many theories of inconsistent concepts include non-standard logics for inconsistent concepts, the potential for treating those who employ inconsistent concepts as inferentially irrational is high.

The third condition is that a theory of inconsistent concepts should allow that inconsistent concepts have intelligible uses. Thus, a theory of inconsistent concepts should not imply that possessing an inconsistent concept is incompatible with rationally using it. In particular, it should not imply that possessing an inconsistent concept requires that one have inconsistent beliefs or some other kind of irrationality. One difficulty that theories of inconsistent concepts face is explaining what happens when one discovers that one's concept is inconsistent. Presumably, this transition will involve a change in beliefs—beliefs in the concept's constitutive principles. It will also probably involve a change in usage—the person will be reluctant to use the concept without some sense of when it gets them into trouble and when it does not. We can think of this condition

as requiring a pragmatics for inconsistent concepts even when the users know that they are inconsistent. There has to be such a thing as a legitimate or felicitous use of an inconsistent concept by someone who knows it is inconsistent. Obviously, a theory of inconsistent concepts should explain why users of inconsistent concepts (both informed and ignorant) do not actually accept contradictions involving the concept (this is called the problem of *discipline*—I discuss it in section four below).[3]

The fourth condition on acceptable theories of inconsistent concepts is that they should apply to both essentially inconsistent concepts and empirically inconsistent concepts. The 'mass' example shows that an account of concepts that are inconsistent by definition is not enough. One must be able to explain concepts that turn out to be inconsistent because of the environment in which they are used.

14.1.2 Theories

With those conditions under our belt, we can turn to some theories of inconsistent concepts. Below are some of the theories of inconsistent concepts that one finds in the literature.[4]

1. *Error theory*: all (perhaps atomic) claims employing the inconsistent concept are false (or indeterminate).[5]

2. *Ambiguity*: inconsistent expressions are ambiguous; whatever semantics is appropriate for ambiguous expressions should be used for inconsistent expressions. A theorist who advocates this position needs a principle of disambiguation that assigns a meaning to the inconsistent expression in each context of use.[6]

3. *Context-dependence*: inconsistent expressions are context-dependent; whatever semantics is appropriate for context-dependent expressions should be used for inconsistent expression.

---

[3] Eklund (2002a).
[4] This list is not meant to be exhaustive and the entries are not meant to be exclusive.
[5] Boghossian (2006) endorses this option.
[6] This option is considered and rejected by Joseph Camp (2002: ch. 5).

A theorist who advocates this position needs a *character* for the inconsistent expression--a principle that assigns a *content* to the inconsistent expression in each context of use.[7]

4. *Dialetheism*: some sentences containing inconsistent expressions are both true and false; a paraconsistent semantics, which uses a non-classical logic, is appropriate for discourse involving inconsistent expressions.[8]

5. *Fictionalism*: inconsistent expressions are part of fictional discourse; whatever semantics is appropriate for fictional discourse should be used for inconsistent expressions.[9]

6. *Supervaluation*: inconsistent expressions are referentially indeterminate—they partially denote several distinct items; one should use a supervaluation semantics for inconsistent expressions (i.e., one should calculate the truth value of sentences containing the inconsistent expression based on the truth value of the sentences that result from replacing the inconsistent expression with expressions for the items it partially denotes; if all resulting sentences have the same truth-value, then the original has that truth value, and if the resulting sentences differ in truth-value, then the original is a truth-value gap).[10]

7. *Weighted majority*: an inconsistent expression has semantic features that make true a weighted majority of the expression's constitutive principles.[11]

8. *Relevance*: sentences containing an inconsistent expression have no truth-value but arguments containing such sentences should be evaluated by an epistemically interpreted relevance logic.[12]

9. *Revision*: an inconsistent expression has semantic features determined by a rule of revision; some sentences containing an inconsistent expression will not have traditional truth-values.[13]

10. *Indeterminate translation*: an inconsistent expression has semantic features that are relative to translation into a set of consistent expressions; there will often be multiple equally good translations (e.g., sometimes it makes sense to translate 'mass' as 'relativistic mass', sometimes as 'proper mass'—since translation is highly context-dependent and interest-relative, the semantic features of an inconsistent expression will be too).[14]

---

[7] This option is considered and rejected by Anil Gupta (1999: 24-29).
[8] This option has been offered by Graham Priest (1979, 2006a, 2006b) and adopted by Jc Beall (2009).
[9] This option is suggested but not evaluated by Joseph Camp (2002: ch. 5); see Yablo (2001), Eklund (2007), Sainsbury (2009), and the papers in French and Wettstein (2001) and Kalderon (2005) for background on fictionalism.
[10] This option is proposed by Hartry Field (1973, 1974); note that Field (2001d) rejects it.
[11] This option is proposed by Matti Eklund (2002a); for discussion see Chapters Two and Twelve.
[12] This option is proposed by Joseph Camp (2002: chs. 11-16; 2007) and endorsed by Scharp (2005, 2008) and rejected in this chapter; see MacFarlane (2007b) and Wilson (2006).
[13] The revision theory is introduced by Gupta (1982); see also Gupta and Belnap (1993). Note that Gupta and Belnap do *not* advocate the revision theory for inconsistent concepts—they think that it is appropriate for circularly defined concepts and that circularly defined concepts are not inconsistent. This option is proposed for inconsistent concepts by Stephen Yablo (1993a; 1993b).
[14] This option is proposed by Hartry Field (1994b, 2001a, 2008a).

11. *Assessment-sensitivity*: an inconsistent expression is assessment sensitive, which means the truth values of sentences that contain it are relative to a context of utterance *and* a context of assessment.[15, 16]

Instead of presenting each of these theories in detail and trying to assess each one (which would be tedious), I want to mention briefly what I take to be the prospects for some of them.

First, many of them (2, 3, 6, 8, 10, 11) require some replacement concepts for the theory to work at all. Thus, many of these views implicitly accept my replacement strategy of basing a descriptive theory on a prescriptive theory.

Second, many of these approaches (1, 4, 6, 7, 8, 11) appeal to a notion of truth. That should not be a surprise since many of the most popular semantic theories appeal to truth and most of these theories of inconsistent concepts were not designed to deal with truth itself as an inconsistent concept. It might seem that this fact disqualifies these theories from serving as descriptive theories of truth since I have taken on a commitment to *not* using truth in a descriptive theory of truth. However, this conclusion is premature. Some of these theories can be reformulated in terms of

---

[15] This option is proposed by John MacFarlane (2007b); see also Pinillos (2010).

[16] Anil Gupta's approach to inconsistent concepts (in Gupta (1999)) is not on this list. Gupta's idea is that a person who employs an inconsistent concept without knowing it is inconsistent privileges certain constitutive principles in certain situations. A *frame* is a way of privileging certain constitutive principles over others when employing an inconsistent concept. For example, in the case of 'up above' (described in Chapter Eleven), the Higherians privilege the perceptual criterion in some cases and they privilege the conceptual criterion in others. Call the former the *perceptual frame* and the latter the *conceptual frame*. One uses a frame to determine the effective semantic and pragmatic features of the Higherians linguistic expressions, mental states, and performances.

It is common to assume that the meaning of a sentence and the context in which it is uttered determine its content, and that the content of a sentence and a possible world determine its truth-value in that world. These are absolute features of the sentence. Likewise, the meaning of a sentence, the context in which it is uttered, and a frame determine its effective content, and the effective content of a sentence and a possible world determine its effective truth-value in that world. One can draw similar distinctions for other semantic and pragmatic concepts. Because Gupta's theory is currently inchoate, it is hard to say much about it. One thing should be clear: Gupta's theory is not a theory of inconsistent concepts—he offers no account of the absolute features of the Higherians' linguistic expressions, mental states, and performances. That is *not* a criticism of Gupta. Indeed, I claim that there is an important place in a theory of inconsistent concepts for a theory of the sort Gupta presents. Gupta's theory is a theory of how those who employ an inconsistent concept actually use it. It seems to me that employers of an inconsistent concept do privilege certain constitutive principles when employing it in a given situation. Gupta gives us the tools to makes sense of this behavior. However, his account is not a theory of inconsistent concepts. Thus, Gupta's theory should not be thought of as a competitor to the other theories on this list.

ascending truth and descending truth (I offer a suggestion for how to do this with the assessment-sensitivity theory later in the chapter).

Third, a couple of these theories (2, 3) are non-starters—they treat would-be inconsistent expressions as if they merely have some hidden semantic feature that has gone unnoticed by those users who run into trouble with the concept in question; that is, they violate the first condition on theories of inconsistent concepts. I consider them in detail below.

Finally, some of these theories (4, 6, 8, 9) require non-classical (or weakly classical) logics, which limits their significance since they imply that inconsistent concepts are not usable in classical contexts; other things being equal, an approach that is compatible with classical logic is preferable.

This survey is just supposed to give the reader a sense of what has been said on this topic. In this chapter I defend the assessment-sensitivity view (11). However, I do discuss many of the others in detail after giving a bit of background on confusion and the assessment-sensitivity view.

14.2 Confusion and Relative Truth

Confusion occurs when a person thinks there is one thing (or one kind of thing), but there are really two (or more). To return to an earlier example, after the late 1600s, but prior to the advent of relativistic mechanics, we thought that there is one physical quantity, mass, that physical objects have. However, in the early twentieth century, we realized that there is no one physical quantity that obeys all of the principles we took mass to obey; instead, there are two physical quantities that are somewhat similar to mass: relativistic mass and proper mass. A person who lived in the 1800s and accepted Newtonian mechanics was confused—he or she thought that there is one quantity, mass, but instead there are two. Throughout the rest of this work, I use 'confusion' in this technical sense, which is considerably more specific than the ordinary sense of the word.

The mass example illustrates a connection between confused concepts and inconsistent concepts. I do not want to say that all confused concepts are inconsistent—perhaps some are not. I also do not think that all inconsistent concepts are confused. However, confused concepts are often inconsistent, as is the case with mass. Moreover, it seems to me that truth is confused as well. If I am correct, there is no normal property of being true—anyone who thinks otherwise is confused.[17] According to my view, such a person is confusing the property of being ascending true and the property of being descending true—he or she thinks there is one thing, being true, when really there are two. *Note well*: even though there is no normal property of being true, there is a concept of truth. We have three concepts: truth, ascending truth, and descending truth.

I mention confusion here because John MacFarlane has suggested that words expressing confused concepts are assessment-sensitive, which is a complex semantic feature associated with relativism. Here is the relevant passage from MacFarlane:

Early in his discussion of the logic of confusion, Camp says:

> When one first thinks about ontological confusion, it is natural and intuitively plausible to talk in terms of perspectival truth. One wants to say: "what the confused person thinks may be true from one perspective but false from another perspective; or it may be true from both perspectives, or false from both." Perspectival truth must replace truth simpliciter when one evaluates a confused belief.[18] (125)

Camp quickly drops the talk of perspectival truth: his four epistemic values, as he notes, aren't truth values at all. But perhaps some kind of perspectival truth is just what is needed here. Think of Camp's "authorities" as occupying different perspectives: from one perspective, Fred is referring to Ant A, from another, he is referring to Ant B. Neither perspective captures the full story about Fred's confused thinking, but that is because there is no way to capture the full story and still think of Fred's thoughts as having truth values. Given that there are no "absolute" truth values for confused claims, only relativized, perspectival truth values, it seems natural to define validity in terms of these, as truth preservation in every perspective.

It seems to me that such an approach might meet Camp's desiderata even better than his

---

[17] Perhaps there is a response-dependent property, however; see the papers in Casati and Tappolet (1993) for more on response dependence. The connection between assessment-sensitivity and response dependence has yet to be explored as far as I know.

[18] Camp (2002: 125).

own multivalued semantics:

1. Because validity is defined in terms of preservation of truth-at-a-perspective, and there is no uniquely appropriate perspective for assessing a confused reasoner, the validity criterion is compatible with Camp's idea that confused thoughts and claims are not true or false simpliciter.

2. The semantic clauses for the logical connectives can be simple and straightforward, no matter how many perspectives are in play. For example, a conjunction is true at a perspective just in case both conjuncts are true at that perspective. There are no anomalies.

3. We achieve complete inferential charity, without embarrassing exceptions like disjunctive syllogism. Since every perspective corresponds to a classical interpretation, all classically valid inferences will be valid on the perspectival-truth semantics as well.

The crucial question is whether the perspectival-truth explication of validity is authoritative for the confused reasoner. Does the confused reasoner have reason to care whether her inferences are valid in this sense? That depends, I think, on what it means to say that a claim is "true at a perspective," and in virtue of what a person "occupies" or "takes up" a particular perspective on a confused reasoner's thought and talk. Unless these questions can be answered, the proposed semantics is of merely technical interest and cannot be authoritative. But I am less pessimistic than most about the prospects for answering them.

In recent work, I have suggested giving significance to perspectival truth by embedding it in a larger theory of language, specifically in a normative account of what it is to make an assertion. I would like to propose, very tentatively, that this kind of framework might be a better home for a "semantics of confusion" than the multivalued, epistemic semantics Camp advocates.[19]

That is the entirety of MacFarlane's proposal, and to my knowledge no one else has advocated anything like this.[20] In the remainder of this chapter, I present a variety of semantic theories, including the assessment-sensitivity view MacFarlane advocates, and I evaluate which of them works best for confused concepts in general, and truth in particular.

14.3 Relative Truth and Formal Semantics

In Chapter Seven, I discussed the most widely accepted semantic theories for natural languages, what Predelli calls *interpretive systems*, and their relation to linguistic practices. For my purposes in this

[19] MacFarlane (2007b).
[20] However, see Pinillos (2010) on duration.

chapter, we need to be much more careful about how interpretive systems work and the options at hand for interpreting natural language utterances.

14.3.1 Presemantic, Semantic, and Postsemantic Theories

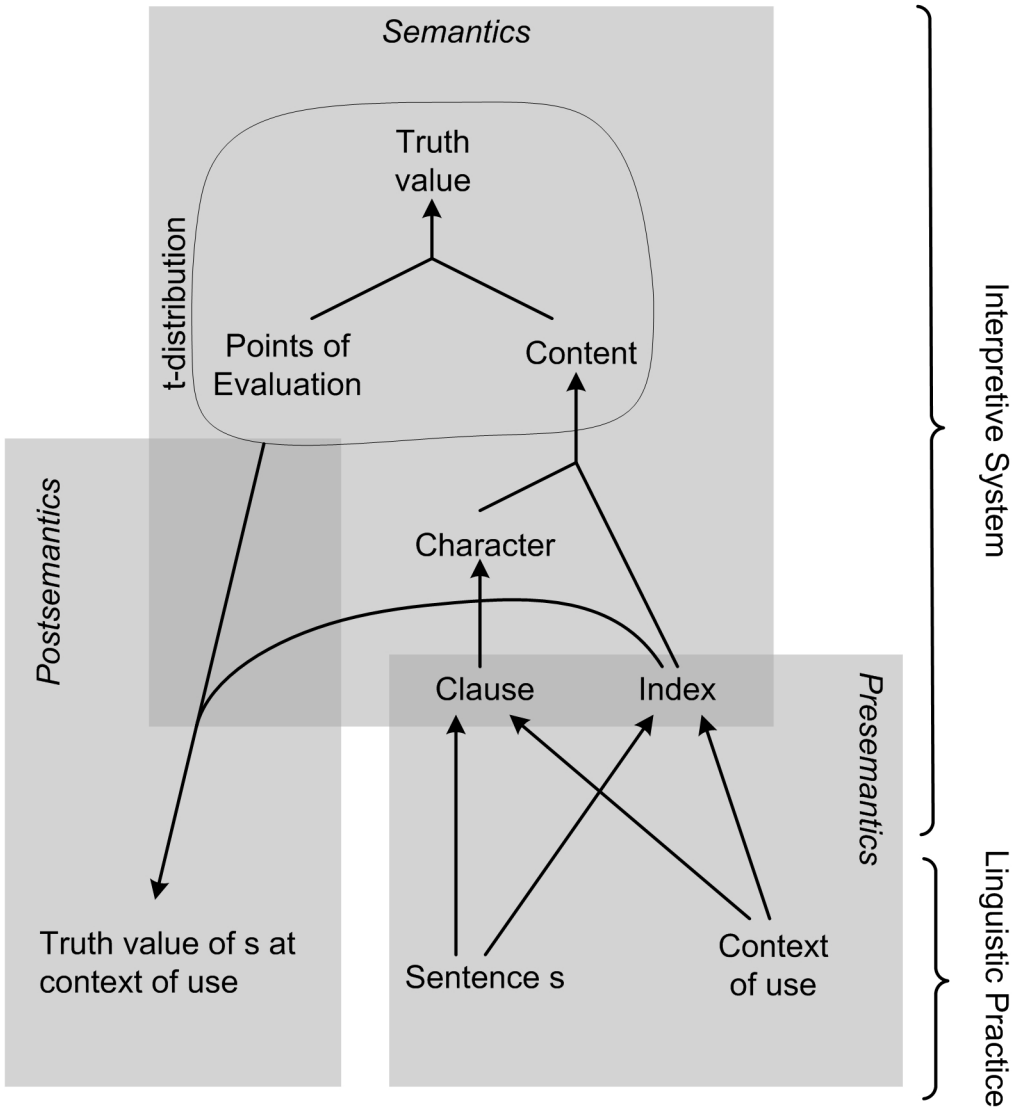
The first issue to clarify is the difference between a presemantic theory, a semantic theory and a postsemantic theory.[21] We have already seen this distinction in Chapter Seven, with the difference between linguistic practices and interpretive systems.

Interpretive systems are semantic theories—they take clause/index pairs as input and produce t-distributions as output. Recall that a *clause* is a reading of a sentence that lays out its categorial or logical form, a *index* is an ordered sequence of information about the context in which the sentence was uttered, and a *t-distribution* is a function from points to truth values, where there is a point for each possible combination of ways of evaluating a clause for truth or falsity. *Presemantic* theories take natural language utterances as input and produce clause/index pairs as output. Thus presemantic theories relate natural language utterances to semantic theory inputs. *Postsemantic* theories take t-distributions as input and produce truth values and truth conditions for natural language utterances. Hence, postsemantic theories relate semantic theory outputs to natural language utterances. In sum, we begin with a natural language utterance, run it through a presemantic theory to arrive at a clause/index pair, then use a semantic theory to compute a t-distribution for that clause/index pair, and finally use a postsemantic theory and that t-distribution to generate truth conditions and a truth value for the natural language utterance with which we began. In what follows, this three-part structure is essential; see Information Box 33 for a handy diagram.[22]

---

[21] I get the term 'presemantics' from Belnap (2005) and 'postsemantic' from MacFarlane (2005a).
[22] Recall that in measurement-theoretic terms, the linguistic practice is the physical structure, and the semantic theory is the combination of the relational structure and the mathematical structure. The presemantic theory and postsemantic theory relate the physical structure and the relational structure.

Presemantic, Semantic, and Postsemantic Theories Information Box 33



Let us now look at semantic theories. There are two main varieties—extensional and intensional. Extensional semantic theories deal only with sentences (clauses), whereas intensional semantic theories posit contents or propositions. There is a large literature on each kind, but the disputes about them do not matter for my purposes here. In order to draw the distinctions we need

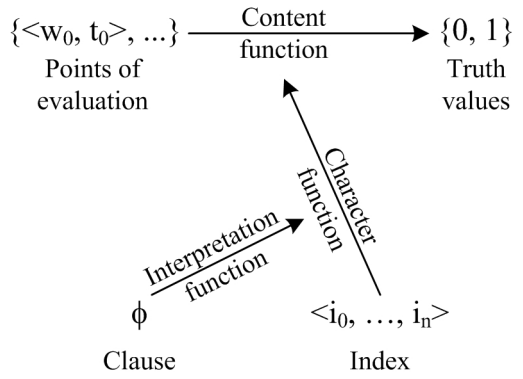
for a descriptive theory of truth, we need to focus on propositions that are the purview of intensional semantic theories.

14.3.2 Intensional Semantics

For intensional semantics, truth values are assigned to propositions instead of clause/index pairs, and intensional semantics introduces points of evaluation, which are usually taken to be just worlds or sometimes world/time pairs (if tense is treated as an intensional operator). A major advantage of intensional semantics is that it provides us with an account of propositions, which can be used in pragmatic theories and theories of propositional attitudes. Moreover, it permits more flexibility in the structure of the semantic theory (e.g., the distinction between indexicalism and non-indexical contextualism discussed below).

Following Kaplan, we distinguish between two levels in the intensional semantic theory. At the first level, expressions are assigned a character, while at the second level, the character/index pair is assigned a content. The output of the intensional semantic theory is an assignment of an extension to this content at every point of evaluation (in the case of sentences, their contents are propositions and their extensions are truth values). As such we can think of characters as functions from indexes to contents, and we can think of contents as functions from points of evaluation to extensions. See Information Box 34 for a diagram.

Intensional Semantics Information Box 34



Clause: a formula of some logical calculus

Index: n-tuple of information from the context of utterance

Point of evaluation: world/time pair

Interpretation: function from clauses to characters

Character: function from indices to contents

Content: function from points to truth values

One motivation for intensional semantics is to make sense of the idea that indexicals and demonstratives have an invariant aspect of their meaning and a variable aspect of their meaning. For example, 'I' usually refers to the speaker—that is an invariant aspect of the meaning of 'I', but different contexts have different speakers, so 'I' refers to different people in different contexts—that is the variable aspect of its meaning. In intensional semantics, indexicals like 'I' are assigned a constant character but variable content. The character of 'I' is always the same, but its content differs from context to context. When we plug in the index, the character of 'I' gives us a content for 'I'. Finally, when we plug in a point of evaluation, the content of each term gives us an extension.

The output of an intensional semantics (a t-distribution) is an assignment of truth values to contents at points of evaluation. To get from this to a truth value and a truth condition for the utterance in question, we need a postsemantic theory. It is standard to use something like:

(1) A sentence p is true at a context c iff the content assigned to the clause that represents p with respect to the index that represents c is true at the point of evaluation that represents the world and time of c.

Note that we have again defined clause truth at an index in terms of the output of the semantic theory—in this case, it is content truth at a point of evaluation.

14.3.3 Varieties of Semantic Phenomena

In an effort to illustrate the ways theorists have tried to accommodate various linguistic phenomena using intensional semantics, let us focus on a particular word, 'fun', and try to give a semantics for it. Imagine we have two conversations, one in which Luke utters 'Whacking Day is fun' and a second in which Mel utters 'Whacking Day is fun'. Here are the options I consider:

(i) 'fun' is univocal and invariant.

(ii) 'fun' is ambiguous.[23]

(iii) 'fun' has an unarticulated constituent.[24]

(iv) 'fun' is use-indexical.[25]

(v) 'fun' is use-sensitive but not use-indexical (i.e., non-indexical contextualism).[26]

(vi) 'fun' is assessment-sensitive (i.e., non-indexical relativism).[27]

---

[23] See Cruse (1986), Atlas (1989), Gillon (2004), Wasow, Perfors, and Beaver (2005), and Kennedy (2010) for discussion of ambiguity.
[24] See Perry (1998), Recanati (2002), Clapp (2002), Marti (2006), Stanley (2007), Cappelen and Lepore (2007), and Hall (2008) for discussion of unarticulated constituents.
[25] See Perry (1979, 2001), Kaplan (1989), Predelli (2005), Stanley (2007), and Cappelen and Hawthorne (2009, forthcoming a, forthcoming b, forthcoming c, forthcoming d) for discussion of indexicals.
[26] See Kölbel (2002, 2003, 2004, 2007), MacFarlane (2008, 2009, forthcoming d), Richard (2004, 2008), Recanati (2007, 2008), Brogaard (2007, 2010), and Cappelen and Hawthorne (2009) on non-indexical contextualism.
[27] See MacFarlane (2003, 2005a, 2005b, 2007a, 2008, forthcoming a, forthcoming b, forthcoming c, forthcoming d), Lasersohn (2005, 2008, 2009), Egan, Hawthorne, and Weatherson (2005), Egan (2006, 2007, forthcoming), Stephenson (2007), Zimmerman (2007), Glanzberg (2007), Stojanovic (2007), von Fintel and Gilles (2007), Cappelen and Hawthorne

(vii) 'fun' is assessment-indexical (i.e., indexical relativism).

In the case (i), the presemantics treats 'fun' as univocal (i.e., not ambiguous) and invariant, which means that since 'fun' is a one-place predicate in the language spoken by Mel and Luke, it is represented by a one-place predicate as its clause and in the semantics, 'fun' is assigned a constant character that delivers the same content for every index. The content of Mel's sentence is the same as the content of Luke's sentence. Assuming their contexts involve the same possible world and time, their contents get the same t-distribution. The postsemantics is the same for each—the sentence they utter is true in their respective contexts iff Whacking Day is fun. Nothing new here.

In case (ii), 'fun' is ambiguous, so the presemantics uses elements of the context in each case to disambiguate. It might be that the clause representing Mel's sentence has one one-place predicate, 'fun_1', while the clause for Luke's has a different one-place predicate, 'fun_2'. Since their clauses are different, the characters, contents, and t-distributions for these clauses might be different as well. In the postsemantics, Mel's sentence is true in his context iff Whacking Day is fun_1, while Luke's sentence is true in his context iff Whacking Day is fun_2. So, even though they uttered tokens of the same sentence type, the presemantics assigns them different clauses; consequently, the semantics and post semantics might be very different. In sum, ambiguity is a *presemantic* phenomenon because the analyses of the two utterances diverge in the presemantics.

In case (iii), we have the unarticulated constituent view for 'fun'.[28] That is, the presemantics assigns the same clause to each sentence, but the clause differs from the surface grammar of the sentences uttered. In this case, the presemantics might assign a clause with a two-place predicate 'x

---

(2009), Saebo (2009), Moltmann (2009), Montminy (2009), Schaffer (forthcoming), Bach (forthcoming), and Greenough (forthcoming) on assessment-sensitivity.

[28] Note that the phrase 'unarticulated constituent' is used in different ways (sometimes by the same author). For example Jason Stanley's definition at Stanley (2007: 47) differs from his definition at Stanley (2007: 183n2). I follow the latter, which reads, "an entity (object, property, or function) e is an unarticulated constituent relative to an utterance u iff (i) e is a constituent of the proposition that a competent, reflective speaker under normal circumstances would intuitively believe to be what is expressed by u, and (ii) e is not the value of any constituent in the expression uttered in u, and (iii) e is not introduced by context-independent composition rules corresponding to the structural relations between elements in the expression uttered." The primary difference between this definition and the one that occurs earlier in the book is that the earlier one has the additional requirement that e is not the semantic value of any constituent of the logical form of the sentence uttered. On my reading, an unarticulated constituent is not the value of any constituent of the sentence uttered, but it can be the value of a constituent of the logical form (i.e., clause).

is fun for y' instead of the one-place predicate 'x is fun'. So, the presemantics treats the logical form of the sentence Mel and Luke utter as different from its surface form. Since they are assigned a clause that has an extra slot, the presemantics uses the context to fill it in. In Mel's case, the presemantics uses the clause 'Whacking Day is fun for Mel', while for Luke, it uses 'Whacking Day is fun for Luke'.[29] Note that this is very different from treating 'fun' as ambiguous; in the case of unarticulated constituents, 'fun' is univocal, but its surface grammar is misleading. Since the presemantics assigns the two sentences different clauses, the semantics might assign them different characters, contents, and t-distributions. The postsemantics yields: Mel's sentence is true in his context iff Whacking Day is fun for Mel, and Luke's sentence is true in his context iff Whacking Day is fun for Luke. In sum: unarticulated constituents are a *presemantic* phenomenon since the analyses of the two utterances diverge in the presemantics.

In case (iv), 'fun' is an indexical. Here we use the term 'use-indexical' to distinguish it from more complex cases below. There are several variants on the use-indexical view, but either way the presemantics assigns their sentences the same clause. It might be 'Whacking Day is fun' or 'Whacking Day is fun-for-me'; the former treats 'fun' as something like a gradable adjective, while the later treats it like a pure indexical (note, the hyphenation indicates that 'fun-for-me' is a one-place predicate, not two-place). Although the clause is the same, the character assigned to it by the semantics will not be. If 'fun' is like a gradable adjective, then the semantics assigns a character to it that has a slot for the standards of the index. On the other hand, if 'fun' is like a pure indexical, then the semantics assigns it a character that has a slot for the speaker. The character assigned to the clause is the same for both utterances. However, since the contexts are different, the content assigned to the clause might be different. On the pure indexical view, for Mel's case, the clause gets assigned the content that Whacking Day is fun-for-Mel, while in Luke's case it gets the content that Whacking Day is fun-for-Luke. On the gradable adjective view, for Mel's case, the clause gets assigned the content that Whacking Day is fun given the standards in Mel's conversation, while in

[29] These are not the only options, but they are the most natural.

Luke's case it gets the content that Whacking Day is fun given the standards in Luke's conversation. Since these contents might differ, they might have different t-distributions as well. The post-semantics might differ since, either way, the semantics might treat them as asserting different propositions. In sum, use-indexicality is a *semantic* phenomenon because the analyses of the two utterances diverge at the semantic level.

Case (v) is where things get interesting. It has come to be known as *non-indexical contextualism*, and it treats 'fun' as use-sensitive (i.e., its extension varies from context to context), but not use-indexical (i.e., its content is the same in every context). The presemantics is the same as in case (i)—both sentences are represented by the same clause; moreover, that clause is given a constant character, so the index has no effect (in the semantics). Thus, the clause expresses the same proposition in each index. However, the non-indexical contextualist adds an extra slot to the points of evaluation. In the case of 'fun' it would most likely be an enjoyment scale for all the objects in the domain. Thus, the clause representing the sentence uttered by Mel and Luke expresses the same content relative to the respective indexes that represent their contexts, and these contents have the same t-distribution. So the semantics treats the two cases in exactly the same way (once the change has been made to the points of evaluation). However, the postsemantics treats the two t-distributions differently. Luke's sentence is true in his context iff the content expressed by the clause assigned to the sentence he uttered is true at the point of evaluation corresponding to Luke's context of utterance. The relevant point of evaluation is the one that uses the enjoyment scale operative in Luke's context. If the proposition in question is true at this point of evaluation, then the sentence he uttered is true in his context; otherwise it is false in his context. The same process is used to determine whether the proposition Mel uttered (which is the same as the one Luke uttered) is true or false. Since their contexts might employ different enjoyment scales, it might be that the sentence Mel uttered is true in his context, but the sentence Luke uttered is false in his context, despite the fact that these sentences express the same proposition in both contexts. In sum, once the change is made to the semantic theory to allow an extra slot in the points of evaluation, non-indexical use-sensitivity is a *postsemantic* phenomenon because that is where the analyses diverge.

Case (vi) is often called *assessment-sensitivity*, but a better name would be *non-indexical relativism* (for compatibility with case (v)). Again, the presemantics is the same as in case (i)—both sentences are represented by the same clause; moreover, that clause is given a constant character, so the index has no semantic effect. Thus, the clause expresses the same proposition in each index. However, the non-indexical relativist also adds an extra slot to the points of evaluation. In the case of 'fun' it would again most likely be an enjoyment scale for all the objects in the domain; however, some non-indexical relativists add a judge slot instead.[30] Thus, the clause representing the sentence uttered by Mel and Luke expresses the same content relative to the respective indexes that represent their contexts, and these contents have the same t-distribution. So the semantics treats the two cases in exactly the same way (once the change has been made to the points of evaluation). However, the relativist suggests a new kind of postsemantics. Instead of defining truth for sentences in contexts (as the standard postsemantic theory does), the relativist defines truth for sentences in contexts of use from contexts of assessment, which is a three-place predicate. If we let u be a context of use, i_u be the index that represents it, a be the context of assessment, and i_a be the index that represents it, then the standard relativist postsemantics is something like:

(2) A sentence p is true in u from a iff the content assigned to the clause that represents p with respect to i_u is true at the point of evaluation <w, t, s>, where w is the world of i_u, t is the world of i_u, and s is the enjoyment scale from i_a.

Notice that the world and time of the point of evaluation encode information about the context of use, whereas the standard of the point of evaluation encodes information about the context of assessment. Thus, even though Luke and Mel assert the same proposition, that proposition might be true at the relevant point of evaluation for Mel's case and false at the relevant point of evaluation for Luke's case. It is hard to say since the scenarios as described do not include information about the context of assessment. Imagine we add this information—say Mel's utterance is assessed by Barbara, who is a participant in the same conversation, and Luke's utterance is assessed by Raghib,

[30] See Lasersohn (2005, 2007, 2009), Egan, Hawthorne, and Weatherson (2005), Egan (2006, 2007, forthcoming), and Stephenson (2007, 2009).

who overhears Luke, but is not is a member of his conversation. Mel's sentence is true in his context of use from Barbara's context of assessment iff the proposition it expresses is true at the point of evaluation consisting of the world from Mel's context, the time from Mel's context, and the standard from Barbara's context. Luke's sentence is true in his context of use from Raghib's context of assessment iff the proposition it expresses is true at the point of evaluation consisting of the world from Luke's context, the time from Luke's context, and the standard from Raghib's context. Since Luke's context, Mel's context, and Raghib's context might differ, Luke's sentence and Mel's sentence might have different truth values even though they express the same proposition. In sum, once the extra slot is added to the points of evaluation, assessment-sensitivity is a *postsemantic* phenomenon.[31]

Finally, in case (vii), we have assessment-indexicalism or indexical relativism. It is like non-indexical relativism except that the context of assessment plays a content-determining role in the semantics; thus, it requires changes to all three theories. The presemantics requires an extra context—the context of assessment. Thus, when Mel and Luke make their utterances, there are two contexts: the context of use and the context of assessment. These might be the same, but they might not. Just as before, the presemantics posits a clause to represent the sentence uttered, and it is the same clause in both cases. However, for indexical relativism, the presemantics needs two indexes—one to represent the context of utterance, the other for the context of assessment. So the semantic theory takes as input a clause/index_u/index_a triple. In Mel's case, index_a represents Mel's context and index_a represents Barbara's context, while in Luke's case, index_a represents Luke's context and index_b represents Raghib's context. In the semantics, a single clause represents both sentences, and the clause gets a character, but the character has two slots, one for each index. The

[31] Another alternative is assessment-indexicality (or indexical relativism), which holds that the context of assessment contributes to the content of certain sentences, not just their truth value; see Weatherson (2009).

two indexes, together with the character, determine a content. That is, the clause expresses a proposition in an index of use from an index of assessment. Different indexes of assessment might take the proposition expressed in a single index of use to be different. Thus, there is no reason to think that the proposition expressed by the clause representing Mel's sentence, as used in his index and assessed from Barbara's index, is the same as the proposition expressed by the clause representing Luke's sentence as use in his index and assessed from Raghib's index. As before, the semantics requires an extra slot, in the points of evaluation, for an enjoyment scale. The t-distribution is an assignment of truth values to the propositions at each point of evaluation. The postsemantics needs to be changed even from the postsemantics in (vi). Since the context of assessment plays a content-determining role here, we need something like (again, we let u be a context of use, f_u be the index that represents it, a be the context of assessment, and f_a be the index that represents it):

(3) A sentence p is true in u from a iff the content expressed by p in f_u *from* f_a is true at the point of evaluation <w, t, s>, where w is the world of f_u, t is the time of f_u, and s is the enjoyment scale from f_a.

The italics indicate what is different from (2). Just as before, the world and time for the relevant point of evaluation comes from the index of use, while the enjoyment scale comes from the index of assessment. According to this option, the proposition expressed by the sentence that Mel and Luke each assert might be different in each case, since it depends on both the context of use and the context of assessment. Even if Mel and Luke occupied the same context of use and uttered the same sentence, they might express different propositions as assessed from distinct contexts. According to assessment-indexicalism, there is no such thing as *the* proposition expressed by a sentence in a context of use—instead, there is the proposition expressed by a sentence in a context of use from different contexts of assessment. In sum, once the changes are made to the

presemantics, the semantics, and the postsemantics, assessment-indexicality is a *semantic* phenomenon.[32]

See Information Box 35 for a diagram of these seven options.

[32] Perhaps Robert Brandom's idea that content specification is intrinsically perspectival should be thought of as global assessment-indexicality; see Brandom (1994).

Varieties of Presemantic, Semantic, and Postsemantic Phenomena

Information Box 35

Invariant/univocal: single clause for s, frame-invariant character, and single content.

Ambiguity: multiple clauses for s (determined by context of use).

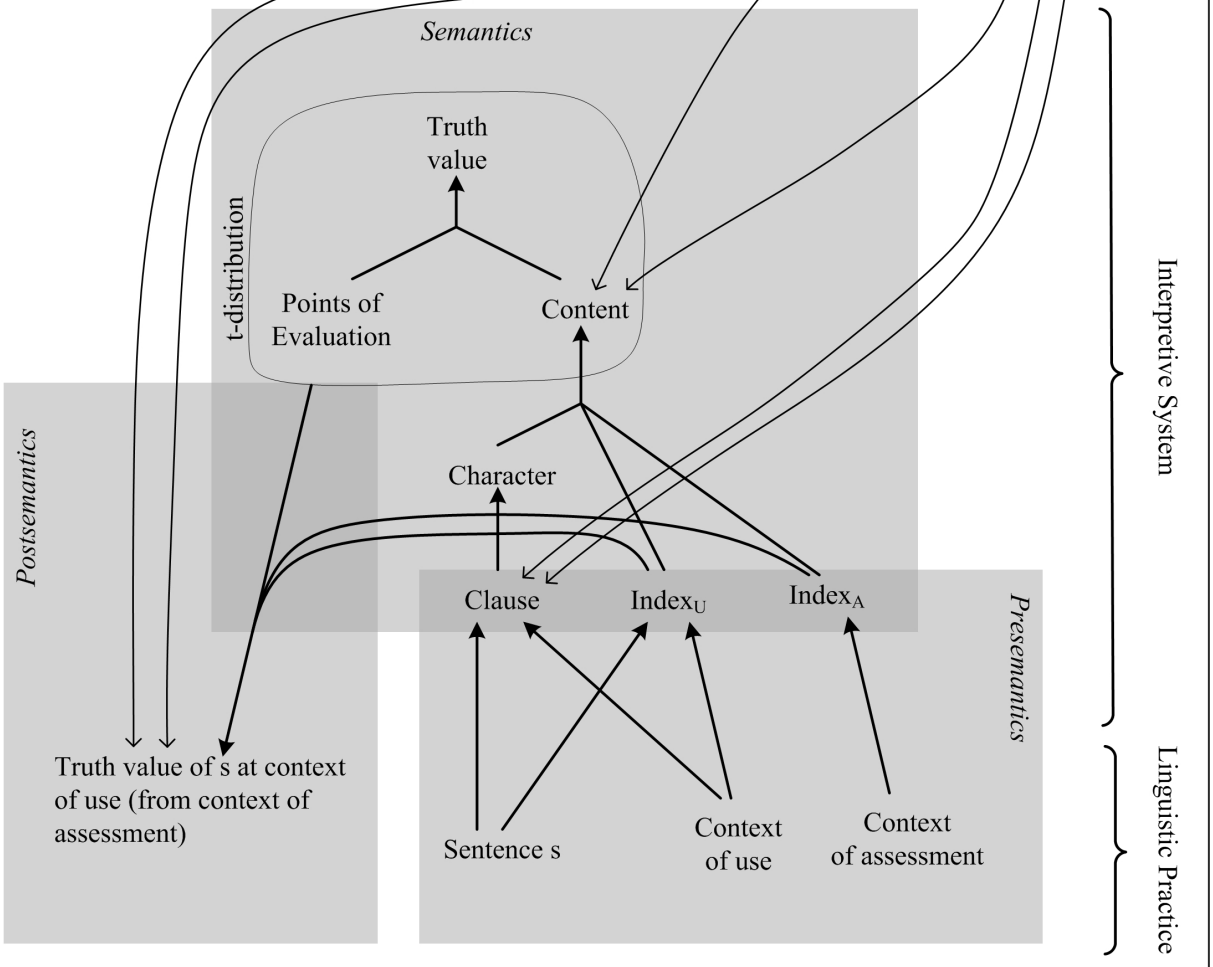
Unarticulated constituents: single clause for s with extra slot (whose value is determined by context of use).

Use-indexicality: single clause for s, but index-variant character, and multiple contents (determined by context of use).

Non-indexical contextualism: single clause for s, index-invariant character, single content, but extra slot in points.

Non-indexical relativism: single clause for s, index-invariant character, single content, but extra slot in points and new postsemantic theory with 's is true at context of use from context of assessment'.

Assessment-indexicality: single clause for s, index-variant character, but extra index that represents context of assessment, and multiple contents (determined by context of use and context of assessment).



14.4 Truth and Assessment-Sensitivity

Now that we have seen the options, let us consider what might work best for inconsistent concepts in general and for truth in particular. Recall the importance of the Gricean Condition defended in Chapter Six and of empirical paradoxicality and empirical unsafety, discussed in Chapters Seven and Thirteen, respectively.

14.4.1 The Options

Here are the options, this time explicitly formulated for truth and assuming the prescriptive theory from Chapter Thirteen:

(i) (univocal/invariant) 'true' purports to denote the property of being true, but since there is no such property, all atomic sentences containing 'true' are false.[33]

(ii) (ambiguity) implies that 'true' is ambiguous and sometimes it means *ascending true*, sometimes *descending true*. The context of use would determine which reading is appropriate.

(iii) (unarticulated constituents) implies that 'true' has the logical form of a two-place predicate; something like 'x true by the y standard', where the options for filling in 'y' are 'ascending' or 'descending'. The context of use would determine which unarticulated constituent is appropriate.[34]

(iv) (use-indexicality) implies that 'true' is an indexical, which has an invariant character, but variable content. In some contexts it would have the same content as 'ascending true' and in others it would have the same content as 'descending true'. The context of use would determine the content.

(v) (non-indexical contextualism) implies that propositions expressed by sentences containing 'true' should be assigned truth values only relative to points of evaluation that contain an extra slot for an aletheic standard (i.e., either the ascending standard or the descending standard). Presumably, the postsemantics would appeal to the context of use in order to determine which standard is relevant.

---

[33] Paul Boghossian defends this view for the inconsistent concept of mass; see Boghossian (2006).
[34] Thomas Hofweber seems to advocate something like this for the inconsistent concept of summer (before we realized that summer is a different time of the year in the northern and southern hemispheres); see Hofweber (1999).

(vi) (assessment-sensitivity) implies that propositions expressed by sentences containing 'true' should be assigned truth values only relative to points of evaluation that contain an extra slot for the aletheic standard. The postsemantics would appeal to the context of assessment in order to determine which standard is relevant.

(vii) (assessment-indexicality) implies that sentences containing 'true' express propositions only relative to contexts of use and contexts of assessment. The semantics would appeal to the context of assessment in order to determine which aletheic standard plays the content-determining role.

Remember, we are doing semantics for a linguistic practice where many (perhaps all) of the participants use 'true', but are unaware that truth is an inconsistent concept and are unfamiliar with ascending truth and descending truth. Given the fact of empirical unsafety, information about whether ascending truth or descending truth is more appropriate might be unavailable in contexts of use. Moreover, since we are taking the Gricean Condition seriously, it would be incorrect to adopt an option that implies that the participants are unable to determine which propositions are expressed by their sentences containing 'true'. Thus, we should not pick an option where content-determining information outruns what is available in a context of use. That point rules out options (ii), (iii), (iv) and (vii).[35] In other words, due to considerations arising from the key ideas defended in Chapters Six and Seven, an adequate theory of truth ought to treat the conceptual inconsistency as a *postsemantic* phenomenon; non-indexical contextualism and assessment-sensitivity are the only ones that comply.[36]

The error theory (i), makes erroneous predictions about arguments containing 'true'. In particular, it treats all arguments consisting of atomic sentences containing 'true' alike—they are all valid (i.e., it is impossible that the premises are true and the conclusion is false, since it is impossible that the premises are true on an error theory).

---

[35] I should add that one or more of these might be acceptable for other inconsistent concepts, but not for truth.
[36] See Pinillos (2010: 6) for a different argument against using option (iv) for inconsistent concepts.

We are left with (v) and (vi)—non-indexical contextualism and assessment-sensitivity. Before assessing them, there is an urgent order of business: in accordance with the replacement policy, we need to revise the presemantics, semantics, and postsemantics so that they do not appeal to the concept of truth. Although I have not discussed how replacing truth would affect our views on other concepts (I leave that task for the next chapter), I feel that I have to say something about it here, at least with respect to formal semantics.

14.4.2 Doing Semantics with Ascending and Descending Truth

The concept of truth plays no role in the presemantics, so it is fine. The semantics assigns truth values to contents as points of evaluation and the postsemantics provides a recursive definition of 'sentence x is true in a context of use' or 'sentence x is true in a context of use from a context of assessment' (depending on whether one adopts non-indexical contextualism or assessment-sensitivity). I claim that truth is not really playing a role in the semantics, but it is in the postsemantics.

Begin with the postsemantics. Although the notions of truth at a context of use or truth at a context of use from a context of assessment are distinct from the concept of truth (i.e., the latter is monadic, while the former two are polyadic), they are still susceptible to paradox. In particular consider the following sentences:

(4) For all contexts of use u, (4) is not true in u.

(5) For all contexts of use u and for all contexts of assessment a, (5) is not true in u from a.

These are both paradoxical (indeed, (4) is routinely used in revenge paradoxes for context dependence approaches to the paradoxes).[37] This is good evidence that these concepts are inconsistent (I am not going to argue this point further).

[37] See Juhl (1997).

In order to avoid using a postsemantic theory that appeals to inconsistent concepts, we need a new postsemantic theory. There are two obvious ways to do this:

(i) a postsemantic theory that takes as input the t-distribution from a semantic theory (whatever that turns out to be) and outputs sentence ascending truth in a context of use and descending truth in a context of use.

(ii) a postsemantic theory that takes as input the t-distribution from a semantic theory (whatever that turns out to be) and outputs sentence ascending truth in a context of use from a context of assessment and sentence descending truth in a context of use from a context of assessment.

If a language contains no assessment sensitive terms at all, then (i) will be fine; otherwise, (ii) is appropriate. Note that these theories make use of predicates that we have not seen yet: 'x is ascending true in context u', 'x is descending true in context u', 'x is ascending true in context u from context a', and 'x is descending true in context u from context a'. ADT, the theory of ascending and descending truth, is a theory of the 1-place predicates 'x is ascending true', and 'x is descending true'; thus, ADT does not serve as a theory of these new polyadic predicates. Although it might be helpful to develop a formal theory of them, I am not going to do so here. Note that one need not adopt a formal theory of 'x is true in context u' to be able to use it in semantics for natural language. MacFarlane appeals to theories of assertion to help readers understand 'x is true in context u from context a', but he does not give even an intuitive theory of it, much less a formal theory.[38] From what has been said so far about ascending truth and descending truth, these new predicates should not pose any real problems in understanding. At worst, one would have to say that they are implicitly defined by the postsemantic theory in question.

So much for the postsemantics; what about the semantic theory? Each of the semantic theories discussed outputs a t-distribution, which is an assignment of truth values to propositions at points of evaluation. Does this use of truth need to be replaced as well? I do not think so. If we look at the way the semantic theory works, we can see that it is powered by truth-in-a-model, not truth. Truth-in-a-model is a mathematical concept, and it is not the same as truth. For discussion, fix a model

[38] MacFarlane (2005a).

$\mathfrak{M}$. The claim that some clause is true-in-$\mathfrak{M}$ is fully representible in set theory as a mathematical function from one set-theoretic entity to another. As long as the relevant mathematical theory (e.g., ZFC) is consistent (and we have no reason to think it is not), there is no problem whatsoever with truth-in-$\mathfrak{M}$. It is not an inconsistent concept.

Still, one might ask, what about paradoxes with truth-in-$\mathfrak{M}$? As long as our language has the relevant mathematical locutions, we can formulate a sentence like:

(6) (6) is not true-in-$\mathfrak{M}$.[39]

However, there is no reason to think that (6) is paradoxical. For one, (T-In) and (T-Out) fail for true-in-$\mathfrak{M}$. That is, there is no reason to think that 'p is true-in-$\mathfrak{M}$' follows from p since there is no reason to think that p is even representable in the language for which $\mathfrak{M}$ is a model, and even if it is representable, there is no reason to think that p would be true-in-$\mathfrak{M}$. For example, $\mathfrak{M}$ might assign some bizarre meaning to p. In addition, 'p is true-in-$\mathfrak{M}$' does not entail p since there is no reason to think that $\mathfrak{M}$ represents the way the world is. Obviously, representing the world is the notion that would be needed to get (T-Out), but the semantics does not appeal to it. What about sentences like:

(7) For all $\mathfrak{M}$, (7) is not true-in-$\mathfrak{M}$.

Does (7) pose a problem? No. Quantifying over set theoretic entities is complicated business, and explaining why (7) is benign would take us too far afield into the technical details of mathematical logic. It should be sufficient to note that (7) is a mathematical claim, pure and simple. As long as set theory (ZFC will do) is consistent, we know that (7) does not pose a problem.

One might wonder, what good is true-in-$\mathfrak{M}$? We have already seen just how important it is; although semantic theories are mathematical theories, they, like many mathematical theories, have important applications. To use this mathematical theory effectively in semantics, one has to give it an empirical interpretation, and that is exactly what the presemantics and postsemantics does.

---

[39] Actually, this is harder than it seems. It is difficult to construct a model for a language that makes one of the singular terms in that language refer to that model. The problem comes in letting the model be a member of the domain that is an element in that very model.

14.4.3 Non-Indexical Contextualism as a Theory of Inconsistent Truth

Let us consider how option (v) handles our inconsistent concept of truth. We have a linguistic practice in which rational agents utter sentences that contain truth predicates. For each one of these utterances, the presemantic theory assigns a clause/index pair. The clause represents the sentence and carries information about the sentence's logical form—it treats 'x is true' as a 1-place predicate; the index represents the context of the utterance. The semantic theory assigns a character to 'true', but it is just a constant character since 'true' is not an indexical. From the character of the expressions in the clause and the index, the semantic theory assigns a proposition to the clause. It then generates a t-distribution for the proposition. Recall that we need world/times/aletheic standard triples as our points of evaluation. The semantic theory assigns a truth value to the proposition at each point of evaluation.

The postsemantic theory takes the t-distribution as input and has two outputs: ascending truth in a context and descending truth in a context. Here are suggestions for how this would work:

(8) A sentence p is *ascending true* at a context c iff the content expressed by p in the index representing c is true at the point of evaluation <w, t, s> where w and t are the world of the index and the time of the index, and s is the *ascending* standard.

(9) A sentence p is *descending true* at a context c iff the content expressed by p in the index representing c is true at the point of evaluation <w, t, s> where w and t are the world of the index and the time of the index, and s is the *descending* standard.

Notice what happens when we substitute the new postsemantics that appeals to ascending truth and descending truth—we do *not* have to take the aletheic standard to be given by the context of use. Instead, the ascending truth conditions are given by considering a point of evaluation that has the ascending standard, while the descending truth conditions are given by considering the point of evaluation that has the descending standard. That is a big difference between this version of non-indexical contextualism and the way it was described in the previous subsections. Note that this would not work in general for other inconsistent concepts since the postsemantic theory would not

be formulated in terms of them. Nevertheless, it offers a very pretty theory of our inconsistent concept of truth without resorting to any kind of relativism (i.e., assessment-sensitivity).

14.4.4 Assessment-Sensitivity as a Theory of Inconsistent Truth

For option (vi), treating 'true' as assessment sensitive, the presemantics is the same as in the non-indexical contextualist option. However, the semantics needs an extra slot in the points of evaluation for aletheic standards in play in the context of assessment. That is, points of evaluation will be world/time/aletheic standard/aletheic standard quadruples since we need the first aletheic standard entry to distinguish between the ascending truth conditions and the descending truth conditions of the sentence in question, and the second is the aletheic standard in play in the context of assessment. Option (iv) also needs a postsemantics that takes t-distributions as input and outputs ascending truth in a context of use from a context of assessment and descending truth in a context of use from a context of assessment. Here is a suggestion for that kind of postsemantic theory (again, letting u be a context of use, a be a context of assessment, f_c be the index representing c, and f_a be the index representing a):

(10) A sentence p is *ascending true* at c from a iff the content expressed by p in f_c is true at the point of evaluation $\langle w, t, s_1, s_2\rangle$ where w and t are the world and time of f_u, s_1 is the *ascending* standard, and s_2 is the aletheic standard from f_a.

(11) A sentence p is *descending true* at c from a iff the content expressed by p in the index representing c is true at the point of evaluation $\langle w, t, s_1, s_2\rangle$ where w and t are the world and time of f_u, s_1 is the *descending* standard, and s_2 is the aletheic standard from f_a.

Notice that even though there are two aletheic-standards slots in the points of evaluation, they do not contribute to the content of the sentence in the context of utterance. Instead, one of them determines which kind of truth conditions are being given (ascending or descending) for the sentence in question and the other determines the aletheic standard in play for the context of

assessment, which determines whether the sentence is ascending true in c from a and whether the sentence is descending true in c from a.

14.5 An Example

To illustrate the two options developed in the last section, I present a toy language with a presemantic theory, the two semantic theories (i.e., one with <w, t, s> points of evaluation and one with <w, t, s_1, s_2> points of evaluation), and the two postsemantic theories (i.e., one with 'x is ascending true in context c' and 'x is descending true in context c', and one with 'x is ascending true in context of use u from context of assessment a' and 'x is descending true in context of use u from context of assessment a').

14.5.1 Syntax for L

Our language, L, has several kinds of basic expressions:

(i) *Constants*: 'I', 'Clancy', 'insanity pepper', 'space coyote', 'the sentence',

(ii) *Predicates*: 'cook(x)', 'eats(x, y)', 'true(x)'.

(iii) *Logical connectives*: '$\sim$', '$\wedge$', '$\vee$', '$\rightarrow$'

(iv) *Operators*: 'now', 'possibly'

The following are the formation rules for L's syntax:

(i) If α and β are a constants, γ is a 1-place predicate, and δ is a 2-place predicate, then '$\gamma(\alpha)$' and '$\delta(\alpha, \beta)$' are sentences.

(ii) If ϕ and ψ are sentences, then '$\sim\phi$', '$\phi\wedge\psi$', '$\phi\vee\psi$', and '$\phi\rightarrow\psi$' are sentences.

(iii) If ϕ is a sentence, then 'nowϕ' and 'possiblyϕ' are sentences.

14.5.2 Semantics for L

A *frame* of L is a 6-tuple $\mathfrak{F}$=<I, D, W, T, S, $\mathfrak{I}$> such that:

(i) I is a non-empty set of indexes where for all i∈I, f=<w, t, d, s, o> where w∈W, t∈T, d∈D, s∈S, and o∈D (i.e., i is a world/time/speaker/aletheic standard/object 5-tuple).

(ii) D is a non-empty set (i.e., the domain), which includes every sentence of L.

(iii) W is a non-empty set (i.e., the worlds).

(iv) T is the set of real numbers (i.e., the times).

(v) S = $\{s_a, s_d\}$, where s_a is the ascending aletheic standard and s_d is the descending aletheic standard.

(vi) $\mathfrak{I}$ is a function that assigns an intension to each constant and predicate (defined below).

Define the following functions on indexes:

(i) W(<w, t, d, s, o>)=w (i.e., the world of the index).

(ii) T(<w, t, d, s, o>)=t (i.e., the time of the index).

(iii) A(<w, t, d, s, o>)=d (i.e., the agent of the index).

(iv) S(<w, t, d, s, o>)=s (i.e., the aletheic standard of the index).

(v) O(<w, t, d, s, o>)=o (i.e., the salient object of the index).

Let $\mathfrak{I}$ assign each expression of L a function from each <w, t, s> triple to an extension in the following way (for the frame $\mathfrak{F}$, index i, world w, time t, and aletheic standard s):

(i) $[[\text{'I'}]]_{\mathfrak{F}, i, w, t, s}$ = A(i)

(ii) $[[\text{'Clancy'}]]_{\mathfrak{F}, i, w, t, s}$ = Clancy (∈D)

(iii) $[[\text{'insanity pepper'}]]_{\mathfrak{F}, i, w, t, s}$ = insanity pepper (∈D)

(iv) $[[\text{'space coyote'}]]_{\mathfrak{F}, i, w, t, s}$ = space coyote (∈D)

(v) $[[\text{'the sentence'}]]_{\mathfrak{F}, i, w, t, s}$ = O(i)

(vi) $[[\text{'cooks(x)'}]]_{\mathfrak{F}, i, w, t, s}$ = {x| x cooks at world w and time t}

(vii) $[[\text{'eats(x, y)'}]]_{\mathfrak{F}, i, w, t, s}$ = {<x, y>| x eats y at world w and time t}

(viii) $[[\text{'true(x)'}]]_{\mathfrak{F}, i, w, t, s}$ = {x| x is true at world w, time t, and aletheic standard s}

Define the following extensions for sentences (for frame $\mathfrak{F}$, index i, world w, time t, and aletheic standard s), where α and β are constants, γ is a 1-place predicate, and δ is a 2-place predicate:

(i) $[[\gamma(\alpha)]]_{\mathfrak{F}, i, w, t, s} = 1$ iff $[[\alpha]]_{\mathfrak{F}, i, w, t, s} \in [[\gamma]]_{\mathfrak{F}, i, w, t, s}$

(ii) $[[\delta(\alpha, \beta)]]_{\mathfrak{F}, i, w, t, s} = 1$ iff $<[[\alpha]]_{\mathfrak{F}, i, w, t, s}\ [[\beta]]_{\mathfrak{F}, i, w, t, s}> \in [[\delta]]_{\mathfrak{F}, i, w, t, s}$

(iii) $[[\sim\phi]]_{\mathfrak{F}, i, w, t, s} = 1$ iff $[[\phi]] = 0$

(iv) $[[\phi\wedge\psi]]_{\mathfrak{F}, i, w, t, s} = 1$ iff $[[\phi]]_{\mathfrak{F}, i, w, t, s} = 1$ and $[[\psi]]_{\mathfrak{F}, i, w, t, s} = 1$

(v) $[[\phi\vee\psi]]_{\mathfrak{F}, i, w, t, s} = 1$ iff $[[\phi]]_{\mathfrak{F}, i, w, t, s} = 1$ or $[[\psi]]_{\mathfrak{F}, i, w, t, s} = 1$

(vi) $[[\phi\rightarrow\psi]]_{\mathfrak{F}, i, w, t, s} = 1$ iff if $[[\phi]]_{\mathfrak{F}, i, w, t, s} = 1$ then $[[\psi]]_{\mathfrak{F}, i, w, t, s} = 1$

(vii) $[[\text{Now}\phi]]_{\mathfrak{F}, i, w, t, s} = 1$ iff $[[\phi]]_{\mathfrak{F}, i, w, t, s} = 1$

(iv) $[[\text{Possibly}\phi]]_{\mathfrak{F}, i, w, t, s} = 1$ iff for some w' $[[\phi]]_{\mathfrak{F}, i, w, t, s} = 1$

Now to define the contents for each sentence, constant, or predicate ϕ:

(i) $\{\phi\}_{\mathfrak{F}, i}$ = the function from <w, t, s> to $[[\phi]]_{\mathfrak{F}, i, w, t, s}$.

Contents are functions from points of evaluation to extensions; in the case of sentences, extensions are truth values.

All that remains is to explain how the aletheic standards work. The basic idea is that the ascending aletheic standard treats the truth predicate in the clause in question as an ascending truth predicate and the descending standard treats the truth predicate in the clause in question as a descending truth predicate. So it seems like a simple substitution would do the trick. However, because some sentences attribute truth to other sentences that themselves contain truth predicates, we need to be a bit more global in our approach. Let L′ be the language that results from substituting 'ascending true' for 'true' in each sentence of L, and let L″ be the language that results from substituting 'descending true' for 'true' in each sentence of L. For each sentence ϕ of L, there is a corresponding sentence ϕ' of L′ and a sentence ϕ'' of L″. If ϕ does not contain a truth

predicate, then $\phi=\phi'=\phi''$. Now, the ascending standard treats the proposition in question, $\{\phi\}_{\mathfrak{F},i}$ as if it were $\{\phi'\}_{\mathfrak{F},i}$, the proposition expressed by the corresponding sentence of L′, and the descending standard treats $\{\phi\}_{\mathfrak{F},i}$ as if it were $\{\phi''\}_{\mathfrak{F},i}$, the proposition expressed by the corresponding sentence of L″.

(i) ϕ is true at world w, time t, and aletheic standard s_a iff ϕ' is true at w and t.

(ii) ϕ is true at world w, time t, and aletheic standard s_d iff ϕ'' is true at w and t.

To implement this strategy, we need extensions for 'descending true' and 'ascending true' at every point of evaluation:

(i) $[[\text{'ascending true(x)'}]]_{\mathfrak{F},i,w,t,s} = \{x \mid x \text{ is ascending true at w and t}\}$

(ii) $[[\text{'descending true(x)'}]]_{\mathfrak{F},i,w,t,s} = \{x \mid x \text{ is descending true at w and t}\}$

So, even though 'ascending true' and 'descending true' are not part of L, they show up in the semantics to handle the aletheic standards. Moreover, the semantic entries for these terms appeal to the 1-place predicates 'ascending true' and 'descending true' defined by ADT.

14.5.3 Presemantics and Postsemantics for L

The semantics presented in the previous subsection is an example of how to apply option (v), the non-indexical contextualist theory of our inconsistent concept of truth. Let us see how it works. Imagine Ned and Clancy are having a conversation using L; in their immediate vicinity is a blackboard they can write on. Assume that Clancy is a cook and that Ned is not a cook throughout their conversation.

14.5.3.1 The Non-Indexical Contextualist Option

At a certain point in their conversation Clancy asserts 'I am a cook'. The presemantic theory posits a clause, 'cook(I)' to represent Clancy's sentence and an index <w, t, d, s, o> to represent the context of Clancy's utterance (w is the world, t is the time, d is the agent, s is the aletheic standard,

and o is the salient object—which will be handy when they start talking about the sentence on the blackboard). It is easy to see that the semantic theory above assigns 1 to the clause relative to i and all the world/time/aletheic standard triples where Clancy is a cook, and 0 otherwise. Our postsemantics will interpret 1 as being true in the model and 0 as false in the model. Our postsemantics implies that Clancy's sentence, 'I am a cook' is ascending true in his context iff Clancy is a cook; it implies that Clancy's sentence 'I am a cook' is descending true in his context iff Clancy is a cook.

At another point in their conversation, Clancy writes 'Clancy is a cook' on the blackboard and utters 'the sentence is true'. The presemantic theory selects 'true(the sentence)' to represent Clancy's sentence and a index to represent his context. In L, 'the sentence' acts like an indexical and it picks out whatever object is salient in the context—here, it is the sentence written on the blackboard.[40] I stipulated that every sentence of L is in the domain, so there is no problem picking out 'Clancy is a cook'. The clause representing Clancy's sentence, 'true(the sentence)', is assigned a proposition, which is then assigned truth values at each point of evaluation. The truth value assigned to it at a point $\langle w, t, s\rangle$ where $s=s_a$ is the truth value that would be assigned to the proposition expressed by the corresponding sentence of L′ at w and t: ascending true(the sentence). That is, the points of evaluation with the ascending aletheic standard treat the proposition in question as if Clancy had uttered a sentence containing 'ascending true' instead of 'true'. Likewise, the points of evaluation with the descending truth standard treat the proposition in question as if Clancy had uttered a sentence containing 'descending true' in place of 'true'. The truth value assigned to it at a point $\langle w, t, s\rangle$ where $s=s_d$ is the truth value that would be assigned to the proposition expressed by the corresponding sentence of L″ at w and t. Since 'Clancy is a cook' is descending true at the world and time in question, the semantic theory assigns 1 to the proposition at both $\langle w, t, s_a\rangle$ and $\langle w, t, s_d\rangle$ where w and t are the world and time of the index representing Clancy's context. Thus, the

[40] Again, this feature is an easy way to generate self-reference, but it is not a realistic take on the semantics for locutions like 'the sentence'.

postsemantic theory implies that 'the sentence is true' is ascending true in Clancy's context and that 'the sentence is true' is descending true in Clancy's context.

Later, Ned writes 'the sentence is not true' on the blackboard. Although we have been considering utterances as verbal performances, let us be a bit more liberal and treat his inscription as an utterance. The presemantic theory selects '~true(the sentence)' to represent Ned's sentence, and the o slot in the index that represents his context picks out '~true(the sentence)' (i.e., the clause of the sentence written on the blackboard). The semantic theory assigns a character to the clause for Ned's sentence. Since it contains an indexical, 'the sentence', the character picks out the proposition that '~true(the sentence)' is not true; call this $\{\phi\}_{\mathfrak{F}, i}$. This proposition is assigned a truth value at every point. For points <w, t, s_a> it assigns a truth value based on whether the corresponding proposition $\{\phi'\}_{\mathfrak{F}, i}$ is true at w and t, and at points <w, t, s_d> it assigns a truth value based on whether the corresponding proposition $\{\phi''\}_{\mathfrak{F}, i}$ is true at w and t. In L′, 'the sentence' refers to '~ascending true(the sentence)', which is ascending true and not descending true at w and t. Hence,

$[[\text{'the sentence'}]]_{\mathfrak{F}, i, w, t, s} \in [[\text{'ascending true'}]]_{\mathfrak{F}, i, w, t, s}$,

so

$[[\text{~ascending true(the sentence)'}]]_{\mathfrak{F}, i, w, t, s} = 1$.

Therefore, $\{\phi\}_{\mathfrak{F}, i}$ is true at <w, t, s_a>. On the other hand, In L″, 'the sentence' refers to '~descending true(the sentence)', which is ascending true and not descending true at w and t. Hence, $[[\text{'the sentence'}]]_{\mathfrak{F}, i, w, t, s} \notin [[\text{'descending true'}]]_{\mathfrak{F}, i, w, t, s}$, so $[[\text{~descending true(the sentence)'}]]_{\mathfrak{F}, i, w, t, s} = 0$. Therefore, $\{\phi\}_{\mathfrak{F}, i}$ is not true at <w, t, s_d>. The postsemantics yields the following results: Ned's sentence, 'the sentence is not true' is ascending true in his context, and his sentence is not descending true in his context. In other words, Ned utters the liar sentence. Since the ascending standard reads his sentence as the ascending liar, it treats Ned's sentence as ascending true in his context. Since the descending standard reads his sentence as the descending liar, it treats Ned's sentence as not descending true in his context.

14.5.3.2 The Assessment-Sensitivity Option

So far, the example illustrates only the non-indexical contextualist option, but it is not hard to see how it would be changed for the assessment-sensitivity option. To do so, redefine $\mathfrak{I}$, the interpretation function, so that it assigns extensions to expressions at $\mathfrak{F}$, i, w, t, s_1, s_2, and add an extra aletheic standard slot to the points of evaluation. Also, we need to change the postsemantics in accord with the description above:

(11) A sentence p is *ascending true* at c from a iff the content assigned to the clause representing p with respect to i_u is true at the point of evaluation <w, t, s_1, s_2> where w and t are the world and time of i_u, s_1 is the *ascending* standard, and s_2 is the aletheic standard from i_a.

(12) A sentence p is *descending true* at c from a iff the content assigned to the clause representing p with respect to i_u is true at the point of evaluation <w, t, s_1, s_2> where w and t are the world and time of i_u, s_1 is the *descending* standard, and s_2 is the aletheic standard from i_a.

Consider again Ned's utterance of 'the sentence is not true', but imagine it is being assessed by Clancy, who decides to use the descending aletheic standard in his context of assessment.

The presemantics is the same: it selects '~true(the sentence)' to represent Ned's sentence, and the o slot in the index that represents his context picks out '~true(the sentence)'. The semantic theory assigns a character to the clause for Ned's sentence, and this character picks out the proposition that '~true(the sentence)' is not true; call this $\{\phi\}_{\mathfrak{F},i}$. But now we have extra options for assigning truth values.

For points <w, t, s_a, s_a> it assigns a truth value based on whether the corresponding proposition $\{\phi'\}_{\mathfrak{F},i}$ is ascending true at w and t, and at points <w, t, s_a, s_d> it assigns a truth value based on whether the corresponding proposition $\{\phi''\}_{\mathfrak{F},i}$ is ascending true at w and t. For points <w, t, s_d, s_a> it assigns a truth value based on whether the corresponding proposition $\{\phi'\}_{\mathfrak{F},i}$ is descending true at w and t, and at points <w, t, s_d, s_d> it assigns a truth value based on whether the corresponding proposition $\{\phi''\}_{\mathfrak{F},i}$ is descending true at w and t. In L′, 'the sentence' refers to '~ascending true(the sentence)', which is ascending true and not descending true at w and t. Hence, $[[\text{'the sentence'}]]_{\mathfrak{F},i,w,}$

$_{t, sa, sd} \in$ [['ascending true']]$_{\mathfrak{F}, i, w, t, sa, sd}$, so [[~ascending true(the sentence)']]$_{\mathfrak{F}, i, w, t, sa, sd} = 1$. Therefore, $\{\phi\}_{\mathfrak{F}, i}$ is true at <w, t, s_a, s_d>. On the other hand, in L″, 'the sentence' refers to '~descending true(the sentence)', which is ascending true and not descending true at w and t. Hence,

[['the sentence']]$_{\mathfrak{F}, i, w, t, sd, sd} \notin$ [['descending true']]$_{\mathfrak{F}, i, w, t, sd, sd}$,

so

[[~descending true(the sentence)']]$_{\mathfrak{F}, i, w, t, sd, sd} = 0$.

Therefore, $\{\phi\}_{\mathfrak{F}, i}$ is not true at <w, t, s_d, s_d>. The second aletheic standard slot produces a reading of the sentence, while the first encodes whether the semantics is assessing it for ascending truth or descending truth.

The postsemantics produces an ascending truth value (based on the second slot) for the sentence in a context of use from a context of assessment (controlled by the first slot), and a descending truth value (based on the second slot) for the sentence in a context of use from a context of assessment (controlled by the first slot). In particular, it yields the following results: Ned's sentence, 'the sentence is not true' is ascending true in his context from Clancy's context, and his sentence is not descending true in his context from Clancy's context. In other words, Ned utters the liar sentence. Since Clancy's context invokes the descending standard, the semantics reads Ned's sentence as the descending liar. Then the ascending standard assesses the descending liar for ascending truth—it is ascending true. Then the descending standard assesses the descending liar for descending truth—it is not descending true.

On the non-indexical contextualist option, the reading of Ned's sentence and the evaluation of that reading are controlled by a single slot, so there are only two options—assessing the ascending liar for ascending truth or assessing the descending liar for descending truth. However, on the assessment-sensitivity option, there are two slots, and hence four options—assessing the ascending

liar for ascending truth and for descending truth, and assessing the descending liar for ascending truth and for descending truth.[41]

14.5.3.3 Other Options

There are other options as well. The semantic theory could use four truth values instead of two: 0=not ascending true, 1=ascending true, 2=not descending true, and 3=descending true. The t-distributions would be assignments of two truth values (0 or 1) and (2 or 3) to a proposition at each point of evaluation. That would eliminate the need for any aletheic-standard slot in the points of evaluation for the non-indexical contextualist option, and it would require only a single aletheic-standard slot in the points of evaluation for the assessment-sensitivity option. In both of these cases, the reading of the sentence Ned utters would be handled by the truth-value assignments, not slots in the points of evaluation. Assigning a sentence 0 or 1 would be reading it in the ascending-truth way, while assigning it 2 or 3 would be reading it in the descending-truth way.

An even more radical departure would be to have the semantic theory generate two distinct models—an ascending model and a descending model. I offer in the next chapter some hints about how this might be done in the discussion of the paradoxes associated with predication and reference. However, the details will have to wait for another time.

14.6 Resolving the Paradoxes

I have not said which option, non-indexical contextualism or assessment-sensitivity, I prefer. There are pros and cons of each. The former is surely simpler (i.e., one degree of variability in the aletheic

[41] There are other options as well. The semantic theory could use four truth values instead of two: 0=not ascending true, 1=ascending true, 2=not descending true, and 3=descending true. The t-distributions would be assignments of two truth values (0 or 1) and (2 or 3) to a proposition at each point of evaluation. An even more radical departure would be to have the semantic theory generate two distinct models—an ascending model and a descending model.

standard), and I would prefer it if the latter were not more versatile (i.e., two degrees of variability). It is not yet clear whether that additional versatility might come in handy. So, I have presented the two postsemantic approaches to truth and the aletheic paradoxes without choosing between them so far.

Note that the two approaches say the same thing about the liar reasoning—it is invalid because it uses (T-In) and (T-Out), both of which have exceptions on these approaches. To justify this claim, we need to take a look at validity.

14.6.1 Validity

In Chapter Five, I mentioned that every logical approach to the paradoxes is inconsistent with the claim that an argument is valid iff it is truth preserving. That might seem to leave us without an account of validity at all. I think this impression is mistaken. The reason has to do with the point made in the last section that truth-in-a-model is a mathematical concept, not affected by the paradoxes associated with truth. The definition of validity is:

(Valid) $\langle \Gamma, \phi \rangle$ is valid iff for every model $\mathfrak{M}$, if all the members of Γ are true-in-$\mathfrak{M}$, then ϕ is true-in-$\mathfrak{M}$.

For our purposes, a model is a point of evaluation in the structure described in the last couple of sections. Thus, an argument $\langle \Gamma, \phi \rangle$ (where ϕ and all the members of $\Gamma$are sentences of L) is valid iff for every point of evaluation e in $\mathfrak{F}$, if for all $\gamma \in \Gamma$, γ is true at e, then ϕ is true at e.

Notice that all the inference rules of classical logic are valid according to (Valid) since every point of evaluation is classical. Notice also that,

(T-In) if ϕ, then $\langle \phi \rangle$ is true, and

(T-Out) if $\langle \phi \rangle$ is true, then ϕ

are false at some points of evaluation. For example, (T-In) is false at some points of evaluation with the descending standard and (T-Out) is false at some points of evaluation with the ascending standard (in the assessment sensitivity approach, which has two aletheic standards at each point, it is the first slot that matters). Thus, according to the categorization of logical approaches to the aletheic paradoxes given in Chapter Three, both postsemantic approaches are classical symmetric—they are fully compatible with classical logic and deny both (T-In) and (T-Out). The associated inference rules,

(T-Intro) $\phi \vdash \langle\phi\rangle$ is true, and

(T-Elim) $\langle\phi\rangle$ is true $\vdash \phi$

are both invalid according to the postsemantic approaches. That is consistent with a classical symmetric logical approach.

In addition, note that on this inconsistency approach, (T-In) and (T-Out) are constitutive of truth, but they are not true in general. That is a crucial feature of any inconsistency approach that avoids dialetheism—since the constitutive principles of an inconsistent concept are inconsistent (possibly with respect to some additional assumptions), they cannot all be true.

14.6.2 The Liar

One point to notice is how this theory deals with paradoxical sentences. For example:

(13) (13) is not true.

The liar argument is below:

1. Assume (13) is true
2. '(13) is not true' is true
3. (13) is not true.

4. Assume (13) is not true.
5. '(13) is not true' is true.
6. (13) is true.
7. Therefore (13) is true iff (13) is not true.

Our language from the last section, L, cannot express this argument since the only way to refer to its sentences is 'the sentence'. Still, it would be easy to add '(13)' as a constant so we could formulate and evaluate this argument (I omit the details).

It is easy to see that the argument is invalid—it fails at two steps. It fails at step 3 because it is not the case that the inference from "(13) is not true' is true' to '(13) is not true' is valid—it fails at points of evaluation with the ascending standard. Likewise, the argument fails at step 5 as well, since this step is invalid due to points of evaluation with the descending standard.

14.6.3 Curry and Yablo

The other aletheic paradoxes are blocked as well for the same reason—neither (T-In) nor (T-Out) is true at every point of evaluation. Since the other two major aletheic paradoxes, Curry's paradox and Yablo's paradox, both require (T-In) and (T-Out), none of the arguments in them is valid according to the postsemantic approaches outlined here.

14.6.4 Montague and McGee

The other two major paradoxes affecting truth (described in Chapter Two) are Montague's paradox and McGee's paradox. I discussed the way in which ascending truth and descending truth avoid each of these in Chapter Thirteen. The postsemantic approaches presented in this chapter avoid them in the same way, which should be obvious by now.

14.6.5 Truth Tellers

Although the truth teller is not paradoxical, thinking about how it is handled by the non-indexical contextualist and assessment-sensitivity approaches provides us with additional data that might be relevant in deciding between them. The truth teller is:

(14) (14) is true.

Again, L cannot express this sentence but it is easy to add '(14)' to L so that it can (I omit the details).

Non-indexical contextualism has only one slot for the aletheic standard, which serves two purposes: a reading of (14) and the entry in the aletheic conditions for (14). As I mentioned in the last chapter, both truth tellers are safe; the descending truth teller is not ascending true and the ascending truth teller is descending true.[42] Thus, the non-indexical contextualist approach has the following consequences for context C: (14) is ascending true in C and (14) is not descending true in C. That is, it has the same aletheic conditions as the liar.

The assessment-sensitivity view differs on the truth teller. Assessment-sensitivity semantics has two slots for aletheic standards; one controls the reading of (14) and the other dictates which aletheic condition is relevant. The assessment-sensitivity approach implies that (as used in context C) (14) is descending true in C from contexts with the ascending standard and (14) is not ascending true in C from contexts with the descending standard. That is significantly different from the status the liar has. Therefore, non-indexical contextualism cannot distinguish between paradoxical sentences like the liar and merely odd sentences like the truth teller, but the assessment-sensitivity approach can. That is a big point in favor of assessment-sensitivity, and for this reason, I tentatively endorse the assessment-sensitivity approach (also known as non-indexical relativism) to the aletheic

[42] Again, this claims are based on the intended xeno model for ADT; I do not know whether they are consequences of ADT.

paradoxes. Nevertheless, it could turn out that the non-indexical contextualist approach is superior after more data come in.

14.7 Problems for the Assessment-Sensitivity Approach

The literature on non-indexical contextualism and assessment-sensitivity is large and growing every week it seems. A brief summary is in order. Below is a list of the views that appeal to assessment-sensitivity or non-indexical contextualism:

(i) *General*: applies language-wide or at least for a large number of linguistic expressions.[43]

(ii) *Predicates of personal taste*: for example, 'tasty', 'disgusting', and 'fun'.[44]

(iii) *Epistemic modals*: for example, 'might' and 'could'.[45]

(iv) *Knowledge*: the word 'knows' and its cognates.[46]

(v) *Future*: all physically possible claims about the future.[47]

(vi) *Morality*: for example, 'good' and 'right'.[48]

(vii) Color: for example, 'red'.[49]

(vii) *Vagueness*: for example, 'bald' and 'heap'.[50]

(viii) *Confusion*: for example, 'mass' and 'Boche'.[51]

(ix) *Relativistic*: for example, 'duration'.[52]

Those seem to be the major ones, but there might be others as well.

---

[43] See Kolbel (2002, 2003, 2004, 2007), Predelli (2005), Recanati (2007, 2008), Predelli and Stojanavic (2008), Einheuser (2008), Egan (2009), Parsons (forthcoming), MacFarlane (forthcoming d).
[44] See Kölbel (2002, 2003), MacFarlane (2005a, 2007a, forthcoming c, forthcoming d), Egan (2006, forthcoming), Lasersohn (2005, 2007, 2009), and Einheuser (2008).
[45] See Kölbel (2002), Egan, Hawthorne, and Weatherson (2005), Egan (2007), Stephenson (2007), MacFarlane (forthcoming a), Einheuser (2008).
[46] See MacFarlane (2005b, forthcoming b), and Brogaard (2009).
[47] See MacFarlane (2003, 2008), and Brogaard (2010).
[48] See Kölbel (2002, 2004, 2007) and MacFarlane and Kolodny (forthcoming).
[49] See Egan (2006, forthcoming) and Brogaard (2010).
[50] See Richard (2004, 2007) and Kölbel (2009).
[51] See MacFarlane (2007b).
[52] See Pinillos (2010).

Most of these theorists argue that their theories capture the linguistic data better than the alternatives. The linguistic data include the surface grammar of the expressions in question, the ways in which speakers take one another to agree on certain points in certain situations and disagree on certain points in other situations, the circumstances under which agents treat one another as having said the same thing, the ways in which speakers treat themselves and one another as authoritative on certain issues, and the ways in which speakers retract claims in light of certain challenges. For example, one might think that if one person asserts that stewed rhubarb is tasty and another asserts that it is not, then they disagree, but neither of them are guilty of some cognitive fault or shortcoming. Instead, this might be a case of faultless disagreement. Some non-indexical contextualists and non-indexical relativists have argued that their views offer the best explanation of faultless disagreement.[53] In addition, some non-indexical relativists argue that their view explains speakers' tendency to retract previous claims better than non-indexical contextualism; for example, if Millhouse at age ten asserts that Squishees are tasty, but then at age twenty denies that they are tasty, he might say that his age ten utterance was mistaken. It is difficult for non-indexical contextualism to explain this behavior since it entails that the sentence Millhouse uttered at age ten is true in that context of use. However, the non-indexical relativist can say that the sentence Millhouse uttered at age ten is true in the age ten context of use from the age ten context of assessment, but the sentence he uttered at age ten is false in the age ten context of use from the age twenty context of assessment.[54]

Non-indexical contextualism and non-indexical relativism have come in for plenty of criticism as well. Below are some of the more prominent objections:

(i) *Self-refutation*: traditional forms of relativism are thought to be self-refuting—some argue that the new forms are as well.[55]

---

[53] See Kölbel (2002, 2003), Lasersohn (2005), Recanati (2007), and MacFarlane (2005a, 2007a, forthcoming d).
[54] See MacFarlane (2005a, 2007a, forthcoming d) for discussion.
[55] See Moruzzi (2008), Wright (2008), and Moruzzi and Wright (2009),

(ii) *Extra parameters*: some argue that it is unclear how to understand the extra parameters in the points of evaluation that are required by non-indexical contextualism and non-indexical relativism.[56]

(iii) *Faultless disagreement*: some argue that the phenomenon of faultless disagreement has been mischaracterized or does not exist.[57]

(iv) *Retraction*: some argue that the retraction data have been mischaracterized or do not exist.[58]

(v) *Representation*: some argue that non-indexical contextualism and non-indexical relativism are incompatible with the claim that the propositions in question are representational.[59]

(vi) *Indexicalism*: some argue that indexicalism offers a better explanation of the phenomena in question.[60]

(vii) *Utterance*: some argue that non-indexical contextualism has counterintuitive consequences for utterance truth.[61]

(viii) *Specific*: some present criticisms that are specific to particular applications—for example, Jason Stanley argues that non-indexical relativism with respect to knowledge entails that 'knows' is not factive.[62]

Since all the non-indexical contextualists and assessment-sensitivity theorists (*postsemantic theorists* hereafter) currently advocate one of these views as descriptive theories of what they take to be *consistent* concepts (except in the case of confusion), most of the objections are irrelevant for my purposes.

In the case of truth, almost all speakers are ignorant of the fact that truth is an inconsistent concept. Thus, speakers use it as if it were consistent; hence, the above kinds of linguistic data with respect to truth are not decisive. Since speakers are ignorant of truth's inconsistency, they are bound to make mistakes with it. We already know that we do not see faultless disagreement or retraction

---

[56] See Glanzberg (2007, 2009), Rosenkranz (2008), and Cappelen and Hawthorne (2009).
[57] See Glanzberg (2007), Zimmerman (2007), Stojanovic (2007), Iacona (2008), Gilles and von Fintel (2008), Garcia-Carpintero (2008), Wright (2008), Cappelen and Hawthorne (2009, forthcoming a, forthcoming b, forthcoming c, forthcoming d), Moltmann (2009), Greenough (forthcoming), and Schaffer (forthcoming).
[58] See Dietz (2008), Gilles and von Fintel (2008), Wright (2008), Cappelen and Hawthorne (2009), Moltmann (2009), and Schaffer (forthcoming).
[59] See Wright (2008). See also Boghossian (2006) and Zimmerman (2007).
[60] See Lopez de Sa (2007), Glanzberg (2007, forthcoming), Capellen (2008), Cappelen and Hawthorne (2009), Gilles and von Fintel (2008), and Schaffer (forthcoming).
[61] Hawthorne and Capellen (2009).
[62] Stanley (2005); see also von Fintel and Gilles (2008) on epistemic modals, and Glanzberg (2007) and Stojanovic (2007) on predicates of personal taste.

data in the case of truth because speakers are unaware that truth is inconsistent, and, hence, they are unaware that 'true' is assessment-sensitive.

Notice, however, that I have not argued for an assessment-sensitive view in the familiar way. Instead, once one accepts an inconsistency approach to the aletheic paradoxes, one must choose presemantic, semantic, and postsemantic theories for truth on the basis of more general principles about language use, like the Gricean Condition defended above. Indeed, the Gricean Condition gives us good reason to think that if truth is an inconsistent concept, then only a postsemantic approach to the aletheic paradoxes is acceptable. The reason is that truth is empirically inconsistent—it is inconsistent by virtue of the empirical context in which it is used—that is, rational agents that reason more or less classically and speak natural languages that have the capacity to represent their own syntax. In this environment, any concept that has (T-In) and (T-Out) as constitutive principles is inconsistent. Had things been different, truth might not have been inconsistent. Therefore, 'true' is assessment-sensitive not entirely because of the ways in which speakers use it. It is assessment-sensitive because of the ways in which speakers use it together with the environment in which it is used. Speakers can be, and often are, ignorant of the fact that this environment is hostile to a concept with (T-In) and (T-Out) as constitutive principles. Thus, speakers are ignorant of the fact that 'true' is assessment-sensitive.

Because of the Gricean Condition, it is unacceptable for us to use a presemantic approach (e.g., 'true' is ambiguous or has unarticulated constituents) or a semantic approach (e.g., 'true' is use-indexical) because these would imply that speakers and hearers are ignorant of the propositions expressed by sentences containing 'true'. In other words, it cannot be that linguistic expressions are ambiguous or indexical or have unarticulated constituents by virtue of the environment in which they are used. However, a linguistic expression can be assessment sensitive by virtue of the environment in which it is used if the concept expressed by that expression is inconsistent in that environment. That, I claim, is exactly the case with truth.

This is a delicate move given the emphasis I have put on accounting for linguistic usage and the criticisms I have leveled at others (e.g., ambiguity approaches, context dependence approaches,

paracomplete approaches, paraconsistent approaches, and Millianism). Both parties claim that some seeming correct uses of natural language truth predicates are unacceptable. However, the difference is that these views run afoul of the Gricean Condition and so make the very use they are trying to explain mysterious. Why would anyone utter sentences that express propositions neither the speaker or the hearer can retrieve? We have made great strides in linguistics and philosophy of language by treating communication as a rational enterprise. Absent some other account that not only does as good a job of explaining the use of language, but also has the same predictive success we have seen in linguistics, it would be insane to abandon it. My view says people make mistakes with truth predicates because they are ignorant of their surroundings (the same goes for 'mass'); the others say people make mistakes because they are ignorant of what their words mean. On my view, communication makes sense and the dominant tools of linguistics are applicable to it; on their view it is at best mysterious and dominant tools of linguistics, at least as they stand, are impotent to explain it.

Some of the other objections to postsemantic theories are relevant. For example, Cappelen and Hawthorne claim that non-indexical contextualism has horribly counterintuitive consequences for the contrast between the truth of utterances and the truth of propositions. Take non-indexical contextualism with respect to 'cold' as an example and assume that Tim arrives in an antechamber from outside and asserts 'the antechamber is not cold', while Crispin walks into the same antechamber from the attached hot baths and asserts 'the antechamber is cold'. Cappelen and Hawthorne claim that the non-indexical contextualist should recommend that Tim respond by asserting 'your assertion is true but the proposition that you are expressing by your assertion is not true'.[63] On the basis of this counterintuitive consequence, Cappelen and Hawthorne conclude: "when the smoke has cleared, we find it hard to see any significant avenues opened up by non-indexical contextualism."[64]

---

[63] Cappelen and Hawthorne (2009: 22).
[64] Cappelen and Hawthorne (2009: 24).

The problem is nicely diagnosed by Brogaard who writes: "Two different notions of truth are in play here. One is monadic utterance truth, the other relative propositional truth. A better-tasting and more easily digestible version […] would be: your [assertion] is true simpliciter but the proposition you are expressing by your [assertion] is not true relative to me as the speaker, though it is true relative to you as the speaker."[65] The "objection" is based on a simple equivocation.

Consider instead a substantive objection: Wright's argument that postsemantic theories are only appropriate for propositions that are non-representational. Although this result would probably be unwelcome for many applications of postsemantic theories, it does not affect mine since I do not think there is any property of truth to be represented by our word 'true' and our concept of truth. Instead, there are two properties, being ascending true and being descending true and anyone who thinks there is a property of being true is confused. So Wright's worry does not affect my proposal.

Finally, I have argued that indexicalism is not a good candidate for 'true' because of the combination of empirical unsafety and the Gricean Condition. So the standard criticism that indexicalism explains the data better than postsemantic theories has no bite in the case of truth.

Under what conditions, on an assessment-sensitivity approach, do speakers make mistakes with 'true'? Here is one obvious proposal. If the difference between ascending truth and descending truth is negligible, then it is legitimate to use 'true'. If it is not, then 'ascending true' and 'descending true' should be used instead. When is the difference negligible? Recall Craige Roberts' pragmatic theory (from Chapter Six). One of her innovations is the idea of a question under discussion (QUD), which guides the conversation, has a role in determining conversational implicatures and presuppositions (because it affects whether participants are following the conversational maxims—quality, quantity, relevance, and manner), and helps explain other phenomena including anaphora, deixis, ellipsis, focus, and prosody. My suggestion is that if it is in the common ground that the common ground entails that an answer to the question under discussion requires a distinction

---

[65] Brogaard (2010: 3).

between ascending truth and descending truth, then the distinction is not negligible. Otherwise it is. Notice the two roles of the common ground here: it determines whether the distinction is needed in order to answer the question under discussion, and it determines whether people know that it makes this determination.[66]

One could also think of this as a suggestion about the standards of precision in the conversation. To go back to our paradigm example, in a conversation about how best to economically design a house so that it is most likely to survive an earthquake of 7.0 or less, the distinction between relativistic mass and proper mass is negligible, since this distinction is not needed in order to answer this question. However, in a conversation about the source of dark flow (i.e., the observed but currently unexplained motion of hundreds of galaxy clusters in the same direction relative to the cosmic background radiation[67]), the distinction is relevant because a failure to distinguish between proper mass and relativistic mass would most certainly prevent the participants from finding the right answer (indeed, it would probably be impossible to even index the question since the distinction is presupposed by ΛCDM—the standard model of cosmology used to demonstrate the phenomenon of dark flow). Likewise, in a conversation about how best to give a semantics for the fragment of language that linguists use in order to do semantics for natural languages, one had better distinguish between ascending truth and descending truth, since a failure to do so would result in an inconsistent theory, which would also fail to answer the question under discussion. However, in a conversation about whether a friend, Jessica, should be trusted, the participants can almost certainly use 'true' without any trouble. Even if they end up uttering or assessing what turn out to be paradoxical sentences, these are "close enough" to non-paradoxical ones for the purposes at hand.

---

[66] This suggestion requires a non-standard notion of common ground since it presupposes that not all logically true propositions are in the common ground. However, it is obvious that we need something like this anyway, otherwise it would be infelicitous to assert the Modularity Theorem even for the first time after proving it!

[67] See Kashlinsky et. al. (2008, 2009) for details.

That is, even if they had distinguished between ascending truth and descending truth, they would have arrived at the same conclusions, but with considerably more effort.

Note that some care must be taken in applying this pragmatic theory for 'true'. Consider a Monty, who lived in 1500 B.C.E. If Monty asserts that the Earth stands still and the Sun revolves around it, then we can hardly fault him for asserting something false since, given his level of intelligence, education, and the state of technology and common knowledge at the time, he is incapable of knowing any better. A proper pragmatic theory ought to reflect these facts. However, if Monty were alive today and asserted that the Earth stands still and the Sun revolves around it, then we should almost certainly say that his assertion is inappropriate because it is not true—and he should know better (unless he was brainwashed by some cult or he grew up in Kansas). I am suggesting that correct assertion should be thought of as dependent on various aspects of context; however, I am not going to delve more deeply into this issue. It should be sufficient to say that, whatever turns out to be the right view of correct assertion, it should handle cases like these; so it should also be able to handle the distinction between cases where people inadvertently use 'true' in situations where the distinction between ascending truth and descending truth is relevant (unbeknownst to them) and in cases where they should know better. The double use of common ground in the suggestion above is meant to be a step in this direction.

In sum, the assessment-sensitivity approach is compatible with the Gricean Condition; with the exception of the non-indexical contextualist approach, the others studied here are not. Thus, it makes sense to say that a linguistic expression can turn out to be assessment-sensitive by virtue of the way it is used and the environment in which it is used (which might be unknown or unrecognized by those who use it), whereas it does not make sense to say this about use-indexicality, ambiguity, or unarticulated constituents. In these cases, the standard objections to assessment-sensitivity views are inapplicable, since these objections presuppose that speakers would be aware of

the fact that their expressions are assessment-sensitive. At some point after the aletheic revolution, when those for whom it is relevant know to distinguish between ascending truth and descending truth, we shall have more data on how these people use 'true', and we might need to reassess whether the standard objections have any bite.

14.8 The Nature of Truth

In the last chapter, I proposed a measurement system for ascending truth and descending truth as an alternative to an analysis or a reductive explanation. As I mentioned in the Introduction, this strategy is in accord with my views on philosophical methodology; I advocate measurement-theoretic methodological naturalism (MTMN) as a philosophical methodology, according to which a philosophical theory of X should have the form of a measurement system for X. It should come as no surprise that I follow the same strategy here. A theory of truth should have the form of a measurement system for truth. Although I am not alone in this view, there are not many of us. The most obvious and famous example is Donald Davidson's theory of truth. His most extended discussion occurs in his 1990 John Dewey lectures "The Structure and Content of Truth". However, even in "Radical Interpretation", published in 1973, Davidson writes:

> Here I would like to insert a remark about the methodology of my proposal. In philosophy we are used to definitions, analyses, reductions. Typically these are intended to carry us from concepts better understood, or clear, or more basic epistemologically or ontologically, to others we want to understand. The method I have suggested fits none of these categories. I have proposed a looser relation between concepts to be illuminated and the relatively more basic. At the centre stands a formal theory, a theory of truth, which imposes a complex structure on sentences containing the primitive notions of truth and satisfaction. These notions are given application by the form of the theory and the nature of the evidence. The result is a partially interpreted theory. The advantage of the method lies not in its free-style appeal to the notion of evidential support but in the idea of a powerful theory interpreted at the most advantageous point. This allows us to reconcile the need for a semantically articulated structure with a theory testable only at the sentential level. The more subtle gain is that very thin evidence in support of each of a potential infinity of points can yield rich results, even with respect to the points. By knowing only the conditions under which

> speakers hold sentences true, we can come out, given a satisfactory theory, with an interpretation of each sentence.[68]

Although Davidson does not explain his views in detail here, when one follows up on the hints in this passage, one arrives at MTMN with respect to truth. As far as I can tell, this kind of theory of truth has not received the attention it deserves in the debates about the nature of truth.[69]

A measurement system for truth based on the assessment-sensitivity view given in the last section has as its physical structure a natural language containing a truth predicate; as its relational structure an artificial language with an assessment-sensitive truth predicate and the theory of that assessment-sensitive truth predicate from the previous section; as and its mathematical structure the model theory for the artificial language. Let us go through these in a little more detail.

The physical structure is just as it was in Chapter Thirteen—a natural linguistic practice with a truth predicate that has a finite lexicon and a recursive syntax. For simplicity, we can assume that it does not contain ascending truth or descending truth predicates (although we can drop this assumption once we see how the truth predicate works).

The relational structure contains a classical first order artificial language with the usual syntax and a truth predicate. As before, the artificial language can contain anything for which we have canonical semantic theories (e.g., names, definite descriptions, mass nouns, adverbs, indexicals, demonstratives, pronouns, and gradable adjectives). The theory of truth for this language takes its truth predicate to be assessment sensitive, as described above.

The mathematical structure consists of the standard set-theoretic structure for modeling assessment-sensitivity, which consists in the standard intensional semantics with possible worlds. However, just as in most model theory, one defines truth-in-a-model, which is a technical,

---

[68] Davidson (1973: 137).
[69] However, see Patterson (2010) for refreshing presentation and defense of a very similar view.

mathematical concept. All of this was described above and illustrated with the example of language L.

As in the measurement system for ascending and descending truth described in the last chapter, the connection between the physical structure and the relational structure is very complex—it consists of translating the natural language into the artificial language (as clauses) in order to arrive at the logical form of the natural language sentences, and representing the contexts of use with indexes. There are many fascinating issues here, but none of them is specific to my project—they all bear on formal semantics in general. I have touched on many of these in this chapter and in Chapter Seven (e.g., Predelli on interpretive systems and linguistic practices and the distinction between presemantics, semantics proper, and postsemantics).

The connection between the relational theory and the mathematical theory is accomplished in the usual way by defining truth-in-a-model. I outlined the way this works above and illustrated it with the example language L, but there are many technical details omitted in the interest of space. The definition of truth-in-a-model requires an account of ascending truth and descending truth, which is provided by ADT in the previous chapter.

Again, this measurement system for truth is an alternative to an analysis of truth or a reductive explanation of truth. Note that no consistent analysis of truth can have (T-In) and (T-Out) as consequences and be compatible with classical logic. When it comes to inconsistent concepts, both conceptual analysis and reductive explanation are non-starters unless one is happy embracing a non-classical logic. Measurement-theoretic methodological naturalism allows us to have a consistent theory of an inconsistent concept without changing our logic. There is much more to be said on this topic, but considerations of length suggest that it should be given an independent treatment.

14.9 A Unified Theory of Truth: CAM

Recall that a unified theory of truth consists of a theory of the nature of truth (which often consists of a conceptual analysis of truth or some other theory specifying what truth is and what kinds of things are true), a philosophical approach to the aletheic paradoxes (which specifies syntactic, semantic, or pragmatic features of natural language truth predicates that are relevant to the aletheic paradoxes), and a logical approach to the aletheic paradoxes (which specifies which aletheic principles truth predicates obey and the strongest logic compatible with those principles). In this chapter and the last, I have presented all the elements of a unified theory of truth.

The measurement system given in the last section is the theory of the nature of truth (together with the claims that truth is an inconsistent concept and words expressing inconsistent concepts are assessment-sensitive). This theory of the nature of truth incorporates both a logical approach to the aletheic paradoxes and a philosophical approach to the aletheic paradoxes.

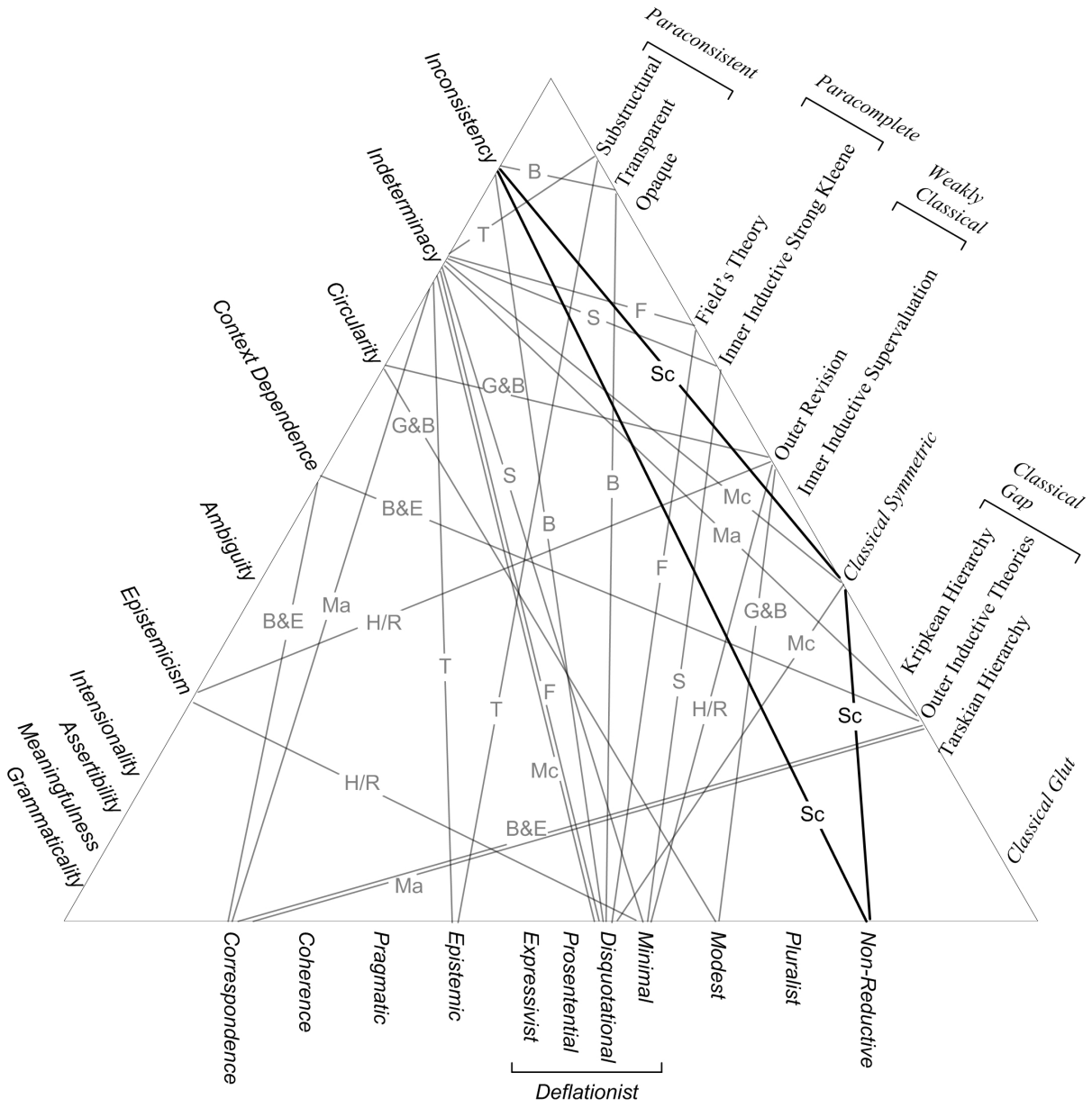
The logical approach is given by the theory of an assessment-sensitive truth predicate—it is a classical theory (i.e., compatible with classical logic) and it takes (T-In) and (T-Out) to hold for most sentences, but they have exceptions. The exceptions are specified by the assessment-sensitive semantics. The philosophical approach is given by the claims that truth is an inconsistent concept and that words expressing inconsistent concepts are assessment-sensitive.

Two aspects of the unified theory of truth deserve mention: since truth is a useful inconsistent concept, it needs to be replaced with one or more consistent concepts (for me, ascending truth and descending truth), and those consistent concepts play an important role in the unified theory of truth—they serve crucial explanatory roles in the assessment-sensitivity semantics for the truth predicate. So, the theory of ascending truth and descending truth (ADT) and its formal semantics (xeno semantics) are incorporated into the unified theory of truth. We can call this a Classical, Assessment-sensitive, Measurement-theoretic unified theory of truth (CAM theory). I repeat a

diagram from Chapter Four showing the other unified theories of truth that have been proposed and how the one I have presented fits into the picture (see Information Box 36).

CAM Theory

Information Box 36



B&E	Barwise & Etchemendy	Ma	Maudlin
G&B	Gupta & Belnap	F	Field
H/R	Horwich / Restall	B	Beall
Mc	McGee	T	Tennant
S	Soames	Sc	Scharp

Note that not only is CAM a unified theory, but it is also an integrated theory in the sense that all the parts fit together naturally in the measurement system for truth. The other unified theories just take a view on the nature of truth (e.g., disquotationalism), a compatible logical approach (e.g., paracomplete), and a compatible philosophical approach (e.g., indeterminacy), and stick them together (often not even explicitly). By contrast, the three components of CAM theory all fall out of the measurement system for truth, which is an integrated whole.

14.10 Key Issues

Presenting the key issues on which the debate about truth should focus is the primary goal of Part II. In this section, I consider CAM in light of those issues.

14.10.1 Expressive Role

Truth predicates play several distinctive expressive roles—they serve as devices of endorsement and as devices of generalization (explained in Chapter Six). How does CAM do in explaining these?

First, according to CAM, (T-In) and (T-Out) are constitutive of the concept of truth, which is the reason why it is an inconsistent concept. Since (T-In) and (T-Out) explain these expressive roles, CAM can make sense of why those who use truth predicates *take them* to serve these expressive roles. Participants in linguistic practices take (T-In) and (T-Out) to be constitutive of truth, which leads them to think these principles are true, which leads them to use truth predicates as devices of endorsement and as devices of generalization. For example, in a conversation between Sherri and Terri, Sherri asserts 'Frink's theory is true'. Both Sherri and Terri take (T-In) and (T-Out) to be constitutive of 'true', so they both take Sherri to have committed herself to the claims that constitute Frink's theory, regardless of the content of those claims, or their levels, or whether Sherri or Terri knows what Frink's theory says, or the language in which Frink's theory is formulated. That

is an example of 'true' being used as a device of endorsement, and CAM explains why people use it that way.

However, according to CAM, (T-In) and (T-Out) have exceptions—there are sentences p such that 'p is true' does not follow from p, and there are sentences q that do not follow from 'q is true' (it turns out that exceptions to one will also be exceptions to the other). Thus, if one calls one of these sentences true, then one has not thereby endorsed that sentence. For example, assume that one of the sentences that make up Frink's theory is an exception. Call it p. When Sherri asserts that Frink's theory is true, she commits herself to all the sentences of Frink's theory that are not exceptions, but she does not commit herself to p since p does not follow from 'p is true'. Even though Sherri and Terri both assume that she has committed herself to p, they are wrong. Thus, CAM predicts that people will use truth predicates as devices of endorsement because of its constitutive principles, but it also predicts that there will be some mistakes about these uses since users do not realize that these constitutive principles are inconsistent (given facts about syntax).

14.10.2 Riskiness

One might think that the above result is not a big deal because the only sentences that are exceptions are paradoxical sentences, and no one would want to endorse them anyway. However, there are empirically paradoxical sentences that are not obviously paradoxical since the empirical facts that render them as such are inaccessible in certain situations. It could be that someone wants to use 'true' to indirectly endorse some sentence, and that sentence turns out to be empirically paradoxical, but the speaker cannot tell that it is paradoxical. Thus, it is possible that someone tries to use 'true' to endorse a sentence and fails because that sentence is, unbeknownst to her, paradoxical. However, the vast majority of these cases occur in informal conversations, the participants are not aware of the problem, and the problem does not have any practical effects; thus,

it is not a worry. In conversations where the problem might have practical effects, the participants should use ascending truth predicates and descending truth predicates instead.

14.10.3 Revenge

How does CAM avoid revenge paradoxes? Consider a sentence that might seem to give rise to a revenge paradox:

(15) (15) is either false or paradoxical.

I have not shown how to introduce 'paradoxical' into our example language, but here is an intuitive way to do it:

A sentence p containing 'true' is *paradoxical* iff (T-In) and (T-Out) fail for p.

In the example language, 'paradoxical' would have as its extension all the sentences that the semantic theory treats as expressing unsafe propositions.

One might try to argue that (15) causes a problem for CAM in the following way:

1. Assume (15) is true.
2. '(15) is either false or paradoxical' is true.
3. (15) is either false or paradoxical.
4. Assume (15) is either false or paradoxical.
5. '(15) is either false or paradoxical' is true.
6. (15) is true.

If this argument were valid, it would be a problem, but it is not. CAM is fully classical, so the logical inferences in this argument are all fine. However, since (15) is paradoxical, the move from 2 to 3 is invalid and the move from 4 to 5 is invalid. The upshot is: CAM implies that (15) is paradoxical, but

it does not imply that '(15) is paradoxical' is true.[70] So, of course, CAM has consequences that it deems untrue. Is this a problem? It would be a problem if CAM implied that truth is a consistent concept, but it does not. Instead, CAM implies that truth is an inconsistent concept that cannot be legitimately applied in every circumstance. Indeed, the theory of ascending truth and descending truth specifies exactly where truth can be used without running into problems (i.e., when the difference between ascending truth and descending truth is negligible). Accordingly, CAM is outside the legitimate scope of truth—in applying truth to it, one gets odd results. Consider the analogy. Even though the concept of mass is inconsistent, it is fine to use it in certain situations (i.e., when the difference between relativistic mass and proper mass is negligible). However, if one tries to use mass outside these situations, say, in calibrating the atomic clocks in GPS satellites, it will not provide accurate predictions. Instead, in these circumstances, one needs to use the replacements.

In the case of truth, one can reasonably ask whether CAM is ascending true and whether it is descending true. All the central principles of CAM are descending true. Recall, however, that descending truth is not preserved under logical consequence. So it could turn out that CAM has some consequences that are not descending true. They would, of course, be ascending true. I have been unable to identify any consequences of CAM that have this feature, but I have not been able to rule it out either. If it does have these kinds of consequences, then it would be in the same boat as ADT. Either way, there are no revenge paradoxes here.

14.10.4 Internalizability

Since CAM does not give rise to revenge paradoxes, it does not need to be restricted in any way. It applies to fully classical languages that contain their own truth predicates and can refer to their own

[70] CAM treats 'paradoxical' defined in this way just as it would treat 'unsafe' when applied to propositions with ascending truth or descending truth as constituents. As such CAM treats '(15) is paradoxical' as an unsafety attribution, and unsafety attributions are unsafe. Thus, it is invalid to infer from '(15) is paradoxical' to "(15) is paradoxical' is true'.

syntax (this is taken to be the gold standard). There are no expressive restrictions on the languages to which CAM applies.

Consider a natural language—call it L—that contains a truth predicate and all the resources we think of today being involved in formal semantics. To see that CAM is internalizable for L, imagine the extension of L that occurs when an ascending truth predicate and a descending truth predicate are added to it. Call that extension L′. The measurement system for ascending and descending truth (which includes ADT) is expressible in L′ and applicable to L′. Likewise, CAM is expressible in L′ and applicable to L′. Thus, CAM is internalizable for L. The same goes for any other natural language.

Chapter 15

The Aletheic Revolution

In this final chapter, I take a step back and consider some of the broader ramifications of the central point of the book, i.e., that we should replace our inconsistent concept of truth with ascending truth and descending truth. I use 'aletheic revolution' as a convenient term for the conceptual revolution pertaining to truth.

15.1 Post-Revolutionary Practice

In Chapter Thirteen, I presented ADT and the measurement system for ascending truth and descending truth. In Chapter Fourteen, I offered CAM, the classical, assessment-sensitive, measurement-theoretic, unified theory of truth. The point of this overall strategy is threefold: (i) recognize that truth is an inconsistent concept and that this feature causes the aletheic paradoxes and the revenge paradoxes that plague attempted solutions to them, (ii) offer a team of consistent concepts that can do the work we require of truth without giving rise to the paradoxes, and (iii) use the replacement concepts as the explanans in a theory of truth. The descriptive theory of truth, CAM, depends on the replacement concepts, ascending truth and descending truth, explained by the prescriptive theory, ADT.

Consider a natural language, like English, with a truth predicate that is used as described in Chapter Six. Now imagine what this language would be like if my advice were heeded. It would still contain a truth predicate; remember, I am **NOT** suggesting that we stop using truth predicates or the concept of truth—truth is a useful inconsistent concept, much like mass. The truth predicate would be treated as assessment-sensitive, as CAM describes it. The language would also contain an

ascending truth predicate and a descending truth predicate. These are not assessment-sensitive, context-dependent, ambiguous, or semantically noteworthy in any way. They are just regular predicates (if there is such a thing). Call this a post-revolutionary linguistic practice.

One might wonder how an assessment-sensitive truth predicate, the ascending truth predicate, and the descending truth predicate interact with one another. First, ascending truth and descending truth are not going to be widely used. In any causal conversation, people will use the truth predicate instead, even when it comes to claims like 'you shouldn't say that if it isn't true'. Just as in casual conversation, people use 'mass' with the understanding that what they are saying might not be, strictly speaking, correct, but it is good enough for the purposes at hand. If a conversational participant wants to insist that the questions under consideration warrant a more precise conceptual framework, then those in the conversation can switch to the more precise terminology of relativistic mass and proper mass. Likewise, if necessary, conversational participants can switch from talk of truth to talk of ascending truth and descending truth.

In cases where the distinction between ascending truth and descending truth matters, people use these terms and stay away from the truth predicate. An example discussed below concerns semantic theories for expressively rich languages. A traditional semantic theory assigns truth values to sentences of the language across a range of conditions, and these are interpreted as truth-conditions. Of course, if the target language is classical, contains a truth predicate (i.e., one that obeys the primary aletheic principles), and the semantic theory treats this truth predicate as consistent, univocal, and invariant, then the semantic theory will be inconsistent. This is a clear case where the replacement concepts play an important role. Instead of attributing truth conditions to sentences, a semantic theory should attribute ascending truth conditions and descending truth conditions to sentences. For the vast majority of sentences, these will be the same, but there will be some for which these differ, and accounting for this difference allows for a consistent semantic theory even

for expressively rich languages. Doing semantics for expressively rich languages is like doing the physics of dark flow in this respect—in both cases, one has to use the replacement concepts in order to avoid problems.

One might find oneself attributing truth to sentences that contain ascending truth predicates or descending truth predicates. In these cases, the two theories, ADT and CAM, work together to specify the results. For example:

(1) Grass is green

(2) (1) is descending true.

(3) (1) is ascending true.

(4) (2) is true.

(5) (3) is true.

Here we have a sentence, (4) that attributes truth to a sentence with a descending truth predicate in it, (2). Since (2) is safe, (T-In) and (T-Out) hold for (4), so it follows from (4) that (1) is descending true. Since (1) is safe as well, it also follows from (4) that grass is green. All the same results hold for (3) and (5), respectively.

Here is another example:

(6) (6) is not true.

(7) (6) is ascending true.

(8) (6) is descending true.

In this example, (6) is paradoxical. (7) says of (6) that it is ascending true, while (8) says of it that it is descending true.

So far, I have not discussed how ascending truth and descending truth apply to sentences containing 'true'. There are at least two options: (i) 'ascending true' and 'descending true' are invariant across the board, and (ii) 'ascending true' and 'descending true' are invariant except when

they are applied to assessment-sensitive sentences, in which case they are assessment-sensitive as well. Note that many who endorse semantic relativism assume that 'true' is assessment-sensitive when applied to sentences that are assessment-sensitive.[1] If they are right and ascending truth and descending truth have the same feature, then we should pick option (ii). However, option (ii) makes the semantic theory considerably more complex; so, in the interest of simplicity, I provisionally adopt option (i) for the purposes of this chapter. Moreover, according to ADT as implemented in Chapter Thirteen, only sentences containing 'ascending true' or 'descending true' are unsafe. So it seems as though (6) should be safe. Notice that (6) is ascending true relative to the ascending standard and ascending true relative to the descending standard and (6) is not descending true relative to the ascending standard and not descending true relative to the descending standard. Is (6) ascending true or descending true? Well, (6) is assessment-sensitive, so it has an ascending truth value only relative to a standard. So, is (7) ascending true? If we follow option (i), then the answer should be no.[2] So, since it is not the case that (6) is ascending true, full stop, (7) is not ascending true. Likewise, (6) is not descending true, so (8) is not ascending true.

Let us look at a converse example.

(9) (9) is not descending true.

(10) (9) is true

(11) (9) is not true.

(9) is unsafe, so it is ascending true and not descending true. How do we evaluate (10) and (11)? CAM says that 'true' is assessment sensitive, so (10) and (11) are assessment sensitive. (10) is ascending true relative to the ascending standard, and not descending true relative to the descending standard. (11) is not ascending true relative to the ascending standard and descending true relative

---

[1] For example, see MacFarlane (forthcoming d).

[2] Notice that our guiding analogy between truth and mass is no help here since 'mass' (the type) does not have relativistic mass or proper mass.

to the descending standard. But is (10) true? Recall that 'true' can only be used in situations where the difference between ascending truth and descending truth is negligible. In this case it is not, so we cannot answer whether (10) or (11) are true. Of course we can consider the further sentence:

(12) (10) is true.

CAM implies that this sentence is assessment-sensitive—(12) has exactly the same status as (10) itself: ascending true relative to the ascending standard, and not descending true relative to the descending standard.

With these in mind, compare the first example to the following one:

(13) Grass is green.

(14) (13) is true.

(15) (14) is descending true.

(16) (15) is ascending true.

Given what we said above, (14) is assessment sensitive, so it is not descending true and it is not ascending true. Thus, (15) is not ascending true and not descending true. Nevertheless, 'grass is green' follows from (14) since it follows from (15), and 'grass is green' follows from (16) as well since (15) is safe.

15.2 Truth and Other Concepts: The Explanatory Role

When we replace a concept we have to reevaluate its connections to other concepts. For example, replacing the concept of mass had implications for how to understand force, momentum, energy, etc. As described in Chapter Five, our concept of truth is a popular explanans—predication, reference, validity, meaning, knowledge, assertion, necessity, and analyticity are all closely related to truth, and popular theories of these concepts explain them by appeal to truth. What is an inconsistency theorist to say about these theories?

For any given concept X that is customarily explained in terms of truth, we have several options for a theory T of X:

(i) X is conceptually tied to one of the replacements (but not the other). *Strategy*: replace 'true' in T by just one of the replacement predicates to get theory T′. *Example*: proof.

(ii) X is conceptually tied to each of the replacements. *Strategy*: replace 'true' in T by one of the replacement predicates to get one theory, T′, and replace 'true' in T by the other to get another theory, T″. *Example*: inquiry.

(iii) X is conceptually tied to both of the replacements. *Strategy*: reformulate T in terms of some combinations of the replacement predicates to get theory T′. *Examples*: objectivity, belief, and meaning.

(iv) X is not conceptually tied to either of the replacements. *Strategy*: explain X in some other way. *Examples*: validity, knowledge, and assertion.

(v) X is inconsistent as well. *Strategy*: if X is useful, then it too should be replaced, and one should search for theories linking X's replacements with truth's replacements. *Examples*: predication and reference.

Perhaps there are other options as well, but these seem to be the primary ones. In this section, there are examples of each of these five options. The discussion in this chapter should answer many questions about what happens to our conceptual scheme after the aletheic revolution. Note that each of these topics is complex, subtle, and has a vast literature; this discussion should be treated as a first step rather than the final word.

15.2.1 Proof

As our first example, consider the concept of proof. There is a remarkably complex mathematical theory of proof that I am not going to consider.[3] However, there is also the intuitive principle that if a proposition or sentence is proven, then it is true. The converse is obviously incorrect since there are many truths yet to be proven. Moreover, even the weaker 'if a proposition or sentence is true,

[3] See Troelstra and Schwichtenberg (2000), Hendricks et al. (2000), and Restall (forthcoming) for more on proof theory.

then it is provable' has to be rejected because of Gödel's incompleteness theorem.[4] Nevertheless, we should consider what happens to our intuitive principle of proof:

(Proof) If a sentence or proposition is proven, it is true.

We might consider replacing 'true' in (Proof) by 'descending true'. However, we have ready-made counterexamples to the resulting principle. For example, ADT proves that descending liars (i.e., sentences that say of themselves that they are not descending true) are ascending true and not descending true. However, that the descending liar is not descending true is the content of the descending liar itself. Thus, ADT proves the descending liar (and proves that it is not descending true). Therefore, proven items need not be descending true.

Instead, we might try ascending truth. The resulting principle would be:

(Proof-A) If a sentence or proposition is proven, it is ascending true.

This principle is fine, and it follows from the definition of ascending truth. That is, if some theory T proves p, then given the definition of ascending truth, it also proves that p is ascending true. Therefore, we should use (Proof-A) as our conceptual connection between proof and the replacement concepts.

15.2.2 Inquiry

Way back in Chapters One and Five, I mentioned that many people take truth to be a goal of inquiry. I say 'a goal' instead of 'the goal' because there are many other goals as well, and truth, by itself, is never thought to be sufficient to justify inquiry. For example, one could inquire into whether there will be a prime number of people alive at the beginning of the next leap second, but

[4] It is actually a much more technical version of this claim that is ruled out by Gödel's theorem, but the details do not matter for my purposes here. See Boolos, Burgess, and Jeffrey (2002) and Smith (2007) and for discussion.

that hardly seems like a worthwhile inquiry, even if it did produce something true. Recall that the following is the way Michael Lynch formulates this principle:

(Inquiry) Other things being equal, true beliefs are a worthy goal of inquiry.[5]

Instead of having to choose between an ascending reading of this principle and a descending reading, we can endorse both. That is, both of the following principles are acceptable:

(Inquiry-A) Other things being equal, ascending true beliefs are a worthy goal of inquiry.

(Inquiry-D) Other things being equal, descending true beliefs are a worthy goal of inquiry.

Of course, if a belief is descending true, then it is ascending true. However, it is interesting that inquiry is tied to each of the replacement concepts in this way. Thus, the concept of inquiry is a good example of one for which option (ii) is appropriate.

15.2.3 Objectivity

The third option is that we might be able to reformulate the philosophical theory in question in terms of ascending truth *and* descending truth. There are several examples of this: objectivity, belief, meaning.

The case of objectivity is extremely complex and I would like to discuss the relation between ascending and descending truth and Crispin Wright's views on objectivity, but I do not have the space.[6] Instead, consider the principle Michael Lynch presents:

> (Objectivity) The belief that p is true if and only if with respect to the belief that p, things are as they are believed to be.[7]

Here we cannot just keep the same principle but substitute 'ascending true' or 'descending true' for

---

[5] Lynch (2009: ch. 1).
[6] Wright (1992, 2003).
[7] Lynch (2009: ch. 1).

'true'. The problem is that 'the belief that p is ascending true if and only if with respect to the belief that p, things are as they are believed to be' is not ascending true since the left-to-right direction fails, while 'the belief that p is descending true if and only if with respect to the belief that p, things are as they are believed to be' is not ascending true because the right-to-left direction fails. Instead, we can have:

> (Objectivity-A) The belief that p is ascending true *if* with respect to the belief that p, things are as they are believed to be.
>
> (Objectivity-D) The belief that p is descending true *only if* with respect to the belief that p, things are as they are believed to be.

Note that these are merely reformulations of the constitutive principles for ascending truth and descending truth, respectively. Once we split the original biconditional into its two component conditionals, we can reformulate each one as a legitimate principle connecting objectivity to a replacement concept.

15.2.4 Belief

Another principle Lynch emphasize is:

> (Norm of Belief) It is prima facie correct to believe that p if and only if the proposition that p is true.[8]

Again, substituting in 'ascending true' or 'descending true' does not work. There are ascending true propositions that it is not prima facie correct to believe (e.g., the ascending liar and the negation of the descending liar), and there are propositions it is prima facie correct to believe that are not descending true (e.g., the descending liar and the negation of the ascending liar). Instead, try:

> (Norm of Belief-A) It is prima facie correct to believe that p *only if* the proposition that p is ascending true.

---

[8] Lynch (2009: ch. 1).

(Norm of Belief) It is prima facie correct to believe that p *if* the proposition that p is descending true.

Again, once we split the biconditional into its component conditionals, we can formulate acceptable principles linking the original concept to the replacement concepts.

15.2.5 Meaning

The connection between meaning and truth is our third example of option (iii). This has been a topic of conversation throughout the book, and encapsulates perhaps the most important explanatory function of truth.[9] There is a huge debate in linguistics and philosophy of language about the extent to which sentences have invariant meanings and the extent to which truth conditions constitute these meanings (if anything does).[10] It is my view that there is a coming revolution in philosophy of language caused by the realization—which is already a dominant view in linguistics—that dynamic semantic theories have tremendous explanatory advantages over their static brethren, and once one makes the transition to dynamic semantics, it is not clear what place truth conditions have in explaining meaning any more.[11] Despite the potentially revolutionary consequences this change will bring for issues in philosophy of language, few, if any philosophers of language seem to have paid it much attention. Nevertheless, in this section I will consider only the traditional and relatively uncontroversial (in philosophy) claim:

(Meaning) The proffered meaning (content) of a sentence is or determines its truth conditions.[12]

---

[9] See the Appendix to Chapter 1 on relations between objections to deflationism for evidence.
[10] See Recanati (2002, 2008, 2010), Carson (2002), King (2003), Lepore and Cappelen (2005), Stanley (2005), Predelli (2005), Ludwig and Lepore (2007), Travis (2008), and Hawthorne and Cappelen (2009) for discussion.
[11] For discussion of dynamic semantic theories, see Kamp (1981), Heim (1982), Groendyke and Stakhoff (1990), Beaver (2001), and Dekker (2010).
[12] The parenthetical addition is meant to cover context-dependent sentences.

Again, it does not seem that simply substituting 'ascending truth' or 'descending truth' would result in a satisfying theory, since each of those would leave something out (namely, the other). Instead, we need something more complex like:

> (Meaning-AD) The proffered meaning (content) of a sentence is or determines its ascending truth conditions and its descending truth conditions.

The main idea is that instead of a single set of conditions (i.e., truth conditions), meaning is explained in terms of dual conditions (i.e., the conditions under which the sentence is ascending true and the conditions under which the sentence is descending true). Of course, for the vast majority of sentences (i.e., all the safe ones), their ascending truth conditions and their descending truth conditions will be the same, so the new principle will not have much effect on them. However, the change makes a huge difference when it comes to sentences like liars. It is impossible for standard truth-conditional theories of meaning to assign truth conditions to liar sentences—trying to results in an inconsistent theory. However, together with the claim that 'true' is assessment-sensitive (from CAM) and the theory of the replacement concepts (ADT), (Meaning-AD) can easily account for the meanings of liar sentences, even empirically paradoxical ones. Moreover, (Meaning-AD) works for ascending liars and descending liars as well.

Again, the connection between truth and meaning is complex and I am unable to give it the space it deserves. Nevertheless, the claims made in this subsection should indicate how I think that conversation should begin.

15.2.6 Validity

Option (iv) is to deny that the concept in question is linked to either of the replacements. As our first example of this option, consider validity. The following is a commonly held principle connecting truth and validity:

(Valid) An argument is valid iff necessarily, it is truth-preserving.

In Chapter Five, we saw that this principle is incompatible with every logical approach to the aletheic paradoxes, so we should not worry if we cannot make room for some variant of it in our new conceptual scheme. Of course, we would still want a theory of validity, but that is another issue.

In Chapter Thirteen, I claimed that validity cannot be defined in terms of either ascending truth or descending truth alone. That is, the class of valid arguments is not identical to the class of those that are necessarily ascending truth-preserving nor is it identical to the class of arguments that are necessarily descending truth-preserving. Therefore, neither option (i) nor option (ii) will work in this case. Moreover, I have been unable to find a way of defining validity in terms of a combination of ascending truth and descending truth, so I have been unable to implement option (iii). Nevertheless, in Chapter Fourteen, I argued that we have a perfectly acceptable notion of truth-in-a-model, which is a mathematical concept not subject to aletheic paradoxes. Using it, we can define validity in the standard way using Kreisel's squeezing argument. Thus, we already have a definition of validity that does not depend on truth, ascending truth, or descending truth.[13]

15.2.7 Knowledge

As another instance of option (iv), consider knowledge.[14] Much ink has been spilled on the topic of whether Plato was right that knowledge is justified true belief[15]; moreover, there has been a considerable amount of recent interest in the so-called *knowledge first* movement, which takes

[13] See also Field (2008a: ch. 2) for discussion.
[14] Some have suggested that knowledge is also an inconsistent concept, and it might well be, but I ignore this view for the purposes of this subsection; see Schiffer (1996) and Weiner (2009).
[15] Plato (1961); see Gettier (1963) for criticism.

knowledge to be a primitive concept incapable of being defined or analyzed.[16] However, in this subsection, I restrict my attention to the Justified True Belief (JTB) analysis of knowledge:

(JTB) S knows that p $=_{df}$ S believes that p, S's belief that p is justified, and $\langle p \rangle$ is true.

It would not work to replace 'true' with 'descending true'. The problem is that I know that ascending liars and descending liars are ascending true and not descending true. However, 'ascending liars and descending liars are ascending true and not descending true' is not descending true—it is ascending true. So replacing 'true' in (JTB) with 'descending true' results in a theory that is too strong. On the other hand, it does not work to replace 'true' with 'ascending true'. The problem in this case is that the analysis no longer implies that knowledge is factive; that is, it could be that S believes that p, S's belief that p is justified, and $\langle p \rangle$ is ascending true, but $\sim p$ (e.g., if $\langle p \rangle$ = the ascending liar).[17] So the result is too weak.

Instead, I claim that we can easily update (JTB) so that it is compatible with our new conceptual scheme. Consider:

(JFB) S knows that p $=_{df}$ S believes that p, S's belief that p is justified, and p.

The 'F' in 'JFB' stands for 'factive'. Here we just replaced '$\langle p \rangle$ is true' with 'p'. The resulting theory works just as well as the JTB theory, but does not rely on the inconsistent concept of truth.

This kind of move is often made by deflationists in the face of explanatory inadequacy objections, which say that a deflationist about truth cannot account for truth's explanatory role in theories of other concepts (discussed in Chapters One and Six). For example, the success objection (it cannot explain why true theories are more likely to give us successful predictions), the assertion objection (it cannot explain why truth is a norm of assertion), the meaning objection (it cannot

---

[16] See Williamson (2000a).

[17] I am not convinced that an analysis of knowledge should imply that the proffered content of 'S knows that p' is factive since factivity projects in these cases and that indicates that it might be a presupposition. However, this is not the place to fight this battle.

explain why the meaning of a sentence determines its truth conditions), the generalization objection (it cannot explain general truths about truth; e.g., a conjunction is true if and only if its conjuncts are true), and the conservativeness objection (it cannot explain why the Gödel sentence is true) can all be thought of in these terms (see Chapter 1 for details). Deflationists often respond to these objections by saying that 'true' does not play an explanatory role in these theories—it is merely serving as a device of generalization. However, in Chapter Six, we saw that this sort of reply runs into lots of problems since the deflationary readings of many of the theories in question are highly implausible. That same lesson carries over to our discussion here. Substituting p for 'p is true' in JTB is unobjectionable, but it would be wrong to substitute 'p' for 'the belief that p is true' in (Objectivity) above since the left-to-right reading of the resulting principle would imply that we believe everything that is the case. The lesson is that deciding what happens to a concept's connections to other concepts post-revolution is to be decided largely on a case-by-case basis.

15.2.8 Assertion

Option (iv) works well in cases where the deflationist response to explanatory challenges works well. However, in the rest, it does not. Moreover, one can use option (iv) even in cases where the deflationist response fails, like the case of assertion. Recall that in Chapter One, one of the principles taken to be a platitude by Crispin Wright is:

(Assertion) To assert something is to present it as true.[18]

We cannot simply insert 'ascending' or 'descending' into this principle and arrive at an acceptable result. Ascending truth is too lax since it would permit the assertion of the ascending liar and the negation of the ascending liar since both of these are ascending true (only the negation of the

[18] Wright (1992: 24).

ascending liar should count as assertible since only it is a consequence of ADT). On the other hand, descending truth is too strong since it would not permit the assertion of the descending liar even though the descending liar is provable from ADT. The problem is that the descending liar is assertible but not descending true. Moreover, it is not clear that one could define assertibility in terms of the two replacement concepts together.

It seems to me that the right way to define assertibility is in two stages. First, for safe sentences—remember, all sentences that do not contain 'ascending true' or 'descending true' are safe—either ascending truth or descending truth is fine since they are coextensive over these sentences. However, for unsafe sentences (notice that these caused the problems in the previous paragraph), assertibility should be defined in terms of consequences of ADT. In particular, if for any acceptable xeno model $\mathfrak{M}$, $\mathfrak{M} \vDash$ p, then p is assertible. From these principles, it follows that the ascending liar is not assertible (though its negation is) and the descending liar is assertible (though its negation is not), just as it should be. There are a host of other issues in this neighborhood, but I cannot address them here.

15.2.9 Predication

We have seen examples of the first four options; in this subsection, I suggest that we should pursue option (v) for the case of predication—that is, it seems to me that predication is an inconsistent concept.

Here is a common principle tying predication (or truth-of) to truth:

(Predication) A subject-predicate sentence 'Ga' is true iff 'G' is true of a.

We cannot just substitute 'ascending true' for 'true', since the negation of the ascending liar is ascending true, but we do not think that 'not ascending true' is true of the ascending liar—indeed,

ADT implies that the negation of the ascending liar is ascending true. Substituting 'descending true' for 'true' will not work either since 'ascending true' is true of the descending liar, but the descending liar is not descending true. I do not see much hope in trying to explain predication in terms of ascending truth and descending truth together. Instead, one might try option (iv) with the following principle:

(Predication-F) Ga iff 'G' is true of a.

This principle seems fine at first, but there is a problem with it. Indeed, this problem is a well-known paradox associated with predication. Consider the predicate, 'heterological', defined in the following way:

(Heterological) a predicate G is heterological $=_{df}$ 'G' is not true of 'G'

The problem is that we can derive a paradox (often called Grelling's paradox or the heterological paradox). Assume that 'heterological' is true of 'heterological'. It follows from (Predication-F) that 'heterological' is heterological. Then from (Heterological) it follows that 'heterological' is not true of 'heterological'. Thus, if 'heterological' is true of 'heterological', then 'heterological' is not true of 'heterological'. On the other hand, assume that 'heterological' is not true of 'heterological'. It follows from (Heterological) that 'heterological' is heterological. And by (Predication-F) it follows that 'heterological' is true of 'heterological'. Thus, if 'heterological' is not true of 'heterological', then 'heterological' is true of 'heterological'. Therefore, 'heterological' is true of 'heterological' iff 'heterological' is not true of 'heterological'. So it seems option (iv) does not get us an acceptable principle for predication.[19]

Note that option (v) only makes sense if there is some independent reason to think that the concept in question is inconsistent—it engenders its own paradoxes. Since it is common knowledge

[19] For more on Grelling's paradox, see Grelling and Nelson (1908), Ryle (1951), Martin (1968), Goldstein (2003), Jacquette (2004), Ketland (2005), Newhard (2005), and Field (2008a).

that predication does give rise to the paradox just presented, there is a good case to be made that predication is an inconsistent concept. It is also a useful concept, and so should be replaced.

We can, and should, replace truth-of with two concepts, ascending truth-of and descending truth-of, which are analogous to ascending truth and descending truth:

(Ascending true-of) If Ga then 'G' is Ascending true-of a

(Descending true-of) If 'G' is Descending true-of a, then Ga

There are many questions this move brings up, and I can only begin to scratch the surface in this chapter. First, let us see how this solves the paradox. We get two Grellings with our replacement concepts:

Ascending Grelling: a predicate 'G' is Aheterological $=_{df}$ 'G' is not Ascending true-of 'G'

Descending Grelling: a predicate 'G' is Dheterological $=_{df}$ 'G' is not Descending true-of 'G'

We can prove the following results:

(i) 'Aheterological' is Ascending true-of 'Aheterological'

(ii) 'Aheterological' is not Descending true-of 'Aheterological'

(iii) 'Dheterological' is Ascending true-of 'Dheterological'

(iv) 'Dheterological' is not Descending true-of 'Dheterological'

The arguments (using the abbreviations 'Dhet', 'Ahet', 'Dtrue-of', and 'Atrue-of') are as follows. Assume (for reductio) that 'Ahet' is not Atrue-of 'Ahet'. By the definition of Atrue-of, it follows that 'Ahet' is not Ahet. By the definition of AGrelling, it follows from the initial assumption that 'Ahet' is Ahet. Contradiction. *Therefore* (by reductio), 'Ahet' is Atrue-of 'Ahet'. Assume (for reductio) that 'Ahet' is Dtrue-of 'Ahet'. From the definition of Dtrue-of, it follows that 'Ahet' is Ahet. From this it follows by the definition of Atrue-of that 'Ahet' is Atrue-of 'Ahet'. Thus, by the definition of AGrelling, 'Ahet' is not Ahet. From this it follows by the definition of Dtrue-of that 'Ahet' is not Dtrue-of 'Ahet'. Contradiction. *Therefore* (by reductio), 'Ahet' is not Dtrue-of 'Ahet'.

Assume (for reductio) that 'Dhet' is not Atrue-of 'Dhet'. It follows from the definition of Atrue-of that 'Dhet' is not Dhet. Hence, by the definition of Dtrue-of, 'Dhet' is not Dtrue-of 'Dhet'. From this, by the definition of DGrelling, we get that 'Dhet' is Dhet. Thus, by the definition of Atrue-of, 'Dhet' is Atrue-of 'Dhet'. Contradiction. *Therefore* (by reductio), 'Dhet' is Atrue-of 'Dhet'. Assume (for reductio) that 'Dhet' is Dtrue-of 'Dhet'. From the definition of Dtrue-of, it follows that 'Dhet' is Dhet. From the initial assumption, it also follows by the definition of DGrelling that 'Dhet' is not Dhet. Contradiction. *Therefore* (by reductio), 'Dhet' is not Dtrue-of 'Dhet'.

We can characterize 'Ahet' and 'Dhet' by saying that they are unsafe where:

(Predicate Safety) A predicate 'G' is safe iff 'G' is Dtrue-of 'G' or 'G' is not Atrue-of 'G'.

Note that the notion of safety associated with ascending truth and descending truth applies to sentences, whereas this notion of safety applies to predicates.

One might object that the above arguments are unconvincing since I have not shown that the principles defining Atrue-of and Dtrue-of are consistent. However, it is easy to define them in terms of ascending truth and descending truth (which we already know to be consistent) in the following way:

(Ascending) ⟨Ga⟩ is ascending true iff ⟨G⟩ is ascending true-of a,

(Descending) ⟨Ga⟩ is descending true iff ⟨G⟩ is descending true-of a.

In addition, one can link the notion of safety for sentences (which is defined in terms of ascending truth and descending truth) to the notion of predicate safety defined above:

(Safety) 'G' is a safe predicate iff "G' is G' is a safe sentence.

Again, this is a major change in our conceptual scheme and I can only give the barest outline of it here.[20]

15.2.10 Reference

Our tour through truth's conceptual connections ends with reference. A traditional link between truth and reference is:

(Reference) 'b' refers to a iff 'a=b' is true.

Again, substituting 'ascending true' or 'descending true' in for 'true' is unacceptable. We might try option (iv), which results in the following principle:

(Reference-F) 'b' refers to a iff a=b

The problem is that this principle is inconsistent. Consider the following situation:

(√) π

(√) 6

(√) the sum of the numbers referred to by ticked expressions in Scharp's *Replacing Truth*

Assume that 'π' refers to π and that '6' refers to 6. Assume as well that 'the sum of the numbers referred to by ticked sentences in Scharp's *Replacing Truth*' (which I abbreviate as 'the Sum') refers to some number; call it k. Now we reason as follows. 'The Sum' refers to k. By (Reference-F), the Sum = k. Therefore, k=π+6+k, which is impossible. Therefore, we should reject the assumption

---

[20] There is another kind of Grelling that deals with extensions: Let $\{\phi(x)\}$ be the extension of $\phi(x)$. We then get 'a $\in$ $\{\phi(x)\}$iff $\phi(a)$'. The connection to 'true of' should be obvious: a $\in$ $\{\phi(x)\}$ $\leftrightarrow$ '$\phi(x)$' is true of a $\leftrightarrow$ '$\phi(a)$' is true $\leftrightarrow$ $\phi(a)$. Now, let H(x) be such that: H($\langle\phi(x)\rangle$) $\leftrightarrow$ $\langle\phi(x)\rangle \notin \{H(x)\}$. Using analogous reasoning to that above, we get a contradiction.

So, let us define $\{\phi(x)\}_D$ and $\{\phi(x)\}_A$ in the following way:

a $\in$ $\{\phi(x)\}_D \rightarrow \phi(a)$

$\phi(a) \rightarrow$ a $\in$ $\{\phi(x)\}_A$

That gives us:

A(ϕ(a)) $\leftrightarrow$ '$\phi(x)$' is Atrue-of a $\leftrightarrow$ a $\in$ $\{\phi(x)\}_A$

D(ϕ(a)) $\leftrightarrow$ '$\phi(x)$' is Dtrue-of a $\leftrightarrow$ a $\in$ $\{\phi(x)\}_D$.

that 'the Sum' refers to some number. However, if we assume that 'the Sum' does not refer, then 'π' and '6' are the only referring ticked expressions in this book. Thus, the sum of the numbers referred to by ticked sentences in Scharp's *Replacing Truth* = π+6. But now, by (Reference-F), we get that 'the Sum' refers to π+6. Therefore, 'the Sum' does refer to some number, which contradicts our assumption.[21]

Note that option (v) makes sense in this case since reference engenders its own paradoxes, and there is a good case to be made that it is an inconsistent concept. It is also a useful concept, and so should be replaced.

We can, and should, replace reference with two concepts, ascending reference and descending reference, which are analogous to ascending truth and descending truth (and ascending truth-of and descending truth-of). We arrive at the following replacement concepts:

(Dreference) If 'b' Descending refers to a, then a=b

(Areference) If a=b, then 'b' Ascending refers to a

Notice the connection between the other replacement concepts for predication and truth:

(A) 'a=b' is ascending true iff 'x=b' is uniquely Atrue of a iff 'b' Arefers to a

(D) 'a=b' is descending true iff 'x=b' is uniquely Dtrue of a iff 'b' Drefers to a

These replacement concepts handle the paradox's of reference in the obvious way. Consider the paradox above. There will be two versions of it, which are listed below:

(√′) π

(√′) 6

[21] This paradox was first formulated by Keith Simmons in Simmons (2003); see also Beall (2003b). There are other famous paradoxes of reference, which include Berry's paradox, Konig's paradox, and Richard's paradox; see Simmons (1994), Chaitin (1995), Uzquiano (2004), and Field (2008a: 291-293) for discussion.

($\surd'$) the sum of the numbers Areferred to by ticked and primed expressions in Scharp's *Replacing Truth*

($\surd''$) π

($\surd''$) 6

($\surd''$) the sum of the numbers Dreferred to by ticked and double-primed expressions in Scharp's *Replacing Truth*

Begin with the first group of three. Assume that 'the sum of the numbers Areferred to by ticked and primed expressions in Scharp's *Replacing Truth*' does not Arefer to any number. Thus the sum of the numbers Areferred to by ticked and primed expressions in Scharp's *Replacing Truth* = π+6. So, by (Areference), 'the sum of the numbers Areferred to by ticked and primed expressions in Scharp's *Replacing Truth*' Arefers to π+6, which contradicts our assumption. Assume, instead, that 'the sum of the numbers Areferred to by ticked and primed expressions in Scharp's *Replacing Truth*' Arefers to π+6. However, we cannot infer from this claim that the sum of the numbers Areferred to by ticked and primed expressions in Scharp's *Replacing Truth* = π+6. Therefore, we cannot derive a contradiction from the assumption by the argument above. On the other hand, assume that 'the sum of the numbers Areferred to by ticked and primed expressions in Scharp's *Replacing Truth*' Drefers to some number, k. By (Dreference), it follows that the sum of the numbers Areferred to by ticked and primed expressions in Scharp's *Replacing Truth* = k. Thus, k=π+6+k, which is impossible, and we have refuted our assumption. Assume, instead, that 'the sum of the numbers Areferred to by ticked and primed expressions in Scharp's *Replacing Truth*' does not Drefer to any number. We infer that the sum of the numbers Areferred to by ticked and primed expressions in Scharp's *Replacing Truth* = π+6. However, we cannot infer from this claim that 'the sum of the

numbers Areferred to by ticked and primed expressions in Scharp's *Replacing Truth*' Drefers to π+6. Therefore, we cannot derive a contradiction from the assumption by the argument above.

Putting these results together we have:

(i) 'the sum of the numbers Areferred to by ticked and primed expressions in Scharp's *Replacing Truth*' Arefers to π+6.

(ii) 'the sum of the numbers Areferred to by ticked and primed expressions in Scharp's *Replacing Truth*' does not Drefer to anything.

(iii) It is not the case that the sum of the numbers Areferred to by ticked and primed expressions in Scharp's *Replacing Truth* = π+6.

Turning now to the second group of three, we assume that 'the sum of the numbers Dreferred to by ticked and double-primed expressions in Scharp's *Replacing Truth*' does not Arefer to any number. Thus the sum of the numbers Areferred to by ticked and double-primed expressions in Scharp's *Replacing Truth* = π+6. So, by (Areference), 'the sum of the numbers Areferred to by ticked and double-primed expressions in Scharp's *Replacing Truth*' Arefers to π+6, which contradicts our assumption. Assume, instead, that 'the sum of the numbers Dreferred to by ticked and double-primed expressions in Scharp's *Replacing Truth*' Arefers to π+6. However, we cannot infer from this claim that the sum of the numbers Areferred to by ticked and double-primed expressions in Scharp's *Replacing Truth* = π+6. Therefore, we cannot derive a contradiction from the assumption by the argument above. On the other hand, assume that 'the sum of the numbers Dreferred to by ticked and double-primed expressions in Scharp's *Replacing Truth*' Drefers to some number, k. By (Dreference), it follows that the sum of the numbers Dreferred to by ticked and primed expressions in Scharp's *Replacing Truth* = k. Thus, k=π+6+k, which is impossible, and we have refuted our assumption. Assume, instead, that 'the sum of the numbers Dreferred to by ticked and primed expressions in Scharp's *Replacing Truth*' does not Drefer to any number. We infer that the sum of the

numbers Dreferred to by ticked and double-primed expressions in Scharp's *Replacing Truth* = π+6. However, we cannot infer from this claim that 'the sum of the numbers Dreferred to by ticked and double-primed expressions in Scharp's *Replacing Truth*' Drefers to π+6. Therefore, we cannot derive a contradiction from the assumption by the argument above. Putting these results together we have:

(i) 'the sum of the numbers Dreferred to by ticked and double-primed expressions in Scharp's *Replacing Truth*' Arefers to π+6.

(ii) 'the sum of the numbers Dreferred to by ticked and double-primed expressions in Scharp's *Replacing Truth*' does not Drefer to anything.

(iii) The sum of the numbers Dreferred to by ticked and double-primed expressions in Scharp's *Replacing Truth* = π+6.

Again, once we distinguish between Areference and Dreference, there is no way to formulate a paradox using them in the above way. The treatment of the other paradoxes of reference is similar.

One can think of the replacements for truth, predication and reference as the basis for a new semantics, call it *AD semantics*. In AD semantics, sentences have ascending truth values and descending truth values, for each predicate there are the items it is ascending true-of and the items it is descending true-of, and each singular term has an ascending reference and a descending reference. There is much more to be said about AD semantics, but it will have to wait for another occasion.[22]

15.3 Objections and Replies

I have dealt with numerous objections all the way through the book. However, in this final section, I consider some objections that might still seem pressing, and offer some replies.

15.3.1 Guide to Objections

[22] I gave a hint on how to implement it with the double-model option in Chapter Fourteen.

Here is a guide to the objections that I have raised and addressed already in the book, along with their locations.

- The argument concerning truthmakers and correspondence theories of truth is unacceptable (§5.2.1).
- The argument concerning the impact of the paradoxes on truth-conditional theories of meaning can be defused by a proper understanding of the relation between theories of truth and meaning theories (§5.2.3).
- The modest attitude of the relation between philosophy and the sciences renders philosophical objections to the science moot (§6.2.4).
- The Gricean Condition is unacceptable because it rules out Millianism and semantic externalism (§6.2.5).
- My criticism of ambiguity, pragmatic, and context-dependence philosophical approaches to the aletheic paradoxes is no better than the semantic blindness objection to epistemic contextualism (§7.4.5).
- The strong internalizability requirement is incompatible with Tarski's indefinability theorem (§9.8.4).
- There are no inconsistent concepts (§11.2).
- There is no way to possess an inconsistent concept (§11.4).
- Inconsistency views require an analytic/synthetic distinction (§11.5).
- Other views explain the aletheic paradoxes better (§12.1).
- There is no need to replace the concept of truth (§12.3).
- A descriptive theory of truth should come before the prescriptive theory (§12.5).
- Ascending truth and descending truth do not serve truth's expressive role (§13.9.1).
- Ascending truth and descending truth give rise to revenge paradoxes (§13.9.3).
- ADT is self-refuting because some of its consequences are not descending true (§13.9.3).
- ADT is not strongly internalizable (§13.9.4).

- The standard objections to assessment-sensitivity views of epistemic modals, knowledge attributions, predicates of personal taste, etc. undermine CAM (§14.7).
- CAM does not respect truth's expressive role (§14.10.1).
- CAM gives rise to revenge paradoxes (§14.10.3).
- CAM is self-refuting because some of its consequences are not true (§14.10.3).
- CAM is not internalizable (§14.10.4).
- Ascending truth and descending truth cannot do the explanatory work of truth (§15.2).

The rest of this section considers a host of other objections that have not been raised yet.

15.3.2 My Uses of 'True'

I use 'true' all the way through the book, so is that inconsistent? No. I have argued over and again that inconsistent concepts can be useful. Truth is one such concept. In order for an objection of this sort to be plausible, one would have to show that I use 'true' in a situation where the distinction between ascending truth and descending truth is relevant. Of course, I use 'true' in describing the views of others as well, but it hardly makes sense to use 'ascending true' or 'descending true' in these cases.

15.3.3 Indispensability

Truth is indispensable. Consider the following passage from Stephen Leeds:

> I think that if we were somehow to become persuaded to use the word 'true' in ways that conflicted with the T-sentences, we would immediately – so important are the disquotational uses of truth in our own language – invent an additional notion of truth – say truth* – that conformed to them; under such circumstances, I think one might as well say that we had never abandoned the T-sentences after all: we had merely decided to rename truth 'truth*' and use the word 'true' to mean something else.[23]

[23] Leeds (1995: 8).

I do not disagree with Leeds on this point. In fact, I think this quote does a good job of explaining why theories of truth that do not treat (T-In) and (T-Out) as constitutive are non-starters. Of course, since we are using a truth predicate that has these as constitutive principles in a linguistic practice that is capable of referring to its own syntax and in which we reason according to certain logical principles, this truth predicate expresses an inconsistent concept. If that makes truth an indispensable inconsistent concept, then so be it. However, that claim has nothing to do with whether truth should be replaced. Being indispensable and being replaceable are compatible. Indeed, we should replace truth exactly because it is inconsistent and indispensible.

15.3.4 Primary Aletheic Principles

In Chapter Six, I criticize approaches to the aletheic paradoxes that do not validate the primary aletheic principles since these views cannot accommodate truth's expressive role. However, since (T-In) and (T-Out) are not true in general for my approach, I cannot accommodate truth's expressive role.

I take this to be one of the biggest problems with CAM. To be sure, I can explain why people treat 'true' as if it serves this expressive role—the reason is that (T-In) and (T-Out) are constitutive of truth. Moreover, I can show, according to a general theory of inconsistent concepts, why these constitutive principles are false; so my rejection of them is not ad hoc. Finally, I have a semantics that specifies exactly when they hold and which sentences are exceptions to them. Nevertheless, it is a consequence of CAM that natural language users are sometimes mistaken in treating certain speakers as committed to certain propositions. Since CAM is a descriptive theory of truth, that is a mark against it. It is no real consolation to note that everyone has this problem—no consistent (or

even inconsistent) descriptive theory of truth can accommodate all our intuitions, since our intuitions trivialize in any logical system.

I think the best thing to say in response to this objection is that the case that I make for the idea that truth is an inconsistent concept is a fortiori a case for the claim that natural language speakers are sometimes mistaken when they use truth predicates. By analogy, the case that Einstein makes for the idea that mass is an inconsistent concept is also a case for the claim that natural language speakers are sometimes mistaken when they use the term 'mass'; e.g., when they use it in situations where the difference between proper mass and relativistic mass is not negligible. I claim that CAM offers the best fit for our intuitions regarding truth predicates, but that does not mean that it is a perfect fit.

15.3.5 Endorsement and Replacement

Descending truth is what is needed for a device of endorsement, since if one calls an unsafe sentence ascending true, that unsafe sentence does not follow from one's claim. Nevertheless, if one calls an unsafe sentence descending true, then the sentence one has uttered is not ascending true. So neither ascending truth nor descending truth functions as a device of endorsement. There is no way to indirectly endorse an unsafe sentence. Moreover, if one wants to say that the ascending liar is unsafe, then what one has said is unsafe. There is no safe way to say that an unsafe sentence is unsafe.

These points are both right, but they do not constitute a legitimate objection. Consider two conversations. In one we have Cletus and Brandine discussing the trustworthiness of their friend, Jessica. In the other, we have Edna and Elizabeth discussing the relative merits of two theories of natural language adjectives. In the first conversation, we can imagine Cletus uttering:

(17) Everything Jessica said at lunch yesterday was true.

In his conversation with Brandine, Cletus uses this as evidence that Jessica should be trusted. Let us imagine that, unbeknownst to Cletus or Brandine, one of the sentences Jessica uttered was paradoxical. Thus, by the assessment-sensitivity semantics presented in Chapter Fourteen, Cletus does not thereby endorse the paradoxical sentence Jessica uttered. Since neither Cletus nor Brandine know that Jessica's sentence is paradoxical, it has no effect on their conversation and the conversational score continues as if it were not paradoxical. They conclude that Jessica is trustworthy, in part on the basis of Cletus's assertion. However, the key point is that even if they had distinguished between ascending truth and descending truth, they would have come to the same conclusion—that Jessica is trustworthy. It makes no difference for their purposes whether she asserted a paradoxical sentence at lunch or whether Cletus committed himself to it by his assertion. These points are irrelevant. Thus, in the case where the distinction between ascending truth and descending truth is negligible, 'true' works fine as a device of endorsement.

In the other conversation, Edna claims that Theory A does a better job of specifying the semantic features for sentences containing adjectives than Theory B. Their conversation turns to 'true', a particularly vexing adjective. Since both Theory A and Theory B specify semantic features of adjectives and 'true' is an adjective, they need to distinguish between ascending truth and descending truth in their conversation or else risk making potentially relevant mistakes in their assessments. At some point in the conversation, Edna wants to endorse an empirical descending liar and reject an empirical ascending liar as part of an argument for the superiority of Theory A. Let the sentences in question be:

(EA) The only sentence on the board is not ascending true

(ED) The only sentence on the projector is not descending true.

It turns out that (EA) is the only sentence on the board in question and (ED) is the only sentence on the projector in question. Thus, (EA) is an empirical ascending liar and (ED) is an empirical

descending liar. Recall that ADT entails the descending liar and the negation of the ascending liar. Let us consider what happens if Edna were to assert one of the following sentences:

(18) (EA) is true.

(19) (EA) is ascending true.

(20) (EA) is descending true.

(21) (EA) is not ascending true.

(22) (EA) is not descending true.

(23) (ED) is not true.

(24) (ED) is not ascending true.

(25) (ED) is not descending true.

(26) (ED) is ascending true.

(27) (ED) is descending true.

Since Edna and Elizabeth are distinguishing between ascending truth and descending truth, it would not make sense for Edna to assert (18) or (23). To *endorse* (ED), it would not do to assert (24) since the negation of (ED) follows from it. Asserting (26) is too weak since (ED) does not follow from it. (ED) does follow from (27), but (27) and (ED) are incompatible; indeed, they are contradictories, so she does not want to assert (27). Instead, she should assert (25) because it follows from (25) and 'the only sentence on the projector=(ED)' that the only sentence on the projector is not descending true, which is exactly what she wants to endorse. So, to endorse (ED), it will not do to call some sentence ascending true or descending true; instead, she needs to say that (ED) is not descending true. On the other hand, to *reject* (EA), it would be a poor choice to assert (20) since that would commit her to (EA). Asserting (22) is fine, but it does not entail the negation of (EA), so is too weak. Given that the only sentence on the board = (EA), it follows from (21) that the only sentence on the board is not ascending true, which is (EA); however, she wants to reject (EA), so asserting

(21) would be a mistake. Instead, asserting (19) is just right—it follows from (19) and 'the only sentence on the board is not ascending true' that the only sentence on the board is ascending true, which is the negation of (EA). Therefore, in asserting (19), she rejects (EA). Therefore, although it is impossible for her to reject the ascending liar by asserting that something is not ascending true or not descending true, she can reject the ascending liar by asserting that it is ascending true. Thus, she should assert (19) and (25).

The result is that Edna can endorse the descending liar and reject the ascending liar by using 'ascending true' and 'descending true' as long as she is careful to endorse the descending liar by saying that it is not descending true and to reject the ascending liar by calling its negation ascending true. Yes, both (19) and (25) are unsafe, but unsafe sentences are assertible if they are consequences of ADT.

15.3.6 Deflationism and Replacement

The replacement strategy depends on truth's explanatory role—otherwise there is no reason to replace truth. So deflationists have no reason to accept this view.

Anyone who thinks that truth predicates perform a useful expressive role has difficulty with an approach to the aletheic paradoxes, and the deflationist is no different. I have argued that deflationism either commits one to non-classical logic (disquotationalism and minimalism) or is antecedently implausible (prosententialism).[24] Moreover, the non-classical approaches to the paradoxes have come in for harsh criticism here—they engender revenge paradoxes and their semantic theories for truth are not internalizable for natural languages. Thus, the deflationist, who not only accepts but also emphasizes truth's expressive role, has good reason to be pushed toward

[24] See Chapter Six for these arguments. I have not addressed expresivism about truth (which is the fourth variety of deflationism).

thinking that truth is an inconsistent concept. So, far from resting my case for replacement on truth's explanatory role, I have stressed truth's expressive role (in Chapter Six) and argued that only an inconsistency theory can do justice to it.

15.3.7 Principle of Uniform Solution

CAM and ADT do not solve the other paradoxes that have often been discussed in conjunction with truth. For example, there is the Sorites Paradox, which affects vague expressions, Russell's paradox and the Burali-Forti paradox, which affect set theory, and Paradoxes of Generality, which affect the concept of a domain or subject-matter of discourse. Some philosophers have claimed that an approach to the liar that cannot be parlayed into an approach to one or more of these other paradoxes is unacceptable. In particular, Jamie Tappenden, Matti Eklund, and Hartry Field have argued that the paradoxes of truth and the paradoxes of vagueness ought to stand or fall together.[25] In addition, Graham Priest has defended a principle of uniform solution, which states that the aletheic paradoxes and the set-theoretic paradoxes all have the same form and should be solved together.[26]

From my point of view, paradoxes that affect a single concept should be solved together. Notice that I give a uniform solution to the liar paradox, Curry's paradox, Yablo's paradox, Montague's paradox, McGee's paradox, and all the revenge paradoxes, but there is good reason to think that these are all symptoms of truth's underlying inconsistency. I also show how to solve the paradoxes of predication and reference by applying what amounts to the same strategy, even though the approach to truth, on its own, does nothing to solve these paradoxes. I feel the same way about the other paradoxes mentioned above. There is no way I would subject the reader to a long and

---

[25] See Tappenden (1993, 1994), Eklund (2002a), and Field (2003b).
[26] Priest (1994b).

tedious discussion of vagueness, set theory, or absolute generality at this point. So let me say that if there is good reason to think that the concepts involved are inconsistent and useful, then a replacement strategy is in order. Otherwise, it is not. However, whether the concepts involved in those phenomena are inconsistent is irrelevant to assessing the merits of CAM and ADT.

Conclusion

I am not one for long conclusions—if you have read this far, then you should know what the book accomplishes. In case you are a bit unclear, I will try to summarize it here. Part I offers, for the first time, a single framework for thinking about work on the nature of truth and work on the aletheic paradoxes. It classifies a wide range of theories, distinguishes between philosophical and logical approaches to the paradoxes, and introduces the idea of a unified theory of truth, which includes all three parts. It also discusses the impact that the aletheic paradoxes have on the prospects for using truth in philosophical explanations of other concepts. Although this framework seems like the best way of representing the literature on truth, it does not highlight what I take to be the most important issues facing those of us who are interested in which unified theory of truth is best.

In Part II, I attempt to reorient the debates about truth around four key issues. In my opinion, theorists working on truth ought to focus on these key issues. After each issue is introduced, I draw out some of its consequences for the theories classified in Part I. Reflection on the significance and scope of these consequences is intended to justify the reorientation I suggest.

Finally, Part III defends an inconsistency view of truth on which truth should be replaced by ascending truth and descending truth. I offer two new theories, ADT, the theory of ascending and descending truth, and CAM, the classical assessment-sensitive measurement theoretic theory of truth.

Work Cited

Aczél, Peter. (1987). *Non-Well-Founded Sets*. Stanford: CSLI.

Alcoff, Linda Martin. (2001). "The Case for Coherence," in Lynch (2001).

Allo, Patrick. (2007). "Logical Pluralism and Semantic Information," *Journal of Philosophical Logic* 36: 659-694.

Almeder, Robert. (2010). *Truth and Skepticism*. New York: Rowman and Littlefield.

Alston, William P. (1996). *A Realist Conception of Truth*. Ithaca: Cornell University Press.

Alwishah, Ahmed and Sanson, David. (2009). "The Early Arabic Liar: The Liar Paradox in the Islamic World from the Mid-Ninth to the Mid-Thirteenth Centuries CE," *Vivarium* 47: 97-127.

Anderson, Alan and Belnap, Nuel. (1975). *Entailment: The Logic of Relevance and Necessity*, vol. 1. Princeton: Princeton University Press.

Anderson, Alan, Belnap, Nuel, and Dunn, Michael. (1992). *Entailment: The Logic of Relevance and Necessity*, vol. 2. Princeton: Princeton University Press.

Anderson, Hanne, Barker, Peter, and Chen, Xiang. (2006). *The Cognitive Structure of Scientific Revolutions*. Cambridge: Cambridge University Press.

Anscombe, G.E.M. (1957). *Intention*. London: Blackwell.

Antonelli, G. Aldo. (1994). "The Complexity of Revision," *Notre Dame Journal of Formal Logic* 35: 67-72.

———. (1996). Review of Simmons (1993), *Notre Dame Journal of Formal Logic* 37: 152-159.

———. (2000). "Virtuous Circles: From Fixed Points to Revision Rules," in Chapuis and Gupta (2000).

Aristotle. (1941). *The Basic Works of Aristotle*, McKeon (ed.), New York: Random House.

Arlo Costa, Horacio and Pacuit, Eric. (2006) "First Order Classical Modal Logic," *Studia Logica* 84: 171-210.

Armour-Garb, Bradley. (2001). "Deflationism and the Meaningless Strategy," *Analysis* 61: 280-289.

———. (2004). "Minimalism, the Generalization Problem, and the Liar," *Synthese* 139: 491-512.

Armour-Garb, Bradley, and Beall, Jc. (2001). "Can Deflationists be Dialetheists?" *Journal of Philosophical Logic* 30: 593-608.

———. (2002). "Further Remarks on Truth and Contradiction," *The Philosophical Quarterly* 52: 217-225.

———. (2003a). "Minimalism and the Dialetheic Challenge," *Australasian Journal of Philosophy* 81: 383-401.

———. (2003b). "Should Deflationists Be Dialetheists?" *Noûs* 37: 303–324.

———. (2005). "Minimalism, Epistemicism, and Paradox," in Beall and Armour-Garb (2005).

Armour-Garb, Bradley, and Woodbridge, James. (2010). "Why Deflationists Should be Pretense Theorists," in Wright and Pedersen (2010).

Armstrong, David. (1973). *Belief, Truth and Knowledge*. Cambridge: Cambridge University Press.

———. (1997). *A World of States of Affairs*. Cambridge: Cambridge University Press.

———. (2002). "Truths and Truthmakers," in Schantz 2002.

———. (2004). *Truth and Truthmakers*. Cambridge: Cambridge University Press.

Asher, Nicholas, Dever, Josh, and Pappas, Chris. (2009). "Supervaluations Debugged," *Mind* 118: 901-933.

Atlas, Jay David. (1989). *Philosophy without Ambiguity: A Logico-Linguistic Essay*. Oxford: Oxford University Press.

———. (2005). *Logic, Meaning, and Conversation: Semantical Underdeterminacy, Implicature, and Their Interface*. Oxford: Oxford University Press.
Austin, J. L. (1950). "Truth," *Proceedings of the Aristotelian Society, Supplementary Volumes* 24: 111-172.
———. (1959). "A Plea for Excuses," in Austin (1979).
———. (1979). *Philosophical Papers*. J. O. Urmson and G. J. Warnock (eds.). Oxford: Oxford University Press.
Ayer , A. J. (1939). *Language, Truth, and Logic*. London: Gollancz.
Azzouni, Jody. (2002). "Truth Via Anaphorically Unrestricted Quantifiers," *Journal of Philosophical Logic* 30: 329-354.
———. (2007). "The Inconsistency of Natural Languages: How We Live with It," *Inquiry* 50: 590-605.
Bach, Kent. (1987a): *Thought and Reference*. Oxford: Oxford University Press.
———. (1987b). "On communicative intentions," *Mind & Language* 2: 141-154.
———. (1994). "Conversational Impliciture," *Mind and Language* 9: 124-162.
———. (2000). "Quantification, Qualification and Context a Reply to Stanley and Szabó," *Mind and Language* 15: 262–283.
———. (2001). "You don't say?" *Synthese* 128: 15-44.
———. (forthcoming). "Perspectives on Possibilities: Contextualism, Relativism, or What?"
———. (MS). "Impliciture vs. Explicature: What's the difference?" http://userwww.sfsu.edu/~kbach/spd.htm.
Bach, Kent, and Harnish, Robert. (1979). *Linguistic Communication and Speech Acts*. Cambridge, MA: MIT Press.
Balog, Katalin. (2001). "Commentary on Frank Jackson's *From Metaphysics to Ethics*," *Philosophy and Phenomenological Research* 62:645–652.
Bar-On, Dorit, Horisk, Claire, and Lycan, William. (2000). "Deflationism, Meaning, and Truth-Conditions," *Philosophical Studies* 101: 1-28.
Barber, Alex. (ed.) (2003). *Epistemology of Language*. Oxford: Oxford University Press.
Barwise, Jon, and Etchemendy, John. (1987). *The Liar: An Essay on Truth and Circularity*. Oxford: Oxford University Press.
Barwise, Jon, and Moss, Lawrence Stuart. (2004). *Virtuous Circles: on the Mathematics of Non-Well-Founded Phenomena*. Stanford: CSLI.
Bays, Timothy. (2009). "Beth's Theorem and Deflationism," *Mind* 118: 1061-1073.
Beall, Jc. (1999). "Completing Sorensen's Menu: A Non-Modal Yabloesque Curry," Mind 108: 737-739.
———. (2000a). "On Mixed Inferences and Pluralism about Truth Predicates," *The Philosophical Quarterly* 50: 380-382.
———. (2000b). "Minimalism, Gaps, and the Holton Conditional," *Analysis* 60: 340-351.
———. (2001a). "Is Yablo's Paradox Non-Circular?" *Analysis* 61: 176-187.
———. (2001b). "A Neglected Deflationist Approach to the Liar," *Analysis* 61: 126-129.
———. (2002). "Deflationism and Gaps: Untying 'Not's in the Debate," *Analysis* 62: 299-305.
———. (2003a). Review of Simmons (1993).
———. (2003b). "On the Singularity Theory of Denotation," in *Liars and Heaps*, Beall (ed.), Oxford: Oxford University Press.
———. (2006). "Truth and Paradox: A Philosophical Sketch," in *Philosophy of Logic*, Jacquette (ed.), Elsevier: Dordrecht.
———. (2007a). *Revenge of the Liar*. Oxford: Oxford University Press.
———. (2007b). "Prolegomenon to Future Revenge," in Beall (2007a).
———. (2009). *Spandrels of Truth*. Oxford: Oxford University Press.

Beall, Jc, and Armour-Garb, Bradley. (2005). *Deflationism and Paradox*. Oxford: Oxford University Press.
Beall, Jc, and Glanzberg, Michael. (2008). "Where the Paths Meet: Remarks on Truth and Paradox," *Midwest Studies in Philosophy* 32: 169-198.
Beall, Jc and Restall, Greg. (2000). "Logical Pluralism," *Australasian Journal of Philosophy* 78: 475-493.
———. (2001). "Defending Logical Pluralism," in *Logical Consequence: Rival Approaches Proceedings of the 1999 Conference of the Society of Exact Philosophy*, ed. J. Woods and B. Brown, Stanmore: Hermes.
———. (2006). *Logical Pluralism*. Oxford: Oxford University Press.
Beall, Jc, and van Fraassen, Bas. (2003). *Possibilities and Paradox*. Oxford: Oxford University Press.
Beaver, David. (2001). *Presupposition and Assertion in Dynamic Semantics*. Stanford: CSLI.
Beebe, Helen, and Dodd, Julian. (2005). *Truthmakers: The Contemporary Debate*. Oxford: Oxford University Press.
Belnap, Nuel. (1961). "Tonk, Plonk and Plink," *Analysis* 22: 130-134.
———. (1982). "Gupta's Rule of Revision Theory of Truth," *Journal of Philosophical Logic* 11: 103-116.
———. (1999). "Truth by Ascent," *Dialectica* 53: 291-306.
———. (2005). "Under Carnap's Lamp: Flat Pre-Semantics," *Philosophical Studies* 80: 1-28.
———. (2006). "Prosentence, Revision, Truth and Paradox," *Philosophy and Phenomenological Research* 73: 705-712.
Bigelow, John. (1988). *The Reality of Numbers. A Physicalist's Philosophy of Mathematics*, Oxford: Clarendon Press.
Bimbo, Katalin. (2007). "Relevance Logics," in *Handbook of the Philosophy of Science*: *Philosophy of Logic*, Jacquette (ed.), Amsterdam: Elsevier.
Bimbo, Katalin and Dunn, Michael. (2008). *Generalized Galois Logics*. Stanford, CA: CSLI.
Blackburn, Simon. (1984). *Spreading the Word*. Oxford: Oxford University Press.
———. (1998). "Wittgenstein, Wright, Rorty, and Minimalism," *Mind* 107: 157-181.
———. (2001). *Think*. Oxford: Oxford University Press.
———. (2005). *Truth: A Guide for the Perplexed*. Oxford: Oxford University Press.
Blamey, Stephen. (2001). "Partial Logic," in *Handbook of Philosophical Logic*, 2nd ed, vol. 5, Gabbay and Guenthner (eds.), Dordrecht: Kluwer.
Blanshard, Brand. (1939). *The Nature of Thought*. London: Allen & Unwin.
Block, Ned, and Stalnaker, Robert. (1999). "Conceptual Analysis, Dualism, and the Explanatory Gap," *Philosophical Review* 108: 1-46.
Boghossian, Paul. (1989). "Content and Self-Knowledge," *Philosophical Topics*, 17: 5–26.
———. (1990a). "The Status of Content," *Philosophical Review* 99: 157-184.
———. (1990b). "The Status of Content Revisited," *Pacific Philosophical Quarterly* 71: 264-278.
———. (1996). "Analyticity Reconsidered," *Noûs* 30: 360-391.
———. (1997). "Analyticity," in *A Companion to the Philosophy of Language*, B. Hale and C. Wright (eds.), Oxford: Blackwell.
———. (2000). "Knowledge of Logic," in *New Essays on the A Priori*, P. Boghossian and C. Peacocke (eds.), Oxford: Oxford University Press.
———. (2001). "How Are Objective Epistemic Reasons Possible?" *Philosophical Studies* 106: 340-380.
———. (2003a). "Blind Reasoning," *Proceedings of the Aristotelian Society, Supplementary Volume* 77: 225-248.
———. (2003b). "Epistemic Analyticity: A Defense," *Grazer Philosophische Studien* 66: 15-35.
———. (2006). "What is Relativism?" in Greenough and Lynch (2006).

Bonevac, Daniel. (1991). "Semantics and Supervenience," *Synthese* 87: 331-361.
Bonnay, Dennis, and Simmonauer, Benjamin. (2005). "Tonk Strikes Back," *The Australasian Journal of Logic* 3: 33-44.
Boolos, George, Burgess, John, and Jeffrey, Richard. (2002). *Computability and Logic*, 4th ed. Cambridge: Cambridge University Press.
Braddon-Mitchell, David, and Nola, Robert (eds.). (2009). Conceptual Analysis and Philosophical Naturalism. Cambridge, MA: MIT Press.
Bradley, F. H. (1914). *Essays on Truth and Reality*. Oxford: Clarendon Press.
Brady, Ross. (ed.) (2003) *Relevant Logics and their Rivals*, vol. 2, Burlington, VT: Ashgate.
———. (2006). *Universal Logic*. Stanford, CA: CSLI.
Brandom, Robert. (1988). "Pragmatism, Phenomenalism, and Truth Talk," *Midwest Studies in Philosophy* 12: 75-93.
———. (1994). *Making It Explicit*. Cambridge, MA: Harvard University Press.
———. (2001). *Articulating Reasons*. Cambridge, MA: Harvard University Press.
———. (2002). "Explanatory vs. Expressive Deflationism about Truth," in Schantz (2002).
———. (2008). *Between Saying and Doing*. Oxford: Oxford University Press.
———. (2009a). *Reason in Philosophy*. Cambridge, MA: Harvard University Press.
———. (2009b). "Reply to Kevin Scharp," in *Reading Brandom*, Weiss and Wanderer (eds.), New York: Routledge.
Brendel, Elke. (2000). "Circularity and the Debate Between Deflationist and Substantive Theories of Truth," in Chapuis and Gupta 2000.
Bricker, Phillip. (2006). "The Relation Between General and Particular: Entailment vs. Supervenience", in *Oxford Studies in Metaphysics*, Volume 2, Zimmerman (ed.), Oxford: Oxford University Press.
Brogaard,Berit. (2007). "Sea Battle Semantics," *The Philosophical Quarterly* 58: 326–335.
———. (2008). "In Defense of a Perspectival Semantics for 'Know' ",*Australasian Journal of Philosophy* 86: 439-459.
———. (2010). "Perspectival Truth and Color Primitivism," in Wright and Pedersen (2010).
Brown, Jessica. (2004). *Anti-Individualism and Knowledge*. Cambridge, MA: MIT Press.
Bueno, Otavio. (2007). 2007). "Troubles with Trivialism," *Inquiry* 50: 655-667.
Burge, Tyler. (1979a). "Semantical Paradox," *The Journal of Philosophy*, 76: 169-198.
———. (1979b). "Individualism and the Mental," *Midwest Studies in Philosophy*, 4: 73-121.
———. (1982a). "Postscript to 'Semantical Paradox'," in Martin 1984.
———. (1982b). "The Liar Paradox: Tangles and Chains," *Philosophical Studies* 41: 353-366.
———. (1986). "Intellectual Norms and the Foundations of Mind," *The Journal of Philosophy* 83: 697-720
———. (1988). "Individualism and Self-Knowledge," *The Journal of Philosophy* 85: 649-663.
———. (1993). "Content Preservation," *Philosophical Review* 103: 457-488.
———. (2009). "Perceptual Objectivity," *Philosophical Review* 118 : 285-324.
———. (2010). *Origins of Objectivity*. Oxford: Oxford University Press.
Burgess, John. (1986). "The Truth is Never Simple," *The Journal of Symbolic Logic* 51: 663-681.
———. (1997). "What is Minimalism about Truth?" *Analysis* 57: 259-267.
———. (2009). *Philosophical Logic*. Princeton: Princeton University Press.
Burton-Roberts, Noel. (1999). "Metalinguistic Negation and Presupposition Cancellation," *Journal of Linguistics 35: 347-364*.
Cameron, Ross. (2008). "How to be a Truthmaker Maximalist," *Nous* 42: 410-421.
Camp Jr., Joseph L. (2002). *Confusion: A Study in the Theory of Knowledge*. Cambridge: Harvard University Press.

Candlish, Stewart, and Damnjanovic, Nic. (2007). "A Brief History of Truth," in *Philosophy of Logic*, Jacquette (ed.), Elsevier: Dordrecht.
Cantini, Andrea. (1995). "Levels of Truth," *Notre Dame Journal of Formal Logic* 36: 185-213.
———. (2009). "Paradoxes, Self-Reference, and Truth in the 20th Century," in *Handbook of the History of Logic*, vol. 5, Gabbay and Woods (eds.), Amsterdam: Elsevier.
Cappelen, Herman. (1999). "Intentions in Words," *Nous* 33: 92-102.
———. (2007). "Shared Content and Semantic Blindness," in Kölbel and Garcia-Carpintero 2008.
Cappelen, Herman, and Hawthorne, John. (2009). *Relativism and Monadic Truth*. Oxford: Oxford University Press.
———. (forthcoming a). "Reply to Mark Richard's 'Relativistic Content and Disagreement'," *Philosophical Studies*.
———. (forthcoming b). "Reply to Peter Lasersohn's 'Context, Relevant Parts and (Lack of) Disagreement Over Taste'," *Philosophical Studies*.
———. (forthcoming c). "Reply to John MacFarlane's 'Simplicity Made Difficult'," *Philosophical Studies*.
———. (forthcoming d). "Reply to Glanzberg, Soames and Weatherson," *Analysis Reviews*.
Cappelen, Herman, and Lepore, Ernie. (2007). "The Myth of Unarticulated Constituents," in *Situating Semantics: Essays on the Philosophy of John Perry*, O'Rourke and Washington (eds.), Cambridge, MA: MIT.
Capria, Marco (ed.). (2005). *Physics Before and After Einstein*. Amsterdam: IOS Press.
Carey, Susan. (2009). *The Origin of Concepts*. Oxford: Oxford University Press.
Carnap, Rudolph. (1928). *The Logical Structure of the World*. Rolf A. George (tr.). University of California Press.
———. (1942). *Introduction to Semantics*. Cambridge, MA: Harvard University Press.
———. (1950). *Logical Foundations of Probability*. University of Chicago Press.
Carruthers, Peter. (1996). *Language, Thought and Consciousness*. Cambridge: Cambridge University Press.
———. (2000). *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge: Cambridge University Press.
Carston, Robyn. (1988). "Implicature, Explicature and Truth-Theoretic Semantics," in *Mental Representations: the Interface between Language and Reality*, Kempson (ed.), Cambridge: Cambridge University Press.
———. (1996). "Enrichment and Loosening: Complementary Processes in Deriving the Proposition Expressed?" *UCL Working Papers in Linguistics* 8: 205-232.
———. (1998). "Negation, 'Presupposition' and the Semantics/Pragmatics Distinction," *Journal of Linguistics*, 34: 309-350.
———. (1999). "Negation 'Presupposition' and Metarepresentation: A Response to Noel Burton-Roberts," *Journal of Linguistics* 35: 365-389.
———. (2002). *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.
———. (2004). "Relevance Theory and the Saying/Implicating Distinction," in *The Handbook of Pragmatics*. WardHorn and Ward (eds.), Oxford: Blackwell.
———. (2005). "Relevance Theory, Grice and the neo-Griceans," *Intercultural Pragmatics* 2: 303-319.
———. (2007). "How Many Pragmatic Systems are There?" in Frapolli,M-J. (ed.) *Saying, Meaning, Referring: Essays on the Philosophy of Francois Recanati*. Palgrave Macmillan.
———. (2008). "Linguistic Communication and the Semantics-Pragmatics Distinction," *Synthese* 165: 321-345.
———. (2009). "The Explicit/Implicit Distinction in Pragmatics and the Limits of Explicit Communication. *International Review of Pragmatics* 1: 35-62.

Carston, Robyn, and Powell, G. (2006). "Relevance Theory - New Directions and Developments," in *Oxford Handbook of Philosophy of Language*, Lepore and Smith (eds.), Oxford University Press.

Carston, Robyn, and Uchida, S. (eds.) (1998). *Relevance Theory: Applications and Implications.* Amsterdam: John Benjamins.

Cartwright, Richard. (1987). "A Neglected Theory of Truth," in *Philosophical Essays*, Cambridge, MA: MIT Press.

Chaitin, Gregory. (1995) "The Berry Paradox," *Complexity* 1: 26-30.

Chalmers, David. (2010). *Constructing the World.* The John Locke Lectures, Oxford, 5 May 2010-9 June 2010.

Chalmers, David, and Jackson, Frank. (2001). "Conceptual Analysis and Reductive Explanation," *Philosophical Review* 110: 315-61.

Chang, Hasok. (2004). *Inventing Temperature.* Oxford: Oxford University Press.

Chapman, Siobhan. (1996). "Some Observations on Metalinguistic Negation', *Journal of Linguistics* 32: 387-402.

Chapuis, A. (1996). "Alternate Revision Theories of Truth," *Journal of Philosophical Logic* 25: 399–423.

Chellas, Brian. (1980). *Modal Logic.* Cambridge: Cambridge University Press.

Chierchia, Gennaro, and McConnell-Ginet, Sally. (2000). *Meaning and Grammar.* 2nd ed. Cambridge MA: MIT Press.

Chihara, Charles S. (1973). "A Diagnosis of the Liar and Other Semantical Vicious-Circle Paradoxes," *The Work of Bertrand Russell.* C. Roberts (ed.) London: Allen and Unwin.

———. (1976). "Truth, Meaning, and Paradox," *Noûs* 10: 305-311.

———. (1979). "The Semantic Paradoxes: A Diagnostic Investigation," *The Philosophical Review* 88: 590-618.

———. (1984). "The Semantic Paradoxes: Some Second Thoughts," *Philosophical Studies* 45: 223-229.

Child, William. (1994). *Causality, Interpretation, and the Mind.* Oxford: Oxford University Press.

Chomsky, Noam. (1986). *Knowledge of Language.* Westport, CT: Praeger.

———. (1995). "Language and Nature," *Mind* 104: 1-61.

Church, Alonzo. (1946). Review of: Koyré, Alexandre (1946), *The Journal of Symbolic Logic* 11: 131.

———. (1976). "Comparison of Russell's Resolution of the Semantical Antimonies with that of Tarski," *The Journal of Symbolic Logic* 41: 747-760.

Cieslinski, Cezary. (2009). "Truth, Conservativeness, and Provability," *Mind* 119: 409-422

Clapp, Lenny. (2002). "What Unarticulated Constituents Could Not Be," in *Meaning and Truth*, Campbell, O'Rourke, and Shier (eds.), New York: Seven Bridges.

Clark, Michael. (1997). "Truth and Success: Searle's Attach on Minimalism," *Analysis* 57: 205-209.

———. (1999). "Recalcitrant Variants of the Liar Paradox," *Analysis* 59: 117-126.

Cohen, L. Jonathan. (1957). "Can the Logic of Indirect Discourse be Formalised?" *The Journal of Symbolic Logic* 22: 225-232.

———. (1961). "Why Do Cretans Have to Say So Much?" *Philosophical Studies* 12: 72-78.

Cohen, S., (1986). "Knowledge and Context", *The Journal of Philosophy* 83: 574-583.

———. (1999). "Contextualism, Skepticism, and The Structure of Reasons", *Philosophical Perspectives 13: Epistemology*: 57-89.

Cook, Roy. (2002). "Counterintuitive Consequences of the Revision Theory of Truth," *Analysis* 62: 16-22.

———. (2003). "Still Counterintuitive: A Reply to Kremer", *Analysis 63*: 257-261.

———. (2005). "What's Wrong with Tonk(?)", *Journal of Philosophical Logic 34*: 217-226.

———. (2008). "Embracing Revenge: On the Indefinite Extensibility of Language," in Beall (2007a).
———. (2010). "Let a Thousand Flowers Bloom: A Tour of Logical Pluralism," *Philosophy Compass* 5: 492-504.
Cruse, D. A. (1986). *Lexical Semantics*. Cambridge: Cambridge University Press.
Culbertson, Jennifer, and Gross, Steven. (2009). "Are Linguists Better Subjects?" *British Journal for the Philosophy of Science* 60: 721-736.
Curry, H. (1942). "The Inconsistency of Certain Formal Logics," *Journal of Symbolic Logic* 7: 115-117.
da Costa, Newton, Bueno, Otavio, and French, Steven. (2005). "A Coherence Theory of Truth," *Manuscrito* 28: 263-290.
Dauer, Francis. (1974). "In Defense of the Coherence Theory of Truth," *Journal of Philosophy* 71: 791-811.
Damjanovic, Nic. (2005). "Deflationism and the Success Argument," *Philosophical Quarterly* 55: 53-67.
Daston, Lorraine and Galison, Peter. (2007). *Objectivity*. New York: Zone.
David, Marian. (1989). "Truth, Eliminativism, and Disquotationalism," *Noûs* 23: 599-614.
———. (1994). *Correspondence and Disquotation: An Essay on the Nature of Truth*. Oxford: Oxford University Press.
Davidson, Donald. (1967). "Truth and Meaning," in Davidson (1984).
———. (1968). "On Saying That," in Davidson (1984).
———. (1969). "True to the Facts," in Davidson 1984.
———. (1970). "Mental Events," in Davidson (1980a).
———. (1973). "Radical Interpretation," in Davidson (1984).
———. (1974a). "On the Very Idea of a Conceptual Scheme," in Davidson (1984).
———. (1974b). "Belief and the Basis of Meaning," in Davidson (1984).
———. (1976a). "Hempel on Explaining Action," in Davidson (1980a).
———. (1976b). "Reply to Foster," in Davidson (1984).
———. (1979). "The Inscrutability of Reference," in Davidson (1984).
———. (1980a). *Essays on Actions and Events*. Oxford University Press.
———. (1980b). "Toward a Unified Theory of Meaning and Action," *Grazer Philosophische Studien* 11 (1980): 1-12.
———. (1982a). "Communication and Convention," in Davidson (1984).
———. (1982b). "A Coherence Theory of Truth and Knowledge." In Davidson (2001).
———. (1984). *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.
———. (1986). "A Nice Derangement of Epitaphs." in *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, E. LePore (ed.), Oxford: Blackwell.
———. (1987). "Knowing One's Own Mind," in Davidson (2001).
———. (1988). "The Myth of the Subjective," in Davidson (2001).
———. (1990). "The Structure and Content of Truth," *The Journal of Philosophy* 87: 279-328.
———. (1992). "The Second Person," in Davidson (2001).
———. (1995). "Could There Be a Science of Rationality?" *International Journal of Philosophical Studies* 3: 1-16.
———. (1996). "The Folly of Trying to Define Truth." *The Journal of Philosophy* 94: 263-278.
———. (1997a). "The Emergence of Thought," in Davidson (2001).
———. (1997b). "Indeterminism and Antirealism." In Davidson (2001).
———. (1999a). "General Remarks." in *Donald Davidson: Truth, Meaning and Knowledge*. Edited by U. Zeglen. London: Routledge, 1999.

———. (1999b). "Reply to John McDowell," in *The Philosophy of Donald Davidson*, Hahn (ed.), Peru, IL: Open Court.
———. (2001). *Subjective, Intersubjective, Objective*. Oxford: Oxford University Press.
———. (2005). *Truth and Predication*. Cambridge: Harvard University Press.
Davis, Lawrence. (1979). "An Alternate Formulation of Kripke's Theory of Truth," *Journal of Philosophical Logic* 8: 289-296.
Davis, Steven, and Gillon, Brendan (eds.). (2004). *Semantics: A Reader*. Oxford: Oxford University Press.
de Caro, Mario, and Macarthur, David (eds.). (2004). *Naturalism in Question*. Cambridge, MA: Harvard University Press.
de Swart, Henriette. (1998). *Introduction to Natural Language Semantics*. Stanford: CSLI.
Debs, Talal and Redhead, Michael. (2007). *Objectivity, Invariance, and Convention*. Cambridge, MA: Harvard University Press.
Dekker, Paul. (2010). "Dynamic Semantics," in *The Handbook of Semantics*, Maienborn, Heusinger, and Portner (eds.).
DeRose, Keith. (1992). "Contextualism and Knowledge Attributions," *Philosophy and Phenomenological Research* 52: 913-929.
———. (1995). "Solving the Skeptical Problem," *Philosophical Review* 10: 1-52.
———. (2002). "Assertion, Knowledge, and Context," *Philosophical Review* 111, 167-203.
———. (2004). "Single Scoreboard Semantics," *Philosophical Studies*, 119: 1-21.
———. (2006). "'Bamboozled by Our Own Words': Semantic Blindness and Some Objections to Contextualism," *Philosophy and Phenomenological Research* 73: 316-338.
Descartes, René. (1641). *Meditations on First Philosophy*. Cottingham (tr.). Cambridge: Cambridge University Press. 1996.
Devitt, Michael. (1984). *Realism and Truth: Second Edition*. Princeton: Princeton University Press.
———. (1990). "Transcendentalism about Content," *Pacific Philosophical Quarterly* 71: 247-263.
———. (2006a). *Ignorance of Language*. Oxford: Oxford University Press.
———. (2006b). "Intuitions in Linguistics," *British Journal for the Philosophy of Science* 57: 481-513.
———. (2009). "Methodology in the Philosophy of Linguistics," *Australasian Journal of Philosophy* 86: 671-684.
———. (2010). "What Intutions are Linguistic Evidence?" *Erkenntnis*.
Devitt, Michael, and Rey, Georges. (1991). "Transcending Transcendentalism: A Response to Boghossian," *Pacific Philosophical Quarterly* 72: 87-100.
Dietz, Richard. (2008). "Epistemic Modals and Correct Disagreement," in Garcia-Carpintero and Kölbel (2008).
Dietz, Richard and Moruzzi, Sebastiano. (eds.) (2010). *Cuts and Clouds*. Oxford: Oxford University Press.
Divers, John, and Miller, Alexander. (1994). "Why Expressivists about Value Should Not Love Minimalism about Truth," *Analysis* 54: 12-19.
Dodd, Julian. (1995). "McDowell and Identity Theories of Truth," *Analysis* 55, pp.160-5.
———. (1996). "Resurrecting the Identity Theory of Truth: A Reply to Candlish," *Bradley Studies* 2: 42-50.
———. (1997). "On a Davidsonian Objection to Minimalism," *Analysis* 57:267-272.
———. (1999a). "There is No Norm of Truth: A Minimalist Reply to Wright," Analysis 59: 291-99.
———. (1999b). "Hornsby on the Identity Theory of Truth," *Proceedings of the Aristotelian Society*, XCIX, pp. 225-32.
———. (2000). *An Identity Theory of Truth*. New York: St. Martin's Press, Inc.

———. (2008). "McDowell's Identity Conception of Truth: a Reply to Fish and Macdonald," *Analysis* 68:76–85.
Dorsey, Dale. (2006). "A Coherence Theory of Truth in Ethics," *Philosophical Studies* 127: 493-523.
Dowty, David, Wall, Robert, and Peters, Stanley. (1981). *An Introduction to Montague Semantics*. D. Reidel.
Dretske, Fred. (1988). *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press.
Dreier, James. (1996). "Expressivist Embedding and Minimal Truth," *Philosophical Studies* 83: 29-51.
Dummett, Michael. (1959). "Truth," *Truth and Other Enigmas*. Cambridge: Harvard University Press: 1-24.
———. (1973). *Frege: Philosophy of Language*. Cambridge, MA: Harvard University Press.
———. (1975). "What is a Theory of Meaning?" in Dummett (1993).
———. (1978). *Truth and Other Enigmas*. Cambridge: Harvard University Press.
———. (1991). *The Logical Basis of Metaphysics*. Cambridge: Harvard University Press.
———. (1993). *The Seas of Language*. Cambridge, MA: Harvard University Press.
———. (1999). "Of What Kind of Thing is Truth a Property?" in *Truth*, Blackburn and Simmons (eds.), Oxford: Oxford University Press..
———. (2000). Elements of Intuitionism, 2nd ed. Cambridge, MA: Harvard University Press.
———. (2002). "The Two Faces of the Concept of Truth," in Schantz 2002.
Dunn, Michael and Hardegree, Gary. (2001). *Algebraic Methods in Philosophical Logic*. Oxford: Oxford University Press.
Dunn, Michael and Restall, Greg. (2001). "Relevance Logic," in *Handbook of Philosophical Logic*, 2nd ed, vol. 6, Gabbay and Guenthner (eds.), Dordrecht: Kluwer.
Earman, John, and Fine, Arthur. (1977). "Against Indeterminacy," *Journal of Philosophy* 74: 535-538.
Ebbs, Gary. (2009). *Truth and Words*. Oxford: Oxford University Press.
Egan, Andy. (2007). "Epistemic Modals, Relativism and Assertion," *Philosophical Studies* 133: 1-22.
———. (2009). "Billboards, Bombs, and Shotgun Weddings," Synthese
———. (forthcoming). "Disputing about Taste,"
Egan, Andy, Hawthorne, John, and Weatherson, Brian. (2005). "Epistemic Modals in Context," in *Contextualism in Philosophy*, Preyer and Peter (eds.), Oxford: Oxford University Press.
Eklund, Matti. (2002a). "Inconsistent Languages," *Philosophy and Phenomenological Research* 64: 251-275.
———. (2002b). "Deep Inconsistency", *Australasian Journal of Philosophy* 80 (2002): 321-31.
———. (2005a). "What Vagueness Consists in", *Philosophical Studies* 125 (2005): 27-60.
———. (2005b). "Fiction, Indifference, and Ontology," *Philosophy and Phenomenological Research* 71: 557-579.
———. (2007). "Meaning-Constitutivity", *Inquiry* 50: 559-74.
———. (2008a). "The Liar Paradox, Expressibility, Possible Languages," in Beall (2007a).
———. (2008b). "Reply to Beall and Priest", *Australasian Journal of Logic* 6 (2008): 94-106.
Einheuser, Iris. (2008). "Three Forms of Truth Relativism," in Garcia-Carpintero and Kölbel (2008).
Engel, Pascal. (2002). *Truth*. Montreal: McGill-Queen's University Press.
Englebretsen, George. (2006). *Bare Facts and Naked Truths*. Ashgate.
Evans, Gareth. (1982). *The Varieties of Reference*. J. McDowell (ed.). Oxford: Oxford University Press.
Everaert, Martin, Lentz, Tom, de Mulder, Hannah, Nilson, Oystein, and Zondervan, Arjen (eds.) (2010). *The Linguistics Enterprise*. Amsterdam: John Benjamins.
Falvey, Kevin and Owens, Joseph. (1994). "Externalism, Self-Knowledge, and Skepticism," *Philosophical Review* 103: 107-37.

Feferman, Solomon. (1982). "Toward Useful Type-Free Theories, I," *The Journal of Symbolic Logic* 49: 75-111.
———. (1991). "Reflecting on Incompleteness," *The Journal of Symbolic Logic* 56: 1-49.
———. (1999). "Logic, Logics, and Logicism." *Notre Dame Journal of Formal Logic* 40: 31–54.
Field, Hartry. (1972). "Tarski's Theory of Truth," in Field 2001.
———. (1973). "Theory Change and the Indeterminacy of Reference," in Field 2001.
———. (1974). "Quine and the Correspondence Theory," in Field 2001.
———. (1977). "Logic, Meaning, and Conceptual Role," *The Journal of Philosophy* 74: 379-409.
———. (1980). *Science without Numbers*. Oxford: Blackwell.
———. (1986). "The Deflationary Conception of Truth," *Fact, Science and Morality*. Graham McDonald (ed.) Oxford: Blackwell Publishers. 55-117.
———. (1994a). "Deflationist Views of Meaning and Content," in Field 2001.
———. (1994b). "Disquotional Truth and Factually Defective Discourse," in Field 2001.
———. (1998). "Some Thoughts on Radical Indeterminacy," in Field 2001.
———. (1999). "Deflating the Conservativeness Argument," *Journal of Philosophy* 96: 533-540.
———. (2000). "Indeterminacy, Degree of Belief, and Excluded Middle," in Field 2001.
———. (2001a). *Truth and the Absence of Fact*. Oxford: Oxford.
———. (2001b). "Attributions of Meaning and Content," in Field 2001a.
———. (2001c). Postscript to Field 1972, in Field 2001a.
———. (2001d). Postscript to Field 1973, in Field 2001a.
———. (2001e). Postscript to Field 1974, in Field 2001a.
———. (2001f). Postscript to Field 1994a, in Field 2001a.
———. (2001g). Postscript to Field 1998, in Field 2001a.
———. (2001h). Postscript to Field 2000, in Field 2001a.
———. (2002). "Saving Truth Schema From Paradox," *Journal of Philosophical Logic* 31: 1-27.
———. (2003). "A Revenge-Immune Solution to the Semantic Paradoxes," *Journal of Philosophical Logic* 32: 139-177.
———. (2003b) "The Semantic Paradoxes and the Paradoxes of Vagueness," in Beall (2003).
———. (2003c). "No Fact of the Matter," *Australasian Journal of Philosophy* 81: 457-480.
———. (2004). "The Consistency of the Naïve theory of Properties," *The Philosophical Quarterly* 54: 78-104
———. (2005a). "Is the Liar Sentence Both True and False?" in Beall and Armour-Garb (2005).
———. (2005b). "Variations on a Theme by Yablo," in Beall and Armour-Garb (2005).
———. (2005c). "Precis of *Truth and the Absence of Fact*," *Philosophical Studies* :41–44
———. (2005d). "Replies," *Philosophical Studies* :105–28.
———. (2006a). "Truth and the Unprovability of Consistency," *Mind* :567–605.
———. (2006b). "Compositional Principles versus Schematic Reasoning," *The Monist* :9–27.
———. (2006c). "Maudlin's Truth and Paradox," *Philosophy and Phenomenological Research* : 713–20.
———. (2007). "Solving the Paradoxes, Escaping Revenge," in Beall 2007a.
———. (2008a). *Saving Truth From Paradox*. Oxford: Oxford University Press.
———. (2008b). *Logic, Normativity, and Rational Revisability,* The John Locke Lectures, Oxford University, 2008.
———. (2009). "What is the Normative Role of Logic," *Proceedings of the Aristotelian Society*, Supplementary Volume 83: 251–68.
———. (2010a). "Precis of Saving Truth from Paradox,"
———. (2010b). "Replies to McGee, Restall, and Shapiro,"
———. (Forthcoming). "Mathematical Undecidables, Metaphysical Realism, and Equivalent Descriptions," in *The Philosophy of Hilary Putnam*, *Library of Living Philosophers*. Forthcoming.

Fine, Kit. (1975). "Vagueness, Truth, and Logic," *Synthese* 54: 235-59.
———. (1982). "First-Order Modal Theories III — Facts," *Synthese* 53: 43-122.
von Fintel, Kai and Gillies, Anthony. (2008). "CIA Leaks," *Philosophical Review* 117:
Fitch, Frederic B. (1946). "Self-Reference in Philosophy," *Mind* 55: 64-73.
———. (1964). "Universal Metalanguages for Philosophy," *Review of Metaphysics* 17: 396-402.
Fitting, Melvin. (1986). "Notes on the Mathematical Aspects of Kripke's Theory of Truth," *Notre Dame Journal of Formal Logic* 27: 75-88.
Fitting, Melvin and Mendelsohn. (1998). *First-Order Modal Logic.* Berlin: Springer.
Fitzgerald, Gareth. (2009). "Linguistic Intuitions," *British Journal for the Philosophy of Science* 61: 123-160.
Fodor, Jerry. (1975). *The Language of Thought.* Cambridge, MA: Harvard University Press.
———. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind.* Cambridge, MA: MIT Press.
———. (1990). *A Theory of Content and Other Essays.* MIT Press.
———. (1998). *Concepts: Where Cognitive Science Went Wrong.* Oxford: Oxford University Press.
———. (2004). "Having Concepts: A Brief Refutation of the Twentieth Century," *Mind and Language* 19: 29-47.
———. (2008). *LOT 2: The Language of Thought Revisited.* Oxford University Press.
Foster, John. (1976). "Meaning and Truth Theory," in *Truth and Meaning*, Evans and McDowell (eds.), Oxford: Oxford University Press.
Fox, John. (2008). "What is at Issue Between Epistemic and Traditional Accounts of Truth?" *Australasian Journal of Philosophy* 86: 407-420.
Franzen, Torkel. (2004). *Inexhaustibility: A Non-Exhaustive Treatment.* Wellesley, Massachusetts: A. K. Peters.
———. (2005).*Gödel's Theorem: An Incomplete Guide to its Use and Abuse.* Wellesley, Massachusetts: A. K. Peters.
Frege, Gottlob. (1892). "On Sense and Reference," in *Translations from the Philosophical Writings of Gottlob Frege*, P. Geach and M. Black (eds. and tr.), Oxford: Blackwell, 1980.
———. (1918). "Thoughts: A Logical Enquiry," in *Translations from the Philosophical Writings of Gottlob Frege*, P. Geach and M. Black (eds. and tr.), Oxford: Blackwell, 1980.
French, Peter, and Wettstein, Howard. (eds.) (2001). Midwest *Studies in Philosophy, vol 25: Figurative Language.* New York: Blackwell.
Friedman, Harvey, and Sheard, Michael. (1987). "An Axiomatic Approach to Self-Referential Truth," *Annals of Pure and Applied Logic* 33: 1-21.
———. (1988). "The Disjunction and Existence Properties for Axiomatic Systems of Truth," *Annals of Pure and Applied Logic* 40: 1-10.
Fumerton, Richard. (2002). *Realism and The Correspondence Theory of Truth.* Boston: Rowman and Littlefield.
Gaifman, Haim. (1983). "Paradoxes of Infinity and Self-Applications, I," *Erkenntnis* 20: 131-155.
———. (1992). "Pointers to Truth," *The Journal of Philosophy* 89: 223-261.
———. (2000). "Pointers to Proposition," in Chapuis and Gupta 2000.
Garcia-Carpenterio, Manuel, and Kolbel, Max (eds.). (2008). *Relative Truth.* Oxford: Oxford University Press.
Garson, James, "Modal Logic," *The Stanford Encyclopedia of Philosophy (Winter 2009 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2009/entries/logic-modal/>.
Gauker, Christopher. (2001). "T-Schema Deflationism versus Gödel's First Incompleteness Theorem," *Analysis* 61: 129-136.

———. (2003). "Truth, Propositions and Context," in *Philosophical Dimensions of Logic and Science*, Rojszczak, Cachro and Kurczewski, (eds.), Kluwer.
———. (2006). "Against Stepping Back: A Critique of Contextualist Approaches to the Semantic Paradoxes," *Journal of Philosophical Logic* 35: .
Gertler, Brie. (2002) "Explanatory Reduction, Conceptual Analysis, and Conceivability Arguments about the Mind," *Noûs* 36: 22-49.
Gettier, Edumund. (1963). "Is Justified True Belief Knowledge?" ***Analysis* 23 (1963): 121-123.** .
Gibbard, Allan. (1990). *Wise Choices, Apt Feelings.* Cambridge: Harvard University Press.
Gillon, Brendan. (2004). "Ambiguity, Indeterminacy, Deixis, and Vagueness, Evidence and Theory," in Davis and Gillon (2004).
Givón, Talmy. (2001). *Syntax: An Introduction.* Amsterdam: John Benjamins.
Glanzberg, Michael. (2001). "The Liar in Context," *Philosophical Studies* 103: 217-251.
———. (2003a). "Against Truth-Value Gaps," in Beall (2003).
———. (2003b). "Minimalism and Paradoxes," *Synthese* 135: 13-36.
———. (2004). "A Contextual-Hierarchical Approach to Truth and the Liar Paradox," *Journal of Philosophical Logic* 33: 27-88.
———. (2005). "Truth, Reflection, and Hierarchies," *Synthese* 142: 289-315.
———. (2007). "Context, Content, and Relativism," *Philosophical Studies* 136: 1-29.
———. (2009). "Semantics and Truth Relative to a World," *Synthese* 166: 281-307.
———. (forthcoming). "Not All Contextual Parameters are Alike,"
Goble, Lou. (ed.) (2001). *The Blackwell Guide to Philosophical Logic.* Oxford: Blackwell.
Gödel, Kurt. (1931), " On Formally Undecidable Propositions of Principia Mathematica and Related Systems I," in *From Frege to Gödel*, Van Heijenoort (ed.), Cambridge: Harvard University Press.
Goldstein (1985). "The Paradox of the Liar — A Case of Mistaken Identity," *Analysis* 45: 9-13.
———. (1986a). "Epimenides and Curry," *Analysis* 46: 117-121.
———. (1986b). "False Stipulation and Semantical Paradox," *Analysis* 49: 192-195.
———. (1992). "'This Statement is not True' is not True," *Analysis* 52: 1-5.
———. (1999). "A Unified Solution to Some Paradoxes," *Proceedings of the Aristotelian Society* 100: 53-74.
———. (2001). "Truth-Bearers and the Liar: A Reply to Alan Weir," *Analysis* 61: 115-126.
———. (2009). "A Consistent Way with Paradox," *Philosophical Studies* 144: 377-389.
Goldstein, Laurence, and Goddard, Leonard. (1980). "Strengthened Paradoxes," *Australasian Journal of Philosophy* 58: 211-221.
Goodman, Nelson. (1955). *Fact, Fiction, and Forecast.* Cambridge: Harvard University Press.
Gottwald, Siegfried. (2001). *A Treatise on Many-Valued Logics.* Hertfordshire: Research Studies Press.
Graff, Delia and Williamson, Timothy. (eds.) (2002). *Vagueness.* Ashgate.
Greenough, Patrick. (2001). "Free Assumptions and the Liar Paradox," *American Philosophical Quarterly* 38: 115-135.
———. (2009). "Truthmaker Gaps and the No-No Paradox," *Philosophy and Phenomenological Research*
———. (2010a). "Deflationism and Truth Value Gaps," in Wright and Pedersen (2010).
———. (2010b). "Truth-Relativism, Norm-Relativism, and Assertion," in *Assertion*, Cappelen (ed.), Oxford: Oxford University Press.
Greenough, Patrick, and Lynch, Michael (eds.). (2006). *Realism and Truth.* Oxford: Oxford University Press.

Grelling, K. and Nelson, L (1908). *Bemerkungen zu den Paradoxien von Russell und Burali-Forti.* In: Abhandlungen der Fries'schen Schule II, Göttingen.

Grice, Paul. (1975). "Logic and Conversation," in *Syntax and semantics*, Cole, P. and Morgan, J. (eds.) vol 3, New York: Academic Press.

———. (1989). *Studies in the Way of Words*. Harvard University Press.

Grim, Patrick. (1991). *The Incomplete Universe: Totality, Knowledge, and Truth.* Cambridge: The MIT Press.

Groenendijk, Jeroen, and Stakhof, Martin. (1991). "Dynamic Predicate Logic," *Linguistics and Philosophy* 14: 39-100.

Gross, Steven. (2005). "Linguistic Understanding and Belief," *Mind* 114: 61-6.

———. (2006). "Can Empirical Theories of Semantic Competence Really Help Limn the Structure of Reality?" *Nous* 40: 43-81.

Grover, Dorothy. (1976). "'This is False' on the Presentential Theory," *Analysis* 36: 80-83.

———. (1977). "Inheritors and Paradox," *The Journal of Philosophy* 74: 500-604.

———. (1992). *A Prosentential Theory of Truth*. Princeton: Princeton University Press.

———. (2001). "The Prosentential Theory: Further Reflections on Locating Our Interest in Truth," in Lynch 2001.

Grover, Dorothy, Camp, Joseph, and Belnap, Nuel. (1975). "A Prosentential Theory of Truth," in Grover (1992).

Gupta, Anil. (1982). "Truth and Paradox," *Journal of Philosophical Logic* 11: 1-60.

———. (1983). "Postscript to 'Truth and Paradox'," in Martin (1984).

———. (1989). "The Liar: An Essay on Truth and Circularity," *Philosophy of Science* 56: 697-709.

———. (1990). "Two Theorems Concerning Stability," in *Truth or Consequences*, J. M. Dunn and A. Gupta (eds.), Dordrecht: Kluwer Academic Publishers.

———. (1993a). "A Critique of Deflationism," *Philosophical Topics* 21: 57-81.

———. (1993b). "Minimalism," *Philosophical Perspecitives* 7: 359-369.

———. (1997). "Definition and Revision: A Response to McGee and Martin," *Philosophical Issues* 8: 419-443.

———. (1999). "Meaning and Misconceptions," in *Language, Logic, and Concepts*. R. Jackendoff, P. Bloom and K. Wynn (eds.), Cambridge: MIT Press.

———. (2000). "On Circular Concepts," in Chapuis and Gupta (2000).

———. (2001). "Truth," in *The Blackwell Guide to Philosophical Logic,* L. Goble (ed.), Oxford: Blackwell.

———. (2002). "Partially Defined Predicates and Semantic Pathology," *Philosophy and Phenomenological Research* 65: 402-409.

———. (2005a). "Do the Paradoxes Pose a Special Problem for Deflationism?" in *Deflationary Truth*, Beall and Armour-Garb (eds.), Chicago: Open Court.

———. (2005b). Postscript to Gupta (1993a), in Beall and Armour-Garb (2005).

———. (2006). "Remarks on a Foundationalist Theory of Truth," *Philosophy and Phenomenological Research* 73: 721-7.

Gupta, Anil, and Belnap, Nuel. (1993). *The Revision Theory of Truth.* Cambridge: MIT Press.

Gupta, Anil, and Martin, Robert L. (1984). "A Fixed Point Theorem for the Weak Kleene Valuation Scheme," *Journal of Philosophical Logi*c 13: 131-135.

Gutting, Gary. (1980). *Paradigms and Revolutions: Applications and Appraisals of Thomas Kuhn's Philosophy of Science*.

Habermas, Jürgen. (1973). "Wahrheitstheorien," in *Wirklichkeit und Reflexion*, H. Fahrenbach (ed.), Pfüllingen: Neske.

Habermas, Jürgen. (2003). *Truth and Justification.* B. Fultner (trans.). Cambridge, MA: MIT Press.

Halbach, Volker. (1994). "A System of Complete and Consistent Truth," *Notre Dame Journal of Formal Logic* 35: 311-327.
———. (1996). "Tarski-hierachies," *Erkenntnis* 43: 339-367
———. (1997). "Tarskian and Kripkean Truth," *Journal of Philosophical Logic* 26: 69-80.
———. (1999). "Disquotationalism and Infinite Conjunctions," *Mind* 108: 1-22.
———. (2000a). "Truth and Reduction," *Erkenntnis* 53: 97-126.
———. (2000b). "Disquotationalism Fortified," in Chapuis and Gupta 2000.
———. (2001). "How Innocent is Deflationism?" *Synthese* 126: 167-194.
———. (2002). "Modalized Disquotationalism," in Halbach and Horsten (2002).
Halbach, Volker, and Horsten, Leon (eds.). (2002). *Principles of Truth*. Munich: Frankfurt am Main.
Halbach, Volker, Leitgeb, Hannes, and Welch, Philip. (2003) "Possible-Worlds Semantics for Modal Notions Conceived as Predicates," *Journal of Philosophical Logic* 32: 179-223.
Hales, Steven. (2006). *Relativism and the Foundations of Philosophy*. Cambridge, MA: MIT Press.
Hall, Alison. (2008). "Free Enrichment or Hidden Indexicals?" *Mind and Language* 23: 426-456.
Halmos, Paul. (1974). *Naïve Set Theory*. Springer.
Hahnle, Reiner. (2001). "Advanced Many-valued Logics," in *Handbook of Philosophical Logic*, 2nd ed, vol. 2, Gabbay and Guenthner (eds.), Dordrecht: Kluwer.
Hanna, Robert. (2006). *Rationality and Logic*. Cambridge, MA: MIT.
Hardy, James. (1997). "Three Problems for the Singularity Theory of Truth," *Journal of Philosophical Logic* 26: 501-520.
Harman, Gilbert. (1986). *Change in View: Principles of Reasoning*. Cambridge: MIT Press.
———. (1995). "Rationality," in *Thinking: Invitation to Cognitive Science*, vol. 3, E. Smith and D. Osherson (eds.), Cambridge: MIT Press.
———. (1999). *Reasoning, Meaning, and Mind*. Oxford: Oxford University Press.
———. (2000). *Explaining Value: And Other Essays in Moral Philosophy*, Oxford: Clarendon Press.
———. (2009). "Field on the Normative Role of Logic," *Proceedings of the Aristotelian Society* 83: 251-268.
Hawthorne, John. (1990). "A Note on 'Languages and Language'," *Australasian Journal of Philosophy* 68:116- 119.
———. (2004). *Knowledge and Lotteries*. Oxford: Oxford University Press.
Hazen, Allen. (1981). "Davis's Formulation of Kripke's Theory of Truth: A Correction," *Journal of Philosophical Logic* 10: 309-311.
———. (1987). "Contra Buridanum," *Canadian Journal of Philosophy* 17: 875-880.
Heil, John. (2003). *From an Ontological Point of View*. Oxford: Oxford University Press.
Heim, Irene. (1983). "File Change Semantics and the Familiarity Theory of Definiteness," in Meaning, Use, and the Interpretation of Language, Bäuerle, Schwarze, (eds.). de Gruyter.
Heim, Irene and Kratzer, Angelika. (1998). *Semantics in Generative Grammar*. New York: Blackwell.
Hendricks, Vincent (ed.) (2000). *Proof Theory: History and Philosophical Significance*, Berlin: Springer.
Hershfield, Jeffrey and Soles, Deborah. (2003). "Reinflating Truth as an Explanatory Concept," *Pacific Philosophical Quarterly* 84: 32–42.
Herzberger, Hans G. (1966). "The Logical Consistency of Language," in *Language and Learning*. Emig, J. A., Fleming, J. T., and Popp, H. M. (eds.) New York: Harcourt, Brace and World: 250-263.
———. (1967). "The Truth-Conditional Consistency of Natural Languages," *The Journal of Philosophy* 64: 29-35.
———. (1970). "Paradoxes of Grounding in Semantics," *The Journal of Philosophy* 67: 145-167.
———. (1973). "Dimensions of Truth," *Journal of Philosophical Logic* 2: 535-556.
———. (1981). "VII. —New Paradoxes for Old," *Proceedings of the Aristotelian Society 81*: 109-123.

———. (1982). "Naïve Semantics and the Liar Paradox," *The Journal of Philosophy* 79: 479-497.
Hill, Christopher. (2002). *Thought and World: An Austere Portrayal of Truth, Reference, and Semantic Correspondence.* Cambridge: Cambridge University Press.
Hinckfuss, Ian. (1991). "Pro Buridano: Contra Hazenum," *Canadian Journal of Philosophy* 21: 389-398.
Hintikka, Jaakko. (1996). *The Principles of Mathematics Revisited.* Cambridge: Cambridge University Press.
Hodges, Wilfred. (1986). "VIII. — Truth in a Structure," *Proceedings of the Aristotelian Society* 86: 135-151.
———. (1997). *A Shorter Model Theory*. Cambridge: Cambridge University Press.
———. (2009). "Logic and Games," Stanford Encyclopedia of Philosophy, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2009/entries/logic-games/>.
Hofweber, Thomas. (1999) "Contextualism and the Meaning-Intention Problem," in *Cognition, Agency and Rationality*, Korta, Sosa, and Arrozola (eds.), Kluwer.
———. (2005). Review of Künne (2003). *The Philosophical Review* 114:136-39.
———. (2007). "Validity, Paradox, and the Ideal of Deductive Logic" in Beall (2007a).
———. (2009). "Ambitious, yet modest, metaphysics" in *Metametaphysics*, D. Chalmers, D. Manley, and R. Wasserman (eds.). Oxford University Press.
———. (forthcoming). "Inferential Role and the Ideal of Deductive Logic," in *Meaning, Understanding, and Knowledge*, Patterson (ed.).
Hohwy, Jakob, and Kallestrup, Jesper (eds.) (2008). *Being Reduced: New Essays on Reduction, Explanation, and Causation*. Oxford: Oxford University Press.
Holton, Richard. (2000). "Minimalism and Truth-Value Gaps," *Philosophical Studies* 97: 137-168.
Hom, Christopher. (2008). "The Semantics of Racial Epithets" *Journal of Philosophy*
Horisk, Claire. (2007). "The Expressive Role of Truth in Truth-Conditional Semantics," *Philosophical Quarterly* 57: 535-557.
Horn, Lawrence. (1984). "Toward a New Taxonomy for Pragmatic Inference: Q-based and R-based Implicature," in *Meaning, Form, and Use in Context: Linguistic Applications,* D. Schiffrin (ed.), Washington: Georgetown University Press.
———. (1985). "Metalinguistic Negation and Pragmatic Ambiguity," *Language* 61: 121-174.
———. (2001). *A Natural History of Negation*, 2nd ed. Stanford: CSLI Publications.
———. (2004). "Implicature," in L. Horn et al. (2006).
———. (2005). "The Border Wars," in *Where Semantics Meets Pragmatics*. Turner et al. (eds.) Elsevie
———. (2006). "More issues in neo- and post-Gricean pragmatics: a reply to Carston," *Intercultural Pragmatics* 3: 81-93.
———. (2007a). "Neo-Gricean pragmatics: a Manichaean Manifesto," in N. Burton- Roberts (ed.), *Pragmatics*, Basingstoke: Palgrave.
———. (2007b). "Toward a Fregean pragmatics: Voraussetzung, Nebengedanke, Andeutung," in *Explorations in Pragmatics: Linguistic, Cognitive, and Intercultural Aspects*, I. Kecskes and L. Horn (eds.), Berlin: de Gruyter.
Horn, Laurence and Gregory Ward (eds.). (2004). *The Handbook of Pragmatics*. Oxford: Blackwell.
Horn, Laurence , Birner, Betty, and Ward, Gregory (eds.). (2006). *Drawing the Boundaries of Meaning.* Amsterdam: John Benjamins.
Hornsby, Jennifer. (1997). "Truth: The Identity Theory", *Proceedings of the Aristotelian Society* 97: 1–24.
———. (2005). "Truth without Truthmaking Entities," in Beebe and Dodd (2005).
Horsten, L. (1995). "The Semantical Paradoxes, the Neutrality of Truth and the Neutrality of the Minimalist Theory of Truth," in *The Many Problems of Realism*, Cortois P. (ed.), Tilburg: University Press.

Horwich, Paul. (1982). "Three Forms of Realims," *Synthese* 51: 181-201.
———. (1994). "The Essence of Expressivism," *Analysis* 54: 19-20.
———. (1997). "Implicit Definition, Analytic Truth and Apriori Knowledge", *Nous* 31(4), 423-440.
———. (1998a). *Truth*. Oxford: Clarendon Press.
———. (1998b). *Meaning*. Oxford: Oxford University Press.
———. (1999). "The Minimalist Conception of Truth", in *Truth: Oxford Readings in Philosophy*, edited by S. Blackburn and K. Simmons, Oxford: Oxford University Press.
———. (2001). "A Defense of Minimalism," Synthese 126: 149-165.
———. (2002). "Norms of Truth and Meaning," in Schantz 2002.
———. (2004). *From a Deflationary Point of View*. Oxford: Oxford University Press.
———. (2005). *Reflections on Meaning*. Oxford: Oxford University Press.
———. (2006). "The Value of Truth", Nous 40: 347-360.
———. (2008). "Being and Truth," *Midwest Studies in Philosophy* 32: 258-273.
———. (2010). *Truth-Meaning-Reality*. Oxford: Oxford University Press.
Hrbacek, Karel, and Jech, Thomas. (1999) *Introduction to Set Theory, 3rd ed.* CRC Press.
Hugly, Philip, and Sayward, Charles. (1980). "Is English Inconsistent?" *Erkenntnis* 15: 343-347.
Iacona, Andrea. (2008). "Faultless or Disagreement," in García-Carpintero and Kölbel (2008).
IAU (2006). "IAU 2006 General Assembly: Result of the IAU Resolution votes". *IAU*. 2006.
Jackson, Frank. (1997). *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Oxford University Press.
———. (2001a) . Precis of *From Metaphysics to Ethics*, *Philosophy and Phenomenological Research* 62: 617-624.
———. (2001b). "Responses," *Philosophy and Phenomenological Research* 62: 653-664.
Jackson, Frank, Oppy, Graham, and Smith, Michael. (1994). "Minimalism and Truth Aptness," *Mind* 103: 287-302.
Jacquette, Dale. (2004). "Grelling's Revenge," *Analysis* 64: 251–56.
James, William. (1907). *Pragmatism*. Cambridge, MA: Harvard University Press.
———. (1909). *The Meaning of Truth*, Cambridge, MA: Harvard University Press.
Jammer, Max. (2000). *Concepts of Mass in Contemporary Physics and Philosophy*. Princeton: Princeton University Press.
Jocahim, Harold. (1906). *The Nature of Truth*. Oxford: Clarendon Press.
Johnston, Mark. (1993). "Objectivity Refigured: Pragmatism without Verificationism," in *Reality, Representation, and Projection*, Haldane and Wright (eds.) Oxford: Oxford University Press.
Jorgensen, Jorgen. (1953). "Some Reflections on Reflexivity," *Mind* 62: 289-300.
———. (1955). "On Kattsoff's Reflections on Jorgensen's Reflections in Reflexivity," *Mind* 64: 542.
Juhl, Cory F. (1997). "A Context-Sensitive Liar," *Analysis* 57: 202-204.
Juhl, Cory, and Loomis, Eric. (2009). *Analyticity*. New York: Routledge.
Kalderon, Mark. (2005). *Moral Fictionalism*. Oxford: Clarendon Press.
Kamp, Hans. (1981). "A Theory of Truth and Semantic Representation". In *Formal Methods in the Study of Language,* Groenendijk, Janssen, and Stokhof (eds.), Mathematical Centre Tracts 135, Amsterdam.
Kaplan, David. (1973). "Bob and Carol and Ted and Alice," in *Approaches to Natural Language*, K. J. J. Hintikka, J. M. E. Moravcsik, and P. Suppes (eds.), Dordrecht: Reidel.
———. (1978). "On the Logic of Demonstratives," *Journal of Philosophical Logic*, 8: 81-98
———. (1989). "Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals," in *Themes From Kaplan*, J. Almog, J. Perry, and H. K. Wettstein (eds.), Oxford: Oxford University Press.

———. (1990). "Words," *Proceedings of the Aristotelian Society Supplement* 64: 93–119.
Kashlinsky et. al. (2008) "A Measurement of Large-Scale Peculiar Velocities of Clusters of Galaxies: Results and Cosmological Implications," *The Astrophysical Journal Letters* 686. doi: 10.1086/592947
———. (2009) "A Measurement of Large-Scale Peculiar Velocities of Clusters of Galaxies: Technical Details," *The Astrophysical Journal Letters* 691. doi: 10.1088/0004-637X/691/2/1479
Kattsoff, L. O. (1955). "Some Reflections on Jorgensen's Reflections in Reflexivity," *Mind* 64: 96-98.
Katz, Jerrold. (ed.) (1985). *The Philosophy of Linguistics*. Oxford: Oxford University Press.
Kearns, John. (1970). "Some Remarks Prompted by van Fraassen's Paper," in Martin (1970).
———. (2007). "An Illocutionary Logical Explanation of the Liar Paradox," *History and Philosophy of Logic*, 28, 31-66.
Keefe, Rosanna and Smith, Peter. (1999). *Vagueness: A Reader*. Cambridge, MA: MIT.
Keene, G. B. (1983). "Self-Referent Inference and the Liar Paradox," *Mind* 92: 430-432.
Kemp, Gary. (1998). "Meaning and Truth-Conditions," *Philosophical Quarterly* 48 (193):483-493.
Kennedy, Chris. (1999). *Projecting the Adjective: The Syntax and Semantics of Gradability and Comparison*. Garland.
———. (2010). "Ambiguity and Vagueness: An Overview," in *The Handbook of Semantics*, Maienborn, Heusinger, and Portner (eds.).
Kenyon, Tim (1999). "Truth, Knowability, and Neutrality," Noûs 33 (1):103-117.
Ketland, Jeffrey. (1999). "Deflationism and Tarski's Paradise," *Mind* 108: 69-94.
———. (2000). "A Proof of the (Strengthened) Liar Formula in a Semantical Extension of Peano Arithmetic," *Analysis* 60: 1-4.
———. (2005a). "Deflationism and Gödel Phenomena – Reply to Tennant," *Mind* 114: 75-88.
———. (2005b). "Jacquette on Grelling's Paradox," *Analysis* 65: 258-260.
———. (2009). "Beth's Theorem and Deflationism – Reply to Bays," *Mind* 118: 1075-1079.
———. (2010). "Truth, Conservativeness, and Provability: Reply to Cieśliński," *Mind* 119: 423-436.
Kindt, Walther. (1978). "The Introduction of Truth Predicates into First-Order Languages," *Formal Semantics and Pragmatics for Natural Languages*. F. Guenthner and S. J. Schmidt (eds.) Holland: D. Reidel Publishing Company.
King, Jeffrey. (1994). "Can Propositions be Naturalistically Acceptable?" *Midwest Studies in Philosophy* 19: 53-75.
———. (2003). "Tense, Modality and Semantic Values," *Philosophical Perspectives volume 17, Philosophy of Language*, J. Hawthorne (ed.).
Kirkham, Richard. (1995). *Theories of Truth*. Cambridge, MA: MIT Press.
Kitcher, Philip. (2002). "On the Explanatory Role of Correspondence Truth', *Philosophy and Phenomenological Research* (66): 346-364.
Kobes, Bernard. (1996). "Mental Content and Hot Self-Knowledge," *Philosophical Topics* 24: 71-99.
Kneale, William C. (1972). "Propositions and Truth in Natural Languages," *Mind* 81: 225-243.
Knobe, Joshua, and Nichols, Shawn (eds.). (2008). *Experimental Philosophy*. Oxford: Oxford University Press.
Kölbel, Max. (2001). "Two Dogmas of Davidsonian Semantics," *Journal of Philosophy* 98: 613-635.
———. (2002). *Truth without Objectivity*. London: Routledge.
———. (2003). "Faultless Disagreement," *Proceedings of the Aristotelian Society* 104: 53–73.

———. (2004). "Indexical Relativism versus Genuine Relativism," *International Journal of Philosophical Studies* 12 (3):297 – 313.
———. (2007). "How to Spell Out Genuine Relativism and How to Defend Indexical Relativism," *International Journal of Philosophical Studies* 15: 281 – 288.
———. (2008a). ""True' as Ambiguous," *Philosophy and Phenomenological Research* 77: 359-384.
———. (2008b). "Truth in Semantics," *Midwest Studies in Philosophy* 32 (1):242-257.
———. (2009). "The Evidence for Relativism," *Synthese* 166: 375-395.
Koons, Robert C. (1992). *Paradoxes of Belief and Strategic Rationality*. Cambridge: Cambridge University Press.
———. (2000b). "Circularity and Hierarchy," in Chapuis and Gupta 2000.
Kraut, Robert. (1993). "Robust Deflationism," *Philosophical Review* 102: 247-263.
Kremer, Michael. (1988). "Kripke and the Logic of Truth," *Journal of Philosophical Logic* 17: 225-278.
———. (2002). "Intuitive Consequences of the Revision Theory of Truth," *Analysis* 62: 330-336.
Kremer, Philip. (1993). "The Gupta-Belnap System $S^{\#}$ and S^{*} are Not Axiomatisable," *Notre Dame Journal of Formal Logic* 34: 583-596.
———. 2000). "On the 'Semantics' for Languages with Their Own Truth Predicates," in Chapuis and Gupta 2000.
Kripke, Saul. (1972). *Naming and Necessity*. Cambridge: Harvard University Press. 1980.
———. (1975). "Outline of a Theory of Truth," *The Journal of Philosophy* 72: 690-716.
———. (1977). "Speaker's Reference and Semantic Reference," in *Studies in the Philosophy of Language*, French, Uehling, and Wettstein (eds.), University of Minnesota Press.
———. (1979). "A Puzzle about Belief," in *Meaning and Use*, A. Margalit (ed.), Reidel: Dordrecht.
———. (1982). *Wittgenstein on Rules and Private Language*. Cambridge: Harvard University Press.
Kuhn, Thomas. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
Kukla, Rebecca, and Lance, Mark. (2008). *'Yo!' and 'Lo!': The Pragmatic Topography of the Space of Reasons*. Cambridge, MA: Harvard University Press.
Künne, Wolfgang. (2002). "Disquotationalist Conceptions of Truth," in Schantz (2002).
———. (2003). *Conceptions of Truth*. Oxford: Clarendon Press.
Kuusela, Oskari. (2008) *The Struggle Against Dogmatism: Wittgenstein and the Concept of Philosophy*. Cambridge, MA: Harvard University Press.
Lance, Mark. (1996). "Quantification, Substitution, and Conceptual Content," *Noûs* 30: 481-507.
———. (1998). "Some Reflections on the Sport of Language," *Philosophical Perspectives* 12: 219-240.
———. (1997). "The Significance of Anaphoric Theories of Truth and Reference," *Philosophical Issues* 8: 181-198.
———. (2001). "The Logical Structure of Linguistic Commitment III: Brandomian Scorekeeping and Incompatibility," *Journal of Philosophical Logic* 30: 439-464.
Lance, Mark, and Kremer, Philip. (1994). "The Logical Structure of Linguistic Commitment I," *Journal of Philosophical Logic* 23: 369–400
Lance, Mark, and Kremer, Philip. (1996). "The Logical Structure of Linguistic Commitment II: Systems of Relevant Commitment Entailment," *Journal of Philosophical Logic* 25 (4).
Larson, Richard, and Segal, Gabriel. (1995). Knoweledge of Meaning. Cambridge, MA: MIT Press.
Lasersohn, Peter. (2005). "Context Dependence, Disagreement, and Predicates of Personal Taste," *Linguistics and Philosophy* 28: 643–686.
———. (2008). "Quantification and Perspective in Relativist Semantics," *Philosophical Perspectives* 22: 305–337.
———. (2009). "Relative Truth, Speaker Commitment, and Control of Implicit Arguments," *Synthese* 166: 359–374
Lavine, Shaughan. (1994). *Understanding the Infinite*. Cambridge, MA: Harvard University Press.

Leeds, Stephen. (1978). "Theories of Reference and Truth," *Erkenntnis* 13: 111-129.
———. (1995). "Truth, Correspondence, and Success," *Philosophical Studies* 79: 1-36.
———. (2000). "A Disquotationalist Looks at Vagueness," *Philosophical Topics* 28:107-128.
———. (2007). "Correspondence Truth and Scientific Realism," *Synthese* 159: 1-21.
Leigh, Graham and Rathjen, Michael. (2010). "An Ordinal Analysis for Theories of Self-Referential Truth," *Archive for Mathematical Logic* 49: 213-247.
Leitgeb, Hannes. (2001a). "Truth as Translation – Part A," *Journal of Philosophical Logic* 30: 281-307.
———. (2001b). "Truth as Translation – Part B," *Journal of Philosophical Logic* 30:309-328.
———. (2005). "What Truth Depends On," *Journal of Philosophical Logic*
———. (2007). "On the Metatheory of Field's 'Solving the Paradoxes, Escaping Revenge'," in Beall (2007a).
Lepore, Ernie, and Cappelen, Herman. (2003). "Context Shifting Arguments," *Philosophical Perspectives* 17: 25–50.
———. (2005). *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism.* Oxford: Blackwell.
Leuenberger, Stephan. (2008). "Supervenience in Metaphysics," *Philosophy Compass* 3/4: 749–762
Levinson, Stephen. (1983). *Pragmatics.* Cambridge: Cambridge University Press.
———. (2000). *Presumptive Meanings.* Cambridge, MA: MIT Press.
Lewis, C. I. and Langford, C. H. (1959). *Symbolic Logic*, 2nd ed. New York: Dover.
Lewis, David. (1969). *Convention.* Cambridge: Harvard University Press.
———. (1970a). "How to Define Theoretical Terms," *Journal of Philosophy*, 67: 427–446.
———. (1970b). "General Semantics," *Synthese*, 22: 18–67.
———. (1974). "Radical Interpretation," *Synthese* 27: 331-344.
———. (1975). "Languages and Language," in *Philosophical Papers* vol. 1, Oxford: Oxford University Press, 1983.
———. (1979a). "Scorekeeping in a Language Game," in *Philosophical Papers*, vol. 1. Oxford: Oxford University Press. 1983.
———. (1980). "Index, Context, and Content," in Stig Kanger and Sven Öhman (eds.), *Philosophy and Grammar*, Dordrecht: Reidel.
———. (1994). "Reduction of Mind," in Samuel Guttenplan (ed.), *A Companion to Philosophy of Mind*, Oxford: Blackwell.
———. (1996). "Elusive Knowledge," *Australasian Journal of Philosophy*, 74: 549–567.
———. (2001a). "Truthmaking and Difference-Making," *Nous* ?: 602-615.
———. (2001b). Forget About the 'Correspondence Theory of Truth'. Analysis 61 (272):275–280.
Lewy, Cassimir. (1947). "Truth and Significance," *Analysis* 8: 24-27.
Longworth, Guy. (2008a). " Linguistic Understanding and Knowledge," *Noûs* 42: 50–79.
———. (2008b). "Comprehending Speech," *Philosophical Perspectives* 22: 339-373.
———. (2009). "A Plea for Understanding," In *New Waves in the Philosophy of Language,* Sarah Sawyer (ed.), New York: Palgrave.
Lopez de Sa, Dan. (2007a). "(Indexical) Relativism about Values," *Analysis*
———. (2007b). "The Many Relativisms and the Question of Disagreement," *International Journal of Philosophical Studies* 15: 269 – 279.
Lowe, E. J. and Rami, A. (eds.). (2009). *Truth and Truthmaking.* Acumen.
Ludlow, Peter and Martin, Norah. (eds.) (1998). *Externalism and Self-Knowledge.* Stanford, CA: CSLI.
Ludwig, Kirk. (2001) "What is the Role of a Truth Theory in a Meaning Theory?" in *Truth and Meaning: Topics in Contemporary Philosophy*, Seven Bridge Press.
———. (2004). "Davidson's Objection to Horwich's Minimalism about Truth," *Journal of Philosophy* 429-437.

Ludwig, Kirk, and Badici, Emil. (2007). "The Concept of Truth and the Semantics of the Truth Predicate," *Inquiry* 50: 622–638.
Ludwig, Kirk, and Lepore, Ernie. (2005). *Donald Davidson: Meaning, Truth, Language, and Reality*. Oxford: Oxford University Press.
———. (2007). *Donald Davidson's Truth Theoretic Semantics*. Oxford: Oxford University Press.
Lycan, William. (forthcoming). "A Truth Predicate in the Object Language,"
Lynch, Michael. (2001). "A Functionalist Theory of Truth," in Lynch 2001.
———. (2004). "Truth and Multiple Realizability," *Australasian Journal of Philosophy* 82: 384 – 408.
———. (2005). *True to Life: Why Truth Matters*. Cambridge, MA: MIT Press.
———. (2008). "Alethic Pluralism, Logical Consequence and the Universality of Reason," *Midwest Studies in Philosophy* 32: 122-140.
———. (2009). *Truth as One and Many*. Oxford: Oxford University Press.
MacFarlane, John. (2003) "Future Contingents and Relative Truth," *The Philosophical Quarterly* 53: 321–36.
———. (2005a). "Making Sense of Relative Truth," *Proceedings of the Aristotelian Society* 105: 321–39.
———. (2005b). "Logical Constants," *Stanford Encyclopedia of Philosophy*.
———. (2005c). "The Assessment Sensitivity of Knowledge Attributions," *Oxford Studies in Epistemology* 1: 197–233.
———. (2007a). "Relativism and Disagreement," *Philosophical Studies* 132: 17–31.
———. (2007b). "The Logic of Confusion," *Philosophy and Phenomenological Research* 74: 700–708.
———. (2008). "Truth in the Garden of Forking Paths," in Kölbel and García-Carpintero (2008).
———. (2009). "Nonindexical Contextualism," *Synthese* 166: 231–250.
———. (forthcoming a). "Epistemic Modals Are Assessment-Sensitive," in *Epistemic Modality*, Weatherson and Egan (eds.), Oxford: Oxford University Press.
———. (forthcoming b). "What Is Assertion?," in *Assertion*, Brown and Cappelen (eds.), Oxford: Oxford University Press.
———. (forthcoming c). "Simplicity Made Difficult," *Philosophical Studies*.
———. (forthcoming d). *Assessment Sensitivity: Relative Truth and Its Applications*.
MacFarlane, John and Kolodny, Nico. (forthcoming). "Ifs and Oughts,"
Machery, Eduard. (2009). *Doing without Concepts*. Oxford: Oxford University Press.
Mackie, J. L. (1973). *Truth Probability and Paradox*: *Studies in Philosophical Logic*. Oxford: Clarendon Press.
———. (1977). *Ethics: Inventing Right and Wrong*. Viking Press.
Mackie, J. L., Smart, J. J. C. (1953). "A Variant of the 'Heterological' Paradox," *Analysis* 13: 61-65.
———. (1954). "A Variant of the 'Heterological' Paradox – A Further Note," *Analysis* 14: 146-149.
Maddy, Penelope. (2007). *Second Philosophy: A Naturalistic Method*. Oxford: Oxford University Press.
Mares, Edwin. (2004). *Relevant Logic: A Philosophical Interpretation*. Cambridge: Cambridge University Press.
Margolis, Eric, and Lawrence, Stephen. (1999). *Concepts: Core Readings*. Cambridge, MA: MIT Press.
Marino, Patricia. (2006). "What Should a Correspondence Theory Be and Do?" *Philosophical Studies* 127: 415-457.
———. (2008). "Toward a Modest Correspondence Theory of Truth," *Dialogue* 47: 81-102.
———. (2010). "Representation-Friendly Deflationism versus Modest Correspondence," in Wright and Pedersen (2010).
Martí, Luisa. (2006). "Unarticulated Constituents Revisited," *Linguistics and Philosophy* 29: 135-166.
Martin, Robert L. (1967). "Toward a Solution to the Liar Paradox," *The Philosophical Review* 76: 279-311.
———. (1968). "On Grelling's Paradox," *The Philosophical Review* 77: 321-331.

Martin, Anthony. (1997). "Revision and its Rivals," *Philosophical Issues* 8: 407-418.
Martinich, A. P. (1983). "A Pragmatic Solution to the Liar Paradox," *Philosophical Studies* 43: 63-67.
Mates, Benson. (1981). "Two Antinomies," *Skeptical Essays*. Chicago: Chicago University Press: 15-57.
Maudlin, Tim. (2004). *Truth and Paradox: Solving the Riddles*. Oxford: Oxford University Press.
———. (2006a). Precis of Maudlin (2004). *Philosophy and Phenomenological Research* : 696-704
———. (2006b). Replies, *Philosophy and Phenomenological Research*: 728-739
———. (2007). "Reducing Revenge to Discomfort," in Beall (2007a).
McCarthy, Timothy. (1985). "Abstraction and Definability in Semantically Closed Structures," *Journal of Philosophical Logic* 14: 255-266.
McDonald, Brian Edison. (2000). "On Meaningfulness and Truth," *Journal of Philosophical Logic* 29: 433-482.
McDowell, John. (1992). "Putnam on Mind and Meaning," *Philosophical Topics* 20: 35-48.
———. (1994). *Mind and World*. Cambridge, MA: Harvard University Press.
———. (1999). "Scheme-Content Dualism and Empiricism", in *The Philosophy of Donald Davidson*, Hahn (ed.), Chicago: Open Court.
———. (2009). "Wittgensteinian Quietism," *Common Knowledge* 15: 365-372.
McGee, Vann. (1985). "How Truthlike Can a Predicate Be? A Negative Result," *Journal of Philosophical Logic* 14: 399-410.
———. (1989). "Applying Kripke's Theory of Truth," *The Journal of Philosophy* 86: 530-539.
———. (1991). *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth*. Cambridge: Hackett Publishing Company.
———. (1992). "Maximal Consistent Sets of Instances of Tarski's Schema (T)," *Journal of Philosophical Logic* 21: 235-241.
———. (1993). "A Semantic Conception of Truth?" *Philosophical Topics* 21: 83-111.
———. (1994). "Afterword: Truth and Paradox," *Basic Topics in the Philosophy of Language*, R. M. Harnish (ed.), Englewood Cliffs: Prentice Hall.
———. (1996). "Logical Operations," *Journal of Philosophical Logic* 25: 567-580.
———. (1997). "Revision," *Philosophical Issues* 8: 387-406.
———. (2000). "The Analysis of 'x is true' as 'for any p, if x = 'p', then p," in Chapuis and Gupta (2000).
———. (2005a). "Afterword: Trying (with Limited Success) to Demarcate the Disquotational/Correspondence Distinction," in Armour-Garb and Beall (2005).
———. (2005b). "Two Conceptions of Truth?" *Philosophical Studies* 124: 71-104.
———. (2010). "Field's Logic of Truth," *Philosophical Studies* 147: 421-432.
McGinn, Colin. (2000). *The Mysterious Flame: Conscious Minds in a Material World*. New York: Basic Books.
———. (2002). "The Truth about Truth," in Schantz 2002.
McGrath, Matthew. (2000). *Between Deflationism and the Correspondence Theory*. New York: Garland.
———. (2002). "Scott Soames: *Understanding Truth*," *Philosophy and Phenomenological Research* 65: 410-417.
———. (2003). "Deflationism and the Normativity of Truth," *Philosophical Studies*, 112: 47-67.
McKinsey, Michael (1991). "Anti-Individualism and Privileged Access," *Analysis*, 51: 9–16.
McLarty, Colin. (1993). "Anti-Foundation and Self-Reference," *Journal of Philosophical Logic* 22: 19-28.
McLaughlin, Brian, and Bennett, Karen. (2005). "Supervenience," *Standford Encyclopedia of Philosophy*.
Melia, Joseph. (2005). "Truthmaking without Truthmakers," in Beebee and Dodd (2005).

Mendelson, Eliot. (2001). *Introduction to Mathematical Logic*, 4th ed. Springer.
Merricks, Trenton. (2007). *Truth and Ontology*. Oxford: Oxford University Press.
Miller, Alexander. (2001). "On Wright's Argument Against Deflationism," *The Philosophical Quarterly* 51: 527-531.
Millikan, Ruth. (1984). *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.
———. (1986). "The Price of Correspondence Truth," *Nous* 20: 453-468.
———. (1990). "Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox," *The Philosophical Review* 99: 323-353.
———. (2000). *On Clear and Confused Ideas: An Essay about Substance Concepts*. Cambridge: Cambridge University Press.
Mills, Andrew P. (1995). "Unsettled Problems With Vague Truth," *Canadian Journal of Philosophy* 25: 103-117.
Mills, Eugene. (1998). "A Simple Solution to the Liar," *Philosophical Studies* 89: 197-212.
Misak, Cheryl. (2004). *Truth and the End of Inquiry*. Cambridge: Cambridge University Press.
Montague, Richard. (1963). "Syntactical Treatments of Modality, with Corollaries on Reflection Principles and Finite Axiomatizability," *Acta Philosophica Fennica* 16: 153-167.
———. (1974). *Formal Philosophy: Selected Papers*. Thomason (ed.). New Haven: Yale University Press.
Montminy, Martin. (2009). "Contextualism, Invariantism and Semantic Blindness," *Australasian Journal of Philosophy* 87: 639-657.
Moore, G. E. (1899). "The Nature of Judgment," *Mind* 8: 176-93.
Moltmann, Friederike. (2009). "Relative Truth and the First Person," *Philosophical Studies*
Montminy, Martin. (2009). "Contextualism, Relativism, and Ordinary Speakers," *Philosophical Studies* 143: 341-356.
Moruzzi, Sebastiano. (2008). "Assertion, Belief, and Disagreement: A Problem for Truth-Relativism," in García-Carpintero and Kölbel (2008).
Moruzzi, Sebastiano, and Wright, Crispin. (2009). "Trumping Assessments and the Aristotelian Future," *Synthese* 166: 301-339.
Moretti, Luca and Ciprotti, Nicola. (2009). "Logical Pluralism is Compatible with Monism About Metaphysical Modality," *Australasian Journal of Philosophy* 82(2).
Murphy, Gregory. (2002). *The Big Book of Concepts*. Cambridge, MA: MIT Press.
Muskens, Reinhard. (1996). "Combining Montague Semantics and Discourse Representation," *Linguistics and Philosophy*, 19: 143-186
Myhill, John. (1975). "Levels of Implication," in *The Logical Enterprise*, A. R. Anderson, R. B. Marcus, and R. M. Martin (eds.), London: Yale University Press.
———. (1984). "Paradoxes," *Synthese* 60: 129-143.
Naess, A. (1938). *'Truth' as Conceived by those who are not Professional Philosophers*. Oslo: I Komisjon Hos Jacob Dybwad.
Narens, Louis. (2002). *Theories of Meaningfulness*. Lawrence Erlbaum Associates.
———. (2007). *Introduction to the Theories of Measurement and Meaningfulness and the Use of Invariance in Science*. Lawrence Erlbaum Associates
Neale, Stephen. (1990). *Descriptions*. Cambridge, MA: MIT Press.
———. (1992). "Paul Grice and the Philosophy of Language," *Linguistics & Philosophy* 15: 509-559.
———. (2001). *Facing Facts*. Oxford: Oxford University Press.
———. (2005). "A Century Later," *Mind* 114: 809-871.
Negri, Sara, and von Plato, Jan. (2001). *Structural Proof Theory*. Cambridge: Cambridge University Press.
Newhard, Jay. (2005). "Grelling's Paradox," *Philosophical Studies* 126: 1-21.

———. (2009). "The Chrysippus Intuition and Contextual Theories of Truth," *Philosophical Studies* 142: 345-352.
Newman, Andrew. (2002). *The Correspondence Theory of Truth*. Cambridge: Cambridge University Press.
Nolt, John. (2008). "Truth as Epistemic Ideal," *Journal of Philosophical Logic* 37: 203-237.
Novek, Ira, and Sperber, Dan. (2004). *Experimental Pragmatics*. Palgrave.
Nozick, Robert. (2001). *Invariances: The Structure of the Objective World*. Cambridge, MA: Harvard University Press.
Nuccetelli, Susana (ed.) (2003). *New Essays on Semantic Externalism and Self-Knowledge*. Cambrdige, MA: MIT Press.
Olsson, Eric. (2005). *Against Coherence: Truth, Probability, and Justification*. Oxford: Oxford University Press.
Pagin, Peter. (2007). "Assertion," *Stanford Encyclopedia of Philosophy*.
Papineau, David. (2007). "Naturalism," *Stanford Encyclopedia of Philosophy*.
Parsons, Charles. (1974). "The Liar Paradox," *Journal of Philosophical Logic* 3: 381-412.
———. (1983). "Postscript to 'The Liar Paradox'," *Mathematics in Philosophy: Selected Essays*. Ithaca: Cornell University Press: 251-267.
Parsons, Josh. (forthcoming). "Assessment-Contextual Indexicals."
Parsons, Terence. (1984). "Assertion, Denial, and the Liar Paradox," *Journal of Philosophical Logic* 13: 137-152.
———. (1990). "True Contradictions," *Canadian Journal of Philosophy* 20: 335-354.
Partee, Barbara. (MS). "Appendix to Lecture 7: Implicatures, Presuppositions, Etc." http://people.umass.edu/partee/MGU_2009/materials/MGU097APPENDIX_2up.pdf.
Patterson, Douglass. (2005). "Deflationism and the Truth Conditional Theory of Meaning", *Philosophical Studies* 124: .
———. (2006). "Tarski, the Liar and Inconsistent Languages", *The Monist* 89: .
———. (2007a). "Inconsistency Theories: The Significance of Semantic Ascent", *Inquiry* 50: .
———. (2007b). "Understanding the Liar," in Beall (2007a).
———. (2009). "Inconsistency Theories of Semantic Paradox", *Philosophy and Phenomenological Research* 79: 387-422.
———. (2010) "Truth as Conceptually Primitive," in Wright and Pedersen (2010).
Peacocke, Christopher. (1976). "What is a Logical Constant?", *Journal of Philosophy* 73: 221-240
———. (1981). "The Theory of Meaning in Analytical Philosophy," in *Contemporary Philosophy*, vol. 1, G. Flöistad (ed.), The Hague: Nijhoff.
———. (1987). "Understanding Logical Constants: A Realist's Account", *Proceedings of the British Academy* 73: 153-200.
———. (1992). *A Study of Concepts*. Cambridge: MIT Press.
———. (2009). "Objectivity," *Mind* 118: 739-769.
Peirce, Charles Sanders. (1877). "The Fixation of Belief," *Popular Science Monthly* 12: 1–15.
———. (1878). "How to Make Our Ideas Clear," *Popular Science Monthly* 12: 286–302.
Peregrin, Jaroslav. (2008). "An Inferentialist Approach to Semantics: Time for a New Kind of Structuralism?" *Philosophy Compass* 3: 1208-1223.
Perry, John. (1977). "Frege on Demonstratives." *Philosophical Review* 86: 474-97.
———. (1979). "The Problem of the Essential Indexical." *Noûs* 13: 3-21.
———. (1998). "Indexicals, Contexts and Unarticulated Constituents," in *Proceedings of the 1995 CSLI-Armsterdam Logic, Language and Computation Conference*. CSLI Publications.
———. (2001). *Reference and Reflexivity*. Stanford: CSLI.
Petkov, Vesselin. (2009). *Relativity and the Nature of Spacetime*. Berlin: Springer.

Pettit, Dean. (2002). "Why Knowledge is Unnecessary for Understanding Language," *Mind* 111: 519-550.
———. (2005). "Belief and Understanding: A Rejoinder to Gross," *Mind* 114: 67-74.
Pettit, Philip. (2004). 'Existentialism, Quietism and the Role of Philosophy', in *The Future for Philosophy*, Leiter (ed.), Oxford: Oxford University Press.
Plantinga, Alvin. (1982). "How to be an Anti-Realist," *Proceedings and Addresses of the American Philosophical Association* 56: 47-70
Plato. (1961). "Theaetetus," in *Plato: Collected Dialogues*, Cornford (tr.), Princeton: Princeton University Press.
Pollock, John L. (1977). "The Liar Strikes Back," *The Journal of Philosophy* 74: 604-606.
Popper, Karl R. (1954). "Self-Reference and Meaning in Ordinary Language," *Mind* 63: 162-169.
Portner, Paul. (2005). *What is Meaning?* New York: Blackwell.
———. (2009). *Modality*. Oxford: Oxford University Press.
Portner, Paul and Partee, Barbara. (eds.) (2002). *Formal Semantics: The Essential Readings*. New York: Blackwell.
Potts, Christopher. (2005). *The Logic of Conventional Implicatures*. Oxford: Oxford University Press.
Predelli, Stephano. (2005). *Contexts: Meaning, Truth, and the Use of Language*. Oxford: Oxford University Press.
Predelli, Stefano, and Stojanovic, Isidora (2008). "Semantic Relativism and the Logic of Indexicals," in García-Carpintero and Kölbel (2008).
Preyer, Gerhard, and Peter, Georg. (2005). Contextualism in Philosophy: Knowledge, Meaning, and Truth. Oxford: Oxford University Press.
Price, Huw. (1988). *Facts and the Function of Truth*. Oxford: Basil Blackwell.
———. (1998). "Three Norms of Assertibility, or how the MOA Became Extinct," *Philosophical Perspectives* 12: 41-54.
———. (2003). "Truth as convenient friction. *Journal of Philosophy* **100**(2003) 167—190.
———. (2010). *Naturalism without Mirrors*. Oxford: Oxford University Press.
Priest, Graham. (1979). "The Logic of Paradox," *Journal of Philosophical Logic* 8: 219-241.
———. (1983). "The Logical Paradoxes and the Law of the Excluded Middle," *Philosophical Quarterly* 33: 160-165.
———. (1984a). "Logic of Paradox Revisited," *Journal of Philosophical Logic* 13: 153-179.
———. (1984b). "Semantic Closure," *Studia Logica* 43 (1-2).
———. (1990). "Boolean Negation and All That," *Journal of Philosophical Logic* 19: 201-215.
———. (1991). "Minimally Inconsistent LP," *Studia-Logica* 50: 321-331.
———. (1993). "Another Disguise of the Same Fundamental Problems: Barwise and Etchemendy on the Liar," *Australasian Journal of Philosophy* 71: 60-69.
———. (1994a). Review of McGee (1991). *Mind* 103: 387-391.
———. (1994b). "The Structure of the Paradoxes of Self-Reference," *Mind* 103: 25-34.
———. (1995). "Gaps and Gluts: Reply to Parsons," *Canadian Journal of Philosophy* 25: 57-66.
———. (1997). "Yablo's Paradox," *Analysis* 57: 236-242.
———. (1998). "What is So Bad About Contradictions," *The Journal of Philosophy* 95: 410-426.
———. (1999). "Semantic Closure, Descriptions and Non-Triviality," *Journal of Philosophical Logic* 28: 549-558.
———. (2000a). "Truth and Contradiction," *The Philosophical Quarterly* 50: 305-319.
———. (2000b). "Could Everything be True?" *Australasian Journal of Philosophy* 78: 189-195.
———. (2001). *An Introduction to Non-Classical Logic*. Cambridge: Cambridge University Press.
———. (2005). "Spiking the Field-Artillery," in Beall and Armour-Garb (2005).

———. (2006a). *In Contradiction: A Study of the Transconsistent, 2nd ed.* Oxford: Oxford University Press.
———. (2006b). *Doubt Truth to be a Liar.* Oxford University Press.
———. (2007). "Revenge, Field, and ZF," in Beall (2007a).
———. (2010). "Hopes Fade for Saving Truth," *Philosophy* 85: 109-140.
Priest, Graham, Beall, Jc, and Armour-Garb, Bradley. (2004). *The Law of Non-Contradiction.* Oxford: Oxford University Press
Prinz, Jesse. (2002). *Furnishing the Mind.* Cambridge, MA: MIT Press.
Prior, A. N. (1958). "Epimenides the Cretan," *The Journal of Symbolic Logic* 23: 261-266.
———. (1960). "The Runabout Inference Ticket," *Analysis* 21: 38-39.
———. (1961). "On a Family of Paradoxes," *Notre Dame Journal of Formal Logic* 2: 16-32.
Putnam, Hilary. (1962). "It Ain't Necessarily So." *Journal of Philosophy* 59: 658-671.
———. (1971). *Philosophy of Logic.* New York: Harper and Row.
———. (1975). "On the Meaning of 'Meaning'," in *Mind, Language and Reality*, Cambridge: Cambridge University Press,
———. (1978). *Meaning and the Moral Sciences.* Boston: Routledge and Kegan Paul.
———. (1981). *Reason, Truth, and History.* Cambridge: Cambridge University Press,
———. (1985). "A Comparison of Something with Something Else." *New Literary History* 17: 61-79.
Quine, W. V. (1948). "On What There Is," in *From a Logical Point of View*. Harvard University Press. 1953.
———. (1951). "Two Dogmas of Empiricism," in *From a Logical Point of View*. Harvard University Press. 1953.
———. (1960). *Word and Object.* Cambridge: MIT Press.
———. (1961). "The Ways of Paradox," in *The Ways of Paradox and Other Essays*, Cambridge: Harvard University Press, 1966.
———. (1970). *The Philosophy of Logic.* Cambridge, MA: Harvard University Press.
Ramsey, F. P. (1926). "Truth and Probability," in *Foundations of Mathematics and Other Logical Essays*, 1931, R. B. Braithwaite (ed.)
———. (1929). "Theories," in *Foundations of Mathematics*, Braithwaite (ed.), London: Routledge and Kegan Paul, 1931.
Ray, Greg. (2002). "Tarski, the Liar, and Truth Definitions," *Blackwell Companion to Philosophical Logic*, Ed. Dale Jacquette. Malden, MA: Blackwell.
———. (2003). "Tarski and the Metalinguistic Liar," *Philosophical Studies* 115 (2003): 55-80.
Rayo, Agustin, and Welch, Philip. (2007). "Field on Revenge," in Beall (2007a).
Read, Stephen. (1988). *Relevant Logic.* Oxford: Blackwell.
———. (2007). "Bradwardine's Revenge," in Beall (2007).
———. (2008a). "The Truth-Schema and the Liar" in *Unity, Truth and the Liar: The Modern Relevance of Medieval Solutions to the Liar Paradox*, Rahman, Tulenheimo and Genot (eds.), Springer Verlag.
———. (2008b). "Further Thoughts on the Truth-Scheme and the Liar," in *Unity, Truth and the Liar: The Modern Relevance of Medieval Solutions to the Liar Paradox*, Rahman, Tulenheimo and Genot (eds.), Springer Verlag.
———. (2009). "Plural Signification and the Liar Paradox," *Philosophical Studies* 145: 363-75.
———. (2010). "Field's Paradox and Its Medieval Solution," *History and Philosophy of Logic* 31: 161-76.
Recanati, François. (2001). "What is Said," *Synthese* 128: 75-91.
———. (2002). "Unarticulated Constituents," *Linguistics and Philosophy* 25: 299-345.

———. (2004). *Literal Meaning*. Cambridge: Cambridge University Press.
———. (2007). *Perspectival Thought*. Oxford: Oxford University Press.
———. (2008). "Contextualism and Relativism," in Kolbel and Garcia-Carpenterio (2008).
———. (2010). *Truth-Conditional Pragmatics*. Oxford: Oxford University Press
Reinhardt, William N. (1986). "Some Remarks on Extending and Interpreting Theories with a Partial Predicate for Truth," *Journal of Philosophical Logic* 15: 219-251.
Resnik, Michael. (1990). "Immanent Truth," *Mind* 99: 405-424.
———. (1997). *Mathematics as a Science of Patterns*. Oxford: Oxford University Press.
Rescher, Nicholas. (1973). *The Coherence Theory of Truth*, Oxford: Oxford University Press.
Restall, Greg. (1996). "Truthmakers, Entailment and Necessity," *Australian Journal of Philosophy,* 74: 331–340.
———. (2000). *An Introduction to Substructural Logics*. New York: Routledge.
———. (2005). "Minimalists about Truth can (and should) be Epistemicists, and it helps if they are Revision Theorists Too," in Beall and Armour-Garb (2005).
———. (2007). "Curry's Revenge: the Costs of Non-classical Solutions to the Paradoxes of Self-Reference," in Beall (2007a).
———. (2008). "Assertion and Denial, Commitment and Entitlement, and Incompatibility (and some consequence)," *Studies in Logic* 1: 26–36.
———. (2009). Appendix to Kukla and Lance (2009).
———. (2010). "What are We to Accept, and What are We to Reject, when Saving Truth from Paradox?" *Philosophical Studies* 147: 433–443.
———. (forthcoming). *Proof Theory and Philosophy*.
Richard, Mark. (1997). "Deflating Truth," *Philosophical Issues* 8: 57-78.
———. (2004). "Contextualism and Relativism," *Philosophical Studies* 119: 215-242.
———. (2008). *When Truth Gives Out*. Oxford: Oxford University Press.
Richards, Thomas J. (1967). "Self-Referential Paradoxes," *Mind* 76: 387-403.
Roberts, Craige. (1996). "Information Structure: Towards an Integrated Formal Theory of Pragmatics," in Yoon and Kathol (eds.) *OSUWPL Volume 49: Papers in Semantics*, 1996.
———. (1998). "Focus, the Flow of Information, and Universal Grammar," in *The Limits of Syntax*, Culicover and McNally (eds.), Academic Press.
———. (2002). "Demonstratives as Definites," in *Information Sharing: Reference and Presupposition in Language Generation and Interpretation*, van Deemter and Kibble (eds.), Stanford: CSLI.
———. (2003). "Uniqueness in Definite Noun Phrases," *Linguistics and Philosophy* 26: 287-350.
———. (2004). "Context in Dynamic Interpretation," in Horn and Ward (2004).
———. (2005). Pronouns as Definites," in *Descriptions and Beyond,* Reimer & Bezuidenhout (eds.) Oxford University Press.
———. (2010). "Retrievability and Definite Noun Phrases," talk given at the philosophy department of The Ohio State University, 5 Feb. 2010.
Rojszczak, Artur. (2005). *From the Act of Judging to the Sentence*. Berlin: Springer.
Rorty, Richard. (1986). "Pragmatism, Davidson, and Truth," in *Objectivity, Relativism, and Truth*, Cambridge: Cambridge University Press, 1991.
———. (1995). "Is Truth a Goal of Inquiry?" in *Truth and Progress*, Cambridge: Cambridge University Press, 1998.
———. (2007). "Naturalism and Quietism," in *Philosophy as Cultural Politics*, Cambridge: Cambridge. University Press.
Rosenberg, Jay. (1974) *Linguistic Representation*, D. Reidel: Dordrecht, Holland.
Ross, Alf. (1969). "On Self-Reference and a Puzzle in Constitutional Law," *Mind* 78: 1-24.
Routley, Richard. (1982). *Relevant Logics and their Rivals*. Atascadero, CA: Ridgeview.

Rozeboom, William W. (1957). "Is Epimenides Still Lying?" *Analysis* 18: 105-113.
Russell, Bertrand. (1905). "On Denoting," *Mind*, 14, 479–493.
———. (1910). "Knowledge by Acquaintance and Knowledge by Description," *Proceedings of the Aristotelian Society*, 11, 108–128.
Russell, Gillian. (2008). *Truth in Virtue of Meaning*. Oxford: Oxford University Press.
Ryle, Gilbert. (1931). "Systematically Misleading Expressions," *Proceedings of the Aristotelian Society* 32: 139-170.
———. (1949). *The Concept of Mind*. London: Hutchinson.
———. (1951). "Heterologicality," *Analysis* 11: 61-69.
Saebo, Kjell. (2009). "Judgment Ascriptions," *Linguistics and Philosophy* 32: 327-352.
Sainsbury, R. M. (2009). *Fiction and Fictionalism*. New York: Routledge.
Salerno, Joe. (2009). *New Essays on the Knowability Paradox*. Oxford: Oxford University Press.
Salmon, Nathan. (1986). *Frege's Puzzle*. Atascadero: Ridgeview.
Savellos, Elias, and Yalçin, Ümit (eds.). (1995). *Supervenience*. Cambridge: Cambridge University Press.
Sawyer, Sarah. (1999). "My Language Disquotes," *Analysis* 59: 206–211.
Schaffer, Jonathan. (2008a). "Truthmaker Commitments," *Philosophical Studies* 141: 7-19.
———. (2008b). "Truth and Fundamental Ontology," *Philosophical Books* 49: 302-316.
———. (2010a). "Monism: The Priority of the Whole," *Philosophical Review* 119: 31-76.
———. (2010b). "The Least Discerning and Most Promiscuous Truthmaker," *The Philosophical Quarterly* 60: 307-324.
———. (forthcoming). "Contextualism for Taste Claims and Epistemic Modals,"
Schantz, Richard (ed.). (2002). *What is Truth?* Berlin: Walter de Gruyter.
Scharp, Kevin. (2005). "Scorekeeping in a Defective Language Game," *Pragmatics and Cognition* 13: 203-226.
———. (2007a). "Aletheic Vengeance," in Beall (2007a).
———. (2007b). "Replacing Truth," *Inquiry* 50: 606-621.
———. (2009a). "Truth's Savior? Critical Study of Field's *Saving Truth From Paradox*," *The Philosophical Quarterly*
———. (2009b). "Truth and Expressive Completeness," in *Reading Brandom*, Weiss and Wanderer (eds.), New York: Routledge.
———. (2010). "Falsity," in Wright and Pedersen (2010).
Schechter, Eric. (2005). *Classical and Non-Classical Logics*. Princeton: Princeton University Press.
Schiffer, Stephen. (1993). "Actual-Language Relations," *Philosophical Perspectives* 7:231-258.
———. (1996). "Contextualist Solutions to Skepticism", *Proceedings of the Aristotelian Society*, 96: 317-333.
———. (2003). *The Things We Mean*. Oxford: Oxford University Press.
———. (2004). "Skepticism and the Vagaries of Justified Belief," *Philosophical Studies* 119 (1-2).
Schlenker, Philipe. (2007). "The Elimination of Self-Reference (Generalized Yablo-Series and the Theory of Truth)," *Journal of Philosophical Logic* 36: 251-307.
———. (forthcoming). "Super Liars," *Review of Symbolic Logic*.
Schroeder, Mark. (2008). Being For. Evaluating the Semantic Program of Expressivism. Oxford: Oxford University Press.
———. (2010). "How to Be an Expressivist About Truth," *New Waves in Truth*, edited by Nikolaj Jang Pedersen and Cory Wright.
Searle, J. R. (1995). *The Construction of Social Reality*, New York: The Free Press.
Sellars, Wilfrid. (1954). "Some Reflections on Language Games," *Philosophy of Science* 21 (1951): 204-28.

———. (1963). *Science, Perception and Reality*. London: Routledge and Kegan Paul.
———. (1969). "Language as Thought and Communication", *Philosophy and Phenomenological Research* 29:
Shapiro, Lionel. (2006). "The Rationale Behind Revision-Rule Semantics," *Philosophical Studies* 129: 477-515.
———. (forthcoming). "Expressibility and the Liar's Revenge," *Australasian Journal of Philosophy*.
Shapiro, Stewart. (1998). "Truth and Proof: Through Thick and Thin," *The Journal of Philosophy* 95: 493-521.
———. (2002). "Incompleteness and Inconsistency," *Mind* 111: 817-832.
———. (2003). "The Guru, the Logician, and the Deflationist: Truth and Logical Consequence," *Noûs* 37:113–132.
———. (2004). "Simple Truth, Contradiction, and Consistency," in Priest et al (2004).
———. (2007). "The Objectivity of Mathematics," *Synthese* 156:
———. (2010). "So Truth is Safe From Paradox: Now What?" *Philosophical Studies* 147: 445-455.
———. (forthcoming). "Truth, Function, and Paradox," *Analysis*.
Shapiro, Stewart, and Taschek, William. (1997). "Institutionalism, Pluralism, and Cognitive Command," *Journal of Philosophy* 93: 74-88.
Sheard, Michael. (1994). "A Guide to Truth Predicates in the Modern Era," *The Journal of Symbolic Logic* 59: 1032-1054.
Sher, Gila. (2003). "A Characterization of Logical Constants Is Possible," *Theoria* 18: 189-97.
Sider, Ted and Braun, David. (2006). "Kripke's Revenge," *Philosophical Studies* 128: 669–682.
Simmons, Keith. (1990). "The Diagonal Argument and the Liar," *Journal of Philosophical Logic* 19: 277-303.
———. (1993). *Universality and the Liar: An Essay on Truth and the Diagonal Argument*. Cambridge: Cambridge University Press.
———. (1994). "Paradoxes and Denotation," *Philosophical Studies* 76: 71-106.
———. (1999). "Deflationary Truth and the Liar," *Journal of Philosophical Logic* 28: 455-488.
———. (2000). "Three Paradoxes: Circles and Singularities," in Chapuis and Gupta 2000.
———. (2003). "Reference and Paradox," in Beall (2003).
Simmons, Keith, and Bar-On, Dorit. (2006). "The Use of Force Against Deflationism: Assertion and Truth," in *Truth and Speech Acts: Studies in the Philosophy of Language* , Greimann and Siegwart (eds.), New York: Routledge.
Simons, Mandy. (2003). "Presupposition and Accommodation: Understanding the Stalnakerian Picture," *Philosophical Studies* 112: 251-278.
Simons, Peter (1982), "Token Resistance," *Analysis*, 42/4: 195-203.
Sinisi, Vito. (1967). "Tarski on the Inconsistency of Colloquial Language," *Philosophy and Phenomenological Research*, Vol. 27: 537-541.
Skinner, R. C. (1959). "The Paradox of the Liar," *Mind* 68: 322-335.
Skyrms, Brian. (1968). "Supervaluations: Identity, Existence, and Individual Concepts," *The Journal of Philosophy* 65: 477-482.
———. (1970a). "VI. Return of the Liar: Three-Valued Logic and the Concept of Truth," *American Philosophical Quarterly* 7: 153- 161.
———. (1970b). "Notes on Quantification and Self-Reference," in Martin 1970.
———. (1982). "Intensional Aspects of Semantical Self-Reference," in Martin 1984: 119-131.
Slezak, Peter. (2009). "Linguistic Explanation and 'Psychological Reality'," *Croatian Journal of Philosophy* 25.
Smith, Michael. (1994). "Why Expressivists about Value should Love Minimalism about Truth," *Analysis* 54: 1-12.

Smith, Nicholas J. J. (2000). "The Principle of Uniform Solution (of the Paradoxes of Self-Reference)," *Mind* 109: 117-122.
Smith, Peter (2007). *An Introduction to Gödel's Theorems*. Cambridge: Cambridge University press.
Soames, Scott. (1984). "What is a Theory of Truth?" *The Journal of Philosophy* 81: 411-429.
———. (1997). "The Truth about Deflationism," in *Philosphical Issues*, vol 8, E. Villanueva (ed.), Atascadero, CA: Ridgeview Press.
———. (1999). *Understanding Truth*. Oxford: Oxford University Press.
———. (2002a). "Précis of *Understanding Truth*," *Philosophy and Phenomenological Research* 65: 397-401.
———. (2002b). "Replies," *Philosophy and Phenomenological Research* 65: 429-451.
———. (2002c). *Beyond Rigidity*. Oxford: Oxford University Press.
———. (2004). "Naming and Asserting," in *Semantics vs. Pragmatics*, Szabo (ed.), Oxford University Press.
———. (2007). "Understanding Assertion," in *Content and Modality: Themes from the Philosophy of Robert Stalnaker*, Thompson and Byrne (eds.), Oxford: Oxford University Press.
———. (2010). *What is Meaning?* Princeton: Princeton University Press.
Sorensen, Roy. (1998). "Yablo's Paradox and Kindred Infinite Liars," *Mind* 107 (425):137-155.
———. (2001). *Vagueness and Contradiction*. Oxford: Clarendon Press.
———. (2003). "A Definite No-No," in *Liars and Heaps*, Jc Beall (ed.), Oxford: Oxford University Press.
———. (2005). "Reply to Critics," *Philosophy and Phenomenological Research* 71: 695-703.
Spade, Paul Vincent. (1988). *Lies, Language and Logic in the Later Middle Ages*. London: Variorum Reprints.
Sperber, Dan and Wilson, Deirdra. (1986). *Relevance: Communication and Cognition*. Oxford: Blackwell.
———. (1993). "Linguistic Form and Relevance," *Lingua* 90: 1-25.
———. (1996). "Fodor's Frame Problem and Relevance Theory: reply to Chiappe & Kukla," *Behavioral and Brain Sciences* 19: 530-532.
———. (1997). "Remarks on Relevance Theory and the Social Sciences," in *Multilingua* 16: 145-51.
———. (1998). "Pragmatics and Time," in *Relevance theory: Applications and implications*. R. Carston & S. Uchida (eds.), Amsterdam: John Benjamins.
———. (2002a). "Truthfulness and Relevance," *Mind* 111, 583.
———. (2002b). "Pragmatics, Modularity and Mind-reading," *Mind and Language* 17: 3-23.
———. (2004). "Relevance Theory," in Horn and Ward (2004).
———. (2006). "Pragmatics," in *Oxford Handbook of Philosophy of Language,* Jackson and Smith (eds.), Oxford: Oxford University Press.
———. (2008). "A Deflationary Account of Metaphor," in *Handbook of Metaphor and Thought*, Gibbs (ed.) . Cambridge: Cambridge University Press.
Sprouse, Jon and Almeda, Diego. (MS). "A Quantative Defense of Linguistic Methodology,"
Stalnaker, Robert. (1970). "Pragmatics," in Stalnaker (1999).
———. (1973). "Presuppositions," in Stalnaker (1999).
———. (1974). "Pragmatic Presuppositions," in Stalnaker (1999).
———. (1978). "Assertion," in Stalnaker (1999).
———. (1987). *Inquiry*. Cambridge: MIT Press.
———. (1998). "On the Representation of Context," in Stalnaker (1999).
———. (1999). *Context and Content*. Oxford: Oxford University Press.
———. (2001). "Metaphysics without Conceptual Analysis," *Philosophy and Phenomenological Research*, 62: 631-636.
———. (2002). "Common Ground," *Linguistics and Philosophy* 25: 701-721.

———. (2004). "Assertion Revisited: On the Interpretation of Two-Dimensional Modal Semantics," *Philosophical Studies* 118: 299-322.
———. (2009). "A Response to Abbott on Presupposition and Common Ground," *Linguistics and Philosophy*, 31: 539-44.
Stanley, Jason. (1997). "Names and Rigid Designation", *A Companion to the Philosophy of Language*, Hale and Wright, ed. (Oxford, Blackwell Press, 1997): 555-585.
———. (2004). "On the Linguistic Basis for Contextualism," *Philosophical Studies*, 119: 119-146.
———. (2005). *Knowledge and Practical Interests*. Oxford: Oxford University Press.
———. (2007). *Language in Context*. Oxford: Oxford University Press.
Stanley, Jason and Szabo, Zoltan. (2000). "On Quantifier Domain Restriction", *Mind and Language* 15: 219-261.
Stebbins, Sarah. (1992). "A Minimal Theory of Truth," *Philosophical Studies* 66: 109-137.
Stenius, Erik. (1972). *Critical Essays*. Amsterdam: North-Holland Publishing Company.
Stephenson (2008). "Judge Dependence, Epistemic Modals, and Predicates of Personal Taste," *Linguistics and Philosophy* 30: 487–525.
———. (2009). "Relativism and the *De Se* Interpretation of PRO," in *Proceedings of the 38th Annual Meeting of the North East Linguistic Society*, A. Schardl, M. Walkow, and M. Abdurrahman (eds.), Amherst, Massachusetts: GLSA.
Stich, Stephen and Weinberg, Jonathan. (2001). "Jackson's Empirical Assumptions," *Philosophy and Phenomenological Research*, 62: f637-643.
Strawson, P. F. (1950). "Truth," *Proceedings of the Aristotelian Society, Supplementary Volumes* 24: 111-172.
———. (1959). *Individuals: An Essay in Descriptive Metaphysics*. London: Methuen.
Stroll, Avrum. (1954). "Is Everyday Language Inconsistent?" *Mind* 63: 219-225.
Stroud, Barry. (1984). *The Significance of Philosophical Skepticism*. Oxford: Oxford University Press
Stojanovic, Isidora. (2007). "Talking about Taste: Disagreement, Implicit Arguments, and Relative Truth," *Linguistics and Philosophy* 30: 691-706.
Suppes, Patrick. (1998). "Theory of Measurement," *Routledge Encyclopedia of Philosophy*, Craig (ed.), London: Routledge.
———. (2002). *Representation and Invariance of Scientific Structures*. Stanford: CSLI.
Suppes, Patrick, Krantz, David, Luce, Duncan, and Tversky, Amos. (1971). *Foundations of Measurement, Vol. I: Additive and Polynomial Representations*. New York: Academic Press.
———. (1989). *Foundations of Measurement, Vol. II: Geometrical, Threshold, and Probabilistic Representations*. New York: Academic Press.
———. (1990). *Foundations of Measurement, Vol. III: Representation, Axiomatization, and Invariance*. New York: Academic Press.
Szabo, Zoltan Gendler. (1999). "Expressions and their Representations," *The Philosophical Quarterly* 49: 145-163.
———. (ed.) (2004). *Semantics vs. Pragmatics*. Oxford: Oxford University Press.
———. (2006a). 'Sensitivity Training,' *Mind and Language*, 21: 31 – 38
———. (2006b). "The Distinction Between Semantics and Pragmatics," in *The Oxford Handbook of Philosophy of Language,* E. LePore and B. Smith (eds.), Oxford: Oxford University Press.
Tappendon, Jamie. (1993). "The Liar and Sorites Paradoxes: Toward a Unified Treatment," *The Journal of Philosophy* 90: 551-577.
———. (1994). Review of McGee (1991), *The Philosophical Review* 103: 142-144.
———. (1999). "Negation, Denial, and Language Change in Philosophical Logic," in *What is Negation?*, D. Gabbay and H. Wansing (eds.), Dordrecht: Kluwer.
———. (2002). "Comments on Soames' *Understanding Truth*," *Philosophy and Phenomenological Research* 65: 418-421.

Tarski, Alfred. (1933). "The Concept of Truth in Formalized Languages," in *Logic, Semantics, Meta-Mathematics*, J. H. Woodger (tr.) and J.Corcoran (eds.), Indianapolis: Hackett Publishing Company, 1983.
———. (1936). "On the Concept of Logical Consequence," in *Logic, Semantics, Meta-Mathematics*, J. H. Woodger (tr.) and J.Corcoran (eds.), Indianapolis: Hackett Publishing Company, 1983.
———. (1944). "The Semantic Conception of Truth," *Philosophy and Phenomenological Research* 4: 341-376
Taylor, Barry. (2006). *Models, Truth, and Realism*. Oxford: Oxford University Press.
Tennant, Neil. (1982). "Proof and Paradox," *Dialectica* 36: 265-296.
———. (1995). "On Paradox without Self-Reference," *Analysis* 55: 199-207.
———. (1995b). "On Negation, Truth, and Warranted Assertibility," *Analysis* 55: 98-104.
———. (1997). *The Taming of the True*. Oxford: Oxford University Press.
———. (2002). "Deflationism and the Gödel Phenomena," *Mind* 111: 551-582.
———. (2005). "Deflationism and the Gödel Phenomena – Reply to Ketland," *Mind* 114: 89-96.
———. (2009). "Deflationism and the Gödel Phenomena – Reply to Cieslinsky," *Mind* 119: 437-450.
———. (forthcoming). *Changes of Mind*.
———. (MS1). "Truth, Provability and Paradox: On some theorems of Lob, Montague and McGee, and a Conjecture about Constructivizability."
———. (MS2). "A New Unified Account of Truth and Paradox,"
———. (MS3). "On Scharpening the Rules of Truth,"
Textor, Mark. (2009). "Devitt on the Epistemic Authority of Linguistic Intuitions," *Erkenntnis* 71: 395-405.
Thagard, Paul. (1992). *Conceptual Revolutions*. Princeton: Princeton University Press.
———. (2007). "Coherence, Truth, and the Development of Scientific Knowledge," *Philosophy of Science* 74: 28-47.
Thomason, Richmond H. (1976). "Necessity, Quotation, and Truth: An Indexical Theory," in *Language in Focus*, A. Kasher (ed.), Holland: D. Reidel.
Tomasi, John. (2006). "Truth, Warrant, and Superassertibility," *Synthese* 148: 31-56.
Toms, Eric. (1956). "The Liar-Paradox," *The Philosophical Review* 65: 542-547.
Travis, Charles. (2008). *Occasion-Sensitivity: Selected Essays*. Oxford University Press.
Troelstra, A. and Schwichtenberg, H. (2000), *Basic Proof Theory*, 2nd ed. Cambridge: Cambridge University Press.
Truncellito, David. (2000). "Which Type is Tokened by a Token of a Word-Type?" *Philosophical Studies* 97: 251-266.
Ushenko, A. P. (1937). "A New 'Epimenides'," *Mind* 46: 549-550.
———. (1955). "A Note on the Liar-Paradox," *Mind* 64: 543.
———. (1957). "An Addendum to the Note on the Liar-Paradox," *Mind* 66: 98.
Urquhart, Alisdair. (2001). "Basic Many-valued Logic," in *Handbook of Philosophical Logic*, 2nd ed, vol. 2, Gabbay and Guenthner (eds.), Dordrecht: Kluwer.
Uzquiano, Gabriel. (2004). "An Infinitary Paradox of Denotation," *Analysis* 64: 128-131.
van Benthem, J. F. A. K. (1978). "Four Paradoxes," *Journal of Philosophical Log*ic 7: 49-72.
van Benthem, Johan and ter Meulen, Alice. (1997). *Handbook of Logic and Language*. Cambridge, MA: MIT Press.
van Dalen, Dirk. (2001). "Intuitionistic Logic," in *Handbook of Philosophical Logic*, 2nd ed, vol. 2, Gabbay and Guenthner (eds.), Dordrecht: Kluwer.
van Eijck, Eric and Visser, Albert. (forthcoming). "Dynamic Semantics," *Stanford Encyclopedia of Philosophy*.

van Fraassen, Bas C. (1968). "Presupposition, Implication, and Self-Reference," *The Journal of Philosophy* 65: 136-152.
———. (1970a). "Inference and Self-Reference," *Synthese* 21: 425-438.
———. (1970b). "Truth and Paradoxical Consequences," in Martin 1970.
———. (2008). *Scientific Representation*. Oxford: Oxford University Press.
van Valin, Robert. (2001). *An Introduction to Syntax*. Cambridge: Cambridge University Press.
von Fintel, Kai. (MS). "What is Presupposition Accommodation?" http://web.mit.edu/fintel/www/accomm.pdf.
Varzi, Achille. (2007). "Supervaluationism and Its Logics," *Mind* 116: 633-676.
Vigano, Luca. (2000). *Labelled Non-Classical Logics*. Dordrecht: Kluwer.
Villalta, Elisabeth. (2003) . "The Role of Context in the Resolution of Quantifier Scope Ambiguities," *Journal of Semantics* 20: 115-162.
Vision, Gerald. (2004). *Veritas: The Correspondence Theory and its Critics*. Cambridge, MA: MIT Press.
Visser, Albert. (1984). "Four Valued Semantics and the Liar," *Journal of Philosophical Logic* 13: 181-212.
———. (2001). "Semantics and the Liar Paradox," *Handbook of Philosophical Logic* 2nd ed., vol. 7, Gabbay and Guenthner (eds.), Amsterdam: D. Reidel.
Walker, Ralph. (1989). *The Coherence Theory of Truth: Realism, Anti-realism, Idealism*. New York: Routledge.
Wansing, Heinrich. (2006). "Connectives Stranger than Tonk," *Journal of Philosophical Logic* 35: 653-660.
Wasow, Thomas, Perfors, Amy, and Beaver, David. (2005). "The Puzzle of Ambiguity," in *Morphology and The Web of Grammar: Essays in Memory of Steven G. Lapointe*, O. Orgun and P. Sells (eds) CSLI.
Weatherson, Brian. (2009). "Conditionals and Indexical Relativism," *Synthese* 166: 333-357.
Wedgwood, Ralph. (1997). "Non-Cognitivism, Truth, and Logic," *Philosophical Studies* 86: 73-91.
Weiner, Matt. (2007). "Norms of Assertion," *Philosophy Compass* 2: 187-195.
———. (MS). "The (Mostly Harmless) Inconsistency of Knowledge Ascriptions."
Weir, Alan. (1996). "Ultramaximalist Minimalism!" *Analysis* 56: 10-22.
———. (2000). "Token Relativism and the Liar," *Analysis* 60: 156-170.
———. (2001). "Rejoinder to Laurence Goldstein on the Liar," *Analysis* 61: 26-34.
Welch, P. D. (2001). "On Gupta-Belnap Revision Theories of Truth, Kripkean Fixed Points, and the Next Stable Set," *The Bulletin of Symbolic Logic* 7: 345-360.
Wetzel, Linda. (1993). "What Are Occurrences of Expressions?" *Journal of Philosophical Logic*, 22: 215-220.
———. (2008). *Types and Tokens: An Essay on Universals*. Cambridge, MA: MIT Press.
Whiting, D. (2007). "Inferentialism, Representationalism and Derogatory Words," *International Journal of Philosophical Studies* 15: 191-205.
Williams, C. J. F. (1969). "What does 'X is true' Say about X?" *Analysis* 29: 113-124.
———. (1976). *What is Truth?* Cambridge: Cambridge University Press.
———. (1992). *Being, Identity, and Truth*. Oxford: Oxford University Press.
Williams, Michael. (1986). "Do We (Epistemologists) Need a Theory of Truth?" *Philosophical Topics* 14: 223-242.
———. (1996). *Unnatural Doubts*. Princeton: Princeton University Press.
———. (1999). "Meaning and Deflationary Truth," *The Journal of Philosophy* 96: 545-564.
———. (2002). "On Some Critics of Deflationism," in Schantz (2002).
Williams, Robert. (2008). "Supervaluationism and Logical Revisionism," *Journal of Philosophy* 105 (4).
Williamson, Timothy. (1994). *Vagueness*. London: Routledge.

———. (1996). "Knowing and Asserting." *Philosophical Review* 105, 489-523.
———. (1997). "Imagination, Stipulation, and Vagueness," in *Philosophical Issues 8: Truth*, E. Villanueva (ed.), Atascadero, CA: Ridgeview Publishing.
———. (2000a). *Knowledge and Its Limits*. Oxford: Oxford University Press.
———. (2000b). "Semantic Paradox and Semantic Change," in *Proceedings of the Twentieth World Congress of Philosophy, vol. 6*, A. Kanamori (ed.), Bowling Green: Philosophy Documentation Center.
———. (2001). "Ethics, Supervenience, and Ramsey Sentences," *Philosophy and Phenomenological Research* 62: 625-630.
———. (2002). "Soames on Vagueness," *Philosophy and Phenomenological Research* 65: 422-428.
———. (2003). "Understanding and Inference," *The Aristotelian Society*, *Supplement* 77: 249-293.
———. (2006). "'Conceptual truth'," *The Aristotelian Society*, *Supplement* 80**:** 1-41.
———. (2008). *The Philosophy of Philosophy*. Oxford: Blackwell.
Wilson, Mark. (2006). *Wandering Significance*. Oxford: Oxford University Press.
Wittgenstein, Ludwig. (1923) *Tractatus Logico-Philosophicus*. C.K. Ogden (trans.), London: Routledge & Kegan Paul.
———. (1953). *Philosophical Investigations*. G.E.M. Anscombe (tr.). Oxford: Blackwell.
Woodruff, Peter W. (1984). "Paradox, Truth and Logic: Part I: Paradox and Truth," *Journal of Philosophical Logic* 13: 213-232.
Woods, John. (2003). *Paradox and Paraconsistency*. Cambridge: Cambridge University Press.
Wormell, C. P. (1958). "On the Paradoxes of Self-Reference," *Mind* 67: 267-271.
Wright, Cory and Pedersen, Nikolaj. (2010). *New Waves in Truth*. New York: Palgrave.
Wright, Crispin. (1975). "On the Coherence of Vague Predicates," *Synthese* 30: 325--65.
———. (1988). "Realism, Anti-Realism, Irrealism, Quasi-Realism," in Wright (2003).
———. (1992). *Truth and Objectivity*. Cambridge: Harvard University Press.
———. (1998). "Comrades against Quietism: Reply to Simon Blackburn on *Truth and Objectivity*," *Mind* 107:183-203.
———. (1999). "Truth: A Traditional Debate Reviewed," in Wright (2003).
———. (2000). "Truth as Sort of Epistemic," in Wright (2003).
———. (2001). "Minimalism, Deflationism, Pragmatism, Pluralism," in Lynch 2001.
———. (2003). *Saving the Differences: Essays on Themes from* Truth and Objectivity. Cambridge, MA: Harvard University Press.
———. (2004a). "Warrant for Nothing (and Foundations for Free)?" *Aristotelian Society Supplementary Volume* 78: 167–212.
———. (2004b). "Intuition, Entitlement and the Epistemology of Logical Laws," *Dialectica* 58: 155–175.
———. (2008). "Relativism about Truth Itself: Haphazard Thoughts about the Very Idea," in García-Carpintero and Kölbel (2008).
Wyatt, Nicole. (2004). "What Are Beall and Restall Pluralists About?" *Australasian Journal of Philosophy* 82: 409 – 420.
Yablo, Stephen. (1982). "Grounding, Dependence, and Paradox," *Journal of Philosophical Logic* 11: 117-137.
———. (1985). "Truth and Reflection," *Journal of Philosophical Logic* 14: 297-349.
———. (1989). "Truth, Definite Truth, and Paradox," *The Journal of Philosophy* 86: 539-541.
———. (1993a). "Definitions, Consistent and Inconsistent," *Philosophical Studies* 72: 147-175.
———. (1993b). "Hop, Skip and Jump: The Agnostic Conception of Truth," *Philosophical Perspectives* 7: 371-396.
———. (1993c). "Paradox Without Self-Reference," *Analysis* 53: 251–52.

———. (1998). "Does Ontology Rest on a Mistake?" *Aristotelian Society Supplementary Volume* 72: 229–262.
———. (2001). "Go Figure: A Path Through Fictionalism," *Midwest Studies in Philosophy* 25: 72-102.
———. (2003). "New Grounds for Naïve Truth Theory," in Beall (2003)
Yi, Byeong-Uk. (1999). "Descending Chains and the Contextualist Approach to Semantic Paradoxes," *Notre Dame Journal of Formal Logic* 40: 554-567.
Yaqub, Aladdin M. (1993). *The Liar Speaks the Truth.* Oxford: Oxford University Press.
Young, J.O., (1995). *Global Anti-realism*, Aldershot: Avebury.
———. (2001). "A Defence of the Coherence Theory of Truth,' *The Journal of Philosophical Research*, 26: 89-101.
Zalta, Edward. (2001). "Fregean Senses, Modes of Presentation, and Concepts", *Philosophical Perspectives*, 15: 333-359
Zangwell, Nick. (1992). "Quietism," *Midwest Studies in Philosophy*
Ziff, Paul. (1960). *Semantic Analysis*. Ithaca: Cornell.
Zimmerman, Aaron. (2007). "Against Relativism," *Philosophical Studies* 133: 313-348.
Zwicky, Arnold and Sadock, Jerrold. (1975). "Ambiguity Tests and How to Fail Them," in *Syntax and Semantics 4*, Reimbald (ed.), New York: Academic Press.