

Brian Andrew Ball

Christ Church

Semantics, Meta-Semantics, and Ontology

A Critique of

The Method of Truth in Metaphysics

D. Phil. Thesis

Faculty of Philosophy

University of Oxford

Trinity Term 2008

Abstract

In this thesis, *Semantics, Meta-Semantics, and Ontology*, I provide *a critique of the method of truth in metaphysics*. Davidson has suggested that we can determine the metaphysical nature and structure of reality through semantic investigations. By contrast, I argue that it is not semantics, but meta-semantics, which reveals the metaphysically necessary and sufficient truth conditions of our claims. As a consequence I reject the Quinean (semantic) criterion of ontological commitment.

In *Part I*, chapter 1, I argue that the metaphysically primary truth bearers are not *propositions*, but rather *concrete representations*, either beliefs or sentences. I show, in chapter 2, that we can give sense to a truth predicate applying to sentences, given a truth operator and quantification into sentence position. I argue that this strategy does not commit us to the existence of propositions serving as truth bearers.

In *Part II* I argue that although we must assign semantic values to sentences and/or predicates, the meaningfulness of these expressions is not thereby explained. In chapter 3 I articulate Davidson's problem of predication and his solution, but argue that he was wrong to attribute this solution to Tarski. In chapter 4 I examine the semantics of modal languages; I conclude that although they require semantic values for predicates and/or sentences we should be instrumentalists about these theories.

In *Part III* I consider the relationship between truth and existence. In chapter 5, I defend Pluralism about truth: in some (though not all) domains of discourse, I claim, semantic reference plays a merely instrumental role in explaining truth. In chapter 6, I show that Hume's Principle, which is committed by the Quinean criterion to the existence of numbers, can be true even though numbers do not exist. In doing so, I appeal to meta-semantic and diachronic considerations.

In the conclusion I compare my views on ontology and commitment to Jody Azzouni's; and in the appendix I suggest how one might pursue diachronic linguistics.

Bibliographic Note

In providing citations I have used the convention of indicating the relevant publication in the text itself by providing the author's name, the date, and where applicable, the page numbers referenced; the full details can then be retrieved from the bibliography. In cases where I cite an article reprinted in a collection of essays, and where I have used that reprinting for my own research, I have tried to indicate the original source, but have given specific page references for the reprinting. It should therefore come as no surprise that the author's name and/or the date will sometimes change between the initial mention of the article and the specific reference. This having been mentioned, however, no confusion should arise.

Acknowledgements

I would like to express my thanks to my supervisors, past and present, for academic advice and support: to James Logue, who supervised my B.Phil., and gave me my first teaching assignment at Oxford; to Bill Brewer, who helped trim my D.Phil. project to a (nearly) manageable size; to John Hawthorne, for valuable comments on Part I; and to Dorothy Edgington, who saw this thesis through from start to finish, granting me enough free reign to pursue issues as I saw fit. Thanks to Stefano Predelli and Gabriel Uzquiano for their comments on the penultimate draft. Many thanks also to Timothy Williamson, who helped out at several crucial junctures, to Dan Isaacson, and to my college advisor, Hugh Rice. Thanks too to those at McGill University who encouraged me to pursue graduate studies, especially Brendan Gillon, Ian Gold, Michael Hallett, Storrs McCall, and Sarah Stroud. Thanks to the Walton (Star and) Crescent; to the members of the Third Man seminar; and to the graduate students, faculty, and staff at the Oxford Faculty of Philosophy.

I would also like to thank Christ Church, Marion Buck, and the Social Sciences and Humanities Research Council of Canada for financing this D.Phil. project.

Finally, I owe personal thanks to my parents, Donna McInnis and Kevin Ball, for sparking my interest in philosophy; to Jennifer Hennigar-Shuh for her confidence in me; to Jason Ball, for showing me the value of persistence in creative endeavour; and to Patrycja Kaszynska, who stood by me for the best part (and the worst parts!) of the last five years while this thesis was being written.

Thank you, friends!

Table of Contents

Introduction	1
I. Truth and Truth Bearers	10
1. Transcendence and Immanence.....	11
Frege's Question.....	11
Transcendence	17
Immanence.....	28
2. Higher-Order Quantification.....	45
Prior on Truth Predications	46
The Strategy of Reductive Elimination	52
The Direct Strategy.....	63
Summary and Conclusion	69
II. Truth and Meaning.....	70
3. Truth Theoretic Semantics.....	71
The Problem of Predication	72
Tarski's Truth Definition.....	76
Davidson's Solution: Truth Theoretic Semantics.....	81
Was Tarski a Davidsonian?.....	91
4. Model Theoretic Semantics.....	101
Introductory Remarks: Etchemendy on Model Theoretic Semantics	101
Representational Model Theoretic Semantics.....	103
Interpretational Model Theoretic Semantics	108
Semantics for Modal Languages (a): Kaplan's Atomism	112
Semantics for Modal Languages (b): Lewis' Molecularism.....	120
A Problem for MTS.....	130
Summary and Conclusion	136

III. Truth and Existence.....	138
5. Truth and Reference.....	139
Two Views of Truth and Reference.....	140
Davidsonian Truth Theory.....	141
Tarskian Truth Definition	144
Connecting Principles	146
The Inscrutability of Reference	148
Empirical vs Mathematical Domains.....	150
Truth as Correspondence and as Coherence	153
6. Hume’s Principle: A Case Study in Ontological Commitment	156
The Criterion of Ontological Commitment	158
Logical Form.....	164
Descriptions.....	172
Two Arguments for the Existence of Numbers.....	184
Frege’s Philosophy of Arithmetic and Neo-Fregeanism.....	187
Quantifier Variance	195
Diachronic Meta-Semantics: A Genealogy of the Concept of Number.....	200
Summary and Conclusion	209
Conclusion.....	211
Appendix: Language Stages	220
Bibliography.....	223

Introduction

Donald Davidson (1977a) has articulated and defended a philosophical methodology which he calls “the method of truth in metaphysics”. He initially motivates, and gives a partial description of, this method as follows:

In sharing a language, in whatever sense this is required for communication, we share a picture of the world that must, in its large features, be true. It follows that in making manifest the large features of our language, we make manifest the large features of reality. One way of pursuing metaphysics is therefore to study the general structure of our language. (Davidson, 1977a: 199)

We may think of this critical approach to philosophy as characterized by transcendental arguments¹ of the following form:

Method of Truth

(P1) Actually, x is true.

(P2) Necessarily, x is true only if p .

Therefore, (C) Actually, p .²

What distinguishes method of truth arguments from other transcendental arguments is the semantic ascent: in the first premise the semantic property of truth is explicitly mentioned, and the second premise is then established through linguistic, and in particular semantic, analysis. Indeed, Davidson says, “the study of truth conditions is the province of semantics.” (Davidson, 1977a: 201)

¹ Transcendental arguments have the form: *Actually P, Necessarily, if P then Q, therefore, Actually Q.*

² Davidson himself would likely not describe his method in this way, sceptical as he is of modality; but given that the claim that x is true only if p is established as a kind of empirical (semantic) *law* (see below), it seems that this claim is as close as Davidson gets to a necessity claim – and so, for those (like myself) who allow modal discourse, this seems a fair interpretation.

This claim articulating a privileged role for semantics within the method of truth is immediately preceded by an account of the manner of justification of first premises in method of truth arguments: Davidson says, “The study of what sentences are true is in general the work of the various sciences.” (Davidson, 1977a: 201) The idea, I suppose, is that physics tells us what is the case physically, and chemistry tells us how things are in the chemical realm, and so on. As a result, we can then, through semantic ascent, arrive at the desired first premise, which interacts with the second, semantic premise to yield the metaphysical conclusion. This sounds like a legitimate description of at least some reasoning which involves the method of truth. For example, the famous Quine-Putnam indispensability argument for the existence of mathematical objects, an instance or variant of which will be the subject of the latter half of the final chapter of this thesis, might be thought to proceed in this way: mathematics delivers the mathematical facts; then semantics tells us what must be the case for those facts to obtain...³ However, Davidson’s account does not seem to be an accurate description of *all* the applications of the method of truth in the literature; in particular it does not cover Davidson’s own (1967a) argument to the effect that events exist,⁴ which may be represented more or less as follows:

Davidson’s Argument for the Existence of Events

(DP1) Actually, “Shem kicked Shaun” is true.

(DP2) Necessarily, “Shem kicked Shaun” is true only if events exist.

Therefore, (C) Events exist.

³ However, this description ignores the holistic account of confirmation appealed to in the indispensability argument; that is, this argument assumes that the various sciences are not entirely autonomous, but are rather interdependent. Confirmation holism is a large can of worms, however, and I will not open it here.

⁴ Or at least, it does not do so if we take the “various sciences” to be the *special* sciences. See below.

Clearly no special science establishes (DP1) – unless we count a very loosely individuated criminology! The point is that in this particular case our grounds for taking the first premise of the argument to be true are quite ordinary. As I understand it, the Davidsonian method of truth in metaphysics often involves such appeals to everyday means of recognizing truths – though it does not exclude recourse to the special sciences.

There is a particular kind of method of truth argument that I want to focus on in which the sentence letter “ p ” (in the schematic representation of the argument form above) is replaced by an existence claim. Indeed, both of the applications of the method of truth that I have mentioned are of this kind: they have ontological conclusions. In order to succeed, such arguments must invoke some semantico-metaphysical principle serving to establish the second premise linking truth and existence. The principle of choice is invariably Quine’s criterion of ontological commitment – a speaker is committed to the existence of F s if what he says can’t, by semantic principles, be true unless the range of the variables of his language includes some F s.

Although the criterion of commitment typically invoked is Quine’s, the standard manner in which it is employed is not. To see this, it may be worth, at this stage, drawing attention to a distinction between two approaches to the philosophy of language. One might study language simply because it is an interesting phenomenon in its own right which stands in need of explanation. Philosophers of this bent will ask such questions as, What are we doing when we produce noises in each other’s vicinity? and, How do we thereby succeed in communicating information to one another? We can call this approach to the philosophy of language “descriptive”. By contrast, we might be interested in language because, as philosophers we trade in ideas. In order to successfully convey our ideas to others, however, we must (in Paul McCartney’s words) indicate precisely what we mean to say; and this can be achieved

if we lay down some strict rules governing our use of linguistic symbols, and then agree, as it were, to talk only in conformity with those rules. Such an approach to the study of language might be called “prescriptive”.

Quine did engage in descriptive philosophy of language; however, as regards the use of his criterion of ontological commitment, he was a prescriptivist. Like Frege, he repudiated natural language as hopelessly vague and imprecise for the purposes of science; he accordingly advocated the use of a tightly regimented replacement language. Thus, on his own view, the criterion of ontological commitment is to be applied only to theories written in an artificially regimented, first-order language. There are, of course, questions that can be raised about Quine’s choice of first-order notation as the language of science: for example, Why not a second-order or modal language instead? I will not address these issues here, however. Quine crucially did not think that we can apply his criterion to our ordinary talk: natural language speakers do not, he thought, undertake (clear) ontological commitments. By contrast, some of Quine’s followers (e.g. Davidson) have thought that it is precisely by making assertions in natural languages like English that we commit ourselves to the existence of entities of one kind or another.

There are a number of reasons for taking greater interest in this latter approach to ontology than in the original Quinean method. For one thing, Quine only refrained from applying his criterion of ontological commitment to natural language assertions out of skepticism: he thought that the semantic properties of natural languages are, for a number of reasons, unclear. However, this skepticism is not shared by contemporary descriptive philosophers of language: the majority of us feel that it should be in principle possible to provide a precise theory of meaning systematically generating truth conditions for natural language sentence tokens. For another thing, since natural languages contain first-order, second-order, modal, etc. fragments, if we employ the method of truth with respect to natural languages, our

conclusions will hold good for whichever restricted part of natural language we take our artificial language of science to reproduce or replace.⁵ Thus, what we might call the “Davidsonian” method is stronger. Finally, we are (collectively) more certain that various sentences of natural language are true than that particular first-order sentences are – not least because more of us understand what, say, English sentences mean. Consequently, the ontological and metaphysical conclusions drawn from the application of the Davidsonian method of truth are, at least *prima facie*, on better epistemic standing than those obtained by the Quinean methodology – they are more intuitive.

Despite this, I want to object to these ontological applications of the Davidsonian method of truth in metaphysics. Indeed, the entire thesis which follows should be regarded as an extended argument against the soundness of this method. In particular, I reject the Quinean criterion of ontological commitment. The dictum, “to be is to be the value of a variable” (Quine, 1961: 15), proposed in 1948, is now not only familiar, but verges on orthodoxy in analytic philosophy. As we have seen, the proposal is meant to give a criterion of ontological commitment and not of existence – being described by scientists is not a necessary condition on being. My aim in this thesis, however, is to challenge the converse claim that being the value of a variable in a true theory is sufficient for existence; and, to a lesser extent, to offer an alternative way of thinking about what there is.

Robert Stalnaker (2003) has drawn a distinction between descriptive semantics on the one hand, and what he calls “meta-semantics” on the other. Roughly, the idea is that the former is a discipline concerned to yield adequate predictions of the meaning behaviour of linguistic expressions, while the latter engages the underlying explanatory questions. Crudely speaking, my complaint about ontological applications

⁵ Indeed, in what follows there are discussions of the semantics of each of these types of language fragments.

of the method of truth is that they leave out of account the meta-semantic principles governing natural language expressions. Slightly less crudely, my strategy will be to argue that there are sentences of natural languages which (i) are true, (ii) have no semantically reductive logical forms, and which, therefore, (iii) are Quinean-committed to the existence of objects of a certain kind,⁶ and yet which (iv) can be true despite the non-existence of that kind of entity. I then give a meta-semantic explanation of these facts.

Let me lay out the plan of the thesis in some detail. In *Part I: Truth and Truth Bearers*, I address Frege's Question, viz., What are the primary truth bearers? In chapter 1, "Transcendence and Immanence", I broach this issue, and distinguish two kinds of primacy: conceptual and metaphysical. I argue that the metaphysically primary truth bearers are concrete representations, and not abstract representational contents, i.e. propositions. This is not only interesting in its own right: it is also important in the context of a discussion of the method of truth since, as we have seen, the first premise in the arguments characteristic of this method involve the claim that a given thing is true; surely, therefore, we should attempt to discover of what *kind* of thing truth is a property. Moreover, some who maintain that propositions are the primary truth bearers might hope to pull themselves up by their own bootstraps (in the form of this assumption) to certain further ontological conclusions.

As an example of this bootstrapping strategy, consider Timothy Williamson's (2002) argument to the effect that everything exists necessarily. Williamson claims that any instance of the schema

- (1) Necessarily, if *a* does not exist then the proposition that *a* does not exist is true

is itself true; and similarly for the schemas

⁶ To be "Quinean-committed" is to be committed by the Quinean criterion of ontological commitment.

(2) Necessarily, if the proposition that a does not exist is true then the proposition that a does not exist exists; and

(3) Necessarily, if the proposition that a does not exist exists, then a exists.

From these he concludes that we must endorse the schema

(4) Necessarily, if a does not exist, then a exists;

and hence also the schema

(5) Necessarily, a exists.

Ignoring the fact that any instance of this conclusion is itself (highly!) contentious, we may remark that each of these five schemas begins with an operator expressing alethic, metaphysical necessity; hence Williamson is committed to the actual truth of each instance of the embedded schemas. So it seems that Williamson will accept the conclusions that, amongst other things, Sherlock Holmes and Pegasus (actually) exist. But these, it seems, are ontological commitments we might hope to avoid.

In the second chapter of *Part I*, “Higher-Order Quantification”, I show that, given a truth operator, together with quantification into sentence position, we can explain the meaning of a truth predicate which applies to sentence tokens. This predicate is, of course, useful for the project of giving truth conditional semantic theories for natural languages. The main work of chapter 2, however, is in showing that quantification into sentence position does not itself commit us to the existence of propositions, which would undermine the conclusion of chapter 1, that it is representations, and not representational contents, that are the metaphysically primary truth bearers. My strongest argument here is based on cardinality considerations: if there were such abstract objects as propositions, there would be too many objects! So there aren’t any such objects – and this despite the intelligibility of higher-order quantification.

In *Part II: Truth and Meaning*, I address the question what form our semantic theory predicting the truth and falsity of natural language sentence tokens should

take. In particular, I attempt to assess the question whether expressions other than (first-order) terms should be assigned entities as meanings, or semantic values. I begin, in “Truth Theoretic Semantics”, with a discussion of Davidson’s “problem of predication”: this enables me to present the truth theoretic approach to semantics, and explain how it avoids the postulation of semantic values for predicates and sentences. Towards the end of this chapter I compare Davidsonian truth theories with Tarskian ones. In the course of this discussion I introduce what will prove to be a crucial distinction; that between semantics and meta-semantics.

Chapter 4, “Model Theoretic Semantics” investigates the prospects of an inference to the best explanation of the semantic phenomena whose conclusion is that predicates and/or sentences really do have entities as meanings after all. I begin by considering Etchemendy’s discussion of the interpretation of model theoretic semantics, in the hope of discovering how such an inference might go. I then examine two model theories for modal languages: I conclude, somewhat anti-climactically, that although our semantic theories must postulate semantic values for predicates and/or sentences, nevertheless these entities are not metaphysically explanatory of the meanings of these expressions.

Part III: Truth and Existence is where I finally tackle the issue of the relationship between semantics and metaphysics, and show the failure of the Davidsonian method of truth. In chapter 5, “Truth and Reference”, I defend what Crispin Wright (1992) has called “pluralism” about truth. I argue that there are domains of discourse in which reference plays a genuine, metaphysical role in explaining truth – that is, where truth consists in correspondence to reality – and domains (including mathematics) in which it does not, and truth is mere coherence.

In the final chapter of the thesis, “Hume’s Principle: A Case Study in Ontological Commitment”, I show why it is that the Quinean approach to ontology fails. I begin by articulating the Quinean condition of commitment. I then show that

simple sentences involving definite descriptions of the form ‘the F ’ are committed by the Quinean condition to the existence of an F ; hence, Hume’s Principle itself, which says that, for any F and G , the number of F s is equal to the number of G s if and only if the F s and the G s can be put into one-to-one correspondence, is committed by Quine’s condition to the existence of numbers. I then articulate a metaphysically invariantist form of contextualism about “ontological” claims. Finally, I give a meta-semantic account of the truth of Hume’s Principle which does not itself mention numbers. I conclude that Hume’s Principle is true, but that numbers do not exist, thereby refuting the Quinean condition of ontological commitment.

I. Truth and Truth Bearers

This first part of the thesis, *Truth and Truth Bearers*, is divided into two chapters. In the first chapter, “Transcendence and Immanence”, I attempt to answer Frege’s Question, viz., What are the metaphysically primary truth bearers? I begin, in an introductory section, by distinguishing metaphysical primacy from conceptual primacy in order to properly frame this question; I then suggest two alternative answers, which I call Transcendence and Immanence, and announce my intention to defend Immanence. In the next section I examine Frege’s three arguments aiming to establish the truth of Transcendence, and provide replies to them on behalf of the defender of Immanence. In the final section of the first part I provide a positive argument in favour of Immanence. Thus I establish that it is concrete representations, rather than abstract representational contents, that are the metaphysically primary truth bearers.

The second chapter of this first part is called “Higher-Order Quantification”. In it I show how we can account for a use of the predicate “true” as applied to sentence tokens given a truth operator and quantification into sentence position. I argue that quantification into sentence position does not commit us to the existence of propositions. I conclude by suggesting that we adopt semantic instrumentalism.

1. Transcendence and Immanence

Frege's Question

Truth plays a central role in our daily economy. We tell the truth, and are told it; that is, we communicate information to one another. We often conceal the truth, perhaps in an attempt to avoid a perceived harm, whether real or emotional. Sometimes we demand the truth from those who conceal it from us; and if we are lucky they reveal it to us. We plan our actions on the basis of what we believe to be the truth: sometimes our efforts are thwarted because we fail to believe the truth (whether by believing a falsehood, or by failing to have considered a given question); if it is not too late, we then redouble our efforts to discover the truth.⁷ We praise those who pursue the truth – scientists, for example, and journalists – especially when they do so at personal cost or risk;⁸ and we condemn those who are unduly negligent in this respect. In short, we value truth.⁹

Truth is, therefore, a notion of fundamental philosophical concern. Philosophy is the love of knowledge and, of course, knowledge is factive; one cannot know something without it being the case that what one knows is true. In this respect philosophy is like other “scientific” disciplines (*wissenschaften*) – its goal is to attain the truth. But more than this, truth is an object, as well as an aim, of philosophical inquiry. Thus, as philosophers we ask, Is there objective truth? If so, what is its nature? Can it be known?

⁷ In 2005, the British pop band Oasis released a record called “Don’t Believe the Truth”. I can’t imagine worse advice, not only from an epistemological point of view, but also from a practical one.

⁸ The recent murder of Anna Politkovskaya, for instance, was cause for mourning (even amongst those unknown to her).

⁹ Bernard Williams (2002) provides an extended discussion of the values associated with truth, or as he puts it, the values of truthfulness. He distinguishes two such values, namely those of accuracy and of sincerity; and he gives a (partly fictional) genealogy of them.

Since Frege, however, philosophers are accustomed to the idea that if we are to pursue such investigations, we must first answer the question, what sort of thing can be true or false? In his paper “Thoughts”, Frege wrote:

Grammatically, the word “true” looks like a word for a property. So we want to delimit... the region within which truth can be predicated, the region in which there is any question of truth. (Salmon and Soames, 1988: 34)

This statement of the issue unfortunately appears to involve a conflation of use and mention. Moreover, attempts to eliminate this confusion reveal an ambiguity. On the one hand, the fact that the abstract noun “truth” is used suggests that Frege’s question is essentially the following: (1) What kind of thing can be true? Alternatively (if we are willing to set aside nominalistic scruples¹⁰), What kind of thing can have, or instantiate, the property *truth*? On the other hand, the use of the verb “predicated” suggests that the word “true” ought to have been *mentioned*; for surely it is predicates, i.e. linguistic expressions of a certain sort, which are predicated. If so, then we may interpret Frege as asking (2) What kind of thing can be properly called “true”? Or, using a verb of indirect discourse, What kind of thing can be correctly said to be true? This second question, which concerns not the particularities of the expressions we employ, but the nature of our linguistic practices involving the word “true” and its cognates, is a question about speech acts. The first question, by contrast, is metaphysical in nature. Although I am principally concerned with the metaphysics of truth, and hence question (1), I will approach the answer to this question via speech act considerations, and question (2) in particular.¹¹

¹⁰ This is something I attempt to do, and for the most part I succeed. Such scruples do, however, occasionally provoke comments in the text that follows.

¹¹ See Richard Kirkham (1992) for a three-way distinction between the metaphysical project, the speech act project, and the justification project, as concerns theories of truth. I am not concerned here with the justification project at all.

In general, if we ask, ‘What kind of thing can properly be said to be F ?’ we may find that there is more than one correct answer.¹² In such cases we can often recognize a certain structure of dependency amongst the various predications; it is clear that some correct uses of a given predicate are more basic than others. When this is the case, there will be an ordering, perhaps partial, on the kinds of things that are correctly said to be F , determined by the relative priority of the kind of saying. Members of any kind which is maximal in such an ordering will be said to be *conceptually primary* F s; and if the G s alone are maximal in an ordering of this type, then the G s will be said to be *the conceptually primary* F s.¹³

If we do recognize that a number of distinct kinds of things may be properly called ‘ F ’ we may want to disquote and say that many kinds of things *are* F .¹⁴ Yet we may also discern a structure of dependency amongst the kinds of things which are F which is not itself to be explained in terms of *conceptual* primacy; rather, the dependency may be one of primacy in the order of metaphysical explanation. If so, there will be another, possibly distinct, possibly partial, ordering on the kinds of F s in terms of this metaphysical primacy. Members of maximally basic kinds will be said to be *metaphysically primary* F s; and if there is a unique kind G of such F s, the G s will be said to be *the metaphysically primary* F s.

¹² Indeed, this may be so for different reasons. If we ask, “What kinds of things may be properly said to be alive?” we might answer (truly) either “Dogs” or “Animals”. On the other hand, if we ask, “What kinds of things may be properly said to be stupid?” we might (arguably) answer correctly either “People” or “Questions” (school teachers who say “There are no stupid questions” notwithstanding). I am not concerned with different correct answers which display the first kind of relationship to one another of inclusion, but rather with the second kind, where there is independence between the kinds in question.

¹³ When a structure of dependency amongst sayings of the kind at issue exists, we will typically be inclined to say that the term in question is used in slightly different senses, or with slightly different meanings, in the different cases (recall the notion of stupidity from footnote 12). Nevertheless, we should be disinclined to say in these cases that the term is lexically *ambiguous* – for the distinct uses of ambiguous terms are *independent*.

¹⁴ Certainly the nominalist will want to say this (barring special considerations) – for he maintains that there is nothing over and above their being called ‘ F ’ which the F s have in common.

Let me illustrate these notions with an example. Consider the question, “What kind of thing can be properly said to be a reason for action?” We might examine the linguistic evidence and discover that we say, perfectly felicitously, of various different sorts of things that they are reasons. Thus, if we ask, “What was Smeagle’s reason for killing his brother?” we might say that the ring of power was his reason; but we might also say that the fact that he couldn’t have the ring without killing his brother was his reason. Since both of these are acceptable things to say, we should conclude that amongst the things which can be properly said to be reasons for action are facts and objects. Nevertheless, we should be willing to recognize that some of these uses of the term “reason” are more basic than others. In this case, it seems clear that reasons are given in a more fundamental manner when presented using a “that” clause, rather than a simple singular noun; facts are conceptually more basic reasons than objects.

On the other hand, there is a debate between so-called “internalists” and “externalists” about reasons. One way of explaining this debate is that it concerns whether or not a subject *S* must believe a proposition in order for it to count as a reason for *S*.¹⁵ The internalist may be motivated by *naturalism*, and maintain that reasons are the springs of action. But if reasons are causally efficacious, they cannot be abstract. Thus, although the internalist will agree that only entities which are typically specified using a “that” clause are reasons, nevertheless, he will claim that metaphysically speaking, it is beliefs themselves which are the basic reasons. By contrast, the externalist may be motivated by *rationalism*. If so, she will argue that reasons must be capable of engaging with a subject’s rational faculties, and so be propositional in form; yet, she will also maintain that they need not *actually* engage with a given subject’s beliefs and desires. Indeed, it is precisely this fact which makes

¹⁵ A slightly different debate between internalists and externalists concerns whether propositions must be true in order to count as reasons – that is, whether it is only facts which can be reasons. Famously, Bernard Williams is an internalist in the sense described in the main body of the text, while Derek Parfit is an externalist in that sense; I do not know what their views on this second issue are.

room for the thought that reasons are normative, and which explains why it is that people sometimes fail to do what they should. For such a philosopher it will seem obvious that the metaphysically primary reasons are also the conceptually primary reasons, viz. abstract propositions. I will not take sides on this debate here, but rest content at remarking that it can be construed as concerning what the metaphysically primary reasons are, *despite agreement about what the conceptually primary reasons are*.

Let's return to the case at hand. We have so far been concerned with Frege's question, which we saw admits of two readings. A number of alternative answers have been given to this ambiguous question, including: (i) propositions – that is, certain “abstract objects representing truth conditions” (Stalnaker, 1999: 32); (ii) beliefs, and other propositional attitudes – i.e., particular mental states of individuals, which are causally explanatory of their behaviour; (iii) sentence tokens – certain noises produced by speakers at particular times and in particular circumstances¹⁶; and there may well be others besides.¹⁷ Moreover, each of these positions has been developed in myriad ways. This diversity of views raises the possibility that different authors have interpreted Frege differently; but it also suggests that there may be more than one correct answer to any given reading of Frege's question – more than one kind of thing for which the question of truth (non-trivially) arises. Thus, I propose to

¹⁶ I here ignore written sentence tokens. Their inclusion would complicate matters (see e.g. MacFarlane (2005)), but not, I believe, irremediably; so my decision to exclude them from consideration is simply the decision not to unduly muddy the waters.

¹⁷ E. J. Lowe acknowledges, in addition to these, the category of statements, where “[a] *statement* is the assertoric use of a sentence by a speaker on a particular occasion” (Lowe, 1995: 881) – that is, a particular speech act. However, John MacFarlane (2005) argues, convincingly to my mind, that these speech acts (which he calls “utterances”) are more properly classified as correct or incorrect, not true or false; and so I will not consider this option further. On the other hand, Quine (1960) famously held that (“eternal”) sentence types are the primary truth bearers. As it turns out, however, I will argue that propositions are nothing other than such types of interpreted sentence – hence this view is covered under option (i).

pursue the distinct, but related, question, What are the *metaphysically primary* truth bearers (MPTBs)?

Despite the plethora of answers to Frege's question, we recognize, following E. J. Lowe, that

a broad division can be drawn between those theories of truth which regard truth as a property of *representations* of some sort (whether linguistic or mental)... and those which regard truth as a property of *propositions*, conceived as items *represented* or *expressed* in thought or speech. (Lowe, 1995: 881)

Indeed, we may, in this manner, distinguish two views concerning the metaphysically primary truth bearers, namely:

(*Immanence*) Particular, concrete representations of some sort (whether mental or linguistic) are the metaphysically primary truth bearers (MPTBs). If they were not true or false, nothing would be.

and its complement:

(*Transcendence*) Representational contents, or propositions, are the metaphysically primary truth bearers (MPTBs). That is, certain abstract objects with intrinsic truth conditions bear the twin properties of truth and falsity; and if they did not, nothing else would.

In what follows, I defend Immanence.

My defence of Immanence is in two parts. I begin by showing that Frege's arguments in favour of Transcendence are not compelling. Then, I argue that propositions, if they exist, depend for their existence on that of certain representations. The upshot is that we have no reason to accept Transcendence, and some reason to doubt it. In the second part of this chapter, I further support Immanence by showing how to give sense to attributions of truth and falsity to attitudes and utterances, given an antecedent use of the word "true" as part of a sentential operator. This account appeals to quantification into sentence position:

accordingly I argue that such quantification does not commit us ontologically to the existence of propositions. It should be clear from my strategy that while I believe representations are the metaphysically primary truth bearers, I take propositions to be truth bearers in a conceptually basic sense. It is no doubt this fact which has erroneously prompted many philosophers to defend Transcendence.

Transcendence

Frege was a proponent of Transcendence; though his terminology was not ours.

He wrote:

Without offering this as a definition, I mean by “a thought” something for which the question of truth can arise at all. (Salmon and Soames, 1988: 36)

Frege’s conception of a thought, as is well known, was not that of a mental particular, “belonging” to an individual thinker; rather, it was that of a publicly accessible abstract object - an inhabitant of a “third realm”, distinct from both the “inner arena” of sense-experiences and the “outer world” of material objects. So Frege held that abstracta of a particular kind, i.e. “thoughts”, or propositions, are truth bearers. Moreover, he argued that anything else that is called “true”, is either true in some different and irrelevant sense, or else is true derivatively; so on his view propositions are the MPTBs. In this section I examine Frege’s arguments for this claim.

Despite thinking it “likely that the content of the word ‘true’ is *sui generis* and indefinable,” (Salmon and Soames, 1988: 35) so that “[t]he meaning of the word ‘true’ is spelled out in the [logical] laws of truth” (Salmon and Soames, 1988: 34) and only in this way, nevertheless, early on in “The Thought” Frege says:

I shall attempt to outline roughly how I want to use ‘true’... so as to exclude irrelevant uses of the word.... What I have in mind is that sort of truth which it is the aim of science to discern. (Salmon and Soames, 1988: 34)

It is then that he raises our question concerning the bearers of this property; and he remarks that “[w]e find truth predicated of pictures, ideas, sentences, and thoughts.” (Salmon and Soames, 1988: 34) He then proceeds to argue that it is thoughts which primarily admit of truth in this intended sense, and sentences (i.e. utterances, or sentence *tokens*) only secondarily, while pictures and ideas do not admit of truth at all in the relevant sense.¹⁸

So we may liken Frege’s strategy to that of Sherlock Holmes: he lists the alternative suspects, then eliminates those which are impossible; whatever remains, however improbable, must be the culprit – in this case, the metaphysically primary truth bearers. The arguments against pictures and ideas as truth-as-the-aim-of-science bearers are exactly parallel. There are two of them occurring more or less explicitly in Frege’s paper, and they run as follows:

The Argument from Absoluteness

(P1.1) Truth, as it applies to pictures and ideas, is (a form of) correspondence.

(P1.2) Correspondence admits of degrees.

(P1.3) Truth (the aim of science) does not admit of degrees.

Therefore, (C1) Truth (the aim of science) does not attach to pictures or ideas.

The Argument from Monadicity

(P2.1) Truth, as it applies to pictures and ideas, is (a form of) correspondence.

(P2.2) Correspondence is a relation.

¹⁸ Frege (or his translator) seems to use the word “idea” in the sense of the British Empiricists, as meaning something like “mental image” – the kind of thing that occurs, for example, in dreams. The Empiricists seem to have thought that such things could constitute beliefs, and other propositional attitudes – Hume, for instance, held that beliefs are simply “vivid” ideas – while we, largely due to Frege, do not. Frege’s discussion, therefore, simply omits consideration of the possibility that it is beliefs (as distinct from ideas) which are true and false. I will not make anything of this in what follows, but will focus on the case of linguistic representations.

(P2.3) Truth (the aim of science) is a property, not a relation.

Therefore, (C2) Truth (the aim of science) does not attach to pictures or ideas.

My aim here is not to challenge the soundness of these arguments – I am not, after all, defending either the view that pictures, or the view that ideas, are the MPTBs. However, I am defending Immanence, and so it is of some interest to discover whether these arguments can be extended to cover not only pictures and ideas, but also sentences.¹⁹ To ascertain this, it will be worthwhile considering their credentials as they stand; in particular, I will consider the case of pictures.

If we are willing to idealize somewhat, there is, I think, a clear sense in which a picture can be said to correspond to a three-dimensional array of things. In a three-dimensional space, a plane P and a point o together determine a projection scheme S . Take any point x on the plane, and draw a line $\ell(x)$ through it and the designated origin point o . Then we may say that the point x *corresponds* under S to point y if, and only if, y is the first point of intersection of $\ell(x)$ and a solid object (on that segment of $\ell(x)$ beginning with o and continuing away from the plane P). Furthermore, we may say that a picture *corresponds* under S to an array of objects just in case each point in the picture is the colour of the point it corresponds to under S .

Clearly, correspondence (under a scheme) so defined – of a picture to an array of objects – is a relation. Moreover, even abstracting away from the fact that colours can match in varying degrees of hue and intensity, it is clear that such correspondence admits of degrees: a picture corresponds to an array of objects better or worse, more or less, as its points match those they correspond to (under the scheme)

¹⁹ I here ignore the possibility that beliefs, or other mental representations, might escape unscathed from these arguments. This is partly because Frege doesn't himself raise the possibility, and partly because as a philosopher of language I am more interested in linguistic representations than mental ones. But notice that ignoring this possibility simply makes my job of defending Immanence more difficult; it is therefore not dialectically inappropriate.

in greater or lesser proportion.²⁰ So premises (P1.2) and (P2.2) are both plausibly true. Moreover, premises (P1.1) and (P2.1) are identical; and the claim they make seems true – at least in the case of pictures, which is under discussion.²¹ What about the major premises of these arguments, Absoluteness (P1.3) and Monadicity (P2.3) themselves?

Take the Argument from Absoluteness first. There are, perhaps, those who would challenge the claim that truth does not admit of degrees - i.e. the major premise (P1.3) of Absoluteness. Consider, for example, the case of baldness. Let's assume that whether someone is, or is not, bald is a matter of the density of hair on his or her head. Some densities of hair on the head are clearly sufficiently low for baldness. Others are clearly sufficiently high for non-baldness. Perhaps, then, the degree of truth of the sentence "He's bald" varies continuously with the density of hair on the head of the guy demonstrated.

Susan Haack (1978) opposes the idea that truth admits of degrees. She suggests that whereas many predicates expressing properties which admit of degrees take adverbial degree modifiers, "true" does not. While some adverbs can be acceptably concatenated with "true", Haack says, it is not clear that when they do so they modify for degree. She writes:

Among the adverbial modifiers which *do* apply to 'true' one has 'quite' and 'very'. Now 'quite' and 'very' apply to predicates of degree, i.e. predicates which denote properties which come in degrees (quite tall, heavy, intelligent..., very tall, heavy, intelligent...) where they indicate possession of the property in, respectively, modest or considerable degree.... But whereas 'quite tall (heavy, intelligent)' can

²⁰ There are of course issues here surrounding measures on infinite sets; but the solution to them will presumably be sufficiently general to cover this case.

²¹ The claim that ideas are true only in the sense that they correspond to aspects of reality is, of course, questionable when read as saying that ideas are true, but they are true only in this sense. The worry with this is due to Berkeley, who famously pointed out that only an idea can resemble an idea; Frege raises it at (Salmon and Soames, 1988: 35). But the original claim might be true if read as saying that ideas are true in the sense that they correspond, *if they are true at all*.

be roughly equated with ‘rather (fairly) tall (heavy, intelligent)’, ‘quite true’ certainly doesn’t mean anything like ‘rather true’ or ‘fairly true’.... In fact, ‘quite true’ can be roughly equated with ‘perfectly true’ or ‘absolutely true’, and (so far from contrasting with it) ‘very true’. (Haack, 1978: 168)

Haack concludes that the linguistic evidence doesn’t support the hypothesis that truth comes in degrees in the way that, say, height does; rather the data favour our Absoluteness.

There is, however, further linguistic evidence which Haack does not consider. For example, suppose you and a friend were at an event together. Now, in the presence of a third party, your friend recounts what happened. When he is done you might say, “Well, that’s *partly* true. Actually, what happened was...”. Or again, suppose that a jury listens to a number of key witnesses’ accounts of the events on a certain day - events that are crucial to the outcome of the case. When later the jury members are deliberating about the verdict, one of them might say, “You know, A’s account of those events strikes me as truer than B’s”; and while this may not sound great to the grammatical purist, nevertheless it is clear enough what is meant. In each of these cases the word “true” appears to be modified for degree.

However, it is possible to explain these cases while maintaining that truth itself does not admit of degrees, as follows. In cases such as these, what is said to be partly true, or claimed to be truer than something else, is a *collection* of claims. None of the particular claims in the collection can be partly true, or more or less true than any other; but the collections can be described in these ways because, in the first case, only some of the claims in the collection are (absolutely) true, and in the second case, a greater proportion of the claims in one collection are (absolutely) true than in the

other. Thus, Absoluteness seems plausible, even if not apodictic; and for our purposes I grant Frege this claim.²²

Let's turn to the Argument from Monadicity; for its major premise, (P2.3), is of much greater interest to our central concerns. According to Frege, truth (that is, the aim of science) is a property, not a relation. This should strike us as a controversial claim: after all, according to a very familiar philosophical position, truth *is* a relation; indeed, it is the very relation Frege cites in the Argument from Monadicity, namely *correspondence*. In particular, according to the Traditional Correspondence Theory, truth is a relation between an object of an appropriate sort and a fact; a truth bearer is true iff there is a fact that it corresponds to.

There are, however, good reasons to reject the Traditional Correspondence Theory. Strawson (1950) argued that facts are not, *pace* Wittgenstein in the *Tractatus* (Wittgenstein, Pears et al., 1974), parts of the world – they don't *exist*. And if there are no facts, then there's nothing for truths to correspond to. So, short of endorsing a universal error theory – nothing we ever say or think is true – we must abandon the claim that truth is correspondence to a fact.²³ Similarly, Davidson argued in his (1969) not that there are *no* facts, but rather that there is *at most one*; consequently, to say that some truth bearer corresponds to the facts is to say no more than that it is true. The Traditional Correspondence Theory is empty, and is therefore best abandoned.²⁴

²² It is perhaps worth noting that Quine's confirmation holism might be viewed as the claim that it is collections of claims (i.e. theories), and not claims themselves (i.e. sentences), which admit primarily of truth and falsity. But this view has trouble explaining the data discussed above: why should it be that individual claims can't be said to be, e.g., *partly true* on this view? Of course, one might be tempted to object that the Fregean cannot gain any argumentative purchase from the fact that the explanation he gives of this phenomenon seem successful, since it presupposes that truth attaches primarily to claims, rather than something else – and this is just what he's arguing *for*. Against this I simply point out that we are here concerned with the truth or falsity of Absoluteness; any argument which bears on this issue will therefore be accepted as worthy of consideration on its own merits.

²³ Perhaps I should say that for Strawson, facts are not independent of truths; hence there's nothing for truths to correspond to. The net effect is the same. See below.

²⁴ Of course, there are those who disagree; my present aim is simply to motivate the search for an alternative theory of truth.

Now, Davidson's argument invokes a principle of extensionality; whereas in the current philosophical climate people are very happy to employ intensional idioms. Thus, one might doubt the Davidsonian result. Indeed, a popular contemporary view of facts is that they are true propositions; and, it is maintained, there are many of these, not one as Davidson claimed, or none as Strawson would have us believe. But even if these theorists are right on this issue, the view they endorse – according to which propositions are the truth bearers – does not provide a relation between truths and facts that correspondence could be.²⁵ The Traditional Correspondence Theory must go.

Davidson (1969) does in the end, however, consider himself to be vindicating what was right about the Correspondence Theory; for he maintains that a sentence (a Davidsonian truth bearer) is true by virtue of the relations obtaining between linguistic items and aspects of the world. But on Davidson's New Correspondence Theory, truth itself is not such a relation; rather it is a relational property. That is, "true" is a monadic predicate which applies, when it does, because some other relation (in particular, the relation of satisfaction) holds between the sentence and some part of the world.²⁶ Somewhat more crudely, a sentence is true, for Davidson, just in case its parts bear some appropriate relation to parts of the world (and to one another). Truth is therefore a relational property, but not a relation, and Monadicity is true.

However, this raises problems for the Argument from Monadicity; for once we recognize that truth is a relational property²⁷ it should be clear that Frege equivocates.

²⁵ One might be tempted to claim that the relevant relation is identity – but this fails to distinguish truths from falsehoods.

²⁶ Or indeed, because satisfaction holds between the sentence and all parts of the world of a given kind – namely all sequences of objects. I discuss Davidson's approach to truth and its relation to satisfaction (or reference) in greater detail in Part III.

²⁷ As indeed it was claimed to be even by traditional correspondence theorists – according to which, recall, a truth bearer is true iff *there is a fact* that it corresponds to.

In the first premise Frege claims that truth is correspondence; and this is true, provided we interpret “correspondence” as (the existentially closed) *correspondence-to-something*. But in the second premise, he says that correspondence is a relation; and the truth of this premise precludes giving “correspondence” the very same interpretation. There are interpretations which make *each* of the premises true, but there is no interpretation which makes *all* of the premises true; thus, the conclusion does not follow.

Can the same be said for the Argument from Absoluteness? Must correspondence be interpreted relationally in order for the claim that correspondence admits of degrees to be true? No. As we have seen, a picture can correspond to an array of objects in various degrees; and here the correspondence relation is “filled” on the array end. Thus this argument can be used to establish that pictures and ideas are not true (in the relevant sense). Yet we have seen reasons to think that the correspondence in terms of which sentence truth is to be understood does not admit of degrees (since sentence truth itself does not); so a modified Argument from Absoluteness does not establish that sentence tokens cannot be true (in the relevant sense).

There is, however, a third argument latent in Frege’s paper which I think he intends should cover *all* the listed alternatives to his own view. It runs as follows:

The Argument from Intrinsicity

(P3.1) A picture/idea/sentence would not be true unless accompanied by an intention.

(P3.2) Anything requiring an accompanying intention in order to be true or false is not intrinsically truth apt.

(P3.3) The primary truth bearers are intrinsically truth apt.

Therefore, (C3) Truth (the aim of science) is not primarily a property of pictures/ideas/sentences.

Frege never explicitly offers this third argument; but here is what he does say:

Is a picture considered as a mere visible and tangible thing really true, and a stone or a leaf not true? Obviously we could not call a picture true unless there were an intention involved. A picture is meant [i.e. intended] to represent something. (Even an idea is not called true in itself, but only with respect to an intention that the idea should correspond to something.) (Salmon and Soames, 1988: 34)

It is clear, then, that he is committed to (P3.1) for both pictures and ideas. And the implication is that since pictures and ideas are not *intrinsically* different from stones or leaves, it would be absurd to think that they were *really* any different in terms of their aptness for truth; in short, the suggestion is that he endorses the conjunction of (P3.2) and (P3.3).²⁸

In the subsequent paragraph, Frege continues:

What is it that we call a sentence? A series of sounds, but only if it has a sense (this is not meant to convey that *any* series of sounds that has a sense is a sentence). And when we call a sentence true we really mean that its sense is true. (Salmon and Soames, 1988: 36)

That is the extent of his argument to the effect that sentences are not the primary truth bearers: more or less a simple assertion. Nevertheless, there is, I think, an argument lurking in the background, which is recoverable from context; it is the Argument from Intrinsicity above. Why is it that a series of sounds is properly called “true” according to Frege? Because, and only because, that series of sounds is accompanied by an intention which gives it sense (this is the relevant instance of

²⁸ Perhaps the Argument from Intrinsicity would be better represented as involving an additional premise to the effect that anything which is really truth apt is intrinsically truth apt, and then modifying (P3.3) to say that the primary truth bearers are really truth apt; but this modified premise is so boring as to be hardly worth mentioning.

(P3.1)); in particular, because it is accompanied by the intention that it should express a certain thought, and that thought may be properly called “true”.

How compelling is this argument? The second premise is a consequence of the truism that anything which is *F* by virtue of standing in a relation *R* to something else is not intrinsically *F*; it is accordingly beyond reproach. Nevertheless, the argument is unsound as it concerns pictures; and its first premise is clearly to blame. This premise seems okay insofar as it relates to drawings and paintings: for a case can be made that the intention of the artist determines the required projection scheme which establishes a correspondence between the picture and the objects depicted. But if we consider the case of photographs, the premise is obviously just plain false: it is the optical facts themselves, and not the photographer’s intentions, which determine the projection scheme needed for representation.

What are we to make of this first premise as it concerns sentences? Suppose one thought that sentential utterances are simply intentional actions of a certain sort. And suppose that one also thought that not all intentional actions are so by virtue of being caused by a distinct, prior intention.²⁹ If so, then one might hold that sentential utterances in particular are intentional acts which are not accompanied by a prior intention. Furthermore, it is commonly held that actions are events,³⁰ and that events have temporal parts.³¹ If this is right, then it might turn out that sentences are acoustic events which are temporal parts of sentential utterances; and so that sentences are not issued with a prior and independent intention. Thus, premise (P3.1) would be false.

²⁹ One case which makes this position plausible is that of Bratman’s (1984) runner, who knows that running a marathon will wear out his shoes, but who intends to run a marathon nevertheless. It seems he wears out the shoes intentionally (certainly, he didn’t do it by accident!); but we are loathe to say he had an intention to wear out his shoes.

³⁰ See Davidson (2001a) for the canonical argument in support of this view.

³¹ For example, Steward (1997) advocates this view.

However, even if (P3.1) should turn out to be true, there is no need for the defender of Immanence to worry about the Argument from Intrinsicity. The reason is that the major premise of Intrinsicity - (P3.3) - is false. This can be seen clearly, now that we have discussed the existence of relational properties in connection with the Correspondence Theory of truth. Being a sibling is a relational property: I have it in virtue of the relations in which I stand to my brother and to my sister;³² other people fail to have it because they fail to stand in these or similar relations to anyone. But there is nothing of which siblinghood is an intrinsic property. Sunburn too is a relational property, rather than a relation: it is a condition of the skin, not of the skin and the sun. But the existence of something (the skin), which has sunburn by virtue of its standing in an appropriate relation to the sun, does not entail the existence of something else which is intrinsically sun-burnt. So why should truth not be a relational property of sentence tokens, just as Davidson maintains? For all that has been said so far, there need not be anything of which truth is an intrinsic property: this leaves it open to reject (P3.3), and to claim that utterances themselves are the MPTBs.

General considerations about properties will not establish the truth of (P3.3); but perhaps there are some special considerations about truth which can be invoked to show that the primary truth bearers must be intrinsically truth apt. If so, I don't know what they could be. After all, being truth apt is just being of the same kind as something which is true or false: and it seems very strange to suppose that something should be true or false without standing in any relations relevant to its assessment as such. I conclude that we have every reason to reject Intrinsicity, and with it the soundness of Frege's third argument. Thus, Frege has not done enough to eliminate e.g. the possibility that sentence tokens are true; and consequently we have no good

³² In fact, siblinghood is even more radically relational than this; for I have it not only in virtue of the relations I stand in to my siblings, but also in virtue of the relations we all stand in to our parents.

reason to accept (the improbable) Transcendence. In the next section I show that we should in fact *reject* Transcendence, endorsing Immanence instead.

Immanence

In the last section we saw that Frege's arguments in favour of Transcendence fail to convince. In this section I provide a positive reason for endorsing Immanence; indeed, I reason as follows. A thing is truth apt only if that thing exists. But propositions, if they exist, do so only derivatively. Since the things upon which their existence is derivative are themselves truth apt, propositions are not the metaphysically primary truth bearers. Regimenting this line of thought, and simplifying somewhat, we get:

The Argument for Immanence

(P1) If the existence of propositions is metaphysically dependent upon the existence of representations, then the truth aptness of propositions is metaphysically dependent upon the truth aptness of representations.

(P2) The existence of propositions is metaphysically dependent upon the existence of representations.

Therefore, (C) The truth aptness of propositions is metaphysically dependent upon the truth aptness of representations.

Since the Argument for Immanence is simply an instance of modus ponens, it is clearly valid. To show that it is sound, we must therefore establish that its premises are true.³³

³³ I should say a few words by way of clarification of the notion of metaphysical dependence that figures in the Argument for Immanence. Crudely speaking, some facts are metaphysically dependent upon some others just in case the obtaining of the latter facts accounts for, or explains, the obtaining of the former. However, this gloss may not be of much help: for one may wonder what it is for some facts

The first premise, (P1), follows from a general principle connecting properties and their bearers, together with certain facts about truth bearers. As a first pass, we may say that the general principle is the following:

(Property Dependence 0) If two kinds of things, Xs and Ys, can both have a property F, then if the existence of the Ys is metaphysically dependent upon that of the Xs, then the F-ness of the Ys is metaphysically dependent upon the F-ness of the Xs.

The relevant instance of this principle is that if representations and propositions are both truth apt, then if the existence of propositions is metaphysically dependent upon that of representations then the truth aptness of propositions is metaphysically dependent upon the truth aptness of representations. Since representations and propositions *are* both truth apt, the premise (P1) can be detached.

However, this general principle (Property Dependence 0) is not true as it stands. To see this, consider the following case, suggested to me by Dorothy Edgington. A cathedral is beautiful. Moreover, its existence is metaphysically dependent upon the existence of the stones which constitute it. Finally, the stones too, we may suppose, are beautiful: they have a particularly nice marbled beige hue. Nevertheless, we do not want to say that the beauty of the cathedral is metaphysically dependent upon the beauty of the stones. After all, the cathedral *could* have been beautiful even if the stones weren't; its beauty consists in the way the stones are arranged, not the intrinsic features of the stones themselves. So we have a

to account for others; moreover, the kind of metaphysical dependence that I mention (arguably) relates entities (objects and properties), not facts. So let me just say, somewhat less crudely, that some entities are metaphysically dependent upon some others just in case facts concerning the latter, but not the former, ultimately serve to causally explain the possibility of talk and thought about entities of the former kind.

counterexample to (Property Dependence 0): both cathedrals and stones can be beautiful, and the existence of cathedrals is metaphysically dependent upon the existence of stones, yet the beauty of cathedrals is not metaphysically dependent upon the beauty of stones.

There is, however, a notable difference between this case and the one with which we are concerned; and this difference will allow us to articulate a principled distinction between those cases, such as the one which is our main concern here, in which (Property Dependence 0) holds, and those like our cathedral case, where it does not. The difference is that although the truth of a proposition is both necessary and sufficient for the truth of a representation with that propositional content, it is not the case that the beauty of a cathedral is both necessary and sufficient for the beauty of a stone constituting it. Accordingly, we can reformulate our general principle as follows:

(Property Dependence) If two kinds of things, Xs and Ys, can both have a property F, and if there is a relation R between the Xs and the Ys such that, for any X and any Y such that the X bears R to the Y, it is both necessary and sufficient for the X to be F that the Y is F, then if the existence of the Ys is metaphysically dependent upon that of the Xs, then the F-ness of the Ys is metaphysically dependent upon the F-ness of the Xs.

I submit that this general principle is true; and since both representations and propositions can be true, and the relation between representations and propositions that obtains when a representation has a proposition as its content is of the required kind, the first premise of the Argument for Immanence, (P1), follows. It remains to establish the truth of (P2). I devote the remainder of this section to this task.

There are three familiar views of the nature of propositions. The first such account is what I will call *The Fregean View*.³⁴ Fregean propositions (i) are structurally complex entities – that is, they have constituent parts which stand in structural relations to one another; and (ii) their constituent parts are, in general, distinct from the worldly objects which they concern. The second view of the nature of propositions is *The Russellian View*. In a famous passage from a letter to Frege, Russell wrote:

I believe that in spite of all its snowfields Mont Blanc itself is a component part of what is actually asserted in the proposition ‘Mont Blanc is more than 4000 meters high’. We do not assert the thought, for this is a private psychological matter: we assert the object of the thought, and this is, to my mind, a certain complex (an objective proposition, one might say), in which Mont Blanc is itself a component part. (Frege, Gabriel et al., 1980: 169)

Thus, Russellian propositions agree with their Fregean counterparts in being structurally complex; but unlike Fregean propositions, Russellian propositions have real world objects, properties, and relations, as constituents. Finally, the third variant of the transcendent account of the primary truth bearers (arguably) has traces in Wittgenstein’s *Tractatus*; I accordingly call it *The Tractarian View*. According to this view, propositions are not structurally complex at all, but are rather (structurally simple) sets of truth supporting circumstances (TSCs). Both Robert Stalnaker and David Lewis have defended the view that propositions are sets of metaphysically possible worlds; but I include under this view accounts according to which the set theoretic members of propositions need be neither maximally specific “worlds”, nor metaphysically possible. I begin my defense of (P2) above, and with it Immanence, by looking more closely at the Russellian notion of a proposition.

³⁴ Let me nip one possible source of objections in the bud: my naming this and the subsequent positions after famous philosophers and their books is meant to be merely suggestive. I am certainly not *claiming* that these philosophers held the view of propositions named after them and theirs – and so I feel no compulsion to *justify* the nomenclature.

One very general worry about the existence of Russellian propositions is the following. If nominalism is true - in the sense that there are no universals, only particulars (whether abstract or concrete) - then there are no Russellian propositions at all. For according to the Russellian, propositions are complex entities with universals as constituent parts; so if there are no constituent parts, there are no wholes. This is not a worry that I will pursue here, for I suspect the Russellian is fully aware of this feature of his view, and is accordingly committed to the falsity of nominalism; but for those who are not so committed, this incompatibility may be thought to tell against the view.

A related line of thought pushes concerns about the existence of Russellian propositions expressed by sentences containing empty names. If I say, "Santa Clause does not exist," it seems quite clear that I have said something true. Yet according to the Russellian, *propositions* are the kinds of things that are true and false - so if he is to accommodate the thought that I have said something true, he must claim that I have expressed a *true proposition*. The problem is that it is hard for the Russellian to give a plausible answer about what proposition that might be. The Russellian will typically claim that a sentence containing a proper name contributes the bearer of that name to the proposition expressed by that sentence. However, in this case there is no such bearer and so no such proposition. It seems that the Russellian must either claim that sentences containing empty names are associated with propositions by different semantic mechanisms than are sentences containing referring names, or else that such sentences can never be uttered truly. Neither is a comfortable position to be in - and so one might prefer to ditch the notion of a Russellian proposition altogether. But this too is a familiar concern, and I won't push it any further.

Defenders of Russellian propositions as the metaphysically primary truth bearers are committed to the existence of objects and universals which may serve as constituents of those propositions; but, as Jeffrey King (2006) has pointed out, they

also need some account of how those constituents are bound together to form a whole. There appear to be two main views on offer about the nature of the relations between the parts of Russellian propositions: the views of Scott Soames, and of King himself. I will argue that Soames' view is inadequate, while on King's view the existence of propositions is metaphysically dependent upon that of sentences.

In his seminal (1987) paper, "Direct Reference, Propositional Attitudes, and Semantic Content", Soames gives a recursive method for associating Russellian propositions with sentences of a familiar first-order language. The base clause is as follows:

The proposition expressed by an atomic formula $[Pt_1, \dots, t_n]$ relative to a context C and assignment f is $\langle \langle o_1, \dots, o_n \rangle, P^* \rangle$, where P^* is the property expressed by P , and o_i is the content of t_i relative to C and f . (Soames, 1987: 224)

It is clear from Soames' discussion that the content of the terms relative to a context and assignment are taken to be objects; moreover, as he offers no indication to the contrary, I assume that the angled brackets are used in their standard manner, i.e. to signify the operation of taking the ordered n -tuple of the items designated by the enclosed expressions. So, for Soames, the proposition expressed by an atomic sentence is the ordered pair whose first member is the n -tuple of the referents of the sentence's n terms, taken in the order in which they occur in the sentence, and whose second member is the n -ary property, or universal, expressed by the n -place predicate of the sentence.

My worry about this position concerns Soames' suggestion that propositions are structured in the same manner as, because are identical with, ordered n -tuples of objects and universals. The problem is that this view is subject to the charge of arbitrariness. Why should the proposition that Cain slew Abel be (ignoring tense) the ordered pair $\langle \langle \text{Cain}, \text{Abel} \rangle, \text{Slaying} \rangle$ rather than $\langle \langle \text{Abel}, \text{Cain} \rangle, \text{Slaying} \rangle$, $\langle \text{Slaying}, \langle \text{Cain}, \text{Abel} \rangle \rangle$, or $\langle \text{Slaying}, \langle \text{Abel}, \text{Cain} \rangle \rangle$? If we were simply to reverse the order of

the elements in Soames' propositions systematically (so that, in general, the i^{th} entity swaps places with the $((n+1)-i)^{\text{th}}$ within an ordered n -tuple) a recursive characterization of truth for propositions would go through just as smoothly as the one Soames (1987: 225-226) gives; so this cannot be a reason for preferring e.g. the first of these ordered pairs over the fourth. It seems that there is nothing that makes one of these ordered pairs a better candidate truth bearer than any other.³⁵ Indeed, we may push the worry further: why are any of these ordered pairs truth apt, while the ordered pair $\langle 1, 2 \rangle$ is not?³⁶

One possible reply to this worry is to claim that, if an ordered pair has appropriate constituents (objects, universals, truth functions, etc.), and if it is structured in such a way that truth can be defined for it, then it is truth apt, and a proposition. Thus, there is nothing arbitrary about which ordered pairs are propositions; all of the ordered pairs in the previous paragraph, except $\langle 1, 2 \rangle$, are propositions. But this reply provides too many propositions: surely there are not four different propositions which are true if and only if Cain stands in the slaying relation to Abel! For if there were, how would we know which of them our sentence "Cain slew Abel" expressed? And if someone were to believe the proposition that Cain slew Abel – which on Soames' relational view of belief they may – which of these propositions would he or she believe? It won't do for Soames to say that the subject believes all of them – that's too many things believed. So this reply won't do.

An alternative response that appears to be available is to claim that the proposition that Cain slew Abel is not any of these ordered pairs in particular; rather, it is the entire equivalence class of those ordered pairs for which truth can be defined

³⁵ This objection to Russellian propositions *a la* Soames is essentially that raised by Benacerraf to the identification of numbers with set theoretic objects of any sort in his (1965) "What Numbers Could Not Be".

³⁶ King raises both facets of this concern in his (2006) entry in the Stanford Encyclopedia of Philosophy for "Structured Propositions".

in the appropriate manner. This will solve the problem of the previous reply - there will be just one proposition to the effect that Cain slew Abel. Unfortunately for the Russellian, however, he has now abandoned his own view: for equivalence classes are not structured entities! I conclude that the Russellian should not adopt Soames' view of propositional structure.³⁷

In a series of publications,³⁸ Jeffrey King has defended a version of Russellianism, which might be thought to provide a solution to the current difficulty. Like Soames, King approaches the notion of a proposition via that of a sentence which expresses it. King makes two assumptions about sentences which are crucial to his view of propositions. First, he endorses the idea of a transformational grammar, so that there is a level of syntactic representation, distinct from surface structure, which serves as the input to semantic theory, and which King calls *SI*; and second, he assumes that lexical items (primitive constituents of *SI*s) have semantic values. He then claims:

Given a sentence *S*, whose *SI* is constituted by lexical items standing in some complex relation *R*, the proposition expressed by *S* consists of the semantic values of those lexical items standing in *the very relation R* (in the way in which the lexical items themselves stand in *R* in the *SI*). If we call the complex relation *R* obtaining between the lexical items in an *SI* associated with a sentence *S* *the sentential relation of S* and call the relation obtaining between the constituents of the proposition *Q* that *S* expresses *the propositional relation of Q*, the view is that the

³⁷ John Hawthorne has suggested to me that a Soamesian might appeal to supervaluationism at this stage. The thought seems to be that on any given interpretation of the expression "proposition", only one of the four candidate ordered pairs is the proposition that Cain slew Abel; yet nothing about how we speak determines which of these we mean by "the proposition that Cain slew Abel". On this view it will be (super)true that propositions are ordered pairs and that there is just one proposition that Cain slew Abel; yet it will be neither (super)true nor (super>false that the proposition that Cain slew Abel is the ordered pair <<Cain, Abel>, Slaying>. It seems quite strange to think that propositions *are* ordered pairs, yet it is indeterminate *which* ordered pairs they are; though of course this does not yet constitute an objection. In any case, there are general concerns about supervaluationist semantics (for example, supervaluationism requires the abandonment of bivalence), the assessment of which are beyond the scope of this thesis.

³⁸ King (1994), (1995), (1996), (2006),(2007).

sentential relation of S *is* the propositional relation of Q. (King, 1996: 498)

I will not here challenge either of King's two assumptions; rather, I will argue that on King's view, propositions exist only if sentences expressing them do. Since I think that (some variant of) King's is the most plausible Russellian view of propositions available, I take this to show that Russellian propositions are not the metaphysically primary truth bearers.

King (1994) briefly considers the objection to his view that such constituents of Russellian propositions as objects and properties couldn't stand in the very syntactic relations that words and other lexical items (e.g. bound morphemes) stand in to one another. He offers a number of responses to this objection, but none of them seems particularly convincing. One of them is given in an endnote - King writes:

I take it that the worry is that words and propositional constituents (individuals, properties, and relations) are so different, that it seems odd to think that both sorts of things could stand in sentential/propositional relations. However, it is far from clear what words are. It may be that words (types) are properties, so that sentences (types) are properties standing in sentential relations. (King, 1994: 75, fn 28)

This, however, does not seem to quell the worry. For consider the sentence "Fido barks". On King's view it (the type) consists of certain properties standing in sentential relations. But the semantic value of "Fido" is an individual – a dog – not a property. So how can a relation which takes a property (the property of being the word "Fido" perhaps) in one of its argument places also take a dog in that same argument place?

A second response King offers is to paraphrase away talk of propositions in terms of sentential relations on lexical items, and the semantic values thereof: "On this view, to say that two sentences express the same proposition is to say no more than that they have SI's whose sentential relations are the same, and which contain lexical items which have the same semantic values in all the same places." (1995a: 72)

If this approach is adopted, then my work is done: since talk of propositions is paraphrased away, it has effectively been conceded that ultimately propositions don't exist; hence they aren't really true or false.

A third response King gives is that “sentential relations are not just grammatical relations, if by this one means relations that can hold only between linguistic items” (King, 1994: 72). But this is hardly convincing either. To see this, it will be worth our while to elaborate on the objection, to show why it might be thought that semantic values can't stand in syntactic relations. For a thing to have *syntactic* properties it must, in addition, have something analogous to phonetic properties; that is, it must have *formal* properties. This is often obscured in discussions of artificial languages, where it is assumed that the syntactic features of an expression are determined by its formal features; in these contexts, for most intents and purposes, such properties might be thought to be identical. But the common phenomenon of ambiguity shows that they are not. Take the homonyms “there” and “their”; their phonetic features are the same, but they belong to different syntactic categories. It follows that syntactic features are not formal features.

In computing we need the assumption that formal features determine syntactic features, since machines are only sensitive to the former, while semantic interpretation relevant to computation is sensitive to the latter. Indeed, this suggests the possibility that the fact that a certain representation has a given syntax is determined by its being processed in a certain way; in short, it suggests that syntactic features are functional properties. But now, being told that sentential relations are not purely grammatical relations doesn't explain how they can relate non-linguistic items; in particular, it doesn't explain how Fido and the property of barking can stand in appropriate functional relations to one another. Indeed, it seems they can't; there is no proposition of the kind King imagines.

Perhaps for these reasons, King takes a slightly different view in his (2006) article on “Structured Propositions”.³⁹ He there claims, not that the proposition Q expressed by a sentence S consists of the semantic values (s.v.’s) of the lexical items of S standing in the sentential relation of S, but rather that it consists of “the s.v.’s of the words in this sentence standing in the complex relation that is the result of composing the sentential relation of the sentence with the semantic relations the words bear to their s.v.’s, existentially generalizing away the words.” (King, 2006: section 3.1) So the semantic values of the lexical items of S are related in the way that things are related when there are words which (i) stand in the sentential relation R of S, and (ii) have those things as their semantic values. Of course, this means that the proposition Q exists only if there are words which stand in the sentential relation R of S; so the proposition is not true or false if there are no sentences. Q.E.D.

One might resist this argument in one of two ways. A first response is suggested by King’s claim, mentioned above, that words are properties. For if one thought that properties could exist independently of their instances, then one might plausibly maintain that words exist necessarily. If so, then the sting is taken out of my contention that propositions are only truth apt if there are sentences, since that condition is met necessarily. However, I do not think this response is adequate. The reason is that if one takes this view, then words aren’t the sorts of things which can stand in syntactic/grammatical relations after all. For as we have seen, syntactic relations are functional relations; but it is word tokens which stand in those functional relations (in virtue of having the formal properties they have), not word types. An alternative response is different in letter, but suffers what is ultimately the same defect. According to this second view, we don’t get the proposition expressed by a sentence by *existentially* generalizing away the words, but rather, as it were, by

³⁹ King formulates his view differently again in his (2007) book, *The Nature and Structure of Content*; however, many of the changes made are terminological, and those which are not are irrelevant to our discussion.

structurally generalizing away the words. In structuralist lingo, we appeal to the office of the word, rather than the officer. In practice what this means is that we replace words by the grammatical role they fill, putting, e.g., the property of being a name in place of the name 'Fido'. It should be clear, however, that if the considerations of the last paragraph are correct, these offices can't be related in the manner required; for again, the relations are functional, and it is the officers which perform the function.

I have argued that King's Russellian propositions exist only if there are sentences which express some of them. The reason is that such propositions are nothing other than types of semantically interpreted sentences;⁴⁰ and this in turn can be seen from the fact that King's account of propositional structure is parasitic on that of sentential structure.⁴¹ Since King's is the best available account of Russellian propositions, I conclude that Russellian propositions are not the metaphysically primary truth bearers.⁴²

If we turn to consider the notion of a Fregean proposition, however, we immediately recognize that the same is true in this case. Like the Russellian, the Fregean needs an account of propositional structure; and the best such account remains that derived from sentential structure. The only significant difference between Fregean and Russellian propositions lies in which semantic features of interpreted sentences are deemed relevant to their typing. Consequently, my

⁴⁰ The relevant typing must ignore formal/phonetic and surface syntactic/structural properties.

⁴¹ In effect, I have argued that the concept of a King-style Russellian proposition is the concept of a possible sentence type; no wonder then that the possibility of thought and talk of propositions is to be causally explained in terms of facts surrounding sentence tokens.

⁴² It is perhaps worth remarking that although King himself does not accept my conclusion that his propositions exist only if there are sentences expressing them, he does claim that such propositions are not truth apt unless people use sentences which express some of them (King, 2006). The reason is that the sentential relations which constitute parts of propositions must be associated with semantic values – and this association is effectuated by human activity. Thus Russellians who follow King (2006), but not my objections to this view, ought to accept that the truth aptness of Russellian propositions is dependent upon that of sentences. (King (2007) appears to have changed this aspect of his earlier (2006) view, so that the proposition now includes as a part the semantic value of the sentential relation.)

conclusions hold good as concerns Fregean propositions too – like Russellian propositions they depend for their existence upon that of linguistic expressions which can be combined in such a way as to form sentences expressing them. And so Fregean propositions are not the metaphysically primary truth bearers either. It remains, therefore, only to consider unstructured, Tractarian propositions.

My argument that Tractarian propositions are not the metaphysically primary truth bearers runs as follows. A Tractarian proposition is nothing other than a set of truth supporting circumstances.⁴³ But circumstances in turn are just⁴⁴ sets of structured propositions. Given the above argument that structured propositions are merely types of sentences (and so metaphysically dependent upon sentences), it follows that Tractarian propositions too are metaphysically dependent upon sentences. So they are not the metaphysically primary truth bearers. In order to establish this conclusion I need only show that circumstances are indeed nothing other than sets of structured propositions; and so this is what I now set out to do.

As we have seen, both David Lewis and Robert Stalnaker are advocates of the Tractarian view of propositions. Indeed, they are the most prominent recent proponents of this view; and so I will consider their views in some detail before considering the more general case. According to both of these philosophers, propositions are sets of metaphysically possible worlds; however, their views on the nature of these possible worlds differ.⁴⁵ Each holds that there is a plurality of worlds, and that the concept of a possible world is theoretically primitive – that is, that possible worlds cannot be reduced to anything more basic. Yet Lewis maintained that each possible world is a *concrete thing* – a spatiotemporally maximal, mereological

⁴³ This terminology is taken from Soames (1987).

⁴⁴ The phrase “is nothing other than” used in the previous sentence is non-committal with respect to whether there are such things as Tractarian propositions. I use “are just” here for variety; but my intention is to be just as non-committal about the existence of circumstances as I was in the previous sentence about propositions. This point applies also to what follows.

⁴⁵ See Lewis (2001), and Stalnaker (2003).

sum of objects and events; while according to Stalnaker, possible worlds are maximal *properties* the concrete world might have, and so *abstract* entities.⁴⁶ I claim that we should side with Stalnaker on this issue; but also that we should oppose both Lewis and Stalnaker in treating worlds as irreducible.

Lewis' extreme modal realism – his combination of the theses of the plurality of worlds and of the concreteness of worlds – is intuitively implausible. The claim that there are material things which are spatiotemporally isolated from us is one which stands in serious need of justification. Lewis does, of course, offer such justification; but as I argued in my B.Phil. thesis, his best argument in favour of the existence of a plurality of concrete worlds (Lewis, 1968) is flawed. In particular, it assumes that his translation scheme, together with his counterpart theory, gives the correct truth conditions for our modal claims; while familiar considerations show that it does not. In particular, on Lewis' view, modal claims about me, for example, are made true by something other than me, namely some (perhaps all) of my counterparts⁴⁷. The pull of intuitive considerations, together with Ockham's razor, leaves as default position the view that there is not a plurality of such things.⁴⁸

Once we side with Stalnaker in recognizing that possible worlds, if they exist, are abstracta, we are naturally tempted to inquire whether the concept of a possible world is not a derived concept, rather than a primitive. Indeed, it is a familiar thought that we can reduce possible worlds (and the like) to sets of propositions.⁴⁹ Here, for example, is how such a reduction might go.⁵⁰ Call a set of (structured) propositions a

⁴⁶ See Stalnaker (2003: chapter 1) for a comparison of his and Lewis' views of possible worlds.

⁴⁷ Kripke (1971), for example, already made this point.

⁴⁸ Lewis also claims that systematic metaphysics goes much more smoothly on the assumption of (extreme) modal realism. The assessment of such a sweeping claim is obviously beyond the scope of this discussion; though perhaps it is needless to say that I am skeptical.

⁴⁹ This idea is in Wittgenstein's (1974) *Tractatus*, and forms the basis of all recent talk of possible worlds. See, for example, propositions 1 "The world is all that is the case", and 1.1 "The world is the totality of facts, not of things", and what follows thereafter.

⁵⁰ This account is inspired in part by Soames' work in his (1987).

(*truth supporting*) *circumstance*. Say that a circumstance is *consistent* iff it does not contain both members of a *contradictory pair* of propositions; it is *complete* iff it contains at least one member of every contradictory pair; and it is *metaphysically possible* iff it is metaphysically possible that all of its members should be true. Call a circumstance a *world* iff it is complete and consistent; a *possible world* is then nothing other than a complete, consistent, metaphysically possible circumstance.

Stalnaker (2003: 33-38) argues, however, that we should not accept attempts such as this to reduce possible worlds to anything else allegedly more basic; and he gives two reasons for this. First, he claims that reducing worlds requires more theoretical primitives than his alternative direction of analysis, which takes possible worlds as basic and defines propositions in terms of them. In particular, the above theory takes as basic the notions of (*structured*) *proposition*, *contradictory pair* of propositions, and *metaphysical possibility*, whereas Stalnaker's theory requires only the notion of a *possible world*.⁵¹ In short, Stalnaker's first point is that his theory is *simpler* than the alternative. Stalnaker's second point is that the theory which reduces worlds to sets of propositions is *weaker* than his own.

In response to this second point, however, it should be noted that many philosophers find Stalnaker's theory *too* strong – in particular, they find it implausible that necessarily equivalent propositions should be identified. Thus it may be an advantage to have the added flexibility of the weaker theory – after all, one can always impose additional constraints on one's theory in those cases where it seems necessary or advantageous to do so. Moreover, there's no reason why we should not simply ignore the complex nature of possible worlds in such cases; we can, albeit temporarily, treat the notion of a possible world as theoretically primitive. As

⁵¹ One might argue that the notion of a *possible world* is really *two* notions: that of a world, and that of *metaphysical possibility*. Be that as it may, it is true despite this that Stalnaker's theory employs fewer primitive notions than the alternative. It is also perhaps worth mentioning that both theories employ the notion of a set.

Stalnaker himself says in another context, “The decision to treat possible worlds... as PRIMITIVE elements... does not require an ontological commitment to possible worlds as basic entities of the universe. Rather, it is a decision to theorize at a certain level of abstraction.” (1999: 79)

As regards the first point, it is important to note that the strength of Stalnaker’s theory is a consequence of its simplicity. One can’t have the simplicity without the strength; and so if one prefers the weaker theory, one must abandon the simplicity, and with it Stalnaker’s first reason for preferring his theory over the alternative. One could, of course, endorse a theory similar to Stalnaker’s, but taking as basic the generic notion of a circumstance, rather than the more specific notion of a possible world; however, the degree of flexibility of the new theory would be proportional to its degree of ideological complexity – that is, we would need new primitive notions in order to explain the differences between kinds of circumstances, and therefore between (Tractarian) propositions.

We have now been moved to consider the more general case of the Tractarian proposition, whose set theoretic elements are not necessarily metaphysically possible worlds, but may be any variety of truth supporting circumstance. Perhaps it will be worth pointing out, therefore, that Stalnaker’s thoughts on the simplicity of a theory are somewhat bizarre; for Stalnaker seems to think that a theory is simpler the fewer primitive notions it employs. But he does not distinguish what we might call *ontological* simplicity from *ideological* simplicity. So while he would count his theory simpler than the alternative I sketched, on the grounds that it employs only one primitive notion, while my alternative employs three, his criterion is insensitive to the fact that two of three primitives of my alternative theory are notions which simply apply, or fail to apply, to entities of the sort delineated by the first primitive. His theory and the alternative I have sketched each appeal to only one *kind* of entity. But now, here is what turns the tables in favour of the direction of analysis I prefer:

we can give an account of the kind of entities my theory speaks of (structured propositions) in terms of entities to which we are already committed, namely linguistic representations, as shown above, whereas we cannot do this for Stalnaker's primitives. So my theory is ontologically more frugal, and therefore simpler, than Stalnaker's. Moreover, this point applies equally well to the generic, more flexible alternatives to Stalnaker's theory which take truth supporting circumstances as basic. I conclude that possible worlds, and other such circumstances, are to be reduced to sets of structured propositions, and hence that Tractarian propositions are nothing other than sets of sets of structured propositions.

In this section I presented the two-premise Argument for Immanence. The argument is valid, and its first premise, (P1) is supported by the truth of the general principle (Property Dependence). In arguing in support of its second premise, (P2), I distinguished three views of the nature of propositions: *The Fregean View*, *The Russellian View*, and *The Tractarian View*. I then argued that both Fregean and Russellian propositions are no more than types of sentences, and therefore metaphysically dependent upon (linguistic) representations. Finally, I claimed that Tractarian propositions are nothing other than collections of collections of propositions of one of the other two kinds – these latter collections being what's meant by a "possible world", or "truth supporting circumstance". Given the preceding argument and the transitivity of metaphysical dependence, it follows that Tractarian propositions are metaphysically dependent upon sentences. Thus, I showed that the existence of propositions, given any available account of their nature, is metaphysically dependent upon the existence of sentences. This established the truth of (P2), and with it the soundness of the Argument for Immanence; hence, therefore, also the truth of Immanence itself.

2. Higher-Order Quantification

In the first chapter I argued, crudely speaking, that although we may think of propositions as the *conceptually* primary truth bearers, it is concrete representations which are the *metaphysically* primary truth bearers.⁵² In this second chapter, I want to say a little more precisely what that amounts to. In particular, in what follows I will defend two claims:

- I. The meaning of a truth predicate applying to sentence tokens can be explained in terms of a truth operator and quantification into sentence position.
- II. Quantification into sentence position does not commit us to the existence of such abstract objects as propositions.

These claims together jointly display the sense in which the *property* of truth attaches to concreta, though truth appears conceptually to be a feature of abstracta.

I should be clear about the first of these claims. It is not meant to be a *descriptive* claim about the nature of our natural language truth ascriptions, to the effect that they are really, at bottom, uses of a truth operator. In order to establish such a strong claim one would need to show that a plausible syntactic theory can generate all the sentences involving the truth predicate from sentences involving only the truth operator and quantifiers binding sentence variables. This is not something I propose to do here. I suggest that claim (I) may instead be read as the *prescriptive* claim that we ought, “for the purposes of science”, to employ the word “true” only as part of a sentential operator, or as a predicate the meaning of which is explained in terms of this operator. This might constitute a kind of Quinean regimentation of our natural discourse.^{53, 54}

⁵² Less crudely speaking, what I argued was that when we say that something is true, the most basic way in which we specify that truth is by using a phrase of the form “that *P*”; but that despite this fact, when it comes to the *property* of *truth*, this attaches to sentences and/or beliefs.

⁵³ See Quine (1960) for the notion of regimentation.

As regards the second claim, it should be plain that it is controversial. A certain sort of Quinean thought has it that all existential quantification commits us to the existence of objects of some sort or other – and there should be no objection, once the existence of some objectual values of the sentence variables is admitted, to calling them “propositions”. So the defence of claim (II) will require making it plausible that quantification into sentence position does not commit us to the existence of such objects as propositions. Indeed, this should also suffice for our purposes; for if propositions are not a special sort of abstract *object*, then it should be clear that the truth operator does not express a *property* of those objects.

The majority of this chapter will be devoted to a defence of claim (II). First, however, I offer a few words in support of claim (I). Since much of what I have to say is inspired by the work of Arthur Prior, I begin with a consideration of his views: this will help to clarify the issues, as well as the exact nature of the project I am pursuing.

Prior on Truth Predications

In *Objects of Thought*, Arthur Prior argues *inter alia* that facts and propositions are not special kinds of objects, parts of the ontological make-up of things, but rather mere “logical constructions”. His defence of this view consists in the claims that (A) facts are no more nor less than true propositions; (B) we can eliminate “true” and “proposition” (and their cognates) from our language without loss of expressive power, provided that we allow quantification into sentential position; and (C) this kind of quantification does not itself ontologically commit us to

⁵⁴ Alternatively, the claim might also be read in the vein of state of nature stories. The idea here is that we can imagine that our ancestors used a language involving quantification into sentence position, and containing “true” only as a part of a sentential operator and never as a predicate. And we can imagine further that, given a population using such a language, it would be possible for that people to extend their language to include a truth predicate applicable to beliefs and sentences in the manner I will suggest. Given this possibility, it does not matter whether the fictional history of the population is accurate; the point is that this can constitute an explanation of the meaning of the truth predicate in terms of that of the truth operator and quantification into sentence position. See Williams (2002) for an insightful discussion of the philosophical purpose of state of nature stories.

the real existence of propositions. Along the way, Prior records a number of proposals regarding logical syntax: these include the suggestions that (i) “true”, in its logically primary usage, is part of a sentential operator; and (ii) the propositional attitude verbs (“believes”, “says”, etc.) likewise form parts of sentential operators (‘S believes that’, ‘S says that’, etc.).

Prior’s strategy for eliminating talk of facts and propositions is, he claims, essentially that of F. P. Ramsey (1927). The idea is that we first show how to do without the relevant vocabulary in basic contexts by invoking the redundancy theory of truth; and then we extend this strategy to more difficult cases involving belief, assertion, and unspecified propositions. Thus, Prior begins:

The basic contexts in which we appear to be talking about propositions and facts are ones in which we ascribe truth or falsehood to the former, and here the elimination of these apparent objects or subjects of discourse is easy.... ‘That grass is green is a true proposition’ = ... ‘It is a fact that grass is green’ = ... ‘It is true that grass is green’ = ... the plain ‘Grass is green’. (Prior, 1971: 11)

I have two comments to make concerning this passage. First, although the end product is a sentence in which “true” does not occur, a crucial step on the way involves a sentence which contains the operator “It is true that”. It is for this reason that I attribute to Prior the view expressed by (i) above, despite the fact that he wants eventually to eliminate “true” altogether. Second, it is unclear what exactly Prior has in mind in using ‘=’ to express the relation that holds between the various sentences mentioned in this passage. (It seems obvious that he can’t mean identity, since the sentences equated are plainly distinct.) I will not engage in serious historical reconstruction of Prior’s view here: I note only that he must intend some fairly strong relation such as synonymy.

So how does Prior generalize this strategy for the elimination of “true” and “proposition” from basic contexts to cases in which these expressions occur in non-

basic contexts? Amongst the crucial cases are sentences ascribing propositional attitudes. Prior says,

[W]e must get away from... the whole idea that 'X fears (thinks) that there will be a nuclear war' *has* to express a relation between X and anything whatever. And we cannot get away from this idea without adverting to the difference between a name (or other designation) and a sentence. (Prior, 1971: 16)

Prior's thought is that, if we insist on viewing expressions of the form 'that P' as singular terms then we will be naturally lead to think of propositions as their designata, and hence as objects. Instead, Prior suggests the following syntax for such locutions:

[E]xpressions like '— fears that —' and '— thinks that —' have [the] function of forming sentences from other expressions of which the first is a name and the second another sentence. They are as it were predicates at one end and connectives at the other. (Prior, 1971: 19)

Recognizing that ascriptions of propositional attitude have this syntax allows us to recast our thought on the matter: we need no longer view propositions as objects of the attitudes. Indeed, it is after this syntactic proposal is put forward that Prior finds himself able to say,

We can now give an account of truth and falsehood as applied not merely to propositions, i.e. to what we believe or assert, but also as applied to particular believings and assertings. 'X's belief (assertion) that there will be a nuclear war is true' means no more and no less than 'X believes (says) that there will be a nuclear war, and there *will* be one'. (Prior, 1971: 21)

The strategy here is clearly an elaboration on that proposed previously to deal with the truth of specific propositions: Prior simply *uses* the sentence which identifies the proposition whose truth is at issue, and thereby avoids *predicating* truth of anything.

The last step is to account for the ascription of truth to unspecified attitudes and propositions. It is here that quantification into sentential position plays a role. Prior says,

Ramsey thought of this [problem] too; his answer to it – the right one, it seems to me – was to move into a slightly more stylised language than ordinary English, with quantifiers binding variables that stand for sentences. (Prior, 1971: 24)

It is important to note here that by “variables that stand for sentences”, Prior means variables that stand *in* for sentences, i.e., occupy sentential position, and not variables whose objective values are sentences. With such quantification in his ideographical repertoire, Prior then suggests that we can simply read such claims as “Everything Cohen believes is true” as

‘For any p , if Cohen believes that p , then it is the case that p ’, or more briefly ‘For any p , if Cohen believes that p , then p ’. Similarly, ‘There are facts which nobody has ever asserted or will ever assert, but which *are* facts all the same’, amounts to ‘For some p , it has never been and never will be asserted that p , but it is the case that p all the same’. (Prior, 1971: 24-25)

Thus Prior concludes his attempt to eliminate “fact”, “true”, and “proposition” from the language, without loss of expressive power.

I am sympathetic to much of Prior’s line of thought. In particular, I agree with (something like) his conclusion, and I also agree that, if each of claims (A) through (C) were correct, these facts would jointly suffice to establish Prior’s conclusion. There is at least one point, however, where my strategy differs from Prior’s. As I said, Prior appeals to the redundancy theory of truth in arguing in support of his principal claims: that is, he thinks that ‘It is true that P ’ is in general synonymous with, or in some other sense strongly equivalent to, ‘ P ’. (It is this which allows him to claim, for instance, that “Everything Cohen believes is true” can be analyzed as “For any p , if Cohen believes that p , then p ”, in which the word “true” has completely disappeared.)

Ian Rumfitt (2003), however, has noted that there are sentential operators O which are not redundant, and yet which validate the schema: $\lceil OP \text{ iff } P \rceil$. One such operator is the “Actually” of modal logics. Although $\lceil \text{Actually } P \text{ iff } P \rceil$ holds in full generality, nevertheless “Actually” is not redundant, since, for example, $\lceil \text{Necessarily, actually } P \text{ iff Necessarily } P \rceil$ is not valid. (The schema - in which “iff” gets wide scope - fails whenever what replaces “ P ” is not itself necessary). Thus, it does not follow from the facts that (a) “It is true that” is an operator, and (b) $\lceil \text{It is true that } P \text{ iff } P \rceil$ holds generally, that “It is true that” is redundant. In what follows, I do not assume the redundancy theory, and thus I remain agnostic regarding Prior’s claim (B).⁵⁵

It is, nevertheless, and despite this agnosticism, quite easy to elucidate a perfectly legitimate predicative use of the word “true” which applies to sentence tokens – or as I will also say here, sentential utterances⁵⁶ - given a prior language containing the operator “It is true that”, as well as quantification into sentence position. Following Prior, we first remark that using only the sentential operator we can account for the truth and falsity of *particular* beliefs and assertions. Schematically, we say:

S ’s belief that P is true iff S believes that P and it is true that P ,

S ’s assertion that P is true iff S asserts that P and it is true that P .

Next, we note that we can define an adverb “truly” which can be used in such contexts. We say:

S believes/says truly that P iff S believes/says that P and it is true that P .⁵⁷

Employing a quantifier binding sentence variables, we may generalize, and say:

⁵⁵ Since (A) is strictly orthogonal to my purposes, and (B) is a conjunctive claim only part of which is at issue, I have framed my position in terms of the endorsement of the claims (I) and (II) from the previous section, rather than directly invoking the claims (A) through (C).

⁵⁶ I am not here coming down on the issue of whether utterances are actions. I am simply using a terminological variant of “sentence token” that proves convenient in this context – and bowing to the Davidsonian semantic tradition.

⁵⁷ This is not to be confused with $\lceil S \text{ truly believes that } P \rceil$ – George W. Bush truly believes that God is on his side; it is entirely another matter whether GWB believes truly that God is on his side.

S speaks truly (simpliciter) iff $\exists P[S \text{ says that } P \text{ and it is true that } P]$.

Finally, we say:

S 's sentential utterance u is true iff S speaks truly in producing u ;

...that is, iff $\exists P[\text{In producing } u \text{ } S \text{ says that } P, \text{ and it is true that } P]$.

Of course, the truth operator figuring in these biconditional schemata may or may not be redundant. If it is, then the conditions given on the right hand side for the use of the left hand side expressions are equivalent to Prior's.

One might argue, however, that I have not yet done enough to fix a genuinely predicational use of the word “true” – for I have not specified general *satisfaction* conditions for this alleged predicate. Instead I have merely given *truth* conditions for (closed) sentences containing an expression that is grammatically (i.e. superficially) similar to a predicate. But this does not determine under what conditions things in general count as “true”. In fact, it is consistent with all that has been said that there are no such things as e.g. beliefs, despite it being correct to make claims of the form ‘ S 's belief that P is true’. Indeed, one might summarize the point with the slogan “Beliefs are mere logical constructions”.

Someone tempted by this line of thought might also note that we cannot solve the problem by simply allowing “existential” generalization on claims of the form ‘ S 's belief that P is true’; for it might be that the quantified claims which result from performing this operation on specific claims of the form in question are not genuinely existential, but ought instead to be given a substitutional interpretation. So the only thing that will alleviate the concern is a specification of full satisfaction conditions for the purported truth predicate.

In response to these concerns, let me say four things, in increasing order of strength. First, I note simply that I am here agnostic about the existence of such things as beliefs and assertions; for I am concerned principally with *utterance* truth – that is, truth considered as a feature of (spoken) *sentence tokens*. Second, the account

I have given of utterance truth at no point appeals to the notion of truth for beliefs or assertions. Although I presented truth conditions for sentences, as it were, ascribing truth to beliefs and assertions prior to giving truth conditions for sentences ascribing truth to utterances, the latter truth conditions are logically independent of the former. Third, let me say that there are, without a doubt, acoustic events which constitute sentence tokens; so this poses no obstacle to regarding the sentences ascribing truth to utterances as genuine predications. And fourth, I can give full satisfaction conditions for “true” in a sense which applies to these things:

$$\forall x (x \text{ is true iff } \exists S [S \text{ produces } x \text{ and } \exists P (S \text{ says that } P \text{ in producing } x \text{ and It is true that } P)])$$

Anything x which fails to meet the condition given on the right hand side of the biconditional is not true;⁵⁸ and thus, I claim, “true” is a genuine predicate. That is, these considerations, I believe, jointly constitute a sufficient defence of the claim (I) above.

The Strategy of Reductive Elimination

What arguments can be given in favour of our claim, (II), that quantification into sentence position is ontologically neutral? Prior views quantification into sentential position as continuous with quantification into predicate position:

For sentences, as Peirce saw, are simply those n -place predicates for which $n = 0$; an n -place predicate is a sentence with n gaps for names to go in, an ‘open’ sentence as it is now excellently called, and an ordinary or ‘closed’ sentence is one with no such gaps left. (Prior, 1971: 33)⁵⁹

⁵⁸ In the relevant sense. There might, of course, be a distinct sense of “true” which applies to e.g. beliefs, for all I say here. Note also, that not being true is not the same as being *false*. A similar satisfaction condition can be given for the predicate “false” – thought of course this one will involve an additional negation. Both truth and falsity as applied to sentential utterances will be very useful notions when it comes to constructing semantic theories in chapter 2.

⁵⁹ I prefer not to use the term “predicate” as Prior does, to speak of open sentences; rather, I reserve it for (not all, but only) those expressions which either do not contain free variables – either because they contain no variables at all, or because those variables are bound, without yet being generalized upon.

This suggests a strategy for showing that quantification into sentence position is not ontologically committing: first, show that quantification into predicate position is ontologically neutral; then argue that if quantification into predicate position is harmless the same holds for quantification into sentential position. Let's explore this possibility.

As we have seen, Prior identifies variable types by the syntactic category of the expressions they stand in for, rather than by the kind of entity they denote. He says, "In all this I cannot see anything mysterious, or anything that need compel us to treat variables that do *not* stand for names of objects as if they did" (Prior, 1971: 37); and he goes on to suggest some idiomatic English renderings of higher-order sentences which he claims are not even *prima facie* ontologically committing.

Thus Prior makes some suggestive comments regarding the interpretation of higher-order quantification; but his followers Stephen Yablo and Agustin Rayo (2001) spell these thoughts out more explicitly, and indeed go beyond them. Using some of Prior's idioms they provide a translation scheme taking sentences of a higher-order formal language into English (supplemented with subscripts so as to indicate anaphoric relations unambiguously). They then give a number of arguments to the effect that the English translations are ontologically innocent.

The first of these arguments may be represented as follows.

The Argument from Instances

(P1.1) The truth of "Roses and sunsets are red" does not require the existence of anything beyond roses and sunsets – in particular, it does not require Redness.

But Prior's comments serve only a suggestive purpose in the text, and so I do not believe anything here hangs on the distinction between predicates and open sentences.

(P1.2) The ontological commitments of a quantified sentence cannot exceed those of its instances.

Therefore, (C1) The truth of “There is something roses and sunsets are” does not require the existence of anything beyond roses and sunsets – in particular, not Redness.

The second argument is really an argument in favour of the second premise of the Argument from Instances. It runs thus:

The Argument from Entailment

(P1.2.1) If the commitments of an existentially quantified sentence could exceed those of its instances, then the former would not follow trivially from the latter.

(P1.2.2) But they do follow trivially; existential generalization *is* valid.

Therefore, (P1.2) The ontological commitments of a quantified sentence cannot exceed those of its instances.

The Argument from Entailment seems sound; and so the second premise (P1.2) of the Argument from Instances is secured.⁶⁰ However, the concatenation of these two

⁶⁰ One might want to deny this - for consider the following parody argument:

(P1.1') The truth of “John got dressed” does not require the existence of anything beyond John.

(P1.2') The ontological commitments of a conclusion in a valid one premise argument cannot exceed those of the premise.

Therefore, (C1') The truth of “John put something on” does not require the existence of anything beyond John.

This argument is terrible: it has an obviously false conclusion, so something must go wrong. But is it that the argument from Entailment is mistaken, or is it that the premise (P1') is false? Certainly “to get dressed” is an intransitive verb – so it's semantic contribution *could* be explained as its simply applying to some things (people) and not others (other people, and more besides). However, it seems clear that analysis of “to get dressed” will reveal that it is to be explained in terms of an existentially closed two place predicate. So I think the argument from Entailment is fine; it is the first premise of the parody argument that is to blame.

arguments establishes the claim that higher-order quantification is not ontologically committing, only if the first premise in the Argument from Instances is true; that is, only if it is true that standard first-order predications are not committed to entities as the values of predicates. It is open to the realist about universals, and indeed about propositions, to reject this claim.⁶¹ Thus, taken together these two arguments are not suasive.

Yablo and Rayo's last two arguments, however, are stronger on this score. Their next argument is as follows:

The Argument from Consistency

(P2.1) The sentence “*a* is something *b* is too, viz. not a member of itself, but we know from Russell's paradox that there is no witnessing set” is not self-undermining (i.e. inconsistent).

(P2.2) It would be if “something” ranged over sets.

Therefore, (C2) “something” does not range over sets.

Yablo and Rayo remark that this argument, suitably modified, would be equally compelling if “object” were to replace “set” throughout.⁶²

Finally, we may consider their last argument:

The Argument from Cardinality

(P3.1) It is not impossible that, take any objects you like, they are something that the rest of the objects are not.

⁶¹ To be fair to Yablo and Rayo, and indeed to Prior, they take their opponents to be Quineans, who grant this claim. It should also be noted, incidentally, that Yablo and Rayo do not explicitly consider the case of 0-place predicate (i.e. sentence) variables.

⁶² This contention cannot be supported by Cantor's Theorem, as it does not concern sets; but it can be supported by what Rayo (2002: 441) calls “Bernays' Principle”. This principle, which is the analogue of Cantor's theorem as concerns objects in general, also seems highly plausible.

(P3.2) It would be impossible if “something” meant *some object*.

Therefore, (C3) “Something” doesn’t mean *some object*.

This argument, like the previous one, rests on cardinality considerations; each provides *prima facie* reason to believe that quantification into predicate position can be construed in a manner that is not ontologically committing.

It is important to note, however, that neither Prior nor Yablo and Rayo are thinking of such quantification as straightforwardly substitutional. Despite the fact that what these higher-order variables “stand for” are expressions of appropriate syntactic type, nevertheless what we can say using such ideography is not to be constrained by the other expressive resources of our language: “Propositions,” Prior says, speaking loosely “and their truth and falsehood, are language-independent.” (Prior, 1971: 12) Strictly speaking, what this means is that whether it is true that *P* or false that *P* is not to be dependent upon what expressions the language we are speaking contains (unless, of course, $\lceil P \rceil$ explicitly concerns this issue); and this must be so even if $\lceil P \rceil$ contains quantification into higher-order position.

It will prove instructive to compare the first-order case. Suppose a substitutional semantic rule were given for “every”: a sentence of the form $\lceil \text{Every } F \text{ } Gs \rceil$ is true just in case every sentence of the form $\lceil n \text{ } Gs \rceil$ is true, where $\lceil n \rceil$ is replaced by a name in the language for an *F*. Then whether “Every dog barks” were true would depend upon whether we had names for certain dogs: it could, for instance, turn out to be true even though some stray dog for which we have no name failed to bark. Intuitively, this is the wrong result. What’s needed instead is as follows: a sentence of the form $\lceil \text{Every } F \text{ } Gs \rceil$ is true just in case, for any object *o* which satisfies $\lceil F \rceil$, $\lceil n \text{ } Gs \rceil$ is, or would be, true in the extension of the language in which $\lceil n \rceil$ names *o*. That is, we want a quantified sentence to be true, not if all of its *actual* substitution instances are true, but rather just in case all of its *possible* (legitimate) substitution instances are

true.⁶³ The same must be said for quantification into higher-order position: a sentence containing a universally (existentially) quantified higher-order variable is true just in case every (respectively, some) possible substitution instance is/would be true.⁶⁴

But now it seems we can mount an argument getting us closer to the conclusion that quantification into sentence position does not commit us to the existence of propositions, on the grounds that there are too many possible substitution instances of monadic predicate variables – or, relatedly, too many *ways* for objects to be. It runs as follows:

The Cardinality Argument Concerning Propositions

(P4.1) There are strictly more pluralities of objects than there are objects.⁶⁵

(P4.2) There are at least as many ways for an object to be as there are pluralities of objects.

(P4.3) For any object, and any way for an object to be, there is the proposition that that object is that way.

Therefore, (C4.1) There are strictly more propositions than there are objects.

Therefore, (C4.2) Propositions are not objects.⁶⁶

⁶³ See Evans (1985: 80-87) for this “Fregean” approach to first-order quantification, and Etchemendy (1999: 55-56) for the need for the qualification “legitimate”; in what follows I omit it, but the reader may take it as tacit.

⁶⁴ I address the issue what a possible substitution instance of predicate and sentence variables is below.

⁶⁵ The phrase “plurality of objects”, though singular, must not be taken to designate a special kind of singular entity, but rather the many objects themselves.

⁶⁶ I should note that the conclusion of this argument, (C4.2), is consistent with the central claim of chapter 1, namely that propositions are metaphysically dependent upon sentence tokens. For starters, metaphysical dependence is not identity, and the concept of a (structured) proposition, I suggested, is that of a *type* of possible sentence token; but, we may suppose, types are not objects (despite being dependent upon them). I should also clarify that what the cardinality argument (if successful) shows, strictly speaking, is that propositions are not *actual* objects. Since there could be more objects than there actually are, the metaphysical dependence of propositions on the possibility of the existence of sentence tokens is in no conflict with (C4.2). Note that a more thorough defence of this point may require a notion of possibility which does not obey the S4 axiom of modal logic.

The conclusion of this argument, (C4.2), is not, of course, our claim (II) itself. And indeed, it seems that before we can get to claim (II), I need to say a few words of clarification.

In fact there are three views available of the ontological commitments of higher-order quantification. They are as follows:

- 1) Quantification into higher-order positions commits us to the existence of objects.
- 2) Quantification into higher-order positions commits us to the existence, or being, of a special logico-metaphysical kind of entity distinct from objects.
- 3) Quantification into higher-order positions does not commit us ontologically to any entities whatsoever.

Crudely speaking, we can associate view 1 with Quine, who thought that typed languages should be understood as multi-sorted first-order languages, and that first-order bound variables are ontologically committing in the only way possible – they commit us to the existence of objects of one kind or another. The type theoretic position 2 may be associated with Frege and Russell who held, for instance, that existence is a property of properties (hence a higher-order property). And finally, the ontologically frugal position 3 may be associated with Boolos and his followers.

I said in the last section that I agree with “something like” Prior’s claim that propositions are not part of the ontological make up of things, but rather “mere logical constructions”. In fact, I do not want to commit myself to a specific view of how it is that propositions fail to exist: I have been, and I will continue to be, arguing against position 1; but I will not decide between positions 2 and 3.

Given these clarifications, it should be clear how we can get from the conclusion of the Cardinality Argument Concerning Propositions to our claim (II) construed as the rejection of position 1. We need only note that quantification into sentence position is intelligible; for, given the argument just reviewed, we know that

it cannot be intelligible by virtue of committing us to the existence of additional objects.

Still, and in light of these recent considerations, one might worry that the Cardinality Argument Concerning Propositions is not sound. In particular, one might reject the second premise, (P4.2), on the somewhat peculiar grounds that it is too *weak*. The thought is that this premise must be strengthened if the argument is to succeed - that there must be *exactly* as many monadic ways as pluralities if we are to avoid the charge of ontological commitment with respect to our quantification into predicate position; for if not, then the existence of the pluralities would not suffice for the “existence” of the ways.

Suppose that there were more ways for objects to be than pluralities of objects. Then while the Arguments from Consistency and from Cardinality appear to establish that ways are not objects, they do not establish that ways are not to be reckoned in our ontology; for it has not been shown that existence – that is, singular being – exhausts reality. Maybe ways are a special kind of entity with their own “general” mode of being. If so, then we might think there is no reason to accept (P4.2) either. Given that the being of the ways is not metaphysically accounted for by the existence of the pluralities, it would seem that nothing prevents some pluralities having *no* corresponding ways. So maybe there are sufficiently few ways for objects to be that the ways can be counted as existing objects after all.⁶⁷

Let me try to draw out the nature of this concern more clearly. Peter Simons, in a (1997) paper on “Higher-Order Quantification and Ontological Commitment” says:

Ontological commitment is a sort of converse to an idea of more recent prominence: truth-making. Whereas when we ask what things are such

⁶⁷ The appearance of soundness in the Arguments from Consistency and Cardinality might then be explained away as a kind of semantic phenomenon: the thought is that all entities are objects, but no one quantifier can range over all objects of all kinds.

that their existence is *necessary* for a sentence to be true, we are asking after its ontological commitments; when we ask what things are such that their existence is *sufficient* for the sentence to be true, we are considering the sentence's *truth-makers*. (Simons, 1997: 265)

The worry, then, is that, if there is no one-to-one correspondence between ways for an object to be and pluralities of objects, then the existence of the objects in the extension of a given predicate will not be sufficient for the truth of a sentence ascribing that predicate to some object in that extension. The ontologically frugal position 3 above will be unavailable, and there will be no reason to favour position 2 over position 1.

Finally, note that for one who held the position just suggested, it would be possible to give a perfectly principled response to Yablo and Rayo's first two arguments purporting to show that higher-order quantification is ontologically innocent. A philosopher of this persuasion would claim that there is no single interpretation of "instance" which renders both premises of the Argument from Instances true. Thus, if a sentence containing "grue," for example, is considered an instance – so that instances are identified formally – then the ontological commitments of quantified sentences do exceed those of their instances; whereas, if the existentially quantified sentences follow trivially from their instances, sentences involving "grue" and the like are not to be counted genuine instances, which are to be identified by their content instead. Either way, the argument fails.

The problem is exacerbated by the following considerations. Concerning the case of monadic ways for objects to be (or "things" that objects are), Yablo and Rayo claim that:

$\forall \mathbf{x} \exists P \forall y (Py \leftrightarrow y \in \mathbf{x}),$
that is, ...

Take any objects_i you like, there is something_j that they_i and only they_i are. (Rayo and Yablo, 2001: 86)⁶⁸

They do not, however, commit themselves to there being at most one such for each plurality - if that's the right way to talk. They allow that different predicates, or predicate variables, true of exactly the same things might embed differently under modal operators; thus, "It could happen," they say,

that... $\exists P \exists Q [\forall x ((Px \leftrightarrow Qx) \ \& \ \blacklozenge \exists y (Py \ \& \ \neg Qy))]$,
comes out true, because...

There is something_j that my cat is – a creature with a kidney – and something_j that my dog is – a creature with a heart – such that everything that is that_i is that_j and vice versa; but there *could* be a thing that was that_i and not that_j. (Rayo and Yablo, 2001: 88)

This does not *by itself* establish that they think there are two things the creatures with a kidney are; we need the additional, though plausible premise that ways for objects to be are distinct whenever they are not necessarily coextensive. In fact, Yablo and Rayo don't even have the expressive resources required to formulate the claim that there are some things_i such that there are two things_(j,k) they_i and only they_i are. The reason is that they have no expression for higher-order identity – and hence no means of speaking of the distinctness of ways either. (Recall that if one wishes to say that there are at least two things in a first-order language, one needs identity for objects: $\exists x \exists y \neg x=y$.)

Prior (1971: 53-54), by contrast, does have a sign for higher-order identity in his preferred ideography: he uses "I" as a two-place sentential connective so that $\lceil Ipq \rceil$ says that the proposition that p is the very same proposition as the proposition that q.⁶⁹ Does his use of a notation for propositional identity mean that Prior is committed to the existence of propositions and universals, whether conceived as objects or as

⁶⁸ In Yablo and Rayo's formal language the bold script indicates that the expression in question is plural; capitals are higher-order expressions; and the symbol "ε" is read "is one of".

⁶⁹ Prior employs Polish notation.

special kinds of higher-order entities? Perhaps we should take Quine's slogan "no entity without identity" as a positive proposal rather than a demand for clarity; thus, "without identity... (you are committed to) no entity" rather than "(you are allowed) no entity without identity". If so, then we might think that Yablo and Rayo are in fact in a better position than Prior is with respect to ontological commitment, precisely because they can't say whether there are some things such that there are two (higher-order) things they and only they are. Indeed, one might be tempted to say that Yablo and Rayo allow not higher-order *quantification*, but higher-order *generalization*; one can't count unless one knows when to add *another* to the tally. But no one ever claimed that mere generalization was metaphysically loaded.

Without going into too many details, let me say that I am not convinced Prior is in a worse position with respect to ontological commitment than Yablo and Rayo, simply by virtue of allowing a symbol for higher-order identity. Recall that their objective, like his, is to show that quantification into higher-order position can be done without ontological commitment. Their opponents - Quineans - purport to find such quantification mysterious. What these Quineans want to know, and claim not to know, is under what conditions such quantified sentences are true. When, in principle, can the search for a verifying or falsifying instance terminate? This is effectively the same as the question which can be posed in our earlier terminology, what is a possible substitution instance? To render plausible the claim that higher-order quantification does not commit us ontologically, the answer had better not be, a possible sentence is a possible substitution instance of a higher-order formula whenever the expression replacing the variable designates a special kind of entity. But Yablo and Rayo, like Prior, allow that two predicates true of exactly the same objects may embed differently under modal operators: so they can't claim either that two possible predicates are identical, or even alike in meaning, whenever they are true of the same objects; nor, therefore, that all possible substitution instances of a sentence

containing a variable for one-place predicates have been checked whenever, for each collection of objects, some predicate true of exactly those objects has been checked. They are left without an informative answer to the Quinean question. Moreover, once they are in a position to give one, they will also be able to introduce a symbol for identity – or at least indiscernability – at the higher-order level.

Nevertheless, on reflection I think we can see that the Cardinality Argument Concerning Propositions is sound. The crucial premise, (P4.2), was that there must be at least one way for an object to be for each collection, or plurality, of objects. But of course there is! For why not just what's expressed by the predicate “is one of them”, where “them” is true of, or denotes, each of the things which constitutes the plurality, and of nothing else? Then, for any object, and any collection of objects, there is the proposition that it (the object) is one of them (the things “in” the collection). So there are strictly more propositions than there are objects. In the next section we will see a couple of semantic theories which interpret quantification into sentence position, and thereby establish conclusively that, Quinean worries notwithstanding, such quantification is intelligible. Claim (II) will thus be vindicated.

The Direct Strategy

Perhaps we can argue directly for the claim that quantification into sentence position is ontologically innocent, without making a detour through quantification into predicate position. Dorothy Grover (1972) gives two accounts of the semantics for a language with quantification into sentence position. By way of explanation of the first, she says:

A mapping I is a *substitution interpretation* if each variable is provided with a substitution range, and I maps the sentence parameters and variables into a (complete) lattice L ;... [moreover] $I(\exists pA(p)) = \vee\{I(A(B)) : B \text{ is in the substitution range of } p\}$... [where, for] $X \subseteq L$, $\vee X$ is the lattice least upper bound of X [and similarly for $I(\forall pA(p))$]. (Grover, 1992: 49)

Grover points out that restrictions on the substitution range of the sentential variables are necessary in this substitution interpretation if circularity is to be avoided. One would not want the interpretation of $\lceil \exists p A(p) \rceil$ to depend upon itself, so this sentence (amongst others) must be excluded from the substitution range of the sentential variable it contains. The second interpretation Grover gives for her language is as follows:

A mapping I is a *domain and values interpretation* if it satisfies the following: (1) I maps the sentence parameters and variables into a lattice L ; (2) $I(\exists p A(p)) = \vee \{I(A(p)) : I' \text{ agrees with } I \text{ except perhaps at } p\}$; (3) [similarly for $I(\forall p A(p))$]. (Grover, 1992: 49-50)

Neither of these semantic accounts yields plausible results for sentences involving intensional or hyperintensional contexts if the elements of the lattice are simply truth and falsity. The problem is that whether $\lceil \delta p \rceil$ is true or false is not a function of whether $\lceil p \rceil$ is true or false when $\lceil \delta \rceil$ is an operator such as “It is necessary that” or “Tom believes that”. Consequently, Grover says, “we may think of the elements of the lattice as propositions.” (Grover, 1992: 49)⁷⁰ This does not seem an auspicious start for an attempted defence of the claim that the use of quantifiers binding sentential variables does not commit one to the existence of propositions. However, the issue is a subtle one, as is Grover’s position on the matter.

Grover (1979) thinks that users of a language with quantifiers binding sentential variables are not committed to the existence of propositions; they do not talk “about” propositions simply by using sentences with bound sentential variables, nor do these variables “range over” propositions.⁷¹ Grover (1992: 62) claims that names and pronouns are not needed to give natural language readings of sentences of her formal language which involve quantification into sentence position; instead we

⁷⁰ Indeed, it is perhaps useful to think of them as sets of possible worlds. Then the lattice least upper bound of a set of propositions, or of sets of worlds, will be the set containing all those worlds which belong to at least one of the propositions in the set, i.e. the generalized union of those sets of worlds.

⁷¹ See especially (Grover, 1992: 140-143).

may use what she calls “prosences” – such expressions as “things are thus and so” and, on her view, “it’s true” and “that’s true”. But speaking of variables “ranging over” the elements of a domain is only appropriate, she thinks, when we do need names of those elements in the relevant natural language renderings. That is, she offers what is effectively the following argument:

The Argument from Natural Language Renderings

(P1.1) Ontological commitment occurs only when variables range over values.

(P1.2) Variables range over values only when pronouns are required for natural language renderings (NLRs).

(P1.3) NLRs can be given for formal sentences involving quantification into sentence position using not pro- *nouns* but pro- *sentences*.

Therefore, (C1) Quantification into sentence position does not incur ontological commitment.

The propositional realist’s reply to this argument seems obvious. The conjunction of (P1.2) and (P1.3) can’t be true, as the variables of the language *do* range over values, and those values are to be thought of as propositions – by Grover’s own admission! So, the Argument from Natural Language Renderings notwithstanding, quantification into sentence position *does* commit us ontologically to the existence of propositions.

Grover’s initial response to this point seems to be that the interpretation of sentences containing quantification into sentence position is no worse (in terms of commitment) than the instances of such sentences which contain no quantification of this kind. Although the interpretation function maps sentences and sentence variables into the elements of a lattice, Grover (1992: 62) points out that this is equally true of the interpretation function for a first-order language. “Therefore,” she concludes, “I need introduce talk of propositions and [of?] formulas ‘ranging over’ propositions in

giving an account of propositional quantification, only if necessary in a domain and values interpretation of first-order logic.” (Grover, 1992: 62) But this response appears inadequate. The strict Quinean does not admit non-extensional language, and so is committed at most to truth values as interpretations of the instances; so Grover will not persuade him in this way to accept the use of her language. By contrast, the propositional realist who does admit intensional or hyper-intensional idioms is happy to recognize propositions as ontological commitments of both instances and quantifications, and so won’t accept her conclusion.

It will prove useful at this stage to introduce a distinction between two notions of ontological commitment. I will say that one is *weakly* ontologically committed to some things by using a certain sentence if that sentence could not be true without those things existing. By contrast, one is *strongly* ontologically committed to some things by using a certain sentence if one could not rationally maintain that that sentence was true without also maintaining that those things exist.⁷² Consider, by way of illustration, Tarskian sequences of objects. The Davidsonian thinks that quantified first-order sentences couldn’t be true if there were no sequences of objects; but he does not think that users of first-order languages are committed to their existence. He has in mind strong commitment. Clearly, however, if Davidsonian semantics is correct, the user of a first-order language *is* weakly committed to sequences of objects.

We can now see more clearly the nature of Grover’s response; but we can also diagnose more accurately its inadequacy. Grover allows that the semantic theorist for her language is committed to propositions. She also allows that users of her language *express* propositions (but, she claims, so do users of other languages). However, she denies that users of her language are committed to propositions. Thus, she advocates

⁷² Compare *It is irrational to think that P and not Q* vs. *It is impossible that P and not Q*. Given Kripke’s “discovery” of the necessary a posteriori it is clear that there are more true substitution instances of the latter claim than of the former. (The former relation is accordingly stronger, since fewer things satisfy it.)

the existence of a *gap* between the commitments of the language user and the semantic theorist. It seems clear, however, that Grover only denies *strong* commitment to propositions on the part of her language users. Yet we want to know whether propositions exist, and are therefore able to have the property of truth – not whether the semantic relation that linguistic expressions bear to them is something other than *expression*. I maintain that propositions don't exist, hence aren't true or false. So I need to deny the claim that users of a language containing quantification into sentence position have a *weak* commitment to the existence of propositions.

The only way around the propositional realist's objection to the Argument from Natural Language Renderings that I can see is to claim that Grover's semantics is not to be viewed as *explanatory* - in the sense of giving the mechanisms of meaning for sentences involving quantification into sentence position – but serves instead as a *heuristic* device. The Quinean, don't forget, purports not to *understand* such quantification; he claims he does not know under what conditions sentences containing it are true or false. If, by speaking as though there were propositions, we can get him to see – even approximately - what is meant by object language sentences containing such quantifiers, then why should we not do so? If, moreover, we can discern a sense in which it is correct to say, while talking as though there were propositions, that object language users are not committed to, and do not talk about, such entities, then it seems we might go on to use this object language without ontological concern, once we abandon the fictional – or theoretical - perspective.

This response requires treating the language containing quantifiers binding sentential variables as primitively understood – our “home language” in Timothy Williamson's phrase. The Quinean claims he does not have such primitive understanding; but this contention might be thought to be disingenuous. As we have seen, what the Quinean wants to know is, when have the possible substitution instances of a higher-order quantification been exhausted? And he may say that, since

the semantics that Grover gives is now claimed to be no more than a heuristic guide, he still does not have anything more than a vague answer. But consider his own treatment of first-order quantification: he says that a sentence of the form $\lceil \text{Everything } Fx \rceil$ is true just in case, for any object o , $\lceil n Fx \rceil$ is true in that extension of the language in which $\lceil n \rceil$ names o . But he can't, by his own lights, mean *any* object o : it is not in general possible to give the semantics of a first-order language without either (i) using higher-order quantification, or (ii) using first-order quantification with variables which range more broadly than in the object language;⁷³ and since he rejects option (i), it seems he's stuck with (ii). So what *exactly* does he mean? He might say "anything o *in the object language domain*"; but this is no more precise than our vague, but heuristic, "any proposition". Of course, this does not prevent us from understanding first-order quantified sentences; but then a certain degree of vagueness does not prevent us from understanding higher-order quantification either.

So perhaps there is a defence available for Grover's Gap Thesis – her claim that the commitments of the semantic theorist are not shared by the language user – a defence which Grover herself does not (explicitly) give. Suppose instrumentalism about semantic theory were true.⁷⁴ Then there might plausibly be thought to be a gap between the ontological commitments undertaken by users of the object language and those undertaken by users of the language of the semantic theory. If so, then we could interpret "commitment" in the Argument from NLRs as *weak* commitment; and the conclusion of this argument would therefore be adequate for our purposes.

Is there any reason to think that semantic instrumentalism is true? Well, perhaps there is. Semantics must be pursued in a meta-language which is strictly

⁷³ See Williamson (2003).

⁷⁴ What exactly instrumentalism amounts to here is not yet clear to me. Some maintain that it is the view that the theory in question is useful but false; however, it might be the view that the theory is true, but not by virtue of corresponding to a further reality – instead its truth *consists* in its usefulness. I discuss different ways of understanding truth in Part III, Chapter 5; this will obviously be germane to the issue at hand.

stronger than the object language, in the precise sense that its variables must range over the elements of a larger domain. This means that we cannot consistently give a semantics for the language we use (without thereby changing languages). This suggests that we ought, strictly, to be quietists about the semantic features of our language. But semantics is useful; so we have a motive for the reification of further entities to enable its pursuit. Recognizing this, we should reject semantic realism, endorsing instrumentalism instead.

Summary and Conclusion

In the first chapter of this part, “Transcendence and Immanence”, I showed that arguments in favour of the view (*Transcendence*) that propositions are the metaphysically primary truth bearers are flawed. I then mounted my own argument in favour of *Immanence* – the view that concrete representations (whether linguistic or mental) are the metaphysically primary truth bearers. In the second chapter, “Higher-Order Quantification”, I argued, first, that one can explain the meaning of a truth *predicate* given a language containing a truth *operator* together with *quantification into sentence position*; and second, that this truth operator does not express a property of propositions (conceived as abstract objects), since quantification into sentence position does not commit us to the existence of any such objects. These considerations now license us to pursue the question of ontological commitment in natural language by attempting to understand how to explain truth for sentence tokens. In *Part II: Truth and Meaning*, I look at what form our best semantic theory should take. Then, in *Part III: Truth and Existence*, I show that we are in a position to reject the Quinean criterion of ontological commitment.

II. Truth and Meaning

In the first part of this thesis, I argued that concrete representations are the metaphysically primary truth bearers, and that the word “true” can be applied to sentence tokens. In this second part, I investigate what form our semantic theory should take in order to predict truth values for these concrete truth bearers. In particular, I attempt to assess Donald Davidson’s contention that the assignment of semantic values to predicates and sentences must be either trivial or useless. I begin, in “Truth Theoretic Semantics”, with a discussion of Davidson’s *Truth and Predication*, in which he presents the problem of predication, and attributes its solution to Tarski. I examine the proposed solution – Davidson’s famous truth theoretic semantics – as well as the contention that Tarski held this view. This leads, in the second division of this part (chapter 4), to a consideration of the rival, model theoretic approach to semantics, and in particular, the semantic theories of David Kaplan and David Lewis. I argue that although the semantic values attributed to predicates and sentences by these theorists are neither trivial nor useless, nevertheless both theories are irremediably philosophically inadequate as metaphysical accounts of meaning.

3. Truth Theoretic Semantics

Truth Conditional Semantics is a research program in linguistics and the philosophy of language - indeed, arguably the dominant such research program. Originating with Frege, its aim is to provide truth conditional semantic theories for specific languages, or language fragments – that is, as a first pass, theories which predict truth conditions for all of the sentences belonging to the language or fragment in question. There are, broadly speaking, two approaches to such Truth Conditional Semantics, which I will call *Truth Theoretic Semantics (TTS)*, and (for want of a better term) *Model Theoretic Semantics (MTS)*. According to truth theoretic semantics, the meaning of any given expression is literally a *condition*; and the meaning of a sentence in particular is a condition of *truth*.⁷⁵ By contrast, model theoretic semantics associates linguistic expressions with certain *entities*, or *semantic values*, which serve as their meanings.⁷⁶ The semantic values associated with full sentences are *propositions*, i.e. “abstract objects representing truth-conditions” (Stalnaker, 1999: 32); and thus, according to Robert Stalnaker, for example, “it is a semantical problem to specify the rules for matching up sentences of a natural language with the propositions they express” (Stalnaker, 1999: 34).

In a famous passage from “Truth and Meaning”, Davidson (1967b) objected to any semantic theory of the second, model theoretic kind. “Paradoxically,” he wrote,

the one thing meanings [i.e. semantic values] do not seem to do is oil the wheels of a theory of meaning – at least as long as we require of such a theory that it non-trivially give the meaning of every sentence of the language. My objection to meanings in the theory of meaning is not that they are abstract or that their identity conditions are obscure, but that they have no demonstrated use. (Davidson, 2001b: 20-21)

⁷⁵ Thus, according to TTS the meaning of an object language sentence *s* is given by a biconditional, one half of which is “*s* is true”; and similarly for other object language expressions.

⁷⁶ This characterization of the debate is due to Mark Sainsbury (2005).

This objection is, I think, best viewed as a dilemma to be faced by any semantic theorist postulating semantic values for predicates or sentences.⁷⁷ On the one hand, it is claimed, if the entities serving as meanings are specified in a non-trivial manner, then they do no semantic work that cannot be done without them. On the other hand, says Davidson, if these meanings are given in such a way as to guarantee that they are semantically efficacious, then it will be impossible to give anything more than a trivial characterization of them; and so they will be of no explanatory value.

The aim of this second part of the thesis is to assess this contention of Davidson's, and to determine whether the meaning of a sentence or predicate can be explained by its being associated with a semantic value. In addressing this issue I will focus on formalized regimentations of fragments of natural languages, rather than the natural languages themselves; however, I will try, from time to time, to keep in view the relations between these formalized languages and the expressions of natural languages of which they are formalizations.

The Problem of Predication

In his posthumously published book, *Truth and Predication*, Davidson claims that Tarski was the first to solve “the problem of predication”. He motivates this problem in connection with “the closely related problem of the unity of the proposition”, claiming that

[t]he puzzle emerges from the fact that once plausible assignments of semantic roles have been made to parts of sentences, the parts do not seem to compose a united whole. (Davidson, 2005: 4)

In particular, if we think of sentences as composed of words, or other constituents, and of those constituents as having semantic values, then it is hard to see how what a sentence expresses could be anything other than a sequence, or list, of those semantic

⁷⁷ All semantic theorists agree that many, if not all, genuinely singular terms (i.e. names) have semantic values, namely the things they are used to refer to – which is to say, their referents.

values. But lists are not evaluable as true or false, whereas sentential utterances are. The problem of the unity of the proposition, then, is to describe the semantic roles of sentential constituents in such a way as to allow (loosely speaking) that what a sentence expresses, a proposition, is a united whole susceptible of evaluation as true or false. The specific problem of predication is to do this for predicates; and this is really the crux of the general problem, “since a predicate is the only part of speech that must appear in every sentence” (Davidson, 2005: 2).

Why, according to Davidson, was Tarski the first to solve this philosophical puzzle? As a first pass we can say that Davidson thinks it is because he, unlike his predecessors, treated predicates as syncategorematic expressions – meaningful only inasmuch as they are constituents of sentences.⁷⁸ But then one might think that Davidson overlooks Frege’s semantic approach to predicates here. Wasn’t it Frege, after all, who articulated the context principle – who pleaded that we ask not for the meaning of an expression in isolation, but always in the context of a sentence? Indeed, wasn’t Frege the first to suggest that not only predicates but also other expressions, such as quantifiers, are syncategorematic?

Frege, of course, never developed a formal semantics for his *Begriffsschrift* or any other language; but he did make a number of precise, informal theoretical suggestions.⁷⁹ Amongst these, he famously maintained that predicates, like names, have referents. Moreover, Frege was aware of the problem of predication, and wanted sentences to function semantically as a united whole: he therefore distinguished the referents of predicates from those of singular terms – that is, he drew an *ontological* distinction between them, calling the former “concepts” and the latter “objects”. Thus,

⁷⁸ Davidson mentions Quine’s application of the term “syncategorematic” to predicates at (Davidson, 2005: 118); that he thinks Tarski treats predicates as meaningful only inasmuch as they contribute to the meaning of sentences in which they occur is clear from the subsequent discussion, for instance at (Davidson, 2005: 155).

⁷⁹ My principal source on all matters Fregean is Dummett (1981); this particular claim is based upon the discussion beginning at (1981: 81).

he took concepts to be not objects, but functions, distinguished from other functions (such as, for example, addition) by their counter-domain, or “range of values” – Fregean concepts are those functions which yield truth values as outputs. This approach to the semantics of predicates acknowledges a privileged role for sentences to play; for it is sentences which are truth apt. The semantic value of a predicate is simply what is needed in order to ensure that the semantic value of a sentence is a truth value, either *Truth* or *Falsity*.

Davidson acknowledges Frege’s unprecedented sensitivity to the problem of predication; nevertheless, he feels that Frege’s attempted solution is inadequate. One reason one might suspect for this is that recognizing the need for a syncategorematic treatment of predicates is not, in itself, to provide such an account; and “incomplete” or “unsaturated” as concepts may be claimed to be, they are nevertheless entities, associated with – and providing an interpretation of – predicates, *independently* of the role these expressions play in the semantic functioning of sentences. Thus, it may seem that sentences end up, even on Frege’s view, as no more than certain grammatically privileged lists.

This objection to Frege’s semantics is not Davidson’s, however, and for good reason. For on the Fregean view, concepts are functions, and the semantic values of singular terms are objects, which can serve as arguments to these functions. Moreover, syntactic concatenation expresses functional application; that is, the meaning of concatenating an n -place predicate with n singular terms is that one should *apply* the concept designated by the predicate to the individuals referred to by the singular terms, taken in the order in which the terms occur in the sentence. The result of this functional application will be a truth value, not a list of $n+1$ entities, amongst which one with an unusual ontological nature.

The source of Davidson’s dissatisfaction with Fregean semantics is not this, but rather the fact that Frege considered truth values to be objects. As a consequence, he

treated sentences as related to truth values in the same way as names are related to objects. This, in Davidson's view, is a mistake; the semantic unity of sentences is not the same as that of names – for sentences, unlike names, can be used to perform speech acts. Once the view that truth values are objects is abandoned, however, there remains no genuine explanatory value in the Fregean suggestion that the semantic values of predicates are functions – after all, functions are entities which yield *objects* as values. So the real problem with Frege's view lies in his associating sentences, and accordingly predicates, with entities serving as their semantic values.

As a second pass, then, we may say that on Davidson's view, Tarski succeeded, where his predecessors failed, in solving the problem of predication because he refrained from assigning semantic values to predicates at all. Tarski's semantics is preferable to Frege's because "it associates no entities which express generality with predicates" (Davidson, 2005: 159): predicates don't *stand* for anything ("saturated" or not), according to Tarski; rather, as Quine has put it, their reference is "divided" over the plurality of things of which they are true. But is Davidson's claim the weak thesis that assigning semantic values to predicates is *insufficient* to explain their semantic role? Or is it the stronger thesis that associating values with predicates is neither sufficient nor *necessary* for their semantic elucidation? Or, finally, is it the yet stronger contention that in order to solve the problem of predication it is necessary *not* to assign semantic values to predicates?

It should be plain that Davidson cannot mean simply to advance the weak thesis – for it seems an obvious and uncontentious point. As we have seen, the Fregean, for instance, acknowledges the importance of semantic composition: one must not only associate a predicate with a function, on Frege's view, but also recognize that in order to ascertain the semantic value of a simple sentence containing that predicate, the function associated with it must be *applied* to the argument or arguments specified by the singular term or terms involved in that

sentence. Indeed, any remotely plausible semantics for a complex language will recognize the semantic importance of syntactic composition. So Davidson must be seen as advancing at least the second, moderately strong thesis. In what follows I endeavor to determine whether he advocates this or the strongest thesis mentioned above, and whether either claim is sustainable. To this end it will be useful to look closely at Tarski's own work, and to compare this with Davidson's commentary.

Tarski's Truth Definition

In "The Concept of Truth in Formalized Languages," which Davidson cites in both "Truth and Meaning" and *Truth and Predication*, Tarski sets out "to construct – with reference to a given language – *a materially adequate and formally correct definition of the term 'true sentence'*" (Tarski, 1956: 152). Roughly speaking, a truth definition is deemed *materially adequate*, if and only if it entails all instances of the schema

(T) x is a true sentence if and only if p

where " x " is replaced with a structural descriptive name (i.e. a name employing only "morphological" terminology) of a sentence of the object language, and " p " is replaced with the meta-language translation of that sentence.⁸⁰ Such a definition is *formally correct* in Tarski's sense, again roughly speaking, just in case it employs only antecedently understood terminology.⁸¹ Indeed, Tarski goes further than is strictly

⁸⁰ It is perhaps worth remarking that a sentence enclosed in quotes is *not* a structural descriptive name; rather, such names consist of descriptions which mention the parts of sentences and the manner in which they are (structurally) related to one another, e.g. the symbol such-and-such, followed by the symbol so-and-so.... Thus, "Snow is white" is true if, and only if, snow is white, is *not* an instance of schema (T), but is instead a biconditional which is used to *motivate* the idea that a definition of truth which entails the instances of (T) is indeed materially adequate (i.e. has the right extension).

⁸¹ More precisely, Wilfrid Hodges says, "The definition of *True* should be 'formally correct'. This means that it should be a sentence of the form For all x , *True*(x) if and only if $\phi(x)$, where *True* never occurs in ϕ ; or failing this, that the definition should be provably equivalent to a sentence of this form. The equivalence must be provable using axioms of the metalanguage that don't contain *True*. Definitions of the kind displayed above are usually called *explicit*, though Tarski in 1933 called them *normal*." (Hodges, Wilfrid, "Tarski's Truth Definitions", *The Stanford Encyclopedia of Philosophy* (Summer 2006

required by the desire for formal correctness, writing, “I shall not make use of any semantical concept if I am not able previously to reduce it to other concepts” (Tarski, 1956: 152-153).

In the third section of the paper, Tarski accomplishes the task he has set himself for “the language of the calculus of classes”. The calculus of classes is the theory of what Stewart Shapiro (1991) has called the “logical sets”; if one’s domain of objects is D , the logical sets based on D are the subsets of D .⁸² The construction of the definition of truth for a language in which this theory can be expressed and developed requires clearly specifying the syntax of the object language. The language Tarski chooses contains expressions of two kinds: constants, and variables. There are four constants: a sign for negation ‘ N ’, a sign for disjunction ‘ A ’, a sign for universal quantification ‘ I ’, and finally, a symbol ‘ I ’, for class inclusion, i.e. the (possibly improper) subclass-of relation. For variables he employs signs consisting of the letter ‘ x ’ followed by any finite natural number, other than 0, of vertical strokes. These exhaust the simple expressions.

Tarski then proceeds to determine which complex expressions the object language contains. In short, any finite sequence of expressions is itself an expression. More interestingly, Tarski then characterizes what might be called the “meaningful” expressions. To this end he gives a recursive definition of “ x is a sentential function” – what we now mean by “ x is a sentence, open or closed”; in doing so he makes it clear that the object language employs Polish construction rules (i.e., the functional expression is always written before the argument expressions in any meaningful

Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/sum2006/entries/tarski-truth/>>.)

⁸² Thus, the Quinean ontological commitments of such a theory are not great. If the members of D are concreta, then the “classes” of members of D could, with the exception of the empty class, be construed as rather the mereological sums of their “members” – or, as this interpretation would have it, their parts. In short, there is no cumulative hierarchy of logical sets, or Tarskian classes – the most objectionable feature of set theory, according to Goodman and Quine (1947).

string). Finally, Tarski defines “ x is a free variable”, and uses this notion, together with that of a sentential function, to define a sentence (i.e. a closed sentence), as any sentential function with no free variables.

Since there are infinitely many sentences in the language of the calculus of classes, a materially adequate definition of the term “true sentence” cannot be given simply as the conjunction of all of the instances of the schema (T);⁸³ such a conjunction would be infinitely long, whereas the expressions of our meta-language are all finite in length. Consequently, Tarski looks to employ recursion, defining the truth of complex sentences in terms of simpler ones. However, as he points out,

in general composite [closed] sentences are in no way compounds of simple [closed] sentences. Sentential functions [i.e. sentences open and closed] do in fact arise in this way from elementary functions, i.e. from inclusions; sentences on the contrary are certain special cases of sentential functions. In view of this fact, no method can be given which would enable us to define the required concept [i.e. truth] directly by recursive means. The possibility suggests itself, however, of introducing a more general concept which is applicable to any sentential function, can be recursively defined, and, when applied to sentences, leads us directly to the concept of truth. (Tarski, 1956: 189)

The concept which Tarski employs for this purpose is that of an infinite sequence of objects (that is, a function from natural numbers greater than zero to objects – which, in this case, are classes of individuals) *satisfying* a given sentential function. He defines this notion as follows:

DEFINITION 22. The sequence f satisfies the sentential function x if and only if f is an infinite sequence of classes and x is a sentential function and these satisfy one of the following four conditions: (α) there exist natural numbers k and l such that $x = t_{k,l}$ [i.e. the symbol ‘ t ’ followed by the k -th and l -th variables] and $f_k \subseteq f_l$; (β) there is a sentential function y such that $x = \bar{y}$ [i.e. the negation of y ⁸⁴] and f does not satisfy the function y ; (γ) there are sentential functions y and z

⁸³ Such a definition could easily be rendered formally correct. See Tarski (1956: 188).

⁸⁴ Tarski writes the bar over, rather than under the letter, to indicate negation.

such that $x = y+z$ [i.e. the disjunction of y and z] and f either satisfies y or satisfies z ; (δ) there is a natural number k and a sentential function y such that $x = \bigcap_k y$ [i.e. the universal quantifier, followed by the k -th variable, followed by y] and every infinite sequence of classes which differs from f in at most the k -th place satisfies the function y . (Tarski, 1956: 193)

These clauses are much as one would now expect; except perhaps the first. This is the clause governing the symbol ' $\vdash$ ': note that the definition does not explicitly associate this primitive expression with a set, or other semantic value. Instead the symbol is simply translated into the meta-language, and rendered as ' $\subseteq$ '. The variables are interpreted as the corresponding members of the given sequence f .

Tarski then notes that whether a given function satisfies a sentential function depends solely upon the values assigned to the free variables occurring in that function. Thus, he is able to define truth for the language of the calculus of classes as follows:

DEFINITION 23. x is a true sentence - in symbols $x \in \text{Tr}$ - if and only if $x \in S$ [i.e. x is a closed sentence] and every infinite sequence of classes satisfies x . (Tarski, 1956: 195)

Tarski goes on to claim (Tarski, 1956: 195) that it is possible to prove that this definition is in fact materially adequate in the sense described above; each of the instances of the schema (T) is derivable from it in the meta-theory.

So does Tarski assign semantic values to the predicates of the language of the calculus of classes? Answering this question is a little more involved than one might expect. In the course of his description of the object language of the calculus of classes, Tarski never once uses the term "predicate": the expression " x is a predicate" is not part of the ideology, primitive or defined, of Tarski's meta-theory. Consequently, an examination of Tarski's work alone cannot tell us whether he assigns semantic values to predicates. We must, in addition, look to Davidson's discussion to determine what is meant by the term "predicate".

On this issue, however, Davidson is not entirely clear. Early on in *Truth and Predication* he gives the following definition: “Any expression obtained from a sentence by deleting one or more singular terms from the sentence counts as a predicate” (Davidson, 2005: 96). However, Tarski’s language contains no names, or definite descriptions. Thus its only “singular terms” are variables; but removing the variables from a closed sentence in which those variables are bound does not yield anything of interest to our concerns. Davidson follows his definition with some examples – “x loves y” is deemed a predicate, as is “x is tall and x is handsome and x is not wealthy”. However, on the previous page, he claims that

We... think of adjectives, like ‘green’, ‘voluptuous’, ‘just’, ‘square’, as parts of predicates, like ‘is green’, ‘is square’, ‘is just’. Common nouns, such as ‘man’, ‘animal’, and ‘skyscraper’, we also treat as inseparable parts of predicates, such as ‘is a man’, ‘is a skyscraper’, and so on. (Davidson, 2005: 95).

These two sets of examples conflict with one another, however, concerning the issue whether predicates contain variables as parts. On page 131, Davidson says, “we should think of the plus sign [and analogously, two place predicates] as containing two spaces... which are really *part* of the expression” (Davidson, 2005: 131, emphasis original); and in a footnote he says, “The letters at the end of the alphabet used to mark the spaces in a predicate should be considered part of the predicate.” (Davidson, 2005: 131, fn 9) However, this footnote continues: “the use of the same letters as the variables of quantification is entirely different.” Davidson never says how these uses differ; and in Tarski’s work nothing suggests that the variables are to be understood differently when they occur in open, rather than closed, sentences. In short, then, Davidson’s use of the term “predicate” leaves it unclear whether the symbol ‘ I ’ of Tarski’s object language should be considered a predicate, or whether such open sentences (i.e. sentential functions) as ‘ Ix_1x_1 ’ are correctly called by this name.

Stipulation is called for. In what follows, I will use the term “predicate” to mean any expression which combines with a finite number of terms to form a sentence;⁸⁵ thus, a predicate is not to be confused with an open sentence. As a result of this terminological decision, we can see that the sole predicate of Tarski’s object language is the simple symbol ‘*I*’. As we have seen, Tarski’s semantics explicitly associates no object with this expression. But how, exactly, does Davidson think this solves the problem of predication?

Davidson’s Solution: Truth Theoretic Semantics

Towards the very end of *Truth and Predication*, in a chapter titled “A Solution”, Davidson describes what he takes to be Tarski’s approach to semantics. The passage begins thus:

Tarski’s essential innovation is to make ingenious use of the idea that predicates are *true of* the entities which are named by the constants that occupy their spaces or are quantified over by the variables which appear in the same spaces and are bound by quantifiers. Because there is no particular limit to the number of free variables in a well-formed open sentence, Tarski introduces infinite sequences of the entities over which the variables range. Since both the sequences and the variables are ordered, any given sequence can be thought of as assigning entities to particular variables, as if those variables were performing the role of names. It is then possible to characterize the circumstances under which a given sequence assigns entities to the variables in a sentence which, were those variables the names of those entities, would create a true sentences. Such sequences are said to *satisfy* the sentence, whether the sentence be open or closed. (Davidson, 2005: 159, emphasis original)

There is, however, something odd about this characterization of Tarski’s approach. For it is no part of the official explanation of the notion of satisfaction that a sequence satisfies a sentence just in case, were the variables of the sentence to name the objects

⁸⁵ This differs from Davidson’s characterization of a predicate as what results from a sentence by removal of singular terms, since Davidson’s characterization includes things which are not expressions.

they are paired with by the sequence, the sentence would be true.⁸⁶ On the contrary, as we have seen, Tarski hoped to define truth in terms of satisfaction, which was taken to be the more basic notion.

Davidson continues his description of the solution-yielding approach to semantics as follows:

Spelling out the satisfaction relation... requires several steps.... First, there is an axiom for each sentence with an unstructured predicate (its spaces filled with variables or names) specifying the conditions under which that sentence is satisfied by a particular sequence. There will be a finite number of such axioms, since the basic vocabulary of any language must be finite. Second, there will be axioms recursively characterizing the satisfaction conditions of sentences built up from simpler sentences by the operations of negation, alternation, and the other sentential connectives, and, of course, the quantifiers. Since closed sentences contain no free variables, true sentences will be satisfied by all sequences, and false sentences by none. (Davidson, 2005: 160)

There are again some odd – if not downright mistaken – features of this description, if it is intended as a description of Tarski's techniques.

First, it should be clear from the above discussion that Davidson is simply wrong to claim that any language whatsoever must have a finite basic vocabulary. Tarski is quite explicit that his object language involves infinitely many distinct variables; indeed, it is for this reason that (just as Davidson says) he employs *infinite* sequences of objects. Perhaps Davidson has in mind that any *natural* language must have a finite basic vocabulary; though it is not clear that this is so either, if we allow that natural languages might themselves contain variables.⁸⁷

⁸⁶ It is arguable whether this may be part of the intuitive motivation for the view.

⁸⁷ Perhaps Davidson does not count the variables amongst the vocabulary, in much the same way that we would not count the letters of a natural language. But again, Tarski, at least, interprets the variables, whereas we don't interpret letters. Alternatively, Davidson might count each of the symbols 'x' and the vertical stroke as primitive vocabulary items. But this too seems strange, since neither is itself interpreted.

This mistake of Davidson's leads to another: if there were indeed one axiom of the theory defining truth for every sentence containing a simple predicate there would be infinitely many axioms in that theory; not because there are infinitely many simple predicates, but because there are infinitely many simple open sentences for each simple predicate, differing only in which variable "fills its gaps". Davidson has argued that if a language is to be learnable, then its semantic theory had better only contain a finite number of axioms. This desideratum could be met by means of a simple modification to the view he espouses in the above quotation: give one axiom for each simple predicate, but generalize over the variables with which it is coupled to form a sentence. This, in effect, is what Tarski did.

Except for one thing. Tarski did not employ any primitive (i.e. undefined) semantic terms.⁸⁸ Thus, he did not invoke *axioms* governing them. Rather, he gave a recursive *definition* of the notion of satisfaction – devoting one *clause* of the definition for each simple predicate of the language (in the case reviewed above, just '*I*') – and then he defined truth in terms of satisfaction. He also remarked that his recursive definition of satisfaction could be converted into a provably equivalent *explicit* definition. However, Davidson acknowledges that his own preferred approach differs from Tarski's in this respect: he says, "I have forsworn the step which yields explicit definitions, and am therefore regarding Tarski's constructions as axiomatizations of the intuitive, and general, concept of truth." (Davidson, 2005: 160)

Let's take stock. Davidson, as we have seen, credits Tarski with the solution to the problem of predication; but now it seems that the position which Davidson himself favours is not exactly, i.e. not in all details, the one Tarski advocated. We are left with the following questions:

- (1) What does Davidson think the solution to the problem of predication is?

⁸⁸ At least, he did not do so when discussing the languages of finite order, and the language of the calculus of classes is one of these.

(2) Is Davidson right to attribute to Tarski what he sees as the solution?

(3) Is what Davidson considers a solution in fact a solution?

The third question is intimately related to the overarching issue with which this chapter deals; the answer to it will therefore have to wait. The second question is taken up in the next section. I begin here by addressing question (1).

I believe that an examination of Davidson's work, both in *Truth and Predication*, and in the various papers collected in *Inquiries into Truth and Interpretation*, licenses the following conclusion. The solution to the problem of predication, as Davidson sees it, lies in giving a truly syncategorematic treatment of predicates. This involves explaining the meaning of each predicate in terms of the satisfaction conditions of open sentences containing it, AND explaining these, in turn, in terms of the truth conditions of closed sentences that result from replacing the variables uniformly by names. According to Davidson, this is just what a "Tarskian" theory of truth for a language L does; and this is why Tarski was the first to solve the problem. Let me unpack this conclusion. There are two features of the Davidsonian solution: I take each in turn.

According to Davidson, a finitely axiomatizable truth theory for a language L can serve as a theory of meaning for L, where a truth theory for a language L is a set of sentences (i) containing axioms from which a *T*-sentence can be derived for each sentence of L, and (ii) closed under derivability.⁸⁹ Now, Tarski showed us how to construct truth theories for a range of languages, and his theories are indeed finitely axiomatizable. But in Tarski's theories the semantic terms are given explicit definitions, and can therefore be eliminated from any theorem (or other sentence) in which they occur. Davidson's proposal is to employ a theory which is formally analogous to Tarski's, but which treats the semantic expressions themselves –

⁸⁹ What is meant by a *T*-sentence is simply an instance of Tarski's schema (T) above. This characterization of a truth theory – and of a *T*-sentence in particular – would need to be modified to account for, e.g., indexical languages; but it will serve our present purposes.

satisfaction, truth, denotation – as undefined primitives. And he claims that such theories can elucidate the meanings of object language expressions.

How can a truth theory of this kind serve as a theory of meaning? First, notice that, as Davidson says, “a satisfactory theory of meaning must give an account of how the meanings of sentences depend upon the meanings of words” (Davidson, 2001b: 17). But if we take the expression “true sentence” as it occurs in the schema (T) at face value, i.e. as meaning *true sentence*, then the instances of this schema give the truth conditions of the sentences of the language L under investigation (the object language); which, according to the underlying assumptions of Truth Conditional Semantics, is to say that these theorems of the truth theory give the meanings of the object language sentences. Thus, if the axioms of the theory concern (a) the lexical primitives of L, and (b) the manners of their composition; and if (c) the *T*-sentences can be derived from the axioms – if all this is true – then a truth theory meets the requirement stated above.

Second, it seems that such a theory, being finitely axiomatizable, could be learned. That is, one could come to know each of the axioms; and then knowledge of the laws of logic would suffice to guarantee (potential) knowledge of the infinitely many theorems. Moreover, if one were to explicitly learn the theory in this way (i.e. by explicitly learning the axioms), it seems that one would have knowledge sufficient for understanding utterances of speakers of the language – or so, at any rate, claims the Davidsonian. For equipped with the knowledge afforded by explicit learning of the theory, one would be in a position to derive a *T*-sentence for any object language sentence a speaker might produce; one would therefore know its truth condition. If one then knew whether that condition obtained, one could assess the utterance as correct or incorrect, according as the sentence uttered was true or not;⁹⁰ and if one did

⁹⁰ See, e.g., Evans (1985: 345-347).

not have this additional knowledge, one could nevertheless tell *what* that speaker was claiming to be the case.

However, notice that we've snuck something illegitimate into our discussion, right at the outset. We've said that the truth theory gives *the* truth condition for each sentence of L; and the use of the definite article implies uniqueness. But the truth theory employs a material conditional, so there are many truth conditions for a given object language sentence – any sentence that is materially equivalent to the OL sentence will serve to give one of its truth conditions. So what guarantee do we have that the truth condition supplied by the theory is the truth condition that proponents of TCS have in mind when they say meaning is truth conditions?

In Tarski's work, we know that the sentence replacing '*p*' in the schema (T) is a *translation* into the metalanguage, of the object language sentence replacing '*s*'. But we can't construct a theory of meaning for a language we don't understand under the simple methodological constraint that we employ translations of its sentences! We need some way of ensuring that we get translations *as a result* of our theory of meaning. Davidson (2001b: 26, fn 11) suggests that a theory of truth discerning semantic structure (as it must if it is to be finitely axiomatized), if confirmed by empirical evidence, would yield *lawlike T-sentences*; and it is for this reason that we may view it as giving information about meaning.

Let's turn now to the second feature of the Davidsonian solution to the problem of predication: the satisfaction relation must be explained in terms of truth. The essential thought is already expressed in the passage quoted at the beginning of this section: the idea is that an open sentence is said to be satisfied by a sequence of objects just in case, were the variables names of the objects with which they are associated by the sequence, the sentence would be true. In order to see what is at issue more clearly, however, it will be worth looking at Gareth Evans' (1977)

discussion of “Fregean” versus “Tarskian” approaches to the semantics of quantification in “Pronouns, Quantifiers, and Relative Clauses (I)”.

Evans points out that there is a problem to be resolved by semanticists, which is posed by the fact that connectives and quantifiers can be used to form both complex sentences and complex “predicates”.⁹¹ Thus, “Snow is white and maple sugar is sweet” is formed by conjunction from simpler sentences, while “Someone is young and conservative” contains a conjunctive predicate. Similarly, “loves someone” is formed by existential generalization from “loves”, and takes a name (or quantifier) to make a sentence, though “Someone waits” is a sentence formed from the simple predicate “waits”. This syntactic fact poses a semantic problem, because the natural way of discussing the semantics of sentences is in terms of truth conditions, and that of predicates in terms of satisfaction conditions; yet it is implausible to suppose that “and” or “someone” is ambiguous in English, having both truth-conditional, and satisfaction-conditional, meaning. Evans claims that there are two ways to resolve this difficulty: one can either assimilate truth-conditions to satisfaction conditions, or vice versa; he calls the former approach “Tarskian”, the latter “Fregean”.

To illustrate the difference between these semantic views, as regards the quantifiers, Evans asks us to consider a language fragment with only monadic predicates; this allows him to simplify matters by treating satisfaction as a relation between sentences and *objects*, rather than sequences thereof.⁹² On the “Tarskian” approach to the semantics of such a language one assigns satisfaction conditions to sentences both open and closed, and defines truth for closed sentences as satisfaction by all objects. Evans writes:

⁹¹ Evans uses the term “predicate” slightly differently again; though I think his usage, as concerns *natural* language, corresponds to the definition given by Davidson and quoted above – it is what results from a sentence by deletion of names.

⁹² The fact that the syntax of such a language does not allow the *formation* of the problematic predicates is irrelevant from the point of view of semantic illustration.

[T]he most natural way of stating the semantic effect of the quantifiers would be in clauses which spoke of truth, along the lines of

(C) A sentence of the form 'Something'^A is true iff something satisfies A.

To give closed sentences the properties Tarski requires, (C) must be replaced by a principle which states the impact of the quantifiers in terms of satisfaction:

(D) An object satisfies 'Something'^A iff something satisfies A.

... [T]his has the form of ' $(x)(Fx \equiv R)$ ', and the effect that a closed sentence is satisfied by all objects iff it is true, and by no objects iff it is not true. (Evans, 1985: 82-83)

Clearly, truth is a derivative notion in a "Tarskian" semantics; it is satisfaction that does the theoretical work.

Whereas the "Tarskian" appeals officially to (D), by contrast, on the "Fregean" approach,

We may use a simple principle for the quantifiers like (C) but the relation of satisfaction which holds between an expression and an object to which that clause directs us is, in the case of a complex predicate, defined in terms of the truth value of the sentence which results when a singular term referring to that object is substituted in the predicate, or, if the language contains no name for the object, in terms of the truth value, in some extension of the language, of a sentence which results when a singular term which refers to that object is substituted in the predicate. (Evans, 1985: 84)

The effect of this is that, at any stage in the evaluation of a quantified (or other complex) closed sentence, one need only look to the evaluations of simpler closed sentences (though not necessarily sentences belonging to the language); consequently, one can appeal, ultimately, to truth conditions, and only truth conditions, in giving the semantics for the language. Indeed, Evans goes on to say, still discussing the "Fregean" semantics:

I shall... collapse the two principles [(C) and the general explanation of the satisfaction relation] into something along the more familiar lines of:

A sentence of the form ‘Something’^A is true iff, upon some extension of the language, there is a substitution instance of the form β^A which is true. (Evans, 1985: 85, fn 9)

Thus truth is the central semantic notion for the “Fregean”, satisfaction for the “Tarskian”.

Davidson prefers the “Fregean” approach. He makes it clear that he believes that truth is “the clearest and most basic semantic concept we have” (Davidson, 2005: 160). Moreover, he claims that the only evidence that can, in principle, be adduced in favour of a truth theory for a language concerns the truth conditions of whole sentences; that is, the *T*-sentences, but not the axioms governing the satisfaction conditions of open sentence, can be confirmed empirically.⁹³ It is perhaps due to this preference for the “Fregean” position that Davidson’s approach to the semantics of natural languages is called “Truth Theoretic Semantics”.⁹⁴

What is relevant for our present purposes, however, is that only those who view the satisfaction relation instrumentally – that is, those who view the semantics of predicates as purely instrumental in giving the truth conditions of closed sentences containing them – count as solving the problem of predication for Davidson. Did Tarski so view them? The nomenclature chosen by Evans in the preceding discussion suggests not; but of course the fact that Evans *calls* a certain view Tarskian does not make it so. We will address issue in the next section.

Before doing so, however, I would like to clarify one point about the solution to the problem of predication, as Davidson sees it. This concerns the fact that Tarski,

⁹³ See, for example, Davidson (1973b) at (2001b: 133-138); (1977b); and finally (1979), where he says that “all the evidence for or against a theory of truth... comes in the form of facts about what events or situations cause, or would cause, speakers to assent to, or dissent from, each sentence in the speaker’s repertoire” (Davidson, 2001b: 230). It is clear that Davidson has in mind *closed* sentences in this passage.

⁹⁴ Evans’ position seems to be somewhere midway between Davidson’s “Fregean” position and the “Tarskian” approach described above: for, like the Fregean, Evans thinks truth is the most basic semantic notion; but unlike Davidson, he believes that there might be evidence confirming not only the *T*-sentences, but also the *axioms* of a truth theory for a natural language. See Evans (1981).

and Davidson following him, assign no entities representing generality to the predicates of the object language. I imagine the following objection. While Tarski has not *explicitly* assigned a semantic value to predicates, or for that matter to sentences, he nevertheless has done so *implicitly*. Consider the case of sentences: Tarski has paired each closed sentence of the language with either the set designated by '*Tr*', or else its complement (relative to the set of sentences of the object language). This is indicated in the symbolic representation of '*x is a true sentence*' by the use of the symbol for set membership. Thus he assigns truth values as the semantic values of sentences. Similarly, though perhaps less obviously, Tarski implicitly associates a set with the predicate '*T*', namely the set of pairs satisfying simple open sentences containing this expression, $\{ \langle x, y \rangle : \exists k \exists l \exists f (f \text{ is an infinite sequence of classes, } k \text{ and } l \text{ are natural numbers greater than } 0, f \text{ satisfies } I_{V_k V_l} \text{ and } (f_k = x \wedge f_l = y)) \}$.

Although Tarski writes in a way that presupposes that '*Tr*' designates an object, this assumption is due to the strength of his meta-theory, and his desire for an explicit definition of truth as a set of object language sentences. Had he been satisfied with an axiomatization of truth, Tarski could just as easily have used a one place predicate '*T*' as the two-place predicate '*∈*' together with a singular term (and similarly for '*x is a closed sentence*').⁹⁵ Had he done so, there would have been no inclination to think that he was associating sentences with semantic values, i.e. with truth values conceived as objects. A similar point applies to the predicate '*T*': Tarski simply translates the object language predicate without assigning it a semantic value; the possibility of then constructing an object which can be associated with it in the meta-theory, or the meta-meta-theory, is irrelevant. It may be worth comparing Davidson's comment in this connection. He says:

If we want to postulate entities such as meanings and propositions, we can explain what these entities are only insofar as we can explain how

⁹⁵ Indeed, we should view a Davidsonian truth theory as doing just this.

singular terms refer to objects and predicates are true of objects. There is no point in supposing we can *first* provide a clear account of meanings and on that basis arrive at an account of naming or reference, and of predication. (Davidson, 2005: 143)

In short, once an adequate theory of meaning is given, we can go on to construct objects which we then associate with the expressions of the language; and we may call them “meanings”, or “semantic values”, if we like. But we should not suppose that in doing so, we have explained the meaningfulness of object language expressions; on the contrary, we have presupposed such an explanation.⁹⁶

These considerations lead us to an answer to our earlier question concerning the strength of Davidson’s contention regarding semantic values for predicates: is it that they are not necessary for an account of the semantics of predication, or is it rather that it is necessary not to have them? I think it should be clear now that only the former claim is plausible;⁹⁷ associating some objects with predicates *post facto* cannot be objectionable in itself; in particular, it cannot impugn the claim of a theory to have explained predication.

Was Tarski a Davidsonian?

There are a number of reasons for not attributing the Davidsonian solution to the problem of predication to Tarski. First, when Tarski discusses the semantics of higher (but finite) order languages, he seems to allow that predicates, and the variables which replace them, are semantically associated with entities. He says, “The first sign of... a complex [meaningful] expression is always the name of a class or

⁹⁶ Two other points are relevant here. One is that the “semantic values” which are claimed to be implicitly associated with predicates and sentences do no semantic work in a truth theory of the sort under consideration, in the precise sense that the semantic evaluation of no expression depends (explicitly) on the semantic value of any other expression. The other point is that, in the case of sentences, the alleged semantic values, i.e. the truth values, are not ontologically independent of the linguistic expressions of which they are (claimed to be) values. But as Kit Fine (2003) points out, that is not how we ordinarily think of *semantic* values.

⁹⁷ Whether it is in fact true is a further question; we will see some reasons below to doubt that it is.

relation or a corresponding variable.” (Tarski, 1956: 213) But, if we ignore the variables, the first sign of such an expression is a monadic or relational predicate: so such expressions “name” certain entities, namely classes and relations. Also, in giving the semantics for these languages, Tarski does not simply translate the predicate variables – how could he? Instead he renders expressions like “ $Xxy...z$ ” as “the I -termed relation X holds between the I individuals $x, y, ..., z$.” (Tarski, 1956: 231) In the meta-language the expression “ X ” is clearly being used as name-like, to stand for an entity. So the fact that, while giving the semantics for the language of the calculus of classes, Tarski assigns no semantic values to predicates might be seen as accidental. Indeed, as we shall see, it is a central feature of Tarski’s model theoretic approach to the study of logical consequence that one assign semantic values to predicates.

Perhaps this evidence is not overwhelming, particularly given Tarski’s view of naming (Tarski, 1956: 194). Writing long before Kripke, Tarski had no notion of a rigid designator, and the standard at the time was not to think of reference as a direct relation, with the referent of an expression constitutive of meaning. (Russell, of course, is a notable exception here.) Thus, it seems that Tarski’s view may well have been that, what a name names is not directly relevant to, because only a by-product of, its meaning. If so, this applies in the case of the naming relation which he took to hold between predicates and the classes or relations they name. Nevertheless, the point remains: Tarski does assign semantic values to predicates of languages of higher, but finite, order.

A second line of thought runs as follows. Even granting that Tarski assigns no semantic value to the sole predicate of the language of the calculus of classes, nevertheless, it is not clear that he treats satisfaction in the way that Davidson suggests one should. Might we not ask whether the defined satisfaction – and in a language with names, the denotation – relations are materially adequate? Hartry Field suggests just this. He writes:

A sentence of the form $(\forall N)(\forall x)[N \text{ denotes } x \equiv B(N,x)]$ *satisfies convention D* if it has as consequences every instance of the schema ' y denotes z ', in which ' y ' is to be replaced by a quotation-mark name for a name N , and ' z ' is to be replaced by (an adequate translation of N into English, i.e.) a singular term of English that contains no semantic terms and that denotes the same thing that N denotes. (Field, 1972: 18, fn 11)

The thought is that a definition of denotation will be materially adequate if it satisfies convention D,⁹⁸ just as a definition of truth will be materially adequate just in case it satisfies Tarski's convention T (i.e., entails all of the T -sentences for the object language L).

Tarski himself does not explicitly ask this question about the denotation relation. However, when he first introduces the notion of satisfaction, he writes as though we already have some grasp of the notion he is concerned with: "Let us try... to make clear by means of some examples the usual meaning of this notion in its customary linguistic usage" (Tarski, 1956: 189). So the question of material adequacy could, in principle, be asked and answered; and this suggests that Tarski does not view the satisfaction and denotation relations instrumentally. If so, then his treatment of predicates is not properly syncategorematic, in the way Davidson requires.

But third, and perhaps most importantly, it is not clear that Tarski is even doing semantics in the sense of attempting to explain the meanings of the object language predicates. Tarski showed us how to construct a *definition* of truth for a given language. But, of course, any defined term can be replaced, whenever it occurs, by the expression which serves to define it: thus, the expression "true sentence" can be eliminated throughout a Tarskian truth theory, and in particular whenever it occurs in a T -sentence. But the primitive vocabulary of Tarski's meta-theory all belongs to two kinds: those capable of expressing logical/set theoretical notions, and

⁹⁸ I should note that *Field* introduces convention D only to argue that its truth would not suffice for a reduction of denotation to non-semantic notions.

those relating to the object language syntax (or as Tarski sometimes calls it, “morphology”). Thus, the sentences which result from the *T*-sentences by substitutions of this kind - and which are therefore equivalent to those *T*-sentences - express truths of logic, set theory and syntax. These, however, are of no immediate interest to the empirical, Davidsonian semanticist.^{99,100}

John Etchemendy (1988) takes this view of Tarski’s work. He claims that the Tarskian *T*-sentences are semantically uninteresting, but that this is obscured by the fact that Tarski gave a *recursive*, rather than an *explicit*, definition of satisfaction. As a result, his truth theory, when supplemented with the claim - which Richard Heck (1997) dubs a “connecting principle” - that the defined term “true sentence” applies to all and only the true sentences of the language, is logically equivalent to a Davidsonian truth theory.¹⁰¹ Thus, it is easy to read the Tarskian theory *as* the homophonous, but semantically substantive, Davidsonian theory.

However, according to Etchemendy, it would be a mistake to read the Tarskian theory in this way. For Tarski was not *trying* to contribute to “formal semantics” in the Davidsonian sense – that is, to empirical semantics. Instead, his aim was to prove the consistency of the notion of truth as it applies to certain languages. This notion was (i) deemed to be of central scientific importance,¹⁰² and (ii) thought to lead to inconsistency in the form of the liar paradox. The provision of an explicit definition of object language truth in the meta-language allowed for a proof of the consistency of the theory containing this notion, relative to that of the original meta-theory. Moreover, says Etchemendy, “there is an important sense in which [Tarski’s]

⁹⁹ This point is well rehearsed: see Field (1972), Putnam (1985), and Etchemendy (1988), and references cited therein. Heck (1997) also states this case very clearly, although he goes on to disagree with it.

¹⁰⁰ This is sometimes illustrated by appeal to the claim that truths of semantics are contingent, while the Tarskian *T*-sentences are necessary.

¹⁰¹ For a qualification of this claim, see Heck (1997: 540, fn 14).

¹⁰² Tarski (1956: 79) claims, for instance, that “[a]s soon as we succeed in showing that an empirical theory contains (or implies) false sentences, it cannot be any longer considered acceptable.” Of course, false sentences are just sentences (of the relevant language) that are not true.

own goal conflicts with that of the other [Davidsonian, semantic] project” (Etchemendy, 1988: 52). The reason he gives is that endorsing the above connecting principle involves “the uneliminated use of a notion of truth” (Etchemendy, 1988: 60).

Heck disagrees with this claim, however, and argues that “Tarski’s work... simultaneously contributes both to empirical semantics and to the resolution of the semantic paradoxes” (Heck, 1997: 534). He points out that Tarski did in fact employ a recursive, rather than an explicit, definition of satisfaction,¹⁰³ and that he insisted on the material adequacy of his definition of truth. Heck takes this last point to mean that Tarski professed belief in the connecting principles which, when conjoined to the Tarskian theory, yield a theory equivalent to the Davidsonian theory. In light of this, Heck claims, it would be at best uncharitable not to credit Tarski with the semantic insight of the Davidsonian project.

It is important to note that Tarski *did* think that his work in “theoretical semantics” would be relevant to “descriptive” empirical linguistics. He wrote:

The relation between theoretical and descriptive semantics is analogous to that between pure and applied mathematics, or perhaps to that between theoretical and empirical physics; the role of formalized languages in semantics can be roughly compared to that of isolated systems in physics. (Tarski, 1956: 77)

These remarks are quite cryptic – How does Tarski view the role of isolated physical systems? What insight is the comparison meant to afford? – but their intent is clear: Tarski was insisting that his work in semantics has empirical applications.

Scott Soames (1984) discusses the philosophical interpretation of Tarski’s work on truth,¹⁰⁴ and gives a reading of it which enables us to resolve some of the issues

¹⁰³ Moreover, he insists that this feature of Tarski’s definition is *essential*; Tarski would not have been able to prove the various laws of truth he goes on to prove, had he given a more minimalist, or list-like definition. See Heck (1997: 548-551).

¹⁰⁴ In addition to “The Concept of Truth in Formalized Languages,” there is also Tarski’s (1956) work on model theory that is relevant here, and in what follows.

broached by the preceding discussion. He begins by giving a taxonomy of theories of truth. He writes:

Generally, theories of truth have tried to do one or the other of three main things:

- (i) to give the meaning of natural-language truth predicates;
 - (ii) to replace such predicates with substitutes, often formally defined, designed to further some reductionist program; or
 - (iii) to use some antecedently understood notion of truth for broader philosophical purposes, such as explicating the notion of meaning or defending one or another metaphysical view.
- (Soames, 1984: 411)

Soames claims that Tarski's theory of truth is of the second kind. It can't be a theory of the first kind, he says, since it does not attempt to give an account of propositional truth. Nor can it be of the third kind, since Tarski views the intuitive notion of truth with suspicion.

How exactly does Tarski's truth theory contribute to a reductionist program? And what program is that? According to Soames, Tarski was a "moderate" physicalist. Thus he believed that all facts are either physical, or logico-mathematical. Accordingly, all scientific descriptive terminology must be explicitly definable in some combination of physical and logico-mathematical terms; but, due to his moderation, the logico-mathematical terms themselves need not be explicitly defined in physical terms. Now, Tarski's truth definitions contribute to the project of reducing all scientific language to a primitive vocabulary of physical and logico-mathematical terms because they allow us to replace, for each object language, the specifically *semantic* vocabulary pertaining to that language with non-semantic vocabulary, as we saw above.

How, though, can Tarski's theoretical semantics be applied on this view? The *T*-sentences, as we have seen, express truths of logic, set theory, and syntax, and would therefore appear to have no bearing whatsoever on the meanings of sounds issuing from the lips of individual speakers: they are not contingent, counterfactual

supporting, law-like claims governing the noises produced in a given population. Soames' solution to this quandary is, following David Lewis (1975), to claim that languages, on Tarski's view, are abstract objects, which have their (Tarskian) semantic properties essentially - a claim which is consistent with the fact that the *T*-sentences, for example, express necessary truths of logic and mathematics. What is contingent is the fact that a given language, rather than some other, is in use in a given population. Soames claims that the question what language a population uses is not a semantic question, but a *pragmatic* one - presumably because it concerns the relation of expressions to their users. While I acknowledge that this question is not strictly semantic, I prefer, following Robert Stalnaker (2003), to call it "meta-semantic".

Stalnaker introduces the distinction between descriptive semantics and meta-semantics as follows:

Descriptive semantic questions ask what the semantic value of certain expressions or certain kinds of expression are. Foundational or metasemantic questions ask what the facts are in virtue of which expressions have the semantic values they have. (Stalnaker, 2003: 15)

Stalnaker's characterization of the distinction between the two levels of what are broadly semantic studies assumes that expressions all have semantic values. But this feature of his way of drawing the distinction is inessential: what is crucial is to keep in mind the difference between the questions, What does a given expression mean? and What explains the fact that a given expression means what it means? By way of illustration of the difference between semantics and meta-semantics, Stalnaker gives the following example:

The thesis that proper names are rigid designators is an answer to a descriptive question about names, while a causal theory of reference is an answer to the foundational or metasemantic question about what it is that explains why names have the referents that they have. (Stalnaker, 2003: 15)

This example involves names, which are agreed to have semantic values; but it should be clear that the issue is about description vs explanation.

I accept this distinction between semantics and meta-semantics. Framing the issue in this way, Soames' contention is that Tarski's languages have their *semantic* properties essentially, hence necessarily; whereas the question which language is in use in a given population is a contingent, empirical one to which Tarski gave no answer. He does not address the issue of when it is that a population P *uses* a language L; in other words, he does not engage in meta-semantics. Whether Tarski's moderate physicalism is to be deemed ultimately successful will depend upon whether an account can be given of this meta-semantic relation that is consistent with this metaphysical view: whether, that is, the terms needed to account for this relation can be explicitly defined using only the terms of physics, logic, and mathematics.

One account one might give of when a population uses a given abstract language is the following: P *uses* L iff, for all sentence types S of L, and sentence tokens u in P, if u is of type S then u is true iff S is L-true. Here "true" is a pre-theoretic, intuitive notion, whereas "L-true" is the defined, Tarskian notion. This is quite similar to the connecting principle considered above (that all and only the Tarski-truths of L are true). Such a connecting principle determines the "semantic" properties of sentence tokens, the products of sentential utterances; and, on such a view, the facts which render the connecting principles true are meta-semantic. We might view one who adopted this account of the use of a language as advocating Heck's reading of Tarski's work.¹⁰⁵ But I think it is clear that Tarski himself would have rejected the connecting principle (despite his insistence on the material adequacy of his definitions), and with it the above account of language use; and this on two grounds. First, it employs the suspect, everyday notion of truth, which, as Soames stresses, Tarski was at pains to eliminate, or at least replace. (This goes some

¹⁰⁵ In fact, Heck (1997) gives two readings of Tarski's work, though I have only discussed one here.

way to vindicating Etchemendy.) And second, it does nothing to further Tarski's reductionist project of moderate physicalism, employing, as it does, a primitive semantic notion. It is safe to conclude, therefore, that Tarski would instead have looked for a different story to tell about when a population uses a language than the one given above.¹⁰⁶ But if this is correct, then Tarski cannot be seen as trying to explain the metaphysics of meaning by relating it to that of truth; and this despite the fact that Tarski's work, *as he himself understood it*, can be given empirical application.

I object, therefore, to Davidson's attribution to Tarski of the solution to the problem of predication sketched in the previous section: for Tarski does not explain predication syncategorematically by its relation to the truth and falsity of sentences. Of course, this complaint does not apply to Davidson's own approach to semantics. Davidson's idea is to employ a primitively understood truth predicate, and relate it to the satisfaction relation in a manner that is structurally analogous to the way in which Tarski does. A theory which does this for an entire language, and which is constructed in accordance with certain empirical methodological constraints, will, according to Davidson, yield meta-language interpretations of all object language sentences without assigning semantic values to them or their constituent predicates. Whether this method is ultimately viable is not something I will attempt to assess immediately; but if it is, it seems Charles Parsons' comment, made in another context (in a footnote of Davidson's book), would be appropriate regarding the solution to the problem of predication: "I would complain that Davidson gives too much philosophical credit to Tarski and too little to... himself" (Davidson, 2005: 152, fn 14).

We have been considering Davidson's discussion of the problem of predication, in the broader context of assessing whether the meaning of predicates and/or sentences can be explained by the assignment to them of semantic values. We have

¹⁰⁶ We will consider the question of when a population may be said to use a given language in more detail in chapter 5.

seen what Davidson regards as the solution to the problem of predication, namely truth theoretic semantics. This consists in the provision of truth theories for natural languages, which assign no entities to serve as meanings for predicates or sentences, but which instead provide necessary and sufficient conditions of truth for sentences, and explicate the semantic contributions of predicates in terms of the satisfaction conditions of open sentences containing them; finally, satisfaction is explained in terms of truth (in such Davidsonian semantics). We have also seen that Tarski did not provide semantic theories of this Davidsonian kind, and so Davidson's attribution to Tarski of this solution to the problem of predication is strictly incorrect. What is more accurate to say is that Tarski provided the formal tools for the solution Davidson favours. Finally, in the course of our discussion of Tarski's views, we encountered the distinction between semantics and meta-semantics, which will play a significant role in the remainder of the thesis.

In the next chapter of *Part II* we will examine model theoretic approaches to semantics. Since these assign semantic values to predicates, and sometimes to sentences as well, this will provide an opportunity for us to assess Davidson's contention that such meanings are either trivial or useless. We will accordingly be in a position to determine whether Davidson's truth theoretic semantics provides the solution to the problem of predication, and the best semantics for natural languages.

4. Model Theoretic Semantics

Introductory Remarks: Etchemendy on Model Theoretic Semantics

I want to turn now to consider a particular brand of truth conditional semantic theory which assigns semantic values to expressions, namely Model Theoretic Semantics (MTS). Following Wilfrid Hodges, we may say broadly speaking that “model theory is the study of the interpretation of any language, formal or natural, by means of set-theoretic structures” (Hodges, 2005: introduction). More precisely, a model theoretic semantics for a language *L* is a theory consisting of two parts: (i) a definition of a class of objects, or structures – the models for the language; and (ii) a recursive characterization of truth relative to a model for sentences of the language.¹⁰⁷

In *The Concept of Logical Consequence*, John Etchemendy claims that there are two alternative philosophical interpretations of MTS available, which he calls *representational* MTS, and *interpretational* MTS. Following Davidson (1973a), Etchemendy remarks that *truth*, as intuitively understood, is a property, while the technical notion of *truth in a model* expresses a relation.¹⁰⁸ We may, of course, connect the notion of absolute truth with that of relative truth characterized by model theory: absolute truth – that is, *truth simpliciter* – is some *specification* of truth in a model; it corresponds to truth in some particular model.¹⁰⁹ The question whose answer divides the two philosophical interpretations of model theory is then, What does it take for a model to be the right, or intended one; for truth in *it* to be, or

¹⁰⁷ Cf. Etchemendy (1999: 1): “In these [introductory] texts [in mathematical logic] we are taught how to delineate a class of models for a simple language and how to provide a recursive characterization of truth in a model – in short, how to construct a simple *model theoretic semantics*.”

¹⁰⁸ To put this in the metaphysically neutral, formal mode: “true” is a monadic predicate, while “true in” is dyadic.

¹⁰⁹ Thus, specification is contrasted with generalization, either existential or universal. (To be a brother is to be a brother of someone; to be benevolent is to be good to everyone.)

represent, *truth simpliciter*? What is it that varies from model to model, such that absolute truth may be thought of as truth in the one particular variation?

According to the first interpretation, we are to think of models as each representing a way the world could be, and truth in a model as akin to truth in a possible world. On this approach, says Etchemendy, models are used to systematically describe how the semantic evaluation of our linguistic expressions depends on how the world is. Absolute truth is captured by the model which accurately represents the world; the one which represents how the world actually is. This is model theory as *representational* semantics.

On the second interpretation of model theory, we think instead of the different models as encoding different ways our linguistic expressions could legitimately be used; that is, as representing different acceptable interpretations of the language under investigation. An interpretation may be considered legitimate just in case the expressions of the language are used in a manner that is consonant with their actual semantic function. Thus, expressions of the language are grouped into semantic categories; a legitimate interpretation of the language is one in which each expression is assigned a semantic value that is appropriate to its category. Models are functions assigning such semantic values to every simple expression of the language. On this way of thinking, absolute truth is represented by truth in the model which interprets our linguistic expressions as they are actually interpreted – the model which assigns to our expressions their actual semantic values. This is model theory as *interpretational* semantics; truth in a model, on this view, is to be understood as truth in a language.

Let's examine in some detail each approach to MTS in turn, in the hope of determining whether either licenses the conclusion that natural language sentences and/or predicates have semantic values.

Representational Model Theoretic Semantics

Model theories have been constructed for a variety of formal languages, and are thought to be of use in the mathematical study of logic. Thus, it is said that a set of premises *entails* a conclusion, and that an argument consisting of those premises and that conclusion is *valid* if, and only if, every model in which the premises are true is also one in which the conclusion is true. A sentence is said to be *logically true* just in case it is true in every model for the language. But Etchemendy claims that we cannot use a representational MTS to analyze the logical notions. There are two reasons for this, which I will present once some groundwork has been laid. The important point for present purposes is that the arguments lead Etchemendy to the following conclusion:

The value of representational semantics does not lie in an analysis of the notions of logical truth and logical consequence, or in the analysis of necessary... truth. Rather, what this approach gives us is a perspicuous framework for characterizing the semantic rules that govern our use of the language under investigation. It should be seen as a method of approaching the empirical study of language, rather than an attempt to analyze any of the concepts involved in that task. (Etchemendy, 1999: 25)

In short, representational MTS is not good for logic, but it is good for the empirical discipline of natural language semantics.

If representational MTS is to be used in the empirical study of natural languages, then it seems that a thorough investigation of this approach is a good place to start the search for evidence that natural language predicates and sentences must, *contra* Davidson, be assigned semantic values. However, I think that Etchemendy is mistaken about the use of representational MTS in the empirical study of the semantic properties of natural languages. My reasons will emerge in the course of this section.

The first thing to notice about a representational MTS is that the models it employs are not themselves (Lewisian) possible worlds (that is, spatio-temporally maximal concrete objects); rather they are *representations*, each of which depicts the world as being some particular way.¹¹⁰ This raises the question: How is it that these models succeed in representing the world at all?¹¹¹ As a general answer, we may say that a representation scheme is established as the product of two factors: (i) an understanding of the primitive (non-logical) expressions of the language for which we are constructing the representational MTS; and (ii) a grasp of the notion of truth, and of the counterfactual conditional. To see this, let's examine in some detail Etchemendy's discussion of the provision of a representational MTS for a simple propositional fragment of English.

At the end of (1999) chapter 2, Etchemendy articulates general "guidelines" for the construction of a representational MTS. Recall that in general a model theoretic semantics consists of two parts – the delineation of a class of models, and the recursive characterization of truth in a model. Etchemendy's guidelines govern each task the semanticist faces. "First," he says,

there is the obvious though rather vague criterion we use in judging the adequacy of our class of models. In a representational semantics the class of models should contain representatives of *all* and *only* intuitively possible configurations of the world. (Etchemendy, 1999: 23, emphasis original)

Given this first guideline, we are in a position to understand Etchemendy's objections to using representational MTS to analyze the logical notions. I present them by concentrating on the notion of logical truth. As we have seen, a logical truth is, according to the purported analysis, any sentence which is true in all models of the language to which it belongs. But first, given that there are, in an adequate

¹¹⁰ Indeed, this is why Etchemendy calls this approach to MTS "representational"; see (Etchemendy, 1999: 20).

¹¹¹ Lewis asks this question regarding what he calls "ersatz possible worlds" at (Lewis, 2001: 141).

representational MTS, models corresponding to all and only the metaphysically possible worlds discernible in the language, it follows from this definition that the logical truths are exactly those sentences of the language which are metaphysically necessary. This, however, is not correct. Consider “ $2+2=4$ ” or “Water is H_2O ”: these are necessary, but not logical, truths. So the attempted analysis overgenerates. The second reason why, according to Etchemendy, this definition fails is that even if one were to accept that the logical truths are the metaphysically necessary truths, one would have to use one’s intuitive understanding of what is, and what is not, metaphysically necessary and/or possible in delineating the class of models; thus the model theoretic semantics can’t afford us insight into the extent of the class of logical truths on this approach.

Etchemendy’s second guideline governs the recursive characterization of truth in a model. He says:

Once we have specified the class of models, our definition of truth in a model is guided by straightforward semantic intuitions, intuitions about the influence of the world on the truth values of sentences in our language. Our criterion here is simple: a sentence is to be *true in* a model if and only if it *would have been true* had the model been accurate – that is, had the world been as depicted by that model. (Etchemendy, 1999: 24, emphasis original)

How are these guidelines to be applied to the case at hand – that of a simple propositional fragment of English?

The propositional language Etchemendy considers contains three atomic English sentences – “Snow is white,” “Roses are red,” and “Violets are blue” – as well as a negation operator and a connective expressing disjunction. Etchemendy says:

A standard representational semantics for [this fragment of English] might proceed in the following way. First we define a class of models that will represent all possible configurations of the world relevant to the truth values of our sentences.... [T]his purpose can be served by the class of functions that assign a truth value, either *true* or *false*, to each of our three atomic sentences....

Our next step is to provide a recursive definition of *S is true in f* for arbitrary sentences *S* and models *f*... [Thus:]

- If *S* is an atomic sentence, then it is true in a model *f* just in case *f* assigns it the value *true*.
- If *S* is the negation of *S'*, then it is true in a model *f* just in case *S'* is not true in *f*.
- If *S* is the disjunction of *S'* and *S''*, then it is true in a model *f* just in case either *S'* is true in *f* or *S''* is true in *f*. (Etchemendy, 1999: 21-22)

This much, no doubt, will strike the reader as utterly familiar. However, Etchemendy continues:

[W]hether an atomic sentence comes out true in a given *model* is determined not by the model itself but by the base clause of our recursive definition, the clause beginning “if *S* is an atomic sentence...” The fact that we took models to be functions that yield the values *true* and *false* is entirely a mnemonic convenience...; any two objects would have worked as well – for example, the numbers zero and one. Indeed, if we had used zero and one, the substantial contribution made by the base clause of our definition would have been highlighted: without the base clause, we would not know whether a model that assigns zero to ‘Snow is white’ represents a world in which snow *is* white, or one in which it is *not*. (Etchemendy, 1999: 22)

Thus, according to Etchemendy, the fact that a model in which “Snow is white” is assigned the value *true* (or zero, or whatever) represents the world as being such that snow is white, is due to the facts (a) that “Snow is white” means, in English, that snow is white, and (b) that “Snow is white” is deemed to be true in models in which it is assigned the value *true* (or zero, or whatever). Given the representational guidelines (in particular, “a sentence is to be *true in* a model if and only if it *would have been true* had the model been accurate”), these facts correspond, respectively, to (i) and (ii) above.

There is, however, something incongruous in the passages I have just cited. For on the one hand, Etchemendy claims that the definition of truth in a model is to be constructed in accordance with “straightforward semantic intuitions”; but on the

other hand, the base clause of this definition is to be regarded as *stipulative*, determining the representational features of the models described by the theory. Surely both can't be the case; if the base clause is a stipulation, it *cannot* be responsive to intuitions of any kind. Something's got to give.

There are two options. We can abandon the guideline governing the recursive definition of truth in a model; that is, we can give up on the thought that this definition is responsive to semantic intuitions, and view it instead as stipulative, determining the representational properties of the various models. Alternatively, we can give up the claim that the images of atomic sentences in the model are arbitrary, insisting rather that they are truth values (rather than, e.g., numbers). This fixes the representational features of the models independently of the recursive definition of truth in a model, and so allows us to view that definition as responding to semantic intuitions.

Let's consider the second alternative first. Can a representational MTS, construed in this way, be of use in empirical semantics? The answer is clearly not. The empirical semanticist wants to be able to describe the meanings of all the expressions of a language with which he or she is unfamiliar. Yet whatever measure of meaning a representational MTS constructed along these lines is intended to offer, it can only be of use to someone who already understands the object language.¹¹² The reason is that to fix the representational features of the model, we must assume (i) an understanding the basic sentences of the language, and (ii) that the images of these sentences under the functions which constitute the model are genuine truth values. The first of these points is sufficient to render the approach useless to anyone engaging in empirical study of the semantic properties of the object language in question.

¹¹² This is crucially different than in the case of Davidsonian truth theoretic semantics, where the semanticist must understand the *meta*-language – which may, of course, but need not, be an extension of the object language.

By contrast, it seems to me that if we opt for the first method of resolving the tension in Etchemendy's discussion, the value we can see in a representational MTS is principally in the understanding it affords us of the notion of a possibility. For the point just raised applies equally well to this approach to representational MTS – the approach assumes a knowledge of the semantic properties of the object language, and so can't be of use in investigating them. But given an understanding of our object language sentences, plus the definition of truth in a model, we are able to construct representations of various possible states of affairs. Some of those representations won't correspond to genuine *metaphysical* possibilities, e.g. those in which "2+2=4" is assigned the value *false*; nevertheless they *do* represent *epistemic* possibilities.¹¹³ These in turn might be useful tools for semanticists, notions they can freely employ in the description of the semantic features of natural languages. However, a representational MTS constructed in this way has no *direct* application to empirical semantics. So, if there is to be an inference to the best explanation, on the basis of a model theoretic semantics, with the conclusion that sentences and/or predicates are meaningful by virtue of having semantic values, then it had better rest upon the provision of an *interpretational* MTS.

Interpretational Model Theoretic Semantics

As the representational approach has not uncovered a need for semantic values for sentences and/or predicates, let us turn to consider the interpretational view of model theoretic semantics. This, recall, treats truth in a model as representative of truth in an acceptable interpretation of the expressions of the language. As Etchemendy stresses, once we adopt this approach,

we are forced to hazard at least a simple theory about the semantic functioning of expressions within a given grammatical category, a

¹¹³ Note, however, that to get this requires abandoning the first guideline Etchemendy imposes on the construction of a representational MTS. So it is a little unclear to me just how stable this first interpretation of MTS is, when all is said and done.

theory of how they each contribute, and differ in their contribution, to the truth values of sentences in which they occur. (Etchemendy, 1999: 40)

The reason is that we must say what the acceptable interpretations of the language are, and this requires specifying the range of permissible values of the model theoretic interpretation functions.

Consider again the simple language discussed above, that containing three atomic sentences and the connectives ‘not’ and ‘or’. In order to provide an interpretational semantics for this language, we must settle the question what the legitimate reinterpretations of the atoms are, as well as what those of the unary operator ‘not’ and the binary connective ‘or’ are. The operator and connective may be taken to express truth functions. As for the atoms, Etchemendy says:

In the present language, we can explain the semantic contribution of any embedded sentence to its embedding sentence in one of two ways: either the component *says something true* or it *says something false*. Thus, we can take the sentence domain to consist of the two truth values, *true* and *false*. (Etchemendy, 1999: 57)

Here it is crucial that the entities in the sentence domain really are truth values, and not, say, the numbers zero and one. Why is this?

The aim of an interpretational semantics is to give, for every (syntactically identified) sentence of the object language a statement of its *material* truth conditions: “[a] sentence is true in a given model if, so to speak, what it would have said on the suggested interpretation is, *in fact*, the case” (Etchemendy, 1999: 61, my emphasis). To achieve this aim, we must derive all the instances of the schema:

‘ $S(\underline{p}, \underline{c}, \underline{o})$ ’ is true in model *fiff* $S(\underline{p}/\underline{s}, \underline{c}/\underline{b}, \underline{o}/\underline{u})$

where the left hand side *mentions* an object language (OL) sentence containing sentences from amongst the collection $\underline{p}$, binary connectives from amongst the collection $\underline{c}$, and unary operators from amongst the collection $\underline{o}$, and the right hand

side *uses* a meta-language (ML) translation of it, with the sentences in $\mathcal{L}$ replacing those in $\mathcal{P}$ uniformly throughout, and so on. Etchemendy says,

The appropriateness of the replacement expressions will now be determined by the following instantiation conditions: sentence s_i must *have the truth value* $f(p_i)$; connective b_i must *express the binary truth function* $f(c_i)$; and operator u_i must *express the unary truth function* $f(o_i)$. (Etchemendy, 1999: 58)

If we allowed the interpretations of object language sentences to be the numbers zero and one, we would have to constrain the translation process by allowing a meta-language sentence to replace an object language one just in case the ML sentence had the number assigned by the model to the OL sentence; but this is absurd – no ML sentence has, in any intuitive sense, a number.

Now, it seems, we might endorse an inference to the best explanation, on the basis of an interpretational semantics, to the claim that sentences have semantic values; in particular, the meanings of sentences are truth values. Or at least, we might accept this inference, if we thought natural languages were as expressively impoverished as the fragment of English we have been considering. In fact, however, there are at least two complications that arise. First, note that even the most basic sentences of natural languages have internal syntactic structure. This structure no doubt contributes to the interpretation of those sentences. Let's look briefly at the prospects for an interpretational MTS for a language whose atomic sentences have internal structure.

Etchemendy considers a language with names and one-place predicates as basic categories of expressions, as well as the truth functional operator “not” and connective “or”. A typical interpretational MTS for such a language assigns objects to the names, and sets of objects to the predicates. The basis of the recursive definition of truth in a model makes it plain that an atomic sentence of this language, consisting of a name concatenated with a predicate, is true in a model f if, and only if, f assigns to

the name an object which is a member of the set that it assigns to the predicate. The recursion then continues, stating that a negation is true in a model *f* iff the embedded sentence is not true in *f*, and a disjunction of two sentences is true in a model *f* iff at least one of the disjuncts is true in *f*. On this approach, truth values are not explicitly assigned as the semantic values of sentences within a model; though, of course, predicates do have model-internal semantic values. The recursion, however, is a definition that characterizes a set of ordered pairs of sentences and models; and in a sufficiently strong meta-theory, this can be transformed into an explicit definition of this set, Truth. The recursion implicitly associates either this set, or its counterpart, Falsehood, with each sentence, relative to a given model; nevertheless, these objects are not explicitly mentioned in the semantics.

An alternative interpretational MTS for a language of this sort again assigns objects to names; but instead of assigning sets as the semantic values of one-place predicates, it associates them with functions taking objects to truth values. Accordingly, sentences have model-internal semantic values on this approach; a sentential SV is a truth value, the result of applying a predicate SV to the SV of a name. On the first approach, the semantic values of subsentential expressions are given independently of those of sentences – those of predicates, for example, are just certain sets; whereas on the second they aren't and can't be – predicates, as we have seen, have as semantic values functions whose values are truth values, i.e. sentential SVs. Co-opting some terminology from Dummett (1978: 222-223), I call the first approach “Atomism”, the second “Molecularism”. Both assign semantic values to *predicates*; yet only the latter, and not the former, assign semantic values to *sentences*. If we are to assess whether sentences have semantic values it seems we will need to choose between atomism and molecularism.¹¹⁴

¹¹⁴ A third approach to MTS is becoming popular: one regards a model as a relation rather than a function. Thus, predicates are not assigned either sets or functions as semantic values; rather each

The second complication preventing us from drawing the immediate conclusion that sentences have truth values as meanings is that natural languages contain non-extensional expressions. If our propositional fragment of English contained the unary sentential operator “It is necessary that”, for example, then we should be able to find object language sentences differing in truth value, but such that the translations of their embedded sentences intuitively have the same truth value. Consider, for instance, the language containing this operator, as well as the atomic sentences “Cicero is Tully” and “Cicero is an orator”. Although these atoms are (ignoring tense) both true, they nevertheless contribute differently to the sentences formed by embedding them under the necessity operator. So it seems we must reassess the semantic values we assign to these atoms. We will consider two approaches to the semantics of modal languages in the upcoming sections. In the next section I examine the atomistic semantics presented in David Kaplan’s (1989) “Demonstratives”; and in the section after that I look at David Lewis’ (1970) molecularist program in “General Semantics”.

Semantics for Modal Languages (a): Kaplan’s Atomism

A semantic theory for a natural language pairs sentences (taken together with representations of contexts) with collections of possibilities. This is (arguably) the Tractarian vision of semantics, and it is endorsed by many contemporary thinkers. Thus, Robert Stalnaker believes that “a proposition is a function from possible worlds to truth-values, or equivalently, a set of possible worlds” (Stalnaker, 1999: 3); and as we have seen, he claims that “it is a semantical problem to specify the rules for matching up sentences of a natural language with the propositions that they express” (Stalnaker, 1999: 34). David Kaplan says “we can represent a content [a kind of

predicate is assigned a number of entities from the domain, each of which is one amongst possibly many semantic values of that predicate. This third alternative approach to MTS somewhat complicates the presentation of the issues I am pursuing here (whether predicates should be assigned semantic values which somehow embody the generality of their application); in what follows I simply ignore it.

meaning] by a function from circumstances of evaluation to an appropriate extension” (Kaplan, 1989: 501-502), which, in the case of sentences is a truth value. And Stefano Predelli, in his (2005) book *Contexts*, claims that semantics aims to characterize

interpretive systems... which, when applied to *clause-index* pairs, yield conclusions of *t-distributions*, i.e. assignments of truth-value at particular *points* of evaluation. (Predelli, 2005: 3)

In short, for all these philosophers, a sentential meaning is, or is represented by, a collection of possibilities, and the task of natural language semantics is to pair sentences (together with indices) with their meanings.

This raises the question, however, *how to effectuate* the semantic pairing of sentences with such propositions. *Atomists* and *molecularists* take different views of this matter. Atomists assign SVs to the primitive expressions of the language immediately, and hope to generate the SVs of complex expressions, including sentences, from these primitive SVs by composition. The guiding intuition is that the basic elements of language serve to represent some aspect of reality, while the entities complex expressions represent are determined as a function of what the parts of those expressions depict. Molecularists, on the other hand, assign semantic values immediately to sentences, and define the semantic values of (some) subsentential expressions in terms of these; here the thought is that one cannot do anything with a linguistic unit smaller than a sentence, so what the meaning of a subsentential expression is cannot, in general, be explained except in terms of its contribution to the meaning of sentences in which it figures.¹¹⁵

The semantics Kaplan (1989) gave for his LD - the language of the logic of demonstratives - is model theoretic, and atomistic. LD differs from the previous

¹¹⁵ The point here is not that atomistic theories are compositional while sententialist theories are not – both are (in some appropriate sense) compositional. The point is rather that, for atomists, the SVs of primitive expressions can (all) be given independently of those of complex expressions in which they figure, while for molecularists the SVs of some primitive expressions can only be given in terms of the SVs of sentences in which they occur.

formal languages we have considered: for starters, it is a modal language. How are we to interpret a model theory for a modal language? Is it to be viewed as representational, in Etchemendy's sense, or interpretational? Recall that this distinction is drawn as follows: in a representational MTS each model is to be thought of as representing a possible world; whereas in an interpretational MTS, each model is, or does the duty of, a language. Moreover, our earlier discussion leads us to believe that only an interpretational MTS was likely to license the conclusion that sentences and/or predicates have semantic values; and it is this question we are still attempting to assess.

In fact, I suspect that model theories for modal languages simultaneously involve both kinds of philosophical interpretation. Typically, a model for a language L containing modal operators is thought to be an ordered triple $\langle W, R, I \rangle$, where W is a set of points (intuitively, the possible worlds), R is a relation on those points (the accessibility relation), and I a function assigning appropriate semantic values to the primitive non-logical expressions of L . An ordered pair $\langle W, R \rangle$ is called a frame; each frame represents a different (epistemically) possible pluriverse, or modal reality. If we fix on the frame which represents the metaphysically real pluriverse, then each model based on that frame represents a different legitimate interpretation of the expressions of L over the pluriverse; that is, the interpretation function is, or represents, the language. If these considerations are correct, then there is still room for an inference to the best explanation, based on a model theory for a modal language, to the conclusion that sentences and/or predicates have semantic values: this conclusion will be legitimate when the interpretation function of the model (or its extension to the complex expressions of the language) assigns entities to the expressions in question.

Let's look in some detail at Kaplan's semantics for the language of the logic of demonstratives. In fact, LD is more complex again than the formal languages we have considered to date: it contains complex singular terms (definite descriptions),

quantifiers, and modal operators, as well as simple singular terms and predicates; moreover, some of its expressions are context dependent. Simplifying somewhat (principally by ignoring the tense logical aspect of Kaplan's system), we may say that an LD structure is an ordered quadruple $\mathbf{U} = \langle C, W, U, I \rangle$, where C is a set of contexts, W a set of worlds, U the universe of discourse, and I the interpretation function.¹¹⁶ The interpretation function I of each model maps the primitive n -place predicates onto intensions; i.e. functions from circumstances (for our purposes, possible worlds – i.e. elements of W) to sets of n -tuples of objects. In terms of these LD structures, Kaplan was able to give recursive definitions of the notions of denotation relative-, and truth relative-, to a context and an assignment function, with respect to a world, for singular terms and sentences (open and closed) respectively; these are written $|\alpha|_{\mathbf{U}_{cfw}}$ for terms α , and $|\models_{cfw}\Phi$, for sentences Φ .¹¹⁷ Now, the crucial point for our purposes is that *only then did Kaplan define the notion of content*, introduced on the back of these more basic ones (Kaplan, 1989: 546):

Where Γ is either a term or a formula, we write: $\{\Gamma\}^{\mathbf{U}_{cf}}$ for The Content of Γ in the context c (under the assignment f and in the structure $\mathbf{U}$).

Definition:

- (i) If Φ is a formula, $\{\Phi\}^{\mathbf{U}_{cf}} =$ that function which assigns to each $w \in W$, Truth, if $|\models_{cfw}\Phi$, and Falsehood otherwise.
- (ii) If α is a term, $\{\alpha\}^{\mathbf{U}_{cf}} =$ that function which assigns to each $w \in W$, $|\alpha|_{\mathbf{U}_{cfw}}$.¹¹⁸

Sentential contents do no formal semantic work for Kaplan; sentences themselves are immediately semantically evaluated, and their contents only derivatively so.

It is important to note that this conclusion holds in spite of the fact that Kaplan's language LD contains the modal operators $\Box$ and $\Diamond$, pronounced "It is necessary that" and "It is possible that" respectively; for it is the semantic explanation

¹¹⁶ The accessibility relation R is suppressed, as it is simply taken to be the universal relation on W .

¹¹⁷ See (Kaplan, 1989: 544).

¹¹⁸ These definitions are modified from those actually given at (Kaplan, 1989: 546) in order to accommodate the simplification mentioned above.

of these expressions that is most commonly thought to require the postulation of propositions.¹¹⁹ It will perhaps come as no surprise that Kaplan succeeds in characterizing sentential truth (relative to a model, etc.) for modal sentences without mentioning the content of the sentences the modal operators embed by quantifying over possible worlds: thus, for instance, $\models_{cfw} \Box \Phi$ just in case, $\forall w' \in W \models_{cfw'} \Phi$; that is, $\lceil \text{It is necessary that } \Phi \rceil$ is true at a world w just in case $\lceil \Phi \rceil$ is true at all worlds.

How far does this go towards validating Davidson's objection, raised in the introductory section, that semantic values are either trivial or useless? One thing which should be mentioned at the outset is that, of course, Kaplan's semantics does assign semantic values to *predicates*, namely, as we have seen, certain intensions, or functions from worlds to sets of n-tuples of objects. These are explicitly mentioned, and therefore play an essential role, in the recursive definition of the central semantic notion (that of truth in a model, relative to a context and an assignment function, and with respect to a world). However, *sentence SVs* – i.e. *propositions* – are assigned to sentences in context only *implicitly*, by recursion; they are never mentioned explicitly in the semantic theory.¹²⁰ It seems Kaplan's atomism does fall prey to Davidson's objection: these sentential SVs are theoretically useless to the semanticist.

Perhaps, though, the need for propositions, or as Kaplan calls them, "contents", can be felt at the meta-semantic level. The thought here is that an account of when it is that a given population uses one language rather than another will appeal to the idea that speakers use sentences to express propositions: thus, they use one language, rather than another, if, in order to convey a given proposition they use a sentence

¹¹⁹ Along with the semantic explanation of the verbs used in attitude ascriptions. I have nothing to say on this subject here.

¹²⁰ The objects implicitly defined by the atomist's recursion are, in the case of a modal language, at least *prima facie*, language independent objects (they are certainly independent of the object language). Whereas Tarski's Truth was simply a set of sentences, Kaplan's contents are sets of possible worlds.

which in the one language, but not in the other, is associated, however implicitly, with that proposition. David Lewis (1980), for example, writes:

Suppose (1) that you do not know whether *A* or *B* or...; (2) that I do know; and (3) that I want you to know; and (4) that no extraneous reasons much constrain my choice of words; and (5) that we both know conditions (1)-(5) obtain. Then I will be truthful and you will be trusting and thereby you will come to share my knowledge.... I will find something to say that depends for its truth-in-English on whether *A* or *B* or... and that I take to be true-in-English; you will trust me to be willing and able to tell the truth-in-English. (Lewis, 1998: 22-23)

According to Lewis, English is in use in our population because we are truthful and trusting in English; and this amounts to our using sentences which depend, for their truth or falsity in English, on the truth or falsity *simpliciter* of appropriately related propositions. Or take Robert Stalnaker's (1999) view, according to which we may represent a speech situation in terms of the set of possible worlds that are compatible with what the speakers presuppose. If one then asserts a proposition – which on both Stalnaker's and Kaplan's views is a set of possible worlds – one updates the representation of the speech situation by intersecting the original set of worlds with the set constituting the proposition asserted. But of course one actually produces a string of noises. One thereby asserts a *proposition* because the sentence one produces is associated, however implicitly, with that proposition by the semantic theory governing one's language. If some account along these lines is correct, then propositions, while strictly useless from a *semantic* point of view, are nevertheless not entirely theoretically useless since they are invoked to explain the meta-semantics of language use.

I will not attempt to assess this line of thought here: the atomist will have to make the case if he or she sees fit. I, on the other hand, see reasons to doubt the truth of atomism. Firstly, and most importantly, atomism is shortsighted. To see this, let's

consider the extensional case again.¹²¹ Recall that the atomist assigned objects to names, sets of objects to one-place predicates, and then recursively characterized truth on the basis of these primitive semantic values: in particular, he did not explicitly assign truth values to sentences. How then did he interpret the operator “not” and the connective “or”? He simply translated these expressions into the meta-language. This strategy was possible, however, only because the atomist’s project was not sufficiently ambitious: he temporarily forgot that these expressions, like any other, can be variously interpreted in different languages. What’s needed is a fully general MTS, which considers the interpretation of all object language expressions, including these. But if one treats “or” and “not” as variable expressions, as this project requires, rather than logical constants, as the atomist does, then one needs to articulate very general constraints on their translation into the meta-language. The constraint Etchemendy suggested was that a meta-language operator or connective is an acceptable translation of an object language expression of the same category just in case it expresses the same truth function. But if these expressions are to express truth functions, then sentences had better have truth values as semantic values – otherwise these functions will not have appropriate arguments to operate on. So it seems that some subsentential expressions – the operators and connectives – have semantic values that are not ontologically independent of the semantic values of sentences. And if this is the case, then atomism cannot be a successful strategy in general.

There is also a second reason to worry about atomism. As we saw, part of the motivation for this view stems from the thought linguistic expressions serve to represent aspects of the world: names represent objects, predicates represent universals, and so on. But it is difficult to fill out the “and so on”: do connectives, for example, really *represent* truth functions? Do sentences *represent* propositions? In

¹²¹ The argument could be made for the intensional case, but its force would be less apparent.

fact, the problem is worse than this; for it is not clear that this talk of representation is correct even in the basic case of predicates.

Perhaps there are some relatively straightforward cases. The intension of a common noun, or predicate, on this view is a function from worlds to sets of individuals. But it is also commonly held that properties and other universals are, or otherwise determine, an intension of this kind. Nathan Salmon (1982: 53), for example, suggests the following: If v is a “general term,” or “common noun,” such as “tiger”, and k is a natural kind, such as the species *Felis Tigris*, then it is both necessary and sufficient for v to *designate* k that the semantic intension of v coincide exactly with the metaphysical intension of k . Indeed, the thought seems to be that the term has the semantic intension it does *because* it designates a kind which has that metaphysical intension.

But what about the predicate “grue”? Presumably this does *not* designate, or otherwise represent, a natural kind; this, at least, is one way of putting the point for which Goodman (1954) introduced the example. Yet “grue” is a perfectly meaningful, syntactically simple predicate, and its semantic behaviour should surely be given the same explanation as that of any other expression of its kind. Or consider vague predicates like “bald”. Can it really be the case that there is a property, *baldness*, which people either have or fail to have, while the word “bald” simply stands for, or otherwise represents, this property? This seems highly implausible, epistemicism notwithstanding; for we simply do not use “bald” in this cut and dry manner. Rather, it is a descriptive term; and while it is designed to carry information, it is not clear that it carries exactly the same information on each occasion of its use. Or again, consider the case of context-sensitive predicates such as, according to some, “tall” or “knows”. Can it be the case that there are whole swathes of universals (properties in the one case, relations in the other), such that uses of these predicates represent different specific universals amongst them on different occasions?

If one begins to worry about these cases, then even the “easy” cases don’t seem so straightforward. Consider the fossilized remains of a large feline dating from some time in the evolutionary past. The phenotype, and even the genotype, of the animal of which the fossil is a remnant may be clear enough; yet we may be unsure whether the cat is of the kind *Felis Tigris*. Might there be no fact of the matter? Or take what might be thought to be the clearest case of a representational term – “blue”. This too exhibits vagueness: which wavelengths of light are to be counted *blue* and which *green*? The problematic phenomena are ubiquitous.

These most recent considerations do not – even jointly – constitute a knock-down objection to atomism in semantics. They do, however, give us reason to consider the alternative. In the next section I examine David Lewis’ semantic molecularism.

Semantics for Modal Languages (b): Lewis’ Molecularism

Lewis’ view in “General Semantics” is that a language must contain at least two basic categories of expression, one of which is always *S* (for sentence); in Lewis’ own sample language there are two others, namely *N* (for name), and *C* (for common noun). These are assigned intensions as semantic values, i.e. functions from indices (n-tuples of contextual features, including at a minimum, a possible world) to appropriate extensions – truth values in the case of sentences, individuals in the case of names, and sets of individuals in the case of common nouns. Other categories of expression are “derived”: for instance, an intransitive verb (of a certain sort) is an *S/N*, since it takes a name and yields a sentence; its semantic value is a function from name intensions to sentence intensions.

It is perhaps worth considering the question whether this Lewisian approach to semantics is in fact model theoretic. According to the rough account given by Hodges, and quoted above, it certainly is: Lewis interprets his languages (with categorially-based transformational grammars) by means of set theoretic structures

(such as sets and functions). Of course, these structures do not belong to the domain of the *pure* theory of sets; however, there is no reason to think that Hodges has in mind that they should. But is Lewis' semantics model theoretic in Etchemendy's sense? Lewis does not explicitly define a class of models and then give a recursive definition of truth relative to a model. Nevertheless, what he does say makes it clear how to go about doing so.

A categorial language consists syntactically of what Lewis calls a "lexicon":¹²² a function which assigns categories to the finitely many *primitive* expressions of that language; at least some of these expressions will be assigned to *derived* categories. (It is important not to confuse primitiveness – a feature of expressions – with basicness – a feature of categories.) A model for such a language will then be an ordered triple consisting of:

- (1) a set of indices;
- (2) a function taking the basic categories to sets of appropriate extensions; and
- (3) a function f taking the primitive expressions to intensions of an appropriate kind.

The first two elements of the model jointly determine the intensions for all the basic and derived categories mentioned in clause (3).

In terms of such models we can recursively define the semantic value of an expression e in a model, which we can call $\llbracket L(e) \rrbracket$: if e is primitive, then $L(e) = f(e)$; if e is a complex expression of category c which results from concatenating an expression e_0 of category $c/c_1, \dots, c_n$ with n expressions $e_1, \dots, e_n$ in order, of category $c_1, \dots, c_n$ respectively, then $L(e) = L(e_0)(L(e_1), \dots, L(e_n))$. Since the intensions appropriate to sentences are functions from indices to truth values, we can give the following definition: sentence-index pair $\langle s, i \rangle$ is true in model M iff $(L(s))(i) = \text{Truth}$.

¹²² It also involves a collection of phrase structure rules "of the form $c \rightarrow c/c_1 \dots c_n + c_1 + \dots + c_n$ " (Lewis, 1983: 192); but these are "common to all categorial grammars" (Lewis, 1983: 192).

So it seems that Lewis can be seen as engaging in Model Theoretic Semantics after all; moreover, his semantics certainly invokes semantic values. The question therefore arises: Can Davidson's objection (from the introduction) be made to work against this view? The objection, as we have seen, is best viewed as a dilemma: either the meanings do no semantic work, or the theory of meaning is trivial and uninformative. It seems clear that Lewis is not caught on the first horn of this dilemma: Lewisian semantic values *do* do semantic work. One applies the function that is the meaning of the embedding word to the functions that are the meanings, taken in order, of the embedded words, repeatedly, until one obtains a function from indices to truth values as output; one then inserts an index and evaluates.

It is, however, a more subtle affair to determine whether Lewis is caught on the second horn of the dilemma: whether, that is, his semantic theory is trivial and uninformative. Let's take a closer look at the Davidsonian objection. After giving a sample derivation within a meaning-postulating theory of meaning, Davidson lodges his complaint:

Ask, for example, for the meaning of 'Theaetetus flies'. A Fregean answer [that is, one postulating semantic values] might go something like this: given the meaning of 'Theaetetus' as argument, the meaning of 'flies' gives the meaning of 'Theaetetus flies' as value. The vacuity of the answer is obvious. We wanted to know what the meaning of 'Theaetetus flies' is; it is no progress to be told that it is the meaning of 'Theaetetus flies'. (Davidson, 2001b: 20)

If all that can be said about the meaning of a particular derived category expression is that it takes inputs to the right outputs, then one might wonder what explanatory gain is achieved by postulating it and other semantic values. This is like being told that a certain drug puts people to sleep because it has a dormative power: since all we know about dormative powers is that they lead to sleep, the phenomena have not been thereby explained, but only re-described. And indeed, concerning the intension

of the expression “alleged” (which belongs to the category Adjective, or C/C), Lewis says the following:

[T]he intension of ‘alleged’ is a function that, when given as argument the intension of ‘Communist’, ‘windshield’, or ‘chipmunk’ yields as value the intension of the compound common noun ‘alleged Communist’, ‘alleged windshield’, or ‘alleged chipmunk’ respectively. (Lewis, 1983: 198)

The parallel between the two passages is remarkable: Davidson would no doubt view the Lewisian theory as vacuous.

Still, it is worth trying to spell out the concern in more detail. James Higginbotham (1999) comments on the above remark of Davidson’s, and attempts an elucidation thereof. It is commonly accepted that if a theory of meaning is to explain the meanings of all of the infinitely many expressions of a given language, then it had better do so in terms of the meanings of the simple parts and the modes of combination of those parts. But then, as Higginbotham points out, the theory must explain not only the meanings of the parts, but *also* the meanings of the modes of composition – that is, it must explain the “combinatorics” of the language. Having noted this, Higginbotham argues that a certain approach to semantics which assigns “concepts” as the semantic values of expressions, fails to achieve such explanation. This theory gives the following account of the semantics of the predication schema (which I call (SPS)):

If **S** is an instance ‘*a Vs*’ of the predication schema, the concept c_1 is the meaning of the instance of *a*, and c_2 is the meaning of the instance of *V*, then the meaning of **S** is (the proposition that) $c_1 c_2$. (Higginbotham, 1999: 675)

However, Higginbotham claims that the “combinatorial semantic principle” given, (SPS),

is like an equation in two unknowns; we know that the meaning of the predication schema is such that it delivers FIDO BARKS given FIDO and BARKS; and we know that FIDO BARKS is the value of the

predication schema for those arguments; but from this single principle we don't know what the predication schema signifies, or what proposition FIDO BARKS may be. (Higginbotham, 1999: 677-678)

He concludes that the semantic theory which embodies (SPS) is of no use.

Higginbotham does not attempt to apply his version of Davidson's objection to Lewis' view in "General Semantics". This is just as well, for had he done so, he would have found that it fails. On Lewis' view, the semantic values of derived category words are certain functions called "compositional intensions". Lewis says:

The general form of the semantic projection rules for an interpreted categorial grammar is implicit in the nature of compositional intensions.... The result of concatenating a $(c/c_1 \dots c_n)$ with intension Φ_0 , a c_1 with intension Φ_1 , ... , and a c_n with intension Φ_n is a c with intension $\Phi_0(\Phi_1 \dots \Phi_n)$. (Lewis, 1983: 198)

Since we know what functional application is, we know what $\Phi_0(\Phi_1 \dots \Phi_n)$ is, provided we know what each of the Φ_i is, for $0 \leq i \leq n$; which, *ex hypothesi*, as Higginbotham formulates the objection, we do. That is, we are taken to know that the semantic value of 'Fido' is FIDO, and that of 'barks' is BARKS; and the semantic projection rule tells us that the meaning of 'Fido' concatenated with 'barks' (i.e. 'Fido barks') is the result of functional application of BARKS to FIDO.

Still, there seems to be room for the thought that the Davidsonian worry should apply to Lewisian semantics, and that it be due to the existence of a kind of equation in two unknowns. On this alternative way of viewing the matter, what's unknown is not the meaning of concatenation, but rather the meanings both of the relatively simple expressions, and of the relatively complex expressions containing those simpler ones: thus, in Davidson's example, "flies" and "Theaetetus flies"; in Higginbotham's, "barks" and "Fido barks"; and in Lewis', "alleged" and, for example, "alleged Communist". In order to spell out this worry more clearly, however, it will be worth considering Dummett's (1981) discussion, in *Frege: Philosophy of Language*, of the hypostasization of meaning.

Dummett distinguishes “three distinct positions which could be adopted about what belongs to the realm of sense” (Dummett, 1981: 155). The first is that senses can be grasped by us directly, by some kind of intellectual intuition. The second is that, although we can only grasp senses as the meanings of certain expressions, a being with superior mental powers could grasp them in some other way. Finally, the third position Dummett considers is that no being could grasp a sense except as the sense of some expression or other. When presenting these alternative positions, Dummett alludes to his earlier discussion of the direction of a line (Dummett, 1981: 155). Frege, of course, defined the direction of a line *a* as the class of lines parallel to *a*.¹²³ thus, since the relation of being parallel is an equivalence relation, directions are equivalence classes of lines – elements of the partition of the class of lines under this relation of parallelity. Dummett takes this to be essentially correct; consequently, he objects to a (hypothetical) philosopher who claims that there are certain objects – directions – which stand in a peculiar and undefinable relation to lines, which we express by saying that the line *has* the direction. Dummett claims that what this philosopher fails to take account of is the fact that one can only grasp what the direction of a line is by grasping it *as* the direction of a line. Dummett seems to think that, due to the very nature of directions, this epistemic limitation will hold for any possible being – the correct view of directions is therefore that which is analogous to the third view of senses above. But Dummett claims that, while the first view of senses is obviously mistaken, just as it is in the case of directions, the second view of senses might be acceptable – and, moreover, it might be Frege’s own.

It is worth noting that although Dummett presents these three views as epistemological positions, there is nevertheless an ontological component implicit in each. Thus, while the third view is that directions and/or senses can’t be *grasped* independently of a line or expression having it as direction or sense, it is clear that the

¹²³ In the *Grundlagen* (Frege and Austin, 1980: section 65).

reason is that directions/senses are taken to be *ontologically* dependent upon lines/expressions. (This is especially clear in the case of lines and directions: if directions are just sets of lines, no wonder we can't grasp them without grasping lines!) On the second view, directions and senses are ontologically independent of lines and expressions, but we suffer an epistemic limitation such that we can't grasp the former without grasping the latter. And finally, on the first view, directions/senses are ontologically independent of lines/expressions, and we suffer no such epistemic limitation.

Now, if one held the third of Dummett's views with regard to meanings, then one might worry about Lewis' view of derived category intensions. For suppose that we can, whether as language users or theorists, grasp the intension of an expression *only as* the intension of an expression: then, in particular, we will only be able to grasp the intension of, say, "Theaetetus flies" as the intension of this, or some synonymous, expression. But since this intension is itself a *part* of the intension of the derived category word "flies" – it is the value of this intension when the argument is the intension associated with "Theaetetus" – then it would seem we cannot grasp the intension of "flies" without already grasping the intension of "Theaetetus flies". However, since one grasps the intension of such a complex expression by grasping the intensions of its parts (and by understanding the combinatorics of the language), we shall not be able to grasp either intension, since grasp of each depends upon grasp of the other. It is considerations along these lines, I suspect, that underlie the triviality objection Davidson raises against the postulation of meanings.

The Lewisian, however, is unlikely to be disturbed by these considerations. Firstly, no *argument* is given against the first position sketched above; Dummett merely *asserts* that it is mistaken. But secondly, and more importantly, Lewis does not seem to be committed to the view that one can grasp an intension directly, by means of some sort of intellectual intuition. Although, on his view, the theorist can *quantify*

over intensions, it is not clear that she needs to be able to “grasp” them – or *demonstratively identify* them in thought, as talk of intellectual intuition suggests. So it seems that Lewis might be advocating a view that is not any stronger than Dummett’s second position; and this Dummett thinks is acceptable.

This response, in assuming that we may quantify over intensions, appears to take for granted that intensions *exist* independently of our grasp of them; and so it seems that it is at least as strong as Dummett’s second position. Accordingly, one who disagreed with this claim, and who advocated Dummett’s third position (on the grounds that intensions are ontologically dependent upon expressions having those intensions), might object to Lewisian semantic theory for this reason. I do not think, however, that any objection along these lines can be made to work. In what follows I will attempt to say why.

The basic concern is that the ontological dependence of the SVs of simple expressions on the SVs of complex expressions is in conflict with the epistemic dependence of our grasp of the SVs of complex expressions on our grasp of the SVs of simple expressions. The dependency is then brought full circle, for the proponent of Dummett’s third view, by the fact that the existence of the SV of any expression (other than a singular term) – hence of any complex expression in particular – is itself dependent upon our grasp of it.

By way of response on Lewis’ behalf we may, to begin with, note the following. The three positions under discussion, regarding the epistemology and ontology of senses, are articulated by Dummett in chapter 6 of his (1981), which is titled “Some Theses of Frege’s on Sense and Reference”. The first of these theses discussed is that “[t]he sense of a complex is compounded out of the senses of the constituents” (Dummett, 1981: 152); and it is in the context of examining this thesis that Dummett comes to consider the reification of meanings. Lewis, however, never claims that the intensions of complex expressions are compounded of out the intensions of its

constituents; indeed, what he does say about intensions directly conflicts with any such claim. However, he draws a distinction between *intensions* on the one hand, and *meanings* on the other; while intensions are the functions we have been talking about so far, meanings are trees, at each node of which there is a pair of a category and an intension. The meaning of a relatively simple expression is a sub-tree of the meaning of a relatively complex expression containing it; thus, meanings of complexes *are* compounded out of those of their constituents. So Lewis has two notions (intension and meaning) where Frege had one (sense); as a result (and temporarily ignoring the acceptance of the third view), there is no ontological interdependence between the intensions of parts and wholes, or between the meanings of parts and wholes. Furthermore, from the fact that one can't grasp the meaning of a complex without grasping the meanings of its constituents (because the former is ontologically dependent on the latter), it simply doesn't follow that one can't grasp the intension of a complex without grasping the intensions of its constituents.

It is true that the Lewisian cannot give a specification of the semantic values (intensions) of certain derived category expressions independently of those of the semantic values (intensions) of expressions of greater complexity: but he may, with some legitimacy, claim that this requirement is too stringent. For one thing, the former semantic values are conceived as *ontologically* dependent on the latter; and so the view that there is epistemic dependence in this direction gains some initial motivation. But there is also another, perhaps more fundamental point. According to Dummett, Frege maintained

that in the order of *explanation* the sense of a sentence is primary, but in the order of *recognition* the sense of a word is primary.... [W]e understand [a] sentence – grasp its sense – by knowing the senses of the constituents, and, as it were, compounding them in a way that is determined by the manner in which the words themselves are put together to form the sentence.... But, when we come to any general explanation of what it is for sentences and words to have a sense, that is,

of what it is for us to grasp their sense, then the order of priority is reversed. (Dummett, 1981: 4)

If Frege was right about this, and in particular about sentential semantic values being primary in the order of explanation, then we should prefer a molecularist semantics to an atomistic one. This is just what Lewis provides. Indeed, if we read the word “sense” in the above passage as meaning *intension* whenever *explanation* is concerned, and *meaning* whenever understanding is at issue, then Lewis’ semantic ontology respects both aspects of the Fregean doctrine. A concomitant effect of this molecularist approach, however, is that the semantic values of certain subsentential expressions, the derived category words, cannot be specified independently of the SVs of sentences containing them. Yet, as we saw in connection with his discussion of the problem of predication, Davidson himself recognizes the importance of explaining what a predicate means in terms of its contribution to what sentences containing it mean. In short, he accepts the fundamental Fregean insight noted above, and its corollary the context principle – never to ask for the meaning of a word in isolation, but always in the context of a sentence in which it occurs.

But are there not problems raised by the acceptance of Dummett’s third view, according to which the meanings of predicates and sentences are ontologically dependent upon our grasp of them? I do not think so. For semantic theorists themselves are full blown *users* of their meta-language. As such, we may assume that they grasp at least the intensions, and perhaps the (Lewisian) meanings as well, of both simple and complex object language expressions. The important point is that this is so, *even if* they grasp them as the SVs of their own meta-language expressions. But then it seems they are in a position to talk about complex object language expression

SVs, and to appeal to them in explaining what the SVs of simple object language expressions are.¹²⁴

The concern was that the semantic values Lewis assigns to sentences and predicates are trivial. According to this objection, the Lewisian theory doesn't tell us enough to determine what the meanings of both simple and complex object language expressions are; nor can it. The thought was that we can't know the intensions of the parts without knowing the intensions of the wholes, since the former are ontologically dependent on the latter; yet we can't know the intensions of the wholes first, since our grasp of them is epistemically dependent upon our grasp of the intensions of the parts. So we can't supplement Lewis' account in a way that solves the problem. But once we are clear about the distinction between meanings and intensions on the one hand, and the distinction between theorist and language user on the other, the problem dissolves. Thus we can see that Davidson's objection does not apply to Lewis' semantics.

A Problem for MTS

There is, however, another worry which I think can be pressed against Lewis' approach to semantics; indeed, as we shall see, it applies equally to any MTS.¹²⁵ The worry in question is one surrounding the semantic value of the English word "set". On Lewis' view, this word is, presumably, a common noun: it belongs to the category C. Thus, according to Lewisian theory its semantic value should be a common noun intension - that is, a function from indices to common noun extensions, which, in turn, are sets of things. Finally, there are surely contexts in which we use the word "set" - for instance, in a set theory course - such that *all* sets are relevant to the truth

¹²⁴ One might worry about how the theorist's - or, for that matter, the user's - knowledge of SVs is *acquired*. But this is another issue again, and one that I shall not address here.

¹²⁵ Or rather, to any MTS subject to the qualification previously made, namely that it contain an interpretation *function* and not merely an interpretation *relation*; in short, any MTS which posits single SVs for non-singular expressions.

or falsity of what we say in that context. If so, then we want the intension semantically associated with the word “set” to map at least some indices (the ones representing the above mentioned contexts) to a set containing all sets. However, we know, by Cantor’s theorem, that there is no such set. So what is the Lewisian semantic value of “set”? It seems there isn’t one that captures the intended English meaning. And since, presumably, every set is a thing, the same problem recurs surrounding the word “thing” too – as well as “object”, “individual”, “entity”, and so on, if these are at least as inclusive as “set”.

The following, then, is an inconsistent triad of claims:

- (A) We sometimes use the word “set”/“thing” to talk about *all* sets/things.
- (B) Whatever we do using the word “set”/“thing” can be semantically described/explained by assigning to this word a Lewisian category C intension.
- (C) There is no set containing all sets/things.

I submit that the offending claim is (B). Lewis, however, is aware of the difficulty which I have articulated as the inconsistency of these three claims; and, of course, he would not accept my diagnosis that claim (B) is false. Since (C) can be rigourously proven, it is non-negotiable: Lewis instead rejects (A). He gives this, his preferred solution, as it concerns the word “thing” (though he gives it in the material mode):

What, then, are things? Of course I want to say, once and for all: *everything* is a thing. But I must not say that. Not all sets of things can be things; else the set of things would be larger than itself. No Carnapian [i.e. basic category] intension can be a thing (unless it is undefined at some indices); else it would be a member of... a member of itself. We must understand the... definitions of extensions, indices, and Carnapian intensions [etc.]... as tacitly relativized to a chosen set of things. (Lewis, 1983: 196 - first ellipsis Lewis'; thereafter mine)

Thus, Lewis suggests that languages in general, hence natural languages in particular, can’t do, semantically, what we thought they could: although we might have assumed that we could talk about all sets, and indeed all things, it turns out that we can’t.

Lewis' strategy for dealing with the trilemma I have framed, then, is to concede that there is no intension which captures what we naively think of as the meaning of "set" or "thing"; but to mitigate this concession by suggesting that sophisticated reflection allows us to relinquish without regret our naïve semantic opinion. He says:

Any language can be treated in a metalanguage in which 'thing' is taken inclusively enough; but the generality of semantics is fundamentally limited by the fact that no language can be its own semantic metalanguage and hence there can be no universal semantic metalanguage. But we can approach generality as closely as we like by taking 'thing' inclusively enough. (Lewis, 1983: 197)

His thought seems to be that all my argument shows is that we can't get for semantics, the truth, the whole truth, and nothing but the truth. But, he suggests, if we can't get the whole truth, two out of three ain't bad, right? After Tarski (who Lewis references) this is the most we should expect: and, given the availability of ever stronger metalanguages, it is, from a theoretical perspective, no more than an inconvenience.

There are, however, a number of problems with this response. Firstly, although Lewisian semantics will, no doubt, yield some correct predictions, there is no guarantee that its proclamations will be restricted to truths. Lewis writes:

[L]et us proceed on the assumption that the set of things has been chosen, almost once and for all, as some very inclusive set: at least as the universe of some intended model of standard set theory with all the non-sets we want, actual or possible, included as individuals. (Lewis, 1983: 197)

The suggestion, of course, is that if the extension of "set" (relative to a given index) is the domain of some standard model of set theory, then there are enough "sets" (relative to that index) to make the claims of set theory true (at that index). However, as Agustin Rayo (2002: 456) points out, there may, for all we know, be set theoretic falsehoods which are true in all standard models of the axioms of set theory (let alone non-standard models); after all, we know that the domain of any such "intended"

model, being a set, fails to include *all* sets. But Lewis' theory may falsely predict that the sentences in question are true.

Secondly, as Timothy Williamson (2003) has pointed out, the contention that our language is expressively limited in the way Lewis requires runs the risk of instability. What is the theorist to say? "There are some things natural languages can't quantify over"? Then she quantifies over them after all. "Some things are not things"? She contradicts herself. Such a theorist, it seems, will be reduced to quietism: whereof one cannot consistently speak, one must remain silent.

This quietism has a further consequence that is relevant to our concerns. We have seen that, on Lewis' view, whatever language one uses in formulating one's semantic theory, the theory in question does not apply to that language. But then one will not be able to endorse an inference to the best explanation whose conclusion is that our predicates and/or sentences have semantic values, and whose major premise is that our best semantic theory for our language posits semantic values for these expressions: we simply fail to have a (good) theory for that language.

Finally, there is a third reason why the Lewisian response to my objection is unsuccessful. Suppose a given Lewisian theory is descriptively adequate for a particular practical purpose: it yields no false predictions, and covers all relevant cases – all utterances produced in a given population, let's say. Nevertheless it is not explanatorily adequate. If it cannot explain how *all* natural language expressions of a given semantic class behave (common nouns, for instance), then the "explanation" it offers of the behaviour of those expressions in the class for which it has a story to tell ("dog", for instance, but not "set") must be viewed as suspect – as mere heuristic explication or elucidation, a predictive tool, and not a metaphysical account of the mechanism of meaning. So once again, but now for a slightly different reason, there can be no inference to the best explanation, based on a Lewisian semantic theory, whose conclusion is that the meaningfulness of the object language sentences and/or

predicates is metaphysically constituted by their having semantic values. Whereas previously the problem was that we had no theory to treat as the best theory, in the case currently under consideration, we have such a theory, but we know it isn't explanatory.

In short, then, there can be no inference to the best explanation of the semantic phenomena based on a Lewisian MTS which licenses the conclusion that non-singular expressions such as common nouns or predicates have semantic values. The problem is that, because there is no set of all sets, any Lewisian semantics for a language which talks of all sets runs the risk of descriptive inadequacy (which in turn engenders explanatory inadequacy). If one then refrains from articulating a semantic theory for those expressively strong languages which allow talk about all sets, one will have no best theory to serve as the basis of one's inference; and if one simply focuses on expressively impoverished languages which cannot be used to speak of all sets, then one cannot avoid the suspicion that one's theory is explanatorily inadequate due to its insufficient generality.

Still, one might suspect that, in a sense, all languages are expressively impoverished – that is, they all fail, intuitively speaking, to permit us to speak about all sets. And yet, this “expressive poverty” might be due to no fault in the languages in question; rather, it might be the world itself that fails to cooperate by failing to provide a determinate totality of all sets. The problem would then be not semantic, but ontological. On this view, it is (A) which is to blame in our triad of claims above, and not (B) as I have claimed; but the reasons for this are not those suggested by Lewis. Indeed, one might think that (A) is to blame not because it is false, but because it makes a less powerful claim than I have been suggesting. Thus although it *appears* inconsistent with (B) and (C), it is in fact perfectly consistent with them. If all of this is right, then my objection fails to go through: there is nothing wrong with Lewisian MTS after all, and we can infer that non-singular expressions do have semantic values.

I think that this response to my objection on behalf of Lewisian MTS fails – though it does draw attention to a simplification I made in presenting my case. For I simply assumed that, as a matter of ontic fact, there is a determinate, non-empty totality of sets. This assumption is, however, controversial. In fact, there are three (epistemic) possibilities to consider. The first is the one we have already examined, in effect that all conceivable sets (whose ur-elements are drawn from the stock of actual objects) actually exist. In this case, however, my arguments go through as above. The second case to consider is that there are in fact no sets whatsoever. It should be clear, though, that if nominalism is true, a theory which posits sets as the semantic values of expressions will not be a metaphysically explanatory theory; and so my conclusion escapes unscathed. Finally, there is a third case, that which was recently raised, namely that although there are some sets, nevertheless there could be more; i.e. that the concept of set is indefinitely extensible, and not all conceivable sets actually exist. It should be noted, however, that if this is the case, then there are no *actual* semantic values for expressions such as “set” in our strongest languages. Yet the mere possibility of the existence of such semantic values cannot metaphysically explain the actual meaningfulness of our predicates and common nouns. So in this case too, there can be no inference to the best explanation of the semantic phenomena, based upon the possibility of a Lewisian MTS for our strongest languages, which yields the conclusion that non-singular expressions have semantic values.

There are just two further points to make. First, note that this objection, although framed against Lewis’ view, applies quite generally to all model theoretic approaches to semantics, provided they take the extensions of common nouns, or predicates, to be sets. In particular, the objection applies to Kaplan’s atomism discussed above. Second, if this assumption is abandoned, and meanings are taken to be objects, but not sets, then we lose the model theoretic account of the meaning of concatenation (be it Lewisian or Kaplanian). Davidson’s objection as interpreted by

Higginbotham then becomes a pressing concern; the theory runs the risk of failing to tell us enough to determine both the meaning of syntactic composition and the meanings of complex expressions.

Summary and Conclusion

Let me briefly summarize what has been done in *Part II: Truth and Meaning*. I began by sketching the distinction between truth theoretic semantics (TTS) and model theoretic semantics (MTS), and by presenting Davidson's objection to MTS. I then examined Davidson's discussion of the problem of predication. I showed that the decision to treat predicates as syncategorematic expressions, and more generally the adoption of TTS, is Davidson's preferred solution to this problem. I also argued that Tarski did not himself advocate Davidson's preferred solution to the problem, despite Davidson's attributing it to him. In the course of this argument I introduced the distinction between semantics and meta-semantics.

In the second chapter of this part of the thesis I proceeded to consider model theoretic approaches to semantics. Seeking a philosophical interpretation of model theoretic semantics, I looked to Etchemendy's (1999) discussion, and followed him in distinguishing between representational MTS and interpretational MTS. I argued that only an interpretational MTS was likely to license the conclusion that sentences and/or predicates have semantic values; but I insisted that even then only an examination of a sufficiently rich fragment of a natural language should be considered relevant. In particular, I suggested that we look at the semantics of modal languages. I also distinguished atomistic and molecularistic approaches to semantics, and then proceeded to examine in some detail both Kaplan's atomistic MTS and Lewis' molecularistic semantic theory. I showed that Davidson's objection to MTS, and to the positing of semantic values in general, cannot be made to work against either view; but I also argued that no MTS can possibly be explanatory of the meaningfulness of sentences or predicates.

These last considerations, raised in the previous section, point to the conclusion that there is no good reason to believe that sentences and/or predicates are meaningful by virtue of having semantic values: on this point Davidson is right. On the other hand, it is not clear that Davidson's own preferred solution to the problem of predication works: for, although I have not shown this here, Davidson fails to address the problems surrounding the semantics of modal expressions. It seems that semantic values for predicates and even sentences are useful for semantic prediction in modal languages, despite the fact that the possession of a semantic value is not in itself explanatory of the way in which these expressions are meaningful. The best resolution of these seemingly contrary facts is to endorse an instrumentalist interpretation of our semantic theory: predicates and sentences have semantic values, but these semantic values are not metaphysically explanatory of the meaningfulness of these expressions.

In the upcoming *Part III: Truth and Existence*, I argue that we should reject the Quinean criterion of ontological commitment. I begin by distinguishing two accounts of the relationship between truth and reference. I then show that where one of these accounts in particular holds, we should not consider ourselves committed to the existence of the "things talked about" by our best theories.

III. Truth and Existence

This third and final part of the thesis, *Truth and Existence*, is itself composed of two chapters. The first (chapter 5) is called “Truth and Reference”; the second (chapter 6) is titled “Hume’s Principle: A Case Study in Ontological Commitment”.

In “Truth and Reference” I argue in favour of metaphysical pluralism about truth. That is, although the concept of truth is univocal, nevertheless, different properties (as it were) answer to that concept in different areas of natural discourse. Or, to put the point in more nominalistically acceptable fashion: what it takes for a claim to be true differs in a systematic way, depending on the subject matter of the claim. In particular, and by way of example, I claim that truth in mathematics consists in mere coherence, while in at least some empirical areas it requires correspondence.

In “Hume’s Principle” I argue against the Quinean approach to ontological investigation. While Quine’s condition of ontological commitment, along with his distinction between ontology and ideology, is largely formal – it relates to the grammatical distinction between singular term and predicate – I suggest instead that we look to the notion of causal explanation in attempting to understand what there is. My own positive view is, however, little more than suggested: the chapter consists principally of a clarification and criticism of the Quinean approach, in the context of an examination of the Neo-Fregean foundations of number theory.

Although the reasoning in “Truth and Reference” may be thought to provide some support for the rejection of the Quinean approach to ontology in “Hume’s Principle”, the conclusions drawn in that fifth chapter – in particular, the possibility of instrumentalism about reference, and hence about referents – are not directly or explicitly appealed to in the arguments of the sixth. The two chapters may therefore be thought of as independent, but complementary.

5. Truth and Reference

In this chapter I defend Pluralism about truth, and in particular, the view that truth is correspondence in some domains of discourse, yet coherence in others. I do so by appeal to a distinction between two ways of viewing the relation between truth and reference, which I call “Bottom Up” and “Top Down”. The view is that where reference plays a real, metaphysical role in explaining truth, Bottom Up holds, and truth consists in correspondence; but where truth instead serves to explain reference, Top Down holds, and truth is no more than coherence. I argue, moreover, that there are areas of our speech governed by each of these relationships, and hence that Pluralism holds.

In the first half of the chapter I attempt a precise account of the difference between Bottom Up and Top Down. Thus, in the first section I give a rough sketch of this distinction and I articulate the thesis of Pluralism. In the second section, I briefly sketch how Davidsonian truth theoretic semantics works; this reveals the formal semantic relationship between truth and reference. In the third section, I discuss Tarski’s definition of truth, and contrast Tarskian truth theories with Davidsonian ones. Finally, in the fourth section I discuss “connecting principles”, which link primitive semantic notions to their defined Tarskian counterparts: this allows me to give a final, precise gloss on the distinction between Bottom Up and Top Down.

In the second half of the chapter I argue for Pluralism. Thus, in the fifth section I present Davidson’s argument for the inscrutability – or instrumentality - of reference: it relies on the assumption of Top Down. In the sixth section I give some examples of the use of open sentences of English, which serve to support Bottom Up; the cases involve discourse in empirical domains. I then argue that no such examples can be found in pure mathematical discourse, where Top Down holds. Finally, in the seventh section I articulate the distinction between truth as correspondence and truth as coherence in terms of the role of reference in one’s explanation of truth. It follows

from this, and from what was said before, that truth in empirical domains is correspondence, whereas truth in mathematics is coherence; Pluralism is thereby upheld.

Two Views of Truth and Reference

There are, represented in the philosophical literature, two conceptions of the relationship between truth and reference, which we might characterize somewhat facetiously as follows: (i) reference is necessary for truth; (ii) truth is sufficient for reference. While these formulations are of course equivalent,¹²⁶ nevertheless they indicate a difference in priority attributed to the two notions being related, the first treating reference as somehow independent of truth, the second regarding truth as the independent notion. Thus, for example, Hartry Field (1972) once wrote, “Tarski succeeded in reducing the notion of truth *to certain other semantic notions*” (Field, 2001: 3, italics original, underlining mine); and those other notions might all be brought under the rubric of “reference”.¹²⁷ This early Field was an advocate of the “bottom up” position (i). On the other hand, according to Crispin Wright, “once it has been settled that a class of expressions function as singular terms by syntactic criteria,” and acknowledged “that appropriate contexts in which they do so figure are true,” “there can be no further question about whether they succeed in objectual reference.” “There is... no *deep* notion of singular reference,” such that expressions meeting the more “superficial” requirements just given, “may nevertheless fail to be in the market for genuine – “deep” – reference.” (Wright, 1992: 28-29) Wright advocates

¹²⁶ One might deny this, claiming that it is not material, but metaphysical, necessity and sufficiency which are at issue. One would then have to give some account of metaphysical necessity and sufficiency; I shall not attempt this here.

¹²⁷ Field himself uses the term “primitive denotation” for this purpose; but Davidson (1977b), writing after him, says, “We may take reference to be a relation between proper names and what they name, complex singular terms and what they denote, predicates and the entities of which they are true” (Davidson, 2001b: 216). This characterization runs roughshod over a number of subtle distinctions; nevertheless, as we shall see, the simplifications involved are philosophically well motivated.

the “top down” position (ii): truth is sufficient for reference (in the only legitimate sense of that term).

Wright also defends the possibility of *pluralism* about truth: “if the only essential properties of a truth predicate are *formal*,” as they are according to Wright’s own minimalism, “a matter of its use complying with certain very general axioms (platitudes)” – in particular, a matter of its satisfying the instances of the Tarskian schema T^{128} – “then such predicates may or may not, in different areas of discourse, have a varying *substance*.” (Wright, 1992: 21, footnote 15, italics original) In short, which *property* plays the role delimited by the truth predicate may change with the domain of discourse. Ironically, I will defend this claim precisely by appeal to the possibility of making out a difference between semantic and metaphysical reference; my contention is that the difference between truth by correspondence, and truth by coherence, (and therefore that between realism and anti-realism) is that between areas of discourse in which the semantic reference “relation” is genuine, and those in which it is not.

Davidsonian Truth Theory

According to Davidson, in order to explain the linguistic behaviour of natural language speakers we should attribute to them tacit knowledge of a truth theory for their language. There are two parts to this explanation: (i) truth theories; and (ii) tacit knowledge. I treat them in order.

(i) A truth theory for a language L is a formal theory, typically, though not necessarily, first-order: that is, it is a set of sentences which is closed under deduction – it contains all of the sentences which are derivable from the sentences which it contains. Moreover, amongst its theorems must be an instance of the schema

¹²⁸ Tarski’s schema T is as follows: s is true iff p , where “ s ” is replaced by a structural description of a sentence, and “ p ” by a (possibly homophonic) translation of the sentence so described.

(T) s is true if, and only if, p

for each sentence s of the language L under investigation, where p is a sentence of the meta-language (the language in which the theory is formulated).

Three constraints jointly require us, in giving a theory of truth, to invoke the relational semantic notions of denotation and satisfaction – in short, the notion of reference. First, natural languages contain a potential infinity of sentences. Second, since it is part of the (touted) explanation of the linguistic behaviour of speakers of L that they have knowledge (albeit tacit) of a truth theory for L , and human beings have finite cognitive capacities, Davidson requires that an acceptable theory of truth for L must be finitely axiomatizable: it must be possible to state a *finite* number of primitive truths of the theory from which all the infinitely many instances of the schema (T) are derivable. This rules out trivial theories of truth for L which take as axioms all instances of the schema (T) . Third, because there are general, or quantified sentences in natural languages – sentences such as “Everything is bald” – we cannot, in this manner, explain the truth of every sentence in terms of the truth of its constituent sentences. If we assume a level of syntactic representation at which a quantified sentence has overt variables, not all of the constituents of (closed) sentences are themselves (closed) sentences – “Everything is bald” is something like “Every x : x is bald”. Since “ x is bald” does not itself have truth conditions, the derivation of truth conditions for the sentence “Everything is bald” can’t appeal to the truth conditions of “ x is bald”.

Consequently, typical axioms for a Davidsonian truth theory will be such claims as the following, where “ σ ” ranges over sequences of objects¹²⁹:

¹²⁹ Sequences are total functions from natural numbers to objects. We appeal to them, rather than to objects directly, because of open sentences such as “ x is bald and y is not bald”, which contain more than one free variable. Since the length of sentences of this kind is unbounded, we cannot appeal to finite sequences of objects, hence the requirement that sequences be total functions with the natural numbers as domain.

(John): $\forall \sigma$ “John” denotes John relative to σ .

(Variables): $\forall \sigma$ “ v_k ” denotes $\sigma(k)$ relative to σ .

(Bald): $\forall \sigma$ σ satisfies “ t is bald” iff the denotation of t relative to σ is bald.

(Not): $\forall \sigma$ σ satisfies “Not P ” iff σ does not satisfy “ P ”.

(And): $\forall \sigma$ σ satisfies “ P and Q ” iff σ satisfies “ P ” and σ satisfies “ Q ”.

(All): $\forall \sigma$ (σ satisfies “ $\forall v_k P$ ” iff $\forall \tau$ differing from σ in at most the k^{th} place, τ satisfies “ P ”).

(Truth): Closed sentences are true iff they are satisfied by all sequences.

From such axioms, an instance of the schema (T) is derivable for each object language sentence s .

There is one complication which arises in constructing truth theories for natural languages, and which I would like to mention here. Natural languages contain so-called “indexical” expressions: expressions which, like “I” contribute differently to the truth conditions of sentences in which they occur depending on the context in which they are uttered. If I say “I am male”, what I say is different from what you say by uttering these same words. Consequently, a Davidsonian truth theory for a natural language will have to govern *utterances* of its sentences – that is, sentence tokens, rather than sentence types. This is, in any case, amenable to the naturalist.

(ii) It is notoriously difficult to say what implicit knowledge is, and I will not attempt to explain this notion here. Instead, I adopt Gareth Evans’ approach (1985: 345-346): we will say that a truth theory for a language L is acceptable if *explicit* knowledge of the theory would equip someone to understand the language; that is, to evaluate utterances as correct or incorrect, true or false, given further knowledge of the facts, i.e. of whether the right hand sides of (T) sentences obtain.¹³⁰

¹³⁰ Evans doesn’t explicitly mention this further, non-linguistic knowledge – but it is clear that it is needed.

Tarskian Truth Definition

Davidson proposed his truth theoretic approach to semantics, after having read Tarski (1956). But Tarski's approach to semantics is in fact subtly different from Davidson's: moreover, this difference makes a considerable impact on the philosophical import of their respective truth theories. Although both Tarski and Davidson devise theories of which all the instances of the schema (*T*) are theorems, nevertheless, what these *T*-sentences state is different in each case: for whereas Davidson's notions of truth and reference are theoretical primitives, Tarski's are defined notions. It will be useful, so as not to confuse the Tarskian and the Davidsonian notions, to have different names for them. Since what Tarski defines in giving a truth theory for a language *L* is a set of sentences of *L*, I will call the Tarskian notion *L*-Truth; similarly, I will speak of *L*-denotation, and *L*-satisfaction.

Tarski's definition of *L*-satisfaction is recursive; it accordingly looks very much like the corresponding Davidsonian axiomatization. But given our notational decision, we can differentiate the clauses of the recursive definition, from the correlative Davidsonian axioms. Thus, a Tarskian theory will contain such claims as:

(John*): $\forall \sigma$ "John" *L*-denotes John relative to σ .

(Not*): $\forall \sigma$ σ *L*-satisfies "Not *P*" iff σ does not *L*-satisfy "*P*".

(Truth*): Closed sentences are *L*-true iff they are *L*-satisfied by all sequences.

Indeed, since these claims serve to define certain sets, we may make them even more distinctive:

(John**): $\forall \sigma \langle \text{"John"}, \text{John}, \sigma \rangle \in \text{L-Den}$

(Not**): $\forall \sigma \langle \sigma, \text{"Not P"} \rangle \in \text{L-Sat}$ iff $\langle \sigma, \text{"P"} \rangle \notin \text{L-Sat}$.

(Truth**): Closed sentences are elements of *L*-Truth iff every ordered pair consisting of a sequence and that sentence are elements of *L*-Sat.

What guarantees the “existence” of the relevant sets is the background theory: its axioms entail that the expressions themselves, as well as the object language entities, exist, and that sequences of object language entities, as well as sets of whatever there is, also exist.

Thus, Tarski gives a recursive definition of the set L-Sat, and uses it to define L-Truth. However, he also provides an explicit definition of L-Sat, which, given the strength of his background theory, is provably equivalent to the recursive one he employs. Of course, any expression which is given an explicit definition can be replaced, wherever it occurs, by the expression serving to define it. For example, if we give the following definition:

(Bachelor): x is a bachelor iff_{df} x is a man and x is unmarried

then whenever “ x is a bachelor” occurs in a statement, we can replace it with “ x is a man and x is unmarried” without change of meaning. In order, then, to see what Tarski’s *T*-sentences say, it is necessary to see what the primitive notions of his theory are, in terms of which he frames his definitions. A careful examination of Tarski’s work reveals that in fact, all of the primitive notions he employs are either (a) logico-mathematical in nature; (b) syntactic; or (c) notions belonging to the object language subject matter. That is, they are notions expressible with phrases like (a) “if...then”, “for all”, “set”; (b) “the k^{th} variable”, “open sentence”; and, in our case, (c) “John”, and “is bald”. Crucially, none of the primitive expressions of Tarski’s theories are quintessential semantic notions like truth and reference. Thus, his *T*-sentences do not make semantic claims in the standard sense; rather, they are abbreviations of truths of logic, mathematics, syntax, and the object language subject matter. Moreover, the occurrence of object language notions is effectively redundant;¹³¹ accordingly the truth of the Tarskian *T*-sentences depends only on issues of logic, maths, and syntax.

¹³¹ John Etchemendy (1988: 57) suggests that when we replace the defined notion of truth in a Tarskian *T*-sentence, we get a claim such as: [(‘S’ = ‘S’ and it is snowing) or (‘S’ = ‘R’ and it is raining)] iff it is snowing.

They are, as a result, necessary truths (or near enough¹³²), and in this respect unlike the claims made by a Davidsonian truth theory. In short, and roughly speaking, a language has its Tarskian semantic properties essentially, whereas it has its Davidsonian semantic properties only contingently – this shows that the two theorists have different properties in mind.

Connecting Principles

A Davidsonian truth theory governing the utterances of a given population of language users makes contingent, genuinely semantic claims, while a Tarskian truth theory for a language makes pseudo-semantic claims, which are essentially no more than thinly veiled truths of logic, mathematics and syntax. Nevertheless, Richard Heck (1997), following up some ideas of John Etchemendy's (1988), has suggested that a Davidsonian theory of truth for a language is logically equivalent to a Tarskian theory, with its definition of L-Truth, together with certain "connecting principles"¹³³ – axioms linking the undefined Davidsonian semantic primitives to the defined Tarskian notions. Since we are not at present concerned with indexical expressions, we may say that a sentence *type* is true if, and only if, each of its *tokens* is true – and similarly for the other semantic primitives. Then, using these semantic primitives governing expression types, we can articulate the following connecting principles:

(CP: Truth) $\forall$ (closed sentences) s (s is true iff $s \in \text{L-Truth}$);

(CP: Satisfaction) $\forall \sigma \forall s$ (open or closed) (σ satisfies s iff $\langle s, \sigma \rangle \in \text{L-Sat}$); and

(CP: Denotation) $\forall \sigma \forall s \forall$ (terms) t (t denotes o relative to σ iff $\langle t, o, \sigma \rangle \in \text{L-Den}$).

¹³² The qualification is due to the fact that the truths of syntax have existential requirements which may not be satisfied.

¹³³ For a crucial qualification of this claim, see Heck (1997: 540, fn 14).

A Tarskian truth theory, supplemented with these further axioms, will allow us to derive all of the Davidsonian *T*-sentences, and other theorems – without allowing the derivation of other sentences expressible in the Davidsonian meta-language.

Heck suggests that this more inclusive theory was what Tarski intended all along, and that it was in fact advocated by the early Davidson (1967b). I don't want to enter into a discussion here of what Tarski's views in fact were on this matter: but I think that Davidson is best interpreted as advocating the theory that results from appending only one of these connecting principles - (CP: Truth) - to the initial Tarskian theory. Let me explain.

Davidson (1977b) recognized that “linguistic phenomena are patently supervenient on non-linguistic phenomena” (Davidson, 2001b: 215). Semantic notions accordingly bear some relation to those from other fields and from everyday life. The description of these relations constitutes an explanation of the supervenience alluded to, and is a task for the philosophically-minded theorist of language. We may therefore agree with Davidson, when, in “Reality Without Reference” he suggests that

the essential question [concerning reference] is whether it is *the*, or at least one, place where there is direct contact between linguistic theory and events, actions or objects described in non-linguistic terms. (Davidson, 2001b: 219).

Davidson's own answer was negative. Reference, unlike truth, he argued, is a purely theoretical notion, bearing no direct or immediate relation to the world pre-theoretically described. We must employ this notion within semantic theory, for without it we have no hope of giving a recursive account of the truth conditions of all the sentences of a language; but we cannot give an analysis, or even an explanation of it in more basic, non-linguistic, terms. Luckily, according to Davidson, we don't require such an explanation: the notion of reference gets all the content it needs from

its relation to that of truth, which may in turn be explicated by its relation to non-semantic notions.

If we take the primitive semantic terms occurring in the connecting principles to express pre-theoretic notions, Davidson cannot be viewed as endorsing the connecting principles governing denotation and satisfaction – for on his view, we have no antecedent concept of reference, and therefore cannot even so much as formulate these principles. This, I think, is a useful way to differentiate proponents of Bottom Up from advocates of Top Down: the former endorse connecting principles concerning reference, and that concerning truth simply falls into place as a result, whereas the latter endorse only (CP: Truth). This approach allows us to explain Davidson's argument for the inscrutability of reference, which I undertake in the next section.

The Inscrutability of Reference

Given similar basic Tarskian theories, which contain only expressions of logic, set theory, and syntax, along with translations of object language expressions, we can define different Tarskian “semantic” properties and relations: L1-Den, L1-Sat, L1-Truth; L2-Den, L2-Sat, L2-Truth; and so on.¹³⁴ These theories will be mathematically different: the members of Li-Den will not, in general, be the same as those of Lj-Den; and similarly for satisfaction. Nevertheless, at the level of the defined notion of *truth*, a wide range of these “semantic” properties will coincide: the closed sentences belonging to L1-Truth, for instance, may be exactly those belonging to L2-Truth. If so, then (by set extensionality) $L1\text{-Truth} = L2\text{-Truth}$, even though $L1\text{-Sat} \neq L2\text{-Sat}$. In such a case, given a connecting principle linking the primitive notion of truth to one

¹³⁴ The meta-language translations of object language expressions will have to be different in order for the defined “semantic” notions to be distinct: nevertheless, if we allow object language expressions to be “translated” by set theoretic expressions (as is standard in model theory), we can employ the same basic vocabulary across these different theories.

of these defined notions, e.g. L1-Truth, primitive truth is *ipso facto* linked to L2-Truth –it’s the same set! However, since Davidson does not endorse a connecting principle governing denotation or satisfaction, he maintains that nothing determines that L1-Den and L1-Sat correspond to the correct semantic relations, rather than L2-Den and L2-Sat – there are no “correct” semantic relations. Thus, Davidson says:

We don’t need a [pre-theoretical] concept of reference; neither do we need reference itself, whatever that may be. (Davidson, 2001b: 224)

This is the thesis of the inscrutability of reference. I will also sometimes describe this position as instrumentalism about reference.

We can put this point another way. If we return to the original formulation of the Davidsonian truth theory, which employs primitive notions of both truth and reference, nothing determines that the correct way to translate this theory into a Tarskian theory plus a version of (CP: Truth) renders “satisfies” and “denotes... relative to” as “L1-Sat” and “L1-Den” rather than “L2-Sat” and “L2-Den”. If explicit knowledge of a Davidsonian truth theory for a language would suffice for understanding utterances of that language, then so too, says Davidson, would explicit knowledge of *either* Tarskian theory plus (CP: Truth).

In order to drive the point home, let me give an example of divergent, but truth-conditionally equivalent theories. Consider discourse about the integers. The sole singular term of the language is “0” and the sole predicate is “even”. One Tarskian theory says:

(L1-0): $\forall \sigma < \text{“0”}, 0, \sigma > \in \text{L1-Den}$

(L1-Variables): $\forall \sigma < \text{“v}_k\text{”}, \sigma(k), \sigma > \in \text{L1-Den}$

(L1-Even): $\forall \sigma < \sigma, \text{“t is even”} > \in \text{L1-Sat}$ iff the $x: < t, \sigma, x > \in \text{L1-Den}$ is even.

Another theory says:

(L2-0): $\forall \sigma < \text{“0”}, -1, \sigma > \in \text{L2-Den}$

(L2-Variables): $\forall \sigma < \text{“v}_k\text{”}, \sigma(k), \sigma > \in \text{L2-Den}$

(L2-Even): $\forall \sigma < \sigma, \text{“t is even”} > \in \text{L2-Sat}$ iff the $x: \langle t, \sigma, x \rangle \in \text{L2-Den}$ is one less than something even.

The truth conditions of closed sentences on either theory will be logically equivalent: the same closed sentences will belong to L1-Truth as to L2-Truth.

The argument for inscrutability cannot be faulted on technical grounds: it is mathematically unassailable. Thus whether we should accept the conclusion depends upon whether we should accept the premise that there is no theory-external notion of satisfaction with which to formulate a true connecting principle governing reference.

What are Davidson’s reasons for thinking that reference is a purely theoretical notion which can receive no immediate explication in more basic terms? His principal line of argument begins with the thought that the only evidence that can be given in support of a semantic theory for a language concerns when (i.e. under what condition) a user of that language holds true a given sentence. Such evidence will be behavioural, and must fit with the project of attempting to rationalize his or her actions; thus it will relate to the ascription to the agent in question of certain propositional attitudes, such as beliefs, which receive expression in the form of whole sentences. In short, one cannot do anything with an expression shorter than a sentence; consequently, a semantic theory can appeal only to evidence relating to the use of whole sentences. In the next section, I respond to this argument, and defend Bottom Up, by appeal to the thought that we make significant linguistic acts using *open* sentences: though as we shall see, my argument only succeeds in empirical domains.

Empirical vs Mathematical Domains

We are looking for a connecting principle linking some pre-theoretic relational notion with the theoretical, language-specific semantic terms: this will constitute a defence of Bottom Up. I will focus on the case of open sentences. It is commonly held that term variables are like pronouns: “x”, “y”, “z” are the formal

counterparts of “he”, “she”, “it”. If that’s right, then open sentences of English will be such expressions as “It is green”, and “It’s alive”. To show that there is a pre-theoretic, utterance counterpart for satisfaction, then, we need to show that these open sentences of English can be evaluated as correct or incorrect, true or false – for it is clear that they aren’t true or false *simpliciter*, but only relative to an assignment of objects to the pronouns.

First example: A psychologist is showing solid colour blocks to a subject, asking “What colour is this?” and then recording the answer. She shows him a green block, and asks the question, to which the subject answers, “It is green”. The psychologist records that the subject’s answer is correct. It seems that in this context the green block is assigned as the value of “it”; the subject’s answer is correct because this block satisfies “It is green”.

But perhaps someone will object that the analysis I have given of this case is underdetermined by the data. It is perfectly consistent with the facts in this case that “It” should be used to refer here to the *colour* green, and “is” express identity. In order to justify the claim that “it” denotes the *block* in this context, we would need to have the psychologist asking the question, “What colour is this block?” But in such a case, the subject’s “it” might be taken to derive its meaning anaphorically from the psychologist’s “this block”, and therefore as somehow descriptive, rather than referential.

My initial response to this objection is to acknowledge that it is perhaps correct. However, if “it” denotes the colour green in this context, rather than the block, the open sentence still seems to be true *of* this colour; hence L-Sat is connected with a pre-theoretic relational notion after all. My second response is to give an example in which it is clear that there is neither indeterminacy nor anaphora.

Second example: Dr. Frankenstein has been working on his monster for months, perhaps even years. Finally, he hooks up the power supply and flips the

switch to “on”. The monster’s eyes open. Dr. Frankenstein says¹³⁵ “It’s alive!” Here it seems the correctness or incorrectness of his utterance turns on whether the monster is alive, i.e. on whether it satisfies “It’s alive”.

Before concluding this section, I would like to suggest that there is a disanalogy between these cases of empirical discourse about a concrete subject matter on the one hand, and mathematical discourse on the other. Again, I focus on open sentences. My opponent claims that empirical and mathematical discourses are exactly analogous. One kind of opponent does not accept that in empirical cases we can use open sentences successfully: I refer him to the previous arguments. The other kind believes that in mathematics we can use open sentences just as successfully as in empirical domains. The idea for him will be to try to create examples in which someone says, for example, “It is prime”, where “it” is supposed to denote some particular, contextually salient number. My counter strategy, however, is to claim that in all such cases we can treat “it” as anaphoric on some description, and therefore not as a case of the use of an open expression after all. Such pronouns occur bound by earlier discourse.

The obvious set up for my opponent will be: You and I are thinking about the number 2. You turn to me and say, “It’s prime”. However, I – as theorist – ask, How did we both get to be thinking about the number 2? Maybe we were considering it as the number of outs in this inning of such and such baseball game; you suggest that it’s prime, in order for me to infer (given that the players are not coming in from the field to bat, i.e. the number of outs is not 3) that there have been 2 outs so far. But then your claim is (materially) equivalent to the one made by an utterance of, e.g., the closed expression, “The number of outs in the third inning of the baseball game is prime”. Or perhaps we are thinking of 2 as the successor of the successor of 0.

¹³⁵ To himself? To his assistant? It doesn’t matter.

Nevertheless, there is an antecedent description here; hence the meaningfulness of “it” can be explained as deriving anaphorically from that closed expression.

Still, this last proposal affords a possibility for stopping somewhere with a non-descriptive expression, namely “0”. Surely, my opponent will say, one can refer to the number 0 using this expression, and not by means of some description. But I reject this claim, and in doing so I appeal to the Fregean analysis of 0. 0 is *the number of things that are not self-identical*. Since it is a description which fixes the meaning of “0”, any apparently open expressions which refer to 0 can be viewed as anaphoric upon this description.

In short, I claim that connecting principles governing reference can be found for empirical aspects of language: Bottom Up is true in these domains. By contrast, no such principles can be found in mathematical discourse: Top Down is true of the language of pure mathematics.

Truth as Correspondence and as Coherence

In an early paper, “True to the Facts”, Davidson (1969) pointed out that the idea of a fact is of no use in clarifying truth. Traditional correspondence theorists of truth had claimed that a sentence or utterance is true just in case there is some sort of complex entity, a fact, to which it bears a particular relation dubbed “correspondence”. But Davidson noted that we have no grasp on what such an entity might be, except as that which is expressed by a true utterance; such an account is clearly of no use in explicating truth.

Davidson’s own suggestion was to appeal to the apparatus of semantic theory in elucidating the metaphysics of truth. His thought was that although (sentential) utterances do not themselves stand in any interesting relation to the world (truth, after all, is a property – or in the formal mode, “is true” is monadic), nevertheless their parts do:

The semantic concept of truth as developed by Tarski deserves to be called a correspondence theory because of the part played by the concept of satisfaction; for clearly what has been done is that the property of being true has been explained, and non-trivially, in terms of a relation between language and something else. (Davidson, 2001b: 48)

I agree that this is so: an utterance may be thought to be true by virtue of corresponding to reality just in case its parts, first, bear interesting relations to parts of the world – they refer to, or are satisfied by them – and second, bear appropriate internal relations to one another – the whole utterance is composed from its parts in accordance with certain rules.

However, the thought that truth by correspondence is sometimes achieved does not sit well with Davidson's instrumentalism about reference (and satisfaction). If reference is not a point of contact between language and the world – that is, if it is not a genuine relation between (parts of sentential) utterances and real objects and events (described in non-linguistic, perhaps natural, terms) – then it seems that the first condition of truth by correspondence specified above is not met.

In a later paper, "A Coherence Theory of Truth and Knowledge", Davidson (1983) argued that "coherence yields correspondence" (Davidson, 2001c: 137). The thought which supports this claim can, I think, already be found in "Reality Without Reference". Davidson (1977b) there wrote:

The theory gives up reference... It can't, however, be said to have given up ontology. For the theory relates each singular term to some object or other, and it tells us what entities satisfy [i.e. L-satisfy] each predicate [i.e. open sentence]. Doing without reference is not at all to embrace a policy of doing without semantics or ontology. (Davidson, 2001b: 223)

The thought seems to be that if we must appeal to reference and satisfaction, even instrumentally, in order to account for the truth conditions of sentences (or sentential utterances), then despite the fact that it doesn't matter which relation (within a

certain range) we appeal to, the fact that there is a relation of the relevant kind is sufficient to establish the need for the entities so related; thus, the theory requires a certain ontology in order to be true.

It is this thought, however, that I wish to challenge. If it is of no consequence *which* objects are related to the open sentences of the language in such a way as to satisfy them, then why think that there need to be objects that are so related at all? That is, if the particularities of one of the relata are irrelevant, then so too, it seems, is that relatum itself. But once we give up the need for this relatum, there is no need to view L-satisfaction as a genuine relation. And if that's right, then L-truth is not a relational property: truth, in such a domain, is not correspondence.

I summarize as follows. Where the truth of a sentence can be explained in terms of the relations it bears to other sentences we can consistently claim that the truth of this sentence consists in its coherence with those other sentences. Even if a semantic theory for the portion of the language containing this sentence appeals to a notion of L-satisfaction, nevertheless, where the particulars of that relation are irrelevant, we do not have a genuine relation. But then coherence does not yield correspondence, and two distinct metaphysical accounts of truth are available.¹³⁶ Moreover, if I am right that we need a non-instrumental notion of reference in explaining ordinary empirical discourse, but we do not in accounting for the truth conditions of pure mathematical statements, then pluralism about truth is sustained: truth in empirical domains is correspondence; truth in mathematics is coherence.

In the next chapter I argue against Quine's criterion of ontological commitment. That I reject this criterion should be unsurprising, given my contention that truth in some domains consists of coherence; yet *how* we are to reject the criterion, and precisely why it fails, remain to be seen.

¹³⁶ It is, of course, a matter of some controversy what coherence itself involves. But this issue need not detain us here: the present point is that however this notion is ultimately cashed out it does not yield correspondence.

6. Hume's Principle: A Case Study in Ontological Commitment

Nominalism is the view that there are no abstract objects. Since numbers are paradigmatic abstracta, the nominalist is committed to the claim that numbers do not exist. However, there are those who claim that number theory expresses truths, and they point out that all of these follow from what is known as Hume's Principle, or HP – the claim that for any pluralities of objects F and G, the number of Fs is equal to the number of Gs iff there is a one to one correspondence between the Fs and the Gs. As a result of this, the nominalist is under some pressure to resist the following:

Argument for the Existence of Numbers

(P1) HP is true.

(P2) HP can't be true unless numbers exist.

Therefore, (C) Numbers exist.

Those who advocate this Argument also offer support for its premises. Typically, (P1) is supported either (i) by *a priori* considerations, or (ii) by holistic, *a posteriori* considerations; and (P2) is supported by Quine's criterion of ontological commitment. Thus, we have the following:

A Priori Argument for the Truth of HP

(P1.0) HP is analytic.

Therefore, (P1) HP is true.

And there is also this:

A Posteriori Argument for the Truth of HP

(P1.1) HP is part of our best scientific theory.

(P1.2) All parts of our best scientific theory are true.¹³⁷

Therefore, (P1) HP is true.

Finally, in support of the second premise of the Argument for the Existence of Numbers, we have this:

Argument for the Commitment to Numbers

(P2.1) HP is Quinean-committed to numbers.

(P2.2) The Quinean criterion of ontological commitment is correct.

Therefore, (P2) HP can't be true unless numbers exist.

The nominalist needs, therefore, not only to say what is wrong with the initial Argument for the Existence of Numbers, but also what is wrong with the argument or arguments in support of the premise of that initial argument that he or she rejects.

Most nominalists reject the first premise, (P1), of the Argument for the Existence of Numbers; accordingly, they need to reject both arguments in its favour. Many philosophers have no time for (P1.0) of the Argument for the Truth of HP; indeed, there are (at least) two arguments against this premise. Some, following Quine, reject the analytic-synthetic distinction, and with it any claim of analyticity. George Boolos (1997), on the other hand, argues that HP can't be analytic, precisely because it is committed to the existence of infinitely many things, i.e. numbers. There are also replies to the A Posteriori Argument for the Truth of HP. Hartry Field (1980), for example, can be read as providing an elaborate defence of the *negation* of (P1.1). Joseph Melia (1995), on the other hand, rejects the premise (P1.2).

¹³⁷ This premise might in turn be supported by appeal to our endorsement of all parts of our best scientific theory, together with the correctness of an appropriate disquotation schema.

I too want to resist the Argument for the Existence of Numbers; but in contrast to the philosophers just mentioned, I reject its second premise, (P2). Accordingly I must also find fault with the Argument for the Commitment to Numbers. I blame premise (P2.2); that is, I reject the Quinean criterion of ontological commitment.

In what follows, I begin by clearly stating the Quinean criterion of ontological commitment; and I modify it - in detail, though not in spirit - in such a way that it can be applied to claims formulated in languages that are not both first-order and extensional. I then proceed to argue that HP is Quinean-committed to the existence of numbers. I first look at the notion of logical form, and articulate some general constraints governing it. I then argue specifically that the logical forms of simple sentences involving definite descriptions are such that those sentences are Quinean-committed to the existence of objects satisfying the descriptions in question. These considerations serve to establish the truth of (P2.1): HP is indeed Quinean-committed to the existence of (objects serving as) numbers.

In the second half of the chapter I show how the Quinean criterion of commitment can be seen to fail. I first look at Ted Sider's (2007) discussion of Neo-Fregeanism and Quantifier Variance. I endorse a modified version of the thesis of Quantifier Variance - which amounts to a kind of contextualism about quantified claims. This semantic position allows me to accept the truth - and indeed, the (local) analyticity - of HP, while denying the existence of numbers. Finally, I give a meta-semantic account of the possibility of HP's meaning what it does, and being true, despite the non-existence of numbers.

The Criterion of Ontological Commitment

"To be," according to Quine (1948), "is to be the value of a variable." (Quine, 1961: 15) Of course, Quine did not suppose that this succinct formula provided an *informative* answer to the ontological question, "What is there?" - or perhaps,

somewhat more explicitly, “What things, or kinds of thing, exist?” – but only that it was a *true* one. In this spirit, we can interpret the dictum as a semantic thesis, concerning how to interpret our variables, rather than as an ontological claim. Thus, in the passage which introduced to the philosophical community the slogan with which we began, Quine himself wrote:

[H]ow are we to adjudicate among rival ontologies? Certainly the answer is *not* provided by the semantical formula “To be is to be the value of a variable”. (Quine, 1961: 15, my emphasis)

Nevertheless, this semantic thesis connecting variables to entities is of some use, on Quine’s view, when we are engaged in ontological investigations. He continued:

We look to bound variables in connection with ontology not in order to know what there is, but in order to know what a given remark or doctrine, ours or someone else’s, *says* there is. (Quine, 1961: 15, emphasis original)

That is, we employ the claim that to be is to be the value of a variable when pursuing ontological investigations in order to arrive at a criterion of *ontological commitment*.

Agustin Rayo thus describes the Quinean approach to ontology as a two step process, writing:

Quine has taught us that ontological inquiry – inquiry as to what there is – can be separated into two distinct tasks. On the one hand there is the problem of determining the ontological commitments of a given theory; on the other, the problem of deciding which theories to accept. The objects whose existence we have reason to believe in are then the ontological commitments of the theories we have reason to accept. (Rayo, 2002: 436)

This is, I believe, quite accurate as a description of the method espoused by Quine.¹³⁸

We might therefore say (if it is not too misleading) that Quine taught that to be is to be the value of a variable *in a true theory* – not because those things which are values

¹³⁸ Though note the use of the factive locution, “has taught us”, which suggests that Rayo thinks Quine got it right; of course, I don’t think so.

of variables in true theories are not also values of variables in false ones,¹³⁹ but rather because this slogan indicates in terse fashion the two stages of the method whereby we ought, according to Quine, to form our ontological beliefs.

Turning our attention to the first of the two distinct tasks enunciated by Rayo we may ask, What exactly is the relevant criterion of commitment? In “On What There Is” Quine says,

we are convicted of a particular ontological presupposition if, and only if, the alleged presuppositum has to be reckoned among the entities over which our variables range in order to render one of our affirmations true. (Quine, 1961: 13)

Elsewhere, he wrote to similar effect:

Entities of a given sort are assumed by a theory if and only if some of them must be counted among the values of the variables in order that the statements affirmed in the theory be true. (Quine, 1961: 103)

As Richard Cartwright (1954) remarked some time ago, these formulations – as well as many others of Quine’s – employ the modal notion of necessity,¹⁴⁰ which, according to Quine himself, is of suspect intelligibility. Of course, most contemporary philosophers are not as suspicious of modality as Quine was; but in any case it will be useful to articulate that criterion in a way that is amenable to the modal sceptic. One reason for doing so is that this is how the criterion in fact seems to be employed by those who follow Quine’s lead; a second, more basic, and no doubt related reason, is that only in this way does the criterion provide a mathematically tractable means for recognizing ontological commitments.

¹³⁹ What would make false theories false if not the very things which are correctly described by true theories? Elsewhere in the same essay Quine was quite clear on this point: “The variables of quantification, ‘something’, ‘nothing’, ‘everything’, range over our whole ontology, *whatever it may be*” (1948: 13, my emphasis); in short, whether our theory of it is correct or not.

¹⁴⁰ In the form of such phrases as “has to be”, and “must be”. As Cartwright (1954) shows, those formulations which do not employ such notions are clearly inadequate.

Cartwright claimed that we may articulate the Quinean criterion of commitment, without appeal to modal notions, as follows:

An elementary theory, T , presupposes [i.e. is committed to] objects of kind K if and only if there is in T an open sentence ϕ having α as its sole free variable such that (i) $\lceil \exists \alpha \phi \rceil$ is a theorem of T ; (ii) it follows from the semantic rules of T that for every x , ϕ is true of x only if x is a member of K . (Cartwright, 1987: 10)

There are a number of points in this account where clarification is in order. First, Cartwright is quite explicit about what he means by “an elementary theory”: in effect, a theory is elementary just in case it is written in an extensional, single-sorted, first-order language with classical rules of inference. Second, the “semantic rules” alluded to here are simply the axioms of the meta-language truth theory T' for the language L of the theory T .

As an interpretation of Quine this is, I think, perfectly apt:¹⁴¹ for, as is well known, Quine advocated the use of classical, extensional, single-sorted, first-order languages for the purposes of science.¹⁴² However, a number of later Quineans have preferred to employ stronger languages of various sorts, and to assess the ontological commitments of theories, or bodies of beliefs, expressed in them. Thus David Lewis, for example, allowed the use of modal languages; George Boolos extolled the virtues of second-order languages; his follower Agustin Rayo has suggested that we employ a special kind of multi-sorted, first-order language, with not only singular, but also

¹⁴¹ Except for the confused presupposition that it is theories, rather than languages, which have “semantic rules”. I remedy this defect below.

¹⁴² Indeed, he thought that one of their advantages was that they brought out the ontological commitments of theories! (Recall his famous remark that second-order logic is “set theory in sheep’s clothing” (Quine, 1970: 66).)

plural, first-order variables;¹⁴³ and Donald Davidson investigated the ontological commitments of speakers of natural languages.¹⁴⁴

In order to formulate a criterion of commitment acceptable to these neo-Quineans, it is clear that Cartwright's attempt will have to be modified. For starters, reference to an "elementary theory" will have to be eliminated – given what is meant by this phrase – and replaced by something more suitable. But second, since we are allowing the possibility of assessing the commitments of e.g. claims expressed in second-order languages, we must take account of the incompleteness of the logic of languages of this kind. Thus, it seems clear that we don't want the first condition to make mention of the proof theoretic notion of a *theorem*; rather, we want to speak of the *consequences* of the claims under ontological scrutiny – that is, of what those claims *entail*. On the other hand, we want to retain the mathematical tractability of the Quinean ontological criterion; so it seems we should prefer to employ the notion of a theorem as regards the meta-theory, even if the logic of that meta-theory should itself be incomplete.

Perhaps the following amendments will suffice to produce a condition of ontological commitment amenable to neo-Quineans:

A set S of sentences of a language L presupposes entities of kind K if it is a theorem of the semantic theory¹⁴⁵ for L that (i) S entails a sentence of the form $\lceil \exists \alpha \phi \rceil$, where α is a variable of first-order, and the sole free

¹⁴³ It should be noted that in Rayo's semantics for these languages the values of variables of both kinds are drawn from a single domain of objects – so "multi-sorted first-order language" might be something of a misnomer. Rayo himself prefers "PFO" (for "plural first-order") and "PFO+".

¹⁴⁴ Of course, Davidson thought that we could account for the semantics of natural language sentences in terms of their logical forms – which he took, as a matter of fact, to be expressible in a multi-sorted first-order language with event quantifiers (Davidson, 1967a). Moreover, it is not clear that he made much philosophical hay of the use of a distinctive sort of variable ranging over events; so perhaps he should not be viewed as diverging that strongly from Quine's position – especially when we bear in mind that Quine (1960) thought we could attribute commitments to speakers *relative to analytical hypotheses*.

¹⁴⁵ I have replaced "truth theory" with "semantic theory" since we are concerned with the notion of logical consequence as well as that of truth condition; and, moreover, it is unclear that a truth theory can account for the semantics of e.g. modal languages.

variable in ϕ , and (ii) for every x , ϕ is true of x only if x is a member of K .¹⁴⁶

I have here articulated not a *criterion* of commitment, stating both necessary and sufficient conditions of ontological commitment, but rather a mere *sufficient condition*. I have done this for two reasons: *first*, because some who accept this condition as sufficient for commitment will not accept it as necessary – for instance, those who allow the use of second-order languages for the articulation of their theories, but who nevertheless (unlike Boolos) think that second-order variables must have sets, classes, or concepts as values;¹⁴⁷ and *second*, because *I intend to argue that it is not sufficient* – the claim that it *is* therefore suffices to differentiate my neo-Quinean opponents from myself.

Let me make my position clear. I reject the application of the above condition of ontological commitment to (sets of) sentences of languages in use – as any language must be if a theory which we accept is expressed in it. Thus, I reject the application of this condition to utterances of natural language sentences. Semantic analysis alone, I claim, does not reveal the metaphysical truth makers – that is, the metaphysically sufficient truth conditions – of our claims; thus, in particular, it does not reveal what must exist in order for our claims to be true. Rather, we must look to the meta-semantics of our language in order to determine the ontological commitments of our claims – or so I will argue.

In order to show that one who holds Hume's Principle to be true is ontologically committed, by the Quinean condition, to the existence of numbers, we

¹⁴⁶ Since the various sentences of an inconsistent theory jointly entail *any* sentence of the language of that theory, it follows from this formulation of the Quinean condition of commitment that such a theory is ontologically committed to the existence of things of *any* kind for which there is vocabulary in the language. This gives perhaps further reason to avoid the acceptance of inconsistent theories.

¹⁴⁷ Stewart Shapiro (1991), for example, seems to be in this camp.

must first investigate the logical form of sentences involving definite descriptions. To this end I turn, in the next section, to consider the notion of logical form in general.

Logical Form

What is logical form? A number of authors have pointed out that when one makes a claim about logical form, there are at least two distinct kinds of claim one might be making. Thus, Mark Sainsbury (2001: Chapter 6, Section 2) notes that Davidson (1967b) drew a distinction, in effect, between giving an analysis of the semantic structure of a sentence on the one hand, and the semantic content of a lexical item, or word, on the other. A correct proposal of the former kind reveals the logical form of a sentence, while a suggestion along the latter lines provides an analysis of the concept expressed by the lexical item in question; or so Davidson thought.

We may illustrate the distinction in question by considering Davidson's paratactic theory of indirect discourse, as presented in "On Saying That". Davidson made two main claims in that (1968) paper. The first was that apparent sentences of the form $[S \text{ said that } p]$ should in fact be recognized to be *two* sentences, the first of which ends after the word "that", and the second of which is simply the "embedded" sentence $[p]$; the occurrence of "that" at the end of the first sentence is taken to be a demonstrative, and is viewed as referring to the utterance of the second sentence. The second claim was that a subject S *says* an utterance u just in case some utterance of S 's makes S and the producer of u *samesayers*; and he proceeded to give a gloss on this neologism, in the hope of rendering this plausible. The first of these claims is, for Davidson, a claim about the logical form of indirect discourse, the second a conceptual analysis of *saying*. Of course, the paratactic theory of indirect discourse is only viable if the logical form claim can be supplemented with a plausible analysis of the concept of saying, construed by it as a relation between subjects and utterances; nevertheless, the two claims are distinct.

Jeffrey King draws a similar distinction to the Davidsonian one between structural and lexical semantic analysis in his (2002) paper, “Two Sorts of Claim about Logical Form”. King’s basic thought is that propositions are structured entities having constituents: thus, in making a claim about the logical form of a proposition, one might be saying something about the structure of that proposition, and identifying its constituents; or, one might be saying something about the internal structure of one of the propositional constituents themselves. King calls the first sort of claim about a proposition’s structure and constituents a *PSC* claim; he calls the latter claims about the nature of propositional constituents *NPC* claims. To my mind, however, the fact that King frames his discussion in terms of the logical forms of *propositions*, muddies the waters.

For starters, King’s view that logical form is a feature of propositions is subject to the objection that the distinction between PSC and NPC claims can’t be maintained after all – for claims about the nature, and in particular the structure, of propositional constituents will turn out to be nothing other than claims about the structure of propositions themselves. If, for instance, one considers the proposition that Andrew is a bachelor, one may be tempted, with King, to view it as a structured entity, whose constituents are Andrew, and the property of being a bachelor. But if one then determines that the property of being a bachelor is itself internally complex, an amalgam of the properties of being male, being adult, and being unmarried,¹⁴⁸ then it will turn out that the proposition that Andrew is a bachelor has more internal structure than our initial analysis suggested. King responds to this objection, saying in effect that it is no more than a terminological suggestion about how to use the term “propositional structure”, and goes on to say that “there are clearly two different sorts

¹⁴⁸ King says, “It might be held, for example, that the property of being a bachelor is complex and has the properties of being unmarried, being adult, and being male as component parts, so that the property of being a bachelor is literally built out of these other properties.” (King, 2002: 124) What does this mean!? See below.

of claim here” (King, 2002: 125). Of course, I agree that there are. The problem, however, is that King’s *account* of the two sorts of claim makes it difficult, perhaps even impossible, to sustain this distinction. If, by contrast, we treat claims about logical form as claims regarding the syntactic structure of *sentences*, then we can maintain the *prima facie* distinction between those two sorts of claim: for on the one hand there are claims about the structures of sentences, and on the other, claims about the definitions (if any) of words or concepts.

Secondly, King wishes to remain neutral about certain issues in metaphysics while drawing his distinction – for instance, whether propositional constituents are Fregean senses, concepts, or properties and objects (King, 2002: 120); moreover, he thinks that those who make PSC claims can remain neutral on metaphysical issues that those who make NPC claims must tackle (King, 2002: 124). But I don’t think he can remain neutral on issues of the first sort; for it is not clear to me that we can come by a suitable notion of internal structure for all possible propositional constituents. Consider the issue about the propositional constituent corresponding to “bachelor” being complex. I don’t think I know what it would take for *properties* to be complex. Mental items, such as concepts, if they exist, might be complex, in much the same way that linguistic expressions are – they might, as it were, have syntax. But properties!? If there are to be any NPC claims, it would seem that – unless my bewilderment is misplaced – propositional constituents had better not be properties after all; King can’t afford to be neutral. And if the first point above is correct, then anyone making any claims about logical form, will turn out to be making PSC claims, and hence will have to take sides on these contentious issues. Finally, it is worth noting that there are those who take propositions to be unstructured sets of possible worlds – or indeed, who don’t think there are propositions at all! How are they to understand and employ King’s distinction?

Lastly, King's decision to frame the discussion of logical form claims in terms of (structured) propositions requires him to rule out "a third sort of claim" which he wants to "set... aside", namely "*syntactic* proposals... about a level of syntax that is intimately connected with semantics" (King, 2002: 119). But these are, it seems to me, the most pertinent logical form claims there are to be made, for two reasons. First, it is agreed by all – those who countenance propositions, and those who don't – that we need a theory of natural language syntax and semantics, and of their interface. Thus, everyone will agree that there *are* questions regarding logical form in this "third" sense. And second, it seems that these will also be epistemically more tractable issues than those surrounding the nature and structure of propositions and their constituents; for we can bring empirical evidence to bear more immediately upon them.

Of course, King has argued elsewhere¹⁴⁹ that propositional structure just *is* the syntactic structure of sentences expressing that proposition. So, if I can draw something like a PSC/NPC distinction in terms of the semantics of sentences, then perhaps that same distinction can be adapted to cover King's propositions. King himself says,

the framework that suffices for distinguishing between our two sorts of claims about logical form is plausible and minimal. We require that propositions are complex entities with constituents "bound together" in a certain way, and that these constituents have some sort of "nature" or "internal structure". I suspect that even more minimal assumptions would suffice for making our distinction or some analogue of it. (King, 2002: 120-121)

I think the distinction King has in mind is just the one Sainsbury finds in Davidson's work between claims of semantic structure (or logical form) and conceptual analysis. This is indeed an interesting and important distinction; it is also one that is drawn on

¹⁴⁹ See e.g. King (1996), (2006), and/or (2007).

the basis of assumptions even more minimal than King's own. Logical form claims concern the semantic features of sentences.

What, then, is the logical form of a sentence? To begin to answer this question, I want to consider briefly Davidson's application of the Quinean condition of ontological commitment to the case of natural language action sentences. Davidson (1967a) remarked that sentences ascribing actions are involved systematically in entailment relations – for example, the sentence “She slurped the soup loudly” entails “She slurped the soup” and is entailed by “She slurped the soup loudly at the counter”; moreover, these entailment relations are obvious to natural language speakers. In order to explain the logical relations between such sentences, Davidson hypothesized that the logical forms of the sentences in question involve explicit existential quantification over events. Since he thought it plain that at least some such action sentences are true, Davidson concluded that events exist.

The conclusion of this line of reasoning is, of course, contentious.¹⁵⁰ I will not attempt to evaluate this claim here, or the argument offered in support of it. However, I will use Davidson's discussion to help sharpen our earlier question. Thus, I ask: What must the logical form of a sentence be if Davidson's argument is to be charitably interpreted?

This much at least seems certain: the logical form of a sentence *S* is a sentence *LF(S)* of a language with a compositional semantics, such that *S* and *LF(S)* are synonymous in the sense that they share truth (and falsity) conditions.¹⁵¹ If we are to apply the Quinean condition of commitment to natural language utterances we must

¹⁵⁰ One point of contention is the taxonomy – or at least the nomenclature: why think that the entities quantified over are events? David Lewis (1986) drew attention to the fact that the “events” needed for semantics may not be the events (if any) needed for metaphysics.

¹⁵¹ There is, in common philosophical discourse, an ambiguity in the expression “logical form”: it is sometimes used to mean a sentence *schema*, sometimes to mean a particular *instance* of that schema. Whereas the former usage is no doubt strictly preferable (in what sense is the thing in question a *form* otherwise?), nevertheless I often use it, as here, to mean the instance.

insist on the truth-conditional equivalence of the sentence of natural language with which we begin and the logical form sentence at which we arrive: otherwise we can derive no interesting metaphysical conclusions from the fact that our natural language utterance is true.

A second requirement on logical form is imposed by some: S must be a syntactic transform of LF(S).¹⁵² On this way of thinking, logical form is the level of syntactic representation at which semantic interpretation occurs.¹⁵³ S and LF(S) are therefore guaranteed to be synonymous since they are two representations of one and the same sentence. Finally, a third suggestion may be considered: the syntax of S' must render perspicuous the entailment relations into which S enters, in the sense that those relations can be recognized by purely syntactic means. Should we accept either of these two further requirements on the relation between natural language sentences and logical forms? I will argue that we should accept the second requirement, but not the third.

Let's look in a little more detail at the Davidsonian strategy. Davidson is committed to the claim that the logical form of "She slurped the soup loudly" is given by the first-order sentence:

$\exists e [\text{Slurped}(\text{she}, e, \text{the soup}) \ \& \ \text{Loud}(e)];^{154}$

¹⁵² What makes one sentence a syntactic transform of another - rather than, say, the image under a translation scheme? The answer, I think, is that a mapping from sentences onto sentences is a syntactic transformation just in case a machine using that language as its machine language, could perform the mapping. In other words, the mapping must be sensitive to the syntax of the sentence, rather than to a representation of the syntax of the sentence.

¹⁵³ Gilbert Harman (1970), for example, identifies Chomskyan "deep structure" with logical form. Jeffrey King (1996) recognizes that this level of syntactic representation exists, and that Chomskyans often call it "LF" for "logical form"; he prefers to call it "SI", which I believe is intended to abbreviate "semantic input". King's terminology may be preferable to that of the Chomskyans; nevertheless I stick with "logical form" in what follows.

¹⁵⁴ Some complications arise surrounding this example. Davidson (2001a: 106-107) argues with respect to the phrase "slow" that it is not clear whether this can be applied to an event except relative to a description: thus, what is slow for a channel crossing may be fast for a swum channel crossing. He might, for a similar reason, object to the application of "loud" directly to events considered

or as some Davidsonians have preferred (ignoring tense):

$\exists e [\text{Slurping}(e) \ \& \ \text{Agent}(\text{she}, e) \ \& \ \text{Patient}(e, \text{the soup}) \ \& \ \text{Loud}(e)]$

If so, and if the logical form of “She slurped the soup” is:

$\exists e [\text{Slurping}(e) \ \& \ \text{Agent}(\text{she}, e) \ \& \ \text{Patient}(e, \text{the soup})]$

then we can readily explain the fact that the former sentence entails the latter; for the inference from the one sentence to the next is nothing more than an instance of and-elimination, book-ended by existential elimination and introduction – all sound formal rules of inference.¹⁵⁵ Indeed, not only can we explain in this manner the fact that these entailment relations hold, we can also thereby explain *how we know them to hold*.

In this connection, however, it is worth reminding ourselves of the existence of incomplete formal systems, e.g. systems of second-order logic. Such systems are governed by relatively well-understood compositional semantic theories, but their consequence relations cannot be syntactically characterized: that is, there is no set of purely formal rules of inference each member of which is sound (i.e. truth preserving), and which collectively suffice to effectively draw out all of the semantic consequences of a given set of sentences of the languages in question. Clearly it would be a mistake to impose the requirement of perspicuity on the logical forms of sentences of any such language: no such requirement could be met. Yet we cannot rule out *a priori* that the semantic structure of natural language is second-order. It

extensionally. In what follows I ignore this; for in any case, there are alternative examples which could be given which are not subject to this concern.

¹⁵⁵ The Davidsonian might even prefer to represent the logical form of action sentences as first-order *open* sentences, with a *free* event variable, to which a contextually salient event is assigned in semantic interpretation. This would allow us to represent the inferences in question as involving just one step without impugning their validity. Moreover, it would still license the metaphysical conclusion that there are events since if an open sentence with *e* free is true relative to an assignment *f*, then so is its existential closure; and an existentially quantified sentence is true iff it is true relative to some assignment. On the other hand, Davidson himself (1967a) is quite explicit that he thinks such claims do not involve *reference* to events – and the assignment of events to free variables might be thought to be just that.

therefore seems possible that the logic of natural language too is incomplete, and that what grounds our knowledge of at least some of the consequence relations of natural language is intuitive, i.e. brutally semantic. Thus, it seems we should not require, as a general principle, that the syntax of LF representations should lay bare the *entailment* relations of a sentence of natural language: this third requirement on logical form is far too stringent.¹⁵⁶

But now, what if we fail to impose the second requirement on logical form? What guarantees that the LF sentences really do share their truth conditions with the natural language sentences of which they are the logical forms? How do we know that they aren't just reasonably good paraphrases? If we claim that we know the truth-conditional equivalence to hold on intuitive grounds, we are surely entitled to ask why we should not think that this same semantic intuition is what grounds our knowledge that the relevant entailment relations hold. What advantage is there in claiming that we have intuitive knowledge of certain equivalences rather than intuitive knowledge of certain entailment relations? The events skeptic may well claim that there is none. If the Davidsonian could show that there is some systematic way to syntactically transform sentences like his alleged first-order logical forms into the relevant sentences of natural language, however, he would have a considerable advantage: he would be able to explain our knowledge of *all* the relevant semantic facts in terms of knowledge of syntactic facts. Unfortunately for the Davidsonian, however, the first-order formalizations Davidson offered do not seem to be likely candidates for inputs to transformation rules with natural language sentence outputs. The question whether our ordinary action sentences commit us, by Quinean

¹⁵⁶ In fact, even first-order languages are not *decidable*: that is, if given a set of claims Γ and a sentence ϕ , we cannot always arrive at answer, using syntactic means alone, as to whether ϕ follows from Γ ; in particular, when it does not, we cannot always determine in this manner *that* it does not. Davidson might contend that it is harder to ascertain when entailment relations fail than to determine when they hold; and this may be psychologically plausible.

standards, to the existence of events therefore turns on whether, in some other syntactically viable logical language, they contain variables ranging over events. I will not address this issue in what follows, but rest content at drawing the conclusion that LF representations must meet the first two, but not the third, of our touted requirements.

Having drawn the distinction between analyses of the semantic structure of a sentence and of the semantic content of a word, Sainsbury points out that Russell and Wittgenstein employed the term “logical form” in such a way that analysis of the latter kind - conceptual analysis - could reveal logical form. He suggests that these early analytic philosophers were unconcerned with any distinction between structural and conceptual analysis since they were principally interested to give the logical forms not of sentences but of thoughts. I have not addressed here questions pertaining to logical features of thought itself. However, I have argued (i) that we ought to regard claims about logical form to be concerned with the semantic features of linguistic expressions not the structural features of propositions; and (ii) that claims of logical form pertain to the semantic structure of sentences rather than the semantic content of words. In short, I side terminologically with Davidson as against King on the one hand, and Russell and Wittgenstein on the other; and anyone hoping to read ontological commitments into natural language sentences should do the same.

Descriptions

In this section I argue that simple sentences involving definite descriptions are Quinean-committed to things satisfying the descriptions in question. To do this, I engage with the question what the logical form of such a sentence is. In particular, I attempt to determine whether descriptions are quantifiers or singular terms.

Bertrand Russell (1905) proposed, in effect, that the logical form of sentences with surface grammatical form ‘The F is G’ is given by the following sentence schema of first-order logic:

$$(LF) \exists x[(\forall y(F(y) \leftrightarrow x=y)) \wedge G(x)].$$

Note that on this view, phrases of the form ‘the F’ are semantically incomplete symbols; at the level of logical form – the level of syntactic representation pertinent to semantic interpretation – there is no syntactic unit corresponding to ‘the F’. As Russell himself put it, “a denoting phrase [such as a definite description] is essentially *part* of a sentence, and does not, like most single words, have any significance on its own account.” (Martinich, 1995: 204) For the sake of definiteness let’s consider Russell’s actual example:

(PKF) The present king of France is bald.

The relevant instance of LF is that in which ‘F’ is read as “is presently king of France”, and ‘G’ as “is bald”, i.e.

$$(LF(PKF)) \exists x[(\forall y(\text{Presently king of France}(y) \leftrightarrow x=y)) \wedge \text{Bald}(x)].$$

This sentence is, of course, true iff there is exactly one present king of France and he’s bald; and false otherwise. Since there is presently no king of France, $LF(PKF)$ is false.

Peter Strawson (1950) was the first to seriously object to Russell’s analysis of definite descriptions: he claimed that descriptions are referring expressions, in the sense that one can use them to *mention* or *refer to* a thing in the course of saying something *about* it; they are accordingly semantically complete expressions. In defending this claim, Strawson famously argued that whereas one who uses $LF(PKF)$, for example, thereby *asserts* that there is a unique present king of France, in uttering PKF itself one merely *implies* or *presupposes* the existence and uniqueness of a king. In fact, claimed Strawson, one who utters the latter in the present political climate does not succeed in stating anything at all. Thus, Strawson thought that while Russell got the truth conditions for sentences containing definite descriptions right, he did not capture their *falsity* conditions.¹⁵⁷ But a sentence must share both truth and falsity

¹⁵⁷ Let me be a little more precise. Strawson explicitly claimed that Russell gave *necessary* conditions for the truth of sentences containing descriptions; however, he was somewhat equivocal on the

conditions with its logical form – particularly given our view of logical form (articulated in the previous section), whereby surface and logical forms are both representations of the same sentence! So Russell’s analysis fails to get the logical form of simple sentences containing descriptions right; or so the objection goes.

This objection does, *prima facie*, have some pull: it *is* intuitively odd to say that one who asserts PKF thereby *claims* that there is exactly one French king; and so it might be thought that Strawson was right to insist that descriptions are referring expressions. On the other hand, Strawson’s proposal requires giving up bivalence for sentence tokens; and this might be thought too high a price to pay. Thus, one might stress, for example, that there is a crucial difference between truth and falsity: truth is the aim of assertion, the object of the act. If falsity receives a recursive semantic characterization parallel to the one given for truth, we do not thereby recognize this crucial difference; by treating falsity as a defective default value, however, we do. Yet this can only be done if we admit that sentence tokens are subject to the law of bivalence.¹⁵⁸ Nevertheless, I do not think these considerations resolve the issue against Strawson: for the liar paradox requires us to abandon bivalence anyway. Nothing rules “This sentence is false” ungrammatical; yet it seems none of its tokens can be true *or* false. Some alternative account must be sought of the difference between truth and falsity than that given by Dummett.

In addition to Strawson’s complaint, there is also another straightforward and seemingly obvious objection to Russell’s logical form claim: no plausible syntactic transformation rule is able to yield standard surface sentences involving definite descriptions from instances of *LF*. For what is meant to have happened to the

uniqueness requirement. Moreover, Strawson did not claim that Russell gave *sufficient* conditions for the truth of the relevant sentences; it seems he thought that in addition to the *existence* and *uniqueness* of e.g. a king of France, the speaker must also secure *reference* to him.

¹⁵⁸ This line of thought is inspired by Michael Dummett’s discussion of Strawson’s views (Dummett, 1978: preface, and chapter 1).

quantifiers, the connectives, and the identity predicate? Indeed, the problem is even more strongly felt if we consider the opposite direction of derivation – that involved in speech comprehension, rather than production. Where have these logical expressions come from? It seems they have materialized out of thin air.¹⁵⁹

Applied to the original Russellian proposal this objection seems quite compelling.¹⁶⁰ However, neo-Russellians Gareth Evans and Stephen Neale have, to my mind, countered this charge successfully. Indeed, Evans (1979) is concerned to address “those who think that the proposal to treat ‘The’ as a quantifier need be accompanied by the butchering of the surface structure of English in which Russell so perversely delighted” (Evans, 1985: 191, fn 20); and Neale, in his (1990) book *Descriptions*, explicitly addresses the objection I have presented. Both arrive at what is in effect the same modified Russellian proposal, which, as Evans puts it, is to “treat ‘The’ as a quantifier; specifically, as a binary quantifier, taking two open sentences to make a sentence.” (Evans, 1985: 191-192)

We can perhaps clarify the suggestion as follows. There is a sentence schema of the Barwise and Cooper (1981) language of “Generalized Quantifiers” which is truth-conditionally equivalent to LF , namely:

$$(LF^*) [\text{The } x: Fx] (Gx).$$

Like the Russellian schema, instances of this schema are true iff there is exactly one F and it is G .¹⁶¹ The neo-Russellian contention is, in effect, that the relevant instance of LF^* gives the logical form of PKF; and, as Neale notes, this logical form representation

¹⁵⁹ There is, of course, a perfectly good *semantic* explanation for their occurrence on the Russellian proposal; the problem is that there is no good *syntactic* explanation.

¹⁶⁰ In effect the complaint is that Russell does not give a good account of the semantic structure of sentences involving definite descriptions. This does not mean that his proposal is no good as a conceptual analysis of “the” (though Strawson’s objection, if successful, does); and given what Russell meant by “logical form” he might not have been bothered by the present objection. (Jeff King (2002) makes a similar point.) Of course, this does not mean that we should not be bothered.

¹⁶¹ Neale treats the sentence as false whenever this condition does not obtain, thus making the schema fully equivalent to the Russellian one. By contrast, Barwise and Cooper treat it as undefined if there is not exactly one F ; this is a concession to Strawson’s objection considered above.

can readily undergo syntactic transformation to yield PKF itself. The reason is that $\lceil \text{the } F \rceil$ is treated as a syntactic unit at the level of logical form, thus obviating the need for additional sentential connectives like $\wedge$ and $\leftrightarrow$; in order to arrive at the surface structure we simply drop (the pronunciation of) the variables.

Still, one aspect of Strawson's objection against Russell's original proposal may seem to be vindicated here. As we saw, Russell claimed that phrases of the form $\lceil \text{the } F \rceil$ are semantically incomplete symbols; and he did so on the grounds that there is no syntactic unit corresponding to $\lceil \text{the } F \rceil$ at the level of logical form. On the neo-Russellian account, however, $\lceil \text{the } F \rceil$ is a syntactic unit at logical form, and hence, it seems, a semantic, or logical unit. So perhaps we can view $\lceil \text{the } F \rceil$ as a referring expression after all.

Neale (1998) thinks that this line of thought is incorrect. He claims that $\lceil \text{the } F \rceil$ is not a semantic unit: rather, he says, "the" expresses a relation between concepts, and $\lceil \text{the } F \rceil$ is just as incomplete as "John admires..."; in doing so he agrees with Evans' claim that "the" is a *binary* quantifier. That this is so on Neale's view can be seen from the fact that he gives a recursive semantic clause for "the": sentences of the form $\lceil [\text{The } x: Fx](Gx) \rceil$ are satisfied by a sequence s iff the sequence s^* differing from s in at most the k^{th} place (where x is the k^{th} variable) which satisfies $\lceil Fx \rceil$ also satisfies $\lceil Gx \rceil$. The phrase $\lceil \text{the } F \rceil$ is of course a *syntactic* unit - its instances can be used, for instance, to answer questions of the form, $\lceil \text{What is } G? \rceil$ - and it is this fact which, according to Neale, lures us into mistakenly regarding it as a semantic unit.

In the original Barwise and Cooper (1981) paper, however - unlike the Neale reworking - $\lceil \text{the } F \rceil$ is regarded as a semantic unit. Quantifiers in general are interpreted as naming families of sets, which such family being a function of two factors: first, which determiner ("the", "every", etc.) is involved, and second, which open sentence the determiner is adjoined to. Predicates are interpreted as naming sets, and quantified sentences are true iff the set designated by the predicate is contained

in the family of sets designated by the quantifier. How does this work in the case of sentences of the form $\lceil \text{The } F \text{ is } G \rceil$? The predicate $\lceil G \rceil$ is interpreted as naming a certain set, as is the predicate $\lceil F \rceil$. There is a recursive clause governing “the”, which says $\lceil \text{the } F \rceil$ designates the family of sets which are supersets of the set which interprets $\lceil F \rceil$ if the cardinality of that set is 1, and the empty family otherwise. This clause of the semantic theory, together with the lexical clause governing the predicate $\lceil F \rceil$ determine a semantic value for $\lceil \text{the } F \rceil$; then the clause governing quantifiers, together with the clause governing the predicate $\lceil G \rceil$ determines a semantic value (true or false) for the sentence.

Note, however, that even if one were to adopt this approach to the semantics of English phrases of the form $\lceil \text{the } F \rceil$, this would not vindicate the suggestion that such phrases are singular terms; for this approach does not distinguish descriptions from quantified noun phrases, e.g. those of the form $\lceil \text{every } F \rceil$ – in fact, it treats them exactly analogously. In order to render plausible the claim that such phrases *are* referring expressions one would have to give another semantic treatment of $\lceil \text{the } F \rceil$ altogether; one on which the contribution it makes to the evaluation of a sentence is the identification of an object. The following should suffice: if $\lceil F \rceil$ is a predicate, then the semantic value of $\lceil \text{the } F \rceil$ is the sole member of the set denoted by $\lceil F \rceil$, if that set has cardinality 1, and is undefined otherwise. Then $\lceil \text{The } F \text{ is } G \rceil$ is true iff the value of $\lceil \text{the } F \rceil$ is in the set designated by $\lceil G \rceil$.

Since there is a semantic theory according to which $\lceil \text{the } F \rceil$ is a singular term – and since this semantics is consistent with the neo-Russellian syntactic proposal – Neale’s mere *assertion* that “the” expresses a relation between concepts is not enough to establish that it is a quantifier; he needs an *argument* that this is so. In his (1990), Neale gives what may be represented as follows:

Neale's Proposition Expressed Argument

(NP1) The proposition semantically expressed by a simple sentence containing a (genuine) singular term or referring expression is *object dependent*.

(NP2) The proposition semantically expressed by a simple sentence containing a description is not object dependent, but rather *object independent*.

Therefore, (NC) Descriptions are not singular terms/referring expressions.

The idea is that a proposition is object dependent just in case its truth at a possible world depends on how things are with respect to the same object at every possible world; otherwise it is object independent,

Neale seems to imagine that the Strawsonian will deny the second premise of this argument, (NP2), claiming that 'The F is G' expresses a proposition which is true at a possible world w iff the actual F is G at w ; but Neale claims that these truth conditions are mistaken. I agree: clearly the truth of 'The F is G' at a world w depends on how things are with respect to the unique F at w , not the actual unique F. But I see no immediate reason for the Strawsonian to deny this.¹⁶² To avoid the conclusion of Neale's Proposition Expressed Argument the Strawsonian should deny not (NP2), but (NP1). To make this denial plausible, the defender of descriptions as singular terms might offer the following claim: the proposition semantically expressed by an atomic sentence - i.e. a sentence consisting of n singular terms and an n -place predicate - is true at a world w iff the things denoted by the n terms relative to w satisfy the predicate relative to w . The difference then between descriptions and names then emerges: it is not that the latter are singular terms while the former are not, but rather that names are what Kripke (1980) called *rigid* designators - they designate the same thing, if any, relative to every possible world - while descriptions designate non-

¹⁶² Nevertheless, as we shall see, Mark Sainsbury (2005) in effect does reject (NP2).

rigidly. Without an argument to the effect that there are no non-rigid designators, Neale's argument fails to go through.

Gareth Evans (1979), however, offers just such an argument in the course of presenting an argument against descriptions as singular terms. This latter argument is similar to Neale's in spirit, but employs some different notions; explaining them will help us to understand Evans' argument against the existence of non-rigid designators. The initial argument runs approximately as follows:

Evans' Polyadicity of Reference Argument

(EP1) The semantic contribution of any genuine singular term can be fully explained in terms of the 2-place *reference* relation.

(EP2) The semantic contribution of a description cannot be fully explained in terms of the 2-place *reference* relation.

Therefore, (EC) Descriptions are not singular terms.

Let me explain how this is meant to work.

It is agreed by all sides in the debate that names are rigid designators; indeed, this is what explains the fact that whether "Aristotle could have been a carpenter" is true depends on whether there is a possible world in which *Aristotle himself* was a carpenter. By contrast, however, Evans notes that "The first man in space might have been American" is ambiguous: some tokens of this sentence are true iff there is a possible world in which Yuri Gagarin was American; others are true iff there is a world *w* in which the first man in space at *w* was American at *w*. Now, Evans argues that if we have a principle governing the truth of simple sentences which runs as follows

$R(t_1 \dots t_n)$ is true at w iff $\langle \text{ref}(t_1), \dots, \text{ref}(t_n) \rangle$ satisfies R at w ¹⁶³

then we can regard names as singular terms whose meaning is explained in terms of the 2-place relation of *reference* holding between terms and things. Moreover, we can capture, in this way, the truth conditions of “The first man in space might have been American” in those cases in which what’s meant is that Gagarin might have been American. But in order to get the second reading of this sentence we would need to appeal to a 3-place denotation relation instead:

$R(t_1 \dots t_n)$ is true at w iff $\langle \text{den}(t_1, w), \dots, \text{den}(t_n, w) \rangle$ satisfies R at w .¹⁶⁴

This would allow the denotation of “the first man in space” to vary from world to world as required. But the theoretical cost is that we must replace the 2-place reference relation with the 3-place denotation relation in our semantic theory; whereas if we regard descriptions as quantifiers we can get by with reference (in conjunction with the world relativized notion of satisfaction). Considerations of theoretical simplicity, says Evans, therefore require us to regard descriptions as quantifiers. There are no non-rigid designators.

Note that Evans’ claim, (EP1), that we can fully explain the semantics of singular terms by appeal to the 2-place reference relation is only strictly correct if we consider expression *tokens*; for a term like “I” is clearly a singular term – yet the conventional semantic contribution of this *type* of expression is to refer *in context* to the speaker in that context. Once we recognize this, however, Evans’ argument becomes even stronger. Evans’ preferred semantic theory would be a Davidsonian truth theory governing expression tokens. In principle, this seems right: for we saw in part I, chapter 1, that sentence tokens are likely candidates for the metaphysically

¹⁶³ The expression “ $\text{ref}(t_i)$ ” refers to the thing related by *reference* to the term t_i ; clearly, then, reference is treated as a 2-place (functional) relation in this principle.

¹⁶⁴ Here “ $\text{den}(t_i, w)$ ” refers to the thing related by *denotation* to the term t_i at w , where *denotation* is a 3-place relation.

primary truth bearers.¹⁶⁵ But there are difficulties in giving such a semantics for a language containing modal operators;¹⁶⁶ so Evans reluctantly speaks of the possible worlds appealed to by model theoretic semanticists in presenting his argument. Once we do this we can, of course, *define* a 3-place denotation relation between terms, objects, and worlds. But we then need some account of why or how it is that names are *rigid* designators. We might at this stage appeal to a meta-semantic relation of *reference*, or even *direct reference*; and we might cash this out further, either in intentional terms, or in causal terms, or perhaps in yet other terms again, depending on our philosophical persuasion.¹⁶⁷ Nevertheless, once we recognize that all of this talk of possible worlds is to be regarded as purely instrumental – a heuristic device which allows us to capture the truth conditions of modal sentence tokens – then we see that reference itself *must* be a 2-place relation between token singular terms and objects. Non-rigid designators are then exposed as what Kaplan (1989) has called “monsters begat by elegance”.

Evans’ Polyadicity of Reference Argument seems quite compelling, particularly in light of this support for its first premise.¹⁶⁸ Mark Sainsbury (2005) has responded by denying the ambiguity Evans imputes for sentences involving descriptions; he says that if we remove our “Russellian earhorns” we will recognize that “The first man in space might have been American” can’t be read in such a way that it is true iff there is a possible world *w* such that the first man in space at *w* is

¹⁶⁵ Strictly speaking, what we saw was that *either* sentence tokens *or* individual beliefs are the metaphysically primary truth bearers.

¹⁶⁶ Peacocke (1978) attempts a homophonic truth theory for a language containing modal operators; however, there are difficulties with his proposal.

¹⁶⁷ I favour the causal theory of reference.

¹⁶⁸ Note also that, since Evans’ argument relies on the denial that there are non-rigid designators, if it is successful then so too, in effect, is Neale’s Proposition Expressed Argument; for it is no longer open to the Strawsonian to deny (NP1). Nevertheless, I find Evans’ argument more compelling than Neale’s, as (EP1) is more straightforward than (NP1); for it was unclarity over the meaning of the woolly phrase “proposition semantically expressed” occurring in (NP1) that enabled the initial Strawsonian reply to Neale’s argument.

American at w. I don't buy this, however: the possibility of reading this sentence in this way seems to me entirely intuitive, and not a theoretically informed judgment. Besides which, Evans points out that if we were to treat descriptions as denoting terms denotation would also have to be relativized to an assignment function (or sequence of objects) in order to account for sentences containing quantified relational descriptions such as "the father of each girl"; and we simply cannot deny the existence of descriptions of this type. By contrast, if we regard such descriptions as quantifiers we need only relativize satisfaction in this way.¹⁶⁹

Peter Millican (1990) has argued that the data surrounding so-called "incomplete" descriptions point decisively in favour of the view that definite descriptions are singular terms, rather than Russellian quantifiers. If so, then there is obviously something wrong with Evans' argument, even if we can't pinpoint the fault. Millican's thought is that we typically use definite descriptions in cases where it is clear that there is more than one thing in the universe satisfying the relevant description; and that, moreover, we do so successfully. He also thinks that the standard moves of attempting to explain these facts by appeal to pragmatics (what's said is *false*, what's *implicated* is true), or by appeal to restricted quantification in some form or another cannot cope adequately with the phenomena without being reduced to triviality or else seeming entirely *ad hoc*. But Millican does not consider the fact that these problems also arise in connection with other expressions which

¹⁶⁹ Indeed, Evans seems to think that we need to relativize truth to a world as well as satisfaction; and he urges that we not also relativize reference. But we need not relativize truth to a world. We can simply give a recursive definition of satisfaction relative to an assignment function, a world, and whatever else we might like (times?), and then say that a closed sentence is true iff it is satisfied by some (all) assignment functions *relative to the actual world*. Truth itself comes out absolute as Davidson, and Evans following him, hoped.

clearly *are* quantifiers.¹⁷⁰ If definite descriptions pattern with other quantifiers in this respect we have, as yet, no reason for rejecting the view that they too are quantifiers.

I conclude that we should accept the neo-Russellian proposal that descriptions are binary quantifiers.¹⁷¹ But then it should be clear that a claim of the form ‘The F is G’ is committed, by the Quinean condition of ontological commitment, to the existence of an F. For as we saw, the logical form of any such sentence is given by:

$$\lceil [\text{The } x: Fx](Gx) \rceil;$$

and this, being a closed sentence, is true iff there is a sequence s of objects which satisfies it; hence, iff it is satisfied by a sequence s^* differing from s at most in what object it assigns to “ x ”; i.e., iff s^* satisfies both ‘F’ and ‘G’; which is to say, iff $s^*(“x”)$ is both F and G, and hence only if $s^*(“x”)$ is F. Indeed, even if the above considerations were mistaken, and the Strawsonian account of descriptions were right, it would still be the case that sentences of the form ‘The F is G’ are Quinean-committed to the existence of an F! For Strawson granted that Russell had articulated *necessary* conditions for the truth of sentences containing descriptions. So on a Strawsonian semantics, ‘The F is G’ *entails* ‘Something is both F and G’; and this in turn is clearly

¹⁷⁰ For example, Millican points out that we can successfully use the description “the door” in a context in which everything in a particular house is in the domain of discourse, even if that house contains more than one door. So straightforward domain restriction cannot salvage the uniqueness claim; and the proposal that there are tacit restrictions in place – what’s claimed to exist uniquely is a *salient* door, or a door *that I have in mind* – seems arbitrary (why one restrictive phrase rather than another?). But first, Stanley (2002) has proposed that the restricting phrase come in the form of a variable, whose *value* is determined by context – and there is nothing arbitrary in this; and second, and more importantly, there are examples of just this sort in connection with quantifiers like “every bottle”. Consider the example: “In most of his classes, John fails exactly three Frenchmen” (Stanley, 2002: 368–371). This could be true even though there were more than three salient Frenchmen!

¹⁷¹ I confess to some residual doubts: it certainly seems, psychologically, that when we say “The man in the yellow hat is George’s friend” we are in some sense talking *about* the man in the yellow hat. Perhaps some secondary notion of aboutness (other than *refers to*) can be secured to account for this (the rejected *denotation relative to a world?*). Russell seems to have thought so when he wrote, “if ‘C’ is a denoting phrase [such as a definite description] it may happen that there is one entity x ... for which the proposition ‘ x is identical to C ’ is true... We may then say the entity x is the denotation of the phrase ‘ C .’” (Martinich, 1995: 204) But then I appear to be agreeing with Russell after all! Might it be that the difference between the Russellian and Strawsonian positions is, in the end, merely terminological?

Quinean-committed to the existence of an F! So sentences of the form ‘The F is G’ provably either are, or entail, sentences with variables whose values must include Fs and Gs if they are to be true. The Quinean can apply his condition of ontological commitment either way to conclude that an F exists.

Two Arguments for the Existence of Numbers

We are now in a position to consider a simple argument, which invokes the Quinean condition of ontological commitment, and which purports to show that numbers, conceived as abstract objects, exist. The first premise is that we often say truly such things as:

(*PLANETS*) The number of solar planets is 8.^{172,173}

It is then claimed that the logical form of *PLANETS* contains variables whose values must include numbers if it is to be true, being something such as:

(*LF(PLANETS)*) [The x: number of solar planets x] (x=8)

The Quinean condition of commitment is then invoked, and the conclusion drawn that numbers exist. For clearly, *LF(PLANETS)* is true iff it is satisfied by some sequence *s*; hence, iff there is a sequence *s** differing from *s* at most in what it assigns to “x” which satisfies “number of solar planets x” and “x=8”; and so only if there is an object *o* = *s**(“x”) which is a number of solar planets; and thus, finally, only if something is a number (of some concept). Thus, we have the following:

¹⁷² There are only eight objects in our solar system which meet the International Astronomical Union’s working definition of “planet”. Pluto, which was previously thought to be the ninth planet, is now classified as a dwarf planet, and not a planet (the meaning of “dwarf planet” does not seem to be compositionally determined); the reason is that it has not cleared its orbit of other massive bodies. (These claims are corroborated on wikipedia, for example.) Thus, the number of planets is 8; from which it follows from the disquotation scheme governing the truth predicate that *PLANETS* is true.

¹⁷³ It might help to imagine someone saying this in response to the question (perhaps issued on a TV quiz show), What is the number of solar planets?

Simple Argument for the Existence of Numbers

(SP1) *PLANETS* is true.

(SP2) *PLANETS* can only be true if a number of solar planets exists.

Therefore, (SC1) A number of solar planets exists.

Therefore (SC2) A number exists.

The conclusion of this argument is obviously not amenable to the nominalist.

The second premise of the Simple Argument is supported by Quine's criterion of ontological commitment. Some have accordingly tried to resist the conclusion by insisting that in fact the logical forms of sentences such as *PLANETS* do *not* contain variables whose values must include numbers. According to this view, the true logical form of a claim like *PLANETS* is:

(Alternative - $LF(PLANETS)$): $\exists_8 x: x$ is a solar planet.¹⁷⁴

Since the variables of this sentence need only range over planets, and not numbers, these philosophers claim that the metaphysically contentious conclusion is thereby avoided. However, given the general constraints on logical form articulated above, and the arguments presented in the previous section concerning the logical form of sentences containing descriptions in particular, I find this suggestion highly implausible. For how could a syntactic transformation rule take this alternative logical form for *PLANETS* onto *PLANETS* itself? Where would the expressions “the” and “number of” come from? To make matters worse for this view, we can consider the sentence:

(*SPE*) The number of solar planets is even.

Given that *PLANETS* is true, *SPE* is also true. But in order to get a logical form proposal for this sentence that is even remotely semantically plausible, yet which does not have a variable which must, intuitively, have a number as value if *SPE* is to

¹⁷⁴ This is to be read: “There are eight $x: x$ is a solar planet”.

be true, we will need to engage in conceptual analysis with respect to the word “even”. The result will be messier yet, with additional expressions to be generated and/or eliminated by syntactic transformation.¹⁷⁵

Others who have hoped to avoid the Platonist’s conclusion that numbers exist have argued that although the logical form of *PLANETS* is indeed given by $LF(PLANETS)$, nevertheless, we need not draw the conclusion that numbers are amongst the things there are, since *PLANETS* is not true. What *is* true, they claim is the sentence “There are eight planets”; and this *does* have the alternative logical form given above for *PLANETS*. But this strategy, to my mind, underestimates the generality of the problem. For there is a strengthened argument which aims to show that not only *PLANETS*, but also a wide range of other claims which, at least superficially, concern numbers, are true. The claims in question are, of course, the various theses of number theory, which follow from the following second-order claim, known in the literature as “Hume’s Principle”:

$$(HP): \forall F \forall G [\#F = \#G \leftrightarrow F \approx G]$$

HP says that for all F and G, the number of Fs is identical to the number of Gs if, and only if, the Fs and the Gs are equinumerous; which is to say, if, and only if, there is a one-one correspondence between the Fs and the Gs. Thus we arrive at the Argument for the Existence of Numbers which I presented in the introduction to this chapter, and which I reproduce here for convenience:

Argument for the Existence of Numbers

(P1) HP is true.

(P2) HP can’t be true unless numbers exist.

Therefore, (C) Numbers exist.

¹⁷⁵ There is a discussion of related issues in the Appendix to this thesis.

The challenge with which this argument presents the nominalist is to explain the truth of HP, or to explain away the appearance of its truth, without appeal to numbers. The Quinean claims that the first alternative is not viable: the truth of HP can't be explained without recognizing the existence of numbers. My strategy in resisting the conclusion of the *Argument for the Existence of Numbers* will be to reject the Quinean condition of ontological commitment.

Frege's Philosophy of Arithmetic and Neo-Fregeanism

In a series of publications in the late 19th and early 20th centuries, Frege articulated and defended a novel philosophy of arithmetic (the science of numbers). This Fregean position comprised two principal claims, one metaphysical, the other epistemological. The metaphysical thesis Frege advocated was arithmetical Platonism: according to Frege, (i) numbers, the subject matter of arithmetic, are mind-independent abstract objects. The main substantive epistemological claim Frege made, on the other hand, was that (ii) the truths of arithmetic can be known, and belief in them ultimately justified, by means of pure thought alone, i.e. independently of intuition.¹⁷⁶

This combination of views – the conjunction of (i) and (ii) – is in tension with Kant's claim that objects are given to the mind only in intuition. Frege accordingly denied this Kantian thesis, maintaining that certain logical objects, though mind-independent, are given to the mind in pure thought, and independently of any intuition. That is, Frege attempted to reconcile his arithmetical Platonism, (i), with the epistemological claim, (ii), that the truths of arithmetic can be known independently of intuition by appeal to *logicism*.

¹⁷⁶ Crispin Wright (1983) similarly characterizes Frege's philosophy of arithmetic as the combination of Platonism with an epistemological thesis; but he misidentifies the epistemological thesis as logicism. Although Frege appeals to logicism in defence of an epistemological view, logicism itself is not epistemological thesis. See the next two paragraphs in the main text, and in particular the Fregean definition of analyticity.

Logicism is the thesis that arithmetic is analytic. In defending this thesis, Frege presented a new definition of the term “analytic”, replacing Kant’s earlier attempt. Whereas Kant held that a truth is analytic iff it is a predication such that the subject concept contains the predicate concept as a part, according to Frege, a truth is analytic iff it follows from the principles of logic and definitions alone; otherwise it is synthetic.¹⁷⁷ Already in 1879’s *Begriffsschrift*, Frege articulated the logicist thesis that arithmetic is analytic, and conjectured that it is true. But he claimed that in attempting to prove what are, in effect, the Dedekind-Peano axioms from the laws of pure thought (together with definitions), he was forced to invent a system of notation so as to avoid the unclarity of natural language expressions; thus his *defence* of logicism had to wait. In the *Grundlagen* of 1884, Frege gave an informal presentation of his principal line of argument in support of this position. He then formalized the mathematical components of this argument in the two volumes of the *Grundgesetze* of 1893 and 1903.

The logical objects of Frege’s system are extensions. Their identity conditions are governed by Basic Law V of *Grundgesetze*:

$$(BLV) \forall F \forall G [\text{Ext}(F) = \text{Ext}(G) \leftrightarrow \forall x (Fx \leftrightarrow Gx)]^{178}$$

From this it follows, for any F , that $\text{Ext}(F) = \text{Ext}(F)$, and hence, by existential generalization, that $\exists x [x = \text{Ext}(F)]$. Thus, in particular, it follows that $\exists x [x = \text{Ext}(\approx G)]$, for any G .¹⁷⁹ Frege is consequently able to define *the number of F s* ($\#F$) as $\text{Ext}(\approx F)$, believing all the while that there is such a thing – for this is guaranteed by his Basic

¹⁷⁷ Frege claims that he is simply clarifying what Kant meant, and this is perhaps plausible; for if one considers that Frege’s logic differs from Kant’s, it might be the case that Kant’s account falls out of Frege’s within the context of the older, weaker logic. Thanks to Michael Hallett for this point.

¹⁷⁸ Strictly speaking this is a special case of Law V, which governs what Frege calls “courses of values” more generally; but no harm will result from this simplification.

¹⁷⁹ This derivation must be quite subtle, and involve further higher-order principles regarded as logical by Frege; for $\approx G$ is a second-order concept! See the next note for further comment.

Law V.¹⁸⁰ He then defines 0 as $\#(x \neq x)$; y *succeeds* x as $\exists F \exists z[(Fz \wedge y = \#F) \wedge x = \#(F$ but not identical to $z)]$; and x *is a natural (i.e. finite) number* as x follows 0 in the successor series.¹⁸¹ He uses these definitions to prove

$$(HP): \forall F \forall G [\#F = \#G \leftrightarrow F \approx G]$$

Finally, he proves, from HP, in his second-order logic, theorems equivalent to the Dedekind-Peano axioms. This last result has become known as “Frege’s Theorem”.

Russell showed, however, that BLV is inconsistent. We may regard BLV as composed of two halves:

$$(BLVa) \forall F \forall G [\text{Ext}(F) = \text{Ext}(G) \rightarrow \forall x (Fx \leftrightarrow Gx)]; \text{ and}$$

$$(BLVb) \forall F \forall G [\forall x (Fx \leftrightarrow Gx) \rightarrow \text{Ext}(F) = \text{Ext}(G)].$$

Russell’s paradox can be derived from these claims as follows.¹⁸² Let’s say that an object x *is Russellian* iff $\exists F(x = \text{Ext}(F) \wedge \neg Fx)$. Since $\forall x(\text{Russellian } x \leftrightarrow \text{Russellian } x)$, by BLVb $\text{Ext}(\text{Russellian}) = \text{Ext}(\text{Russellian})$, and by existential generalization $\exists x(x = \text{Ext}(\text{Russellian}))$. Call this extension r .

Suppose r is Russellian. Then $\exists F(r = \text{Ext}(F) \wedge \neg Fr)$. But by the definition of r , $r = \text{Ext}(\text{Russellian})$. So $\text{Ext}(F) = r = \text{Ext}(\text{Russellian})$. By BLVa, $\forall x(Fx \leftrightarrow \text{Russellian } x)$. But $\neg Fr$; hence r is not Russellian. So r is Russellian and r is not Russellian. This is a contradiction. So suppose instead that r is not Russellian. Then $\neg \exists F(r = \text{Ext}(F) \wedge \neg Fr)$. But $r = \text{Ext}(\text{Russellian})$. So, by second-order existential generalization $\exists F(r = \text{Ext}(F) \wedge \neg Fr)$ after all, and r is Russellian. So again we get a contradiction: r is Russellian and r is not Russellian. Assuming that r exists, either r is Russellian, or r is not Russellian. Either way, in the presence of BLVa we get a contradiction. So either BLVa is false or

¹⁸⁰ Perhaps Frege intended that the number of F s should be the extension of the concept *extension of a concept equinumerous to F* . This would avoid the difficulty raised in the previous footnote. Moreover, there are places in the *Grundlagen* where Frege seems equivocal about the distinction between concepts and their extensions; hence the difference between saying that numbers are extensions whose members are concepts, and saying that numbers are extensions whose members are the extensions of those same concepts may not have struck Frege as important.

¹⁸¹ Some serious work goes into giving a second-order logical definition of following in a relation series.

¹⁸² The derivation which follows draws heavily on Boolos’ presentation (Boolos, 1998: 150).

r does not exist. Yet BLVb entails the existence of r. So either BLVa or BLVb is false; either way, we must reject BLV.

It turns out, however, that BLV was used by Frege only in order to derive HP; in particular, the proof of Frege's Theorem never appeals to this inconsistent principle. Moreover, HP itself is consistent.¹⁸³ This has led some, starting with Crispin Wright (1983), to abandon the letter, but not the spirit, of Frege's philosophy of arithmetic. These Neo-Fregeans hold on to both of Frege's principal philosophical claims about arithmetic, (i) and (ii). But on their view number theory consists of the principles of a second-order logic (such as the one in the *Begriffsschrift*) supplemented not with BLV but instead with Hume's Principle. Thus, they maintain that arithmetic is analytic since its claims follow from HP, which they contend is a conceptual truth – that is, in particular, a truth revealed by the analysis of the concept of number.¹⁸⁴ It is this variant of the logicist thesis to which the Neo-Fregeans appeal in easing the tension between (i) and (ii).

It is perhaps worth noting that, given plausible assumptions, the Neo-Fregean is indeed committed, by the Quinean condition I have articulated, to the existence of numbers. The reasoning is as follows. The term “number”, according to the Neo-Fregeans, is a defined term: *x is a number* (more specifically, *x is a cardinal number*), they claim, is to mean the same as $\exists F[x = \#F]$.¹⁸⁵ But, of course, the sentence “ $\exists x \exists F[x = \#F]$ ” follows readily from HP: “ $\#F = \#F \leftrightarrow \forall x(Fx \leftrightarrow Fx)$ ” can be obtained from HP by

¹⁸³ See, for instance, Boolos' discussion of “Gottlob Frege and the Foundations of Arithmetic” (Boolos, 1998: 143-154) for claims to this effect.

¹⁸⁴ More precisely, HP must be regarded by the Neo-Fregean as providing a conceptual analysis of the relational concept of *numbering* – a relation which, as we shall see, they take to hold between certain objects (numbers) and concepts.

¹⁸⁵ Indeed, Frege himself put forward this definition in section 72 of the *Grundlagen*. Note, however, that Frege defined $\#F$ as $\text{Ext}(\approx F)$, so for Frege himself this definition of *x is a number* is equivalent to $\exists F[x = \text{Ext}(\approx F)]$. By contrast, the Neo-Fregean does not appeal to any notion of extension, and the definition of *x is a number* must essentially be made in terms of the symbol ‘#’ which is governed by HP.

(second-order) universal instantiation (twice), and “ $\forall x(Fx \leftrightarrow Fx)$ ” is a logical truth, from which we can infer “ $\#F=\#F$ ”, and hence finally, by existential generalization, “ $\exists x\exists F[x = \#F]$ ”. By the Neo-Fregean definition of number, this is equivalent to: “ $\exists x(x \text{ is a number})$ ”. The only difficulty in showing that this sentence is Quinean-committed to the existence of a number lies in the fact that the phrase “ x is a number” which occurs in it will not be given a homophonic semantic clause since it is a defined expression. As a consequence, we must unpack the definition to reveal the Quinean-commitments of this sentence. The phrase “ x is a number” was defined in terms of the ‘ $\#$ ’ symbol, which in turn we may think of as abbreviating the definition “The x : x *numbers* F ”.¹⁸⁶ Thus, all that can be strictly shown is that HP is committed to the existence of objects which number concepts. If the Neo-Fregean analysis of number is then accepted, we can infer that HP is Quinean-committed to the existence of numbers. Still, it seems reasonable to accept that if there are things which number concepts they are indeed numbers.¹⁸⁷ So HP is Quinean-committed to the existence of a number.

In fact, HP can be shown to be Quinean-committed to two distinct numbers since, for example, the concept *not self-identical* and the concept *identical to zero* are not equinumerous. More generally, HP is Quinean-committed to the existence of n distinct numbers *for any finite* n . Indeed, as George Boolos (1997) has noted, and as

¹⁸⁶ Two points. First, on this approach ‘ $\#$ ’ is not the only symbol which occurs in HP as an abbreviation of something more explicit – for recall that “ $F \approx G$ ” is short for a sentence of second-order logic beginning with an existential quantifier binding a relation variable. Second, reading “The x : x numbers F ” as *the number of Fs* is as natural as reading the description “The y : y fathered x ” as *the father of x*.

¹⁸⁷ We can also show, in exactly the same manner, that “ $\exists x(x \text{ is a number})$ ” is a theorem of Frege’s system: however, the chain of definitions reveals that this commits Frege to the existence of certain extensions, and given the inconsistency of BLV it would not be wise for us, as theorists, to accept that these are to be identified with numbers. Of course, introducing a *primitive* expression into Frege’s language for numbers would *ipso facto* render his theory Quinean-committed to numbers, since an inconsistent theory is Quinean-committed to the existence of things of any kind for which it has a word.

Wright himself was aware, HP appears to require the existence of at least a denumerable infinity of objects. Boolos argued as follows:

HP entails... that there is a partition of concepts [values of the second-order variables “F” and “G”] into equivalence classes, in which two concepts belong to the same class if and only if they are equinumerous. If there are k objects, k a finite number, then... there will be $k + 1$ equivalence classes, *viz.* a class containing each concept under which zero objects fall, a class containing each concept under which one object falls,..., and a class containing each concept under which all k objects fall.... Thus, if there are only k objects, there is no function mapping concepts to objects that takes nonequidistributed concepts to different objects, for there won’t be enough objects around to serve as values of the function, since $k + 1$ are needed. So if HP holds... there must be infinitely many objects. (Boolos, 1998: 305-306)

It is perhaps unsurprising that HP, which is Quinean-committed to the existence of numbers, appears to require the existence of infinitely many things – for surely if there are numbers there are infinitely many of them! Thus we can think of Boolos as arguing that HP is committed to the existence of enough objects for there to be numbers.¹⁸⁸

¹⁸⁸ Note that commitment to infinitely many things cannot be neatly represented as a Quinean-commitment according to the formulation I gave; for only singular variables are claimed to engender commitments on that account, and no single thing is infinite in number (unless, perhaps, we count sets which have infinitely many members; but HP is not committed to the existence of sets). However, (1) as Frege notes in section 84 of the *Grundgesetze der Sprache*, there is a number associated with the concept *natural number*, and that number is not itself a natural number (since it is its own successor). If we invoke this criterion to define *infinite number*, then HP is Quinean-committed to the existence of an infinite number which numbers the natural numbers. Moreover, (2) Agustin Rayo (2002) has made a proposal which is Quinean in spirit. He has suggested that we read ontological commitments off of sentences of what he calls PFO+ languages – languages, that is, with plural variables and (collective) plural predicates. He also suggests a translation scheme from second-order sentences onto PFO+ sentences; thus HP can be rendered in such a language. Rayo writes, “a theory is committed to the existence of infinitely many things if some objects satisfying the plural predicate ‘**InfiniteInNumber**(xx)’ must be admitted as the possible values of a plural variable [such as ‘xx’] in order for the theory to be true.” (Rayo, 2002: 451) It seems to me likely that ‘ $\exists xx$ **InfiniteInNumber**(xx)’ will be derivable from the translation of HP in the logic appropriate to PFO+ languages, and hence that HP will be Rayo-committed to infinitely many things.

Of course, the Neo-Fregean accepts such realism about numbers, and the concomitant Platonism. In order to see why, it will be worth examining Hartry Field's discussion, in his (1984) review of Wright's (1983) *Frege's Conception of Numbers as Objects*, of ontological reductionism. Field considers a reductionist about directions who holds that "the direction of c_1 = the direction of c_2 " is logically equivalent to " c_1 is parallel to c_2 "; and similarly for any other sentences purporting to mention directions. The reductionist argues that since the latter sentence does not require the existence of directions according to the Quinean condition of ontological commitment, neither does the former. Analogously, the reductionist about numbers claims that 'The number of Fs = the number of Gs' is logically equivalent to 'The Fs and the Gs are equinumerous'; from which he concludes that the former sentence does not require the existence of numbers for its truth, since the latter is not committed, by the Quinean condition, to numbers.

There are a couple of things that might be said in response to this reductionist line of thought. Firstly, an ontological inflationist might, it seems, equally well argue that since the first members of the relevant pairs of sentences *are* Quinean-committed to directions or numbers as the case may be, so too are the second members of these pairs. The reductionist has said nothing to privilege his position over this inflationary alternative. Secondly, there is Field's own preferred reply. Field claims that the two members of each pair are simply *not* logically equivalent; rather, they are *provably* (materially) equivalent *within a theory containing axioms postulating the existence of directions or numbers*. I will return to consider the first reply below; here I want to focus on the second reply.

The idea of the ontological reductionist is that claims of the form 'The number of Fs = the number of Gs' aren't really identity claims, and numerical expressions ('The number of Fs') aren't genuine descriptions. Of course, from the point of view of *syntax* the claims consist of two descriptions flanking an identity sign; but from a

semantic perspective they are mere claims of equinumerosity. In short, the various expressions occurring in claims of the form ‘The number of Fs = the number of Gs’ might be thought to be semantically syncategorematic. The truth conditions of the left-hand sides of instances of HP are fixed by the corresponding right-hand side sentences; thus the two are logically equivalent.

There is, however, a problem lurking here for the foundations of number theory. If we do not regard numerical expressions (‘The number of Fs’) as genuine descriptions, then we cannot prove the crucial theorem that every number has a successor – which, in the presence of other theorems concerning the successor relation, in turn shows that there are infinitely many numbers. For in proving this theorem, Frege essentially invoked such predicates as “is identical to 0”, where “0” is itself defined as “The number of things which are not self-identical”. Thus, the Neo-Fregean cannot restrict the occurrence of number terms to linguistic contexts of the form ‘The number of Fs = the number of Gs’. Instead, he must regard the number terms as genuine descriptions; for they must also occur in the predicates which replace the variables “F” and “G” on the right hand side of HP. This in turn requires that such right-hand side sentences are themselves meaningful; yet their truth conditions can’t be fixed in the way suggested above for sentences of the form ‘The number of Fs = the number of Gs’. So, just as Field says, the left-hand sides of HP instances can’t be logically – that is, semantically – equivalent to the corresponding right-hand sides.¹⁸⁹ The reductionist position is misguided, and the Neo-Fregean rightly rejects it.

Nevertheless, there remains a tension in the Neo-Fregean position. For the Neo-Fregean holds that HP, and the various theorems of arithmetic, can be known to hold independently of any intuition; yet at the same time he maintains that HP is

¹⁸⁹ This, Heck (2003) suggests, may be the essence of the Julius Caesar problem. If so, then Frege’s concern is mathematical, rather than purely philosophical.

committed to the existence of numbers. But how can it be known that there are *any things whatsoever* independently of intuition? The Neo-Fregean's appeal to the analyticity of HP does not seem to help here: for whatever is meant by "analytic", it is implausible to think that being analytic is consistent with being knowable independently of intuition *and* being committed to the existence of even one thing. Indeed, the same point might have been applied directly to the original Fregean account: the laws of pure thought alone, it seems, are consistent with an empty universe. Logic does not carry ontological commitment.

Quantifier Variance

Ted Sider, in a recent and very interesting paper (2007), has suggested that Neo-Fregeans should endorse what he calls "Quantifier Variance". Quantifier Variance itself is a kind of semantic thesis that Sider develops, not because he endorses it, but because he wants to be able understand what his opponents – people he calls "Neo-Carnapians" – believe. Neo-Carnapians, unlike Carnap, reject positivism: they believe that the truth and falsity of our claims can outstrip our capacity to know. However, like Carnap, they are ontological deflationists in the sense that they believe that (i) quantificational expressions of "rival" theories can be interpreted so that those theories come out true; and (ii) none of these interpretations is any more natural, basic, or fundamental than any other. In short, they are deflationists about the *significance* of ontological debates.¹⁹⁰

Sider articulates Neo-Carnapian Quantifier Variance as effectively the thesis that there is a class of meanings *C* *with more than one member* which contains the maximally natural candidate quantifier meanings; the idea is that all of the "rival" theories will come out true on some interpretation in *C*. The problem for the Neo-

¹⁹⁰ We should not confuse this kind of ontological deflationism with the view (reductionism) which is opposed to the ontological *inflationism* discussed above in connection with Hartry Field.

Carnapians is to say what the candidate quantifier meanings are in such a way that they do not thereby undermine their own claim that none is more natural than any other. For if they simply speak of domains of discourse, says Sider, they will have to quantify over all of the members of all of those domains, thereby making all but one of those domains mere restrictions on a more natural, more inclusive domain.¹⁹¹ The key idea of Quantifier Variance, however, is that the domain of discourse can be expanded in such a way that this is not merely the removal of restrictions.

In order to make sense of his opponents, Sider develops what he calls an “algebraic” approach to quantifier meanings – saying what they must do, rather than what they are.¹⁹² His theory involves the following entities, and notions:

- *meanings*, with a relation $\leq$ of naturalness on them;
- *contexts* which may *belong* to meanings;
- *quantifiers*, which are identified with meaning-context pairs; and
- *models*, which *depict* quantifiers.

The idea here is that models, which are the familiar objects from semantic theory, consisting of a domain plus an interpretation function, are representations (or depictions) of the world as it appears semantically from the perspective of a given quantifier. Sider then defines truth relative to a meaning and a context:

- ϕ is true at m and c iff ϕ is true in some model which depicts $\langle m, c \rangle$

He also defines a number of other notions which are relevant for our purposes. Thus:

- M is a (proper) *supermodel* of M' iff (i) the domain of M' is a (proper) subset of the domain of M , and (ii) the descriptive expressions common to M and M' receive the same interpretation in both.

¹⁹¹ Strictly speaking, this won't be true if domains are thought of as sets, for there is no set of all sets. For this point to hold good, then, we must think of domains simply as pluralities of objects.

¹⁹² He points out that it is consistent with what he says that quantifier meanings should be translation manuals, sets of possible worlds, *sui generis* entities, or something else besides.

- Quantifier q (properly) *expands* q' iff some model that depicts q is a (proper) supermodel of some model that depicts q'
- $q' = \langle m', c' \rangle$ is a (proper) *restriction* of $q = \langle m, c \rangle$ iff $m = m'$ and q (properly) expands q'
- q is *unrestricted* iff there exists no proper restriction of it.

The thought of the Neo-Carnapian is that there will be more than one unrestricted, and therefore equally natural, quantifier.

So why, according to Sider, should the Neo-Fregean endorse Quantifier Variance? Because, he says, Neo-Fregeans want to *stipulate* both of the following claims to be true:

- (1) The symbol “#” is to be understood in such a way that HP comes out true; and
- (2) “#F” is to be regarded semantically as a genuine description.¹⁹³

The difficulty is that these can only both be true, as we have seen, if there are infinitely many objects – and this, says Sider, is not something that one can stipulate to be the case. Sider’s thought is that if the Neo-Fregean endorses Quantifier Variance then it is open to her to say that the joint stipulation of (1) and (2) forces an interpretation of the quantifiers on which HP is true; but that this meaning is no more natural than one on which the variables of the language don’t range over numbers, which is why it is open to stipulation.

I accept most, though not all, of the semantic position Sider suggests for the Neo-Fregean. Sider tacitly assumes that every quantifier whose meaning is maximally natural under $\leq$ is equally natural. He then views the issue which separates his own ontologically realist position from the ontologically deflationist Neo-Carnapian position as the question whether the class C' of maximally natural quantifiers has one member, or more than one. The position that I wish to defend, however, is that there

¹⁹³ More precisely, as I suggested above, it should be regarded as *abbreviating* a genuine description, viz. “The x : x numbers F ”. Alternatively, the Neo-Fregean could regard ‘#F’ as a singular term: Sider’s point would not be affected.

are a number of *locally* maximally natural quantifiers – these being all those which are unrestricted – yet such that one amongst these is *most* natural. I am therefore not an ontological deflationist in Sider’s sense: I think that there are facts about what exists and what does not. Nevertheless, if C’ contains all of the unrestricted quantifiers, then it has more than one member; and HP is true with respect to some quantifier in C’. Of course, the metaphysical Platonist might agree with me about all of this. What distinguishes me from him is that I think the single *most* natural quantifier meaning is not depicted by a model containing numbers.

To get clearer on what my view is, and how it differs from both Sider’s own, and that of his Neo-Carnapian opponent, it will be worth considering a standard form of semantic contextualism about the quantifiers. Everyone can agree that we sometimes say, for example, “There’s no beer” and thereby communicate that there’s no beer in the house; or that we convey that everyone at a particular party is having a good time by saying “Everyone is having a good time”. Some claim that these phenomena are to be accounted for by pragmatic mechanisms.¹⁹⁴ The thought is that what we say is literally false, though we succeed in communicating a related truth which is nevertheless not semantically expressed by the sentence we utter. Yet many theorists are happy to seek a semantic explanation. According to one simple such semantic account of these phenomena, what varies from one linguistic context to another is the domain of discourse. If the only things that count as pertinent in a given context are the things in a particular house, then “There’s no beer” will be true iff there’s no beer in that house. In another context the things in a given supermarket might constitute the salient domain, and “There’s no beer” will be true iff there’s no beer in the supermarket.

Whether this is the correct account of the semantic mechanism behind the context-sensitivity of our quantified claims is not important for our present

¹⁹⁴ E.g. Kent Bach (1994), (2000).

purposes.¹⁹⁵ What is important is that even if it is correct, the Quinean still has an interpretive, philosophical problem regarding ontology. For there are two alternative positions open to him: one might, as it were, privilege some contexts as ontologically more significant, or one might not. The first of these approaches combines semantic contextualism about the quantifiers with ontological absolutism; the second embraces the context relativity of ontology.

Robert Stalnaker (2003) puts forward what seems to be a relativist suggestion in ontology. He writes:

One may think of quantification theory as a framework for representing, once and for all, one's total ontological commitment. To do this is to assume that there is one big domain of discourse – the domain of what there is – which gives the intended interpretation of quantification theory. To be, then, is to be a value of a bound variable, under this interpretation. But one can reject this metaphysical interpretation and still accept the referential semantics for quantification theory. One can deny that there is a domain of all there is, of which all the domains of discourse in particular contexts are subdomains. One can deny that it makes sense to ask ontological questions outside of a particular context. (Stalnaker, 2003: 39)

It seems clear that Stalnaker's sympathies lie with the second option, which, it seems, makes not only quantification, but also ontology, context relative. This ontological relativity, however, seems too much to swallow: How can what there is depend upon the linguistic context in which we find ourselves?

It should be clear that this second, relativistic position that I am rejecting is just that of the Neo-Carnapian. Like Sider, I accept the combination of semantic contextualism with ontological absolutism.¹⁹⁶ But thanks to Sider's work in setting out the semantic thesis of Quantifier Variance, there is no pressure on me to regard more

¹⁹⁵ Jason Stanley (2002) argues that it is not, and provides an alternative theory (see also (Stanley and Szabo, 2000)); Timothy Williamson (2003) in turn argues, *inter alia*, that Stanley's account is inadequate.

¹⁹⁶ I assume here that Sider goes for a semantic account of the context-sensitivity of quantified claims.

inclusive domains as more natural than sparser domains. Thus, as a contextualist I can consistently maintain (a) that mathematicians speak truly when they say, e.g., “There are prime numbers greater than 10”; (b) that this entails “There are numbers”, in the sense that the latter sentence is true in any context in which the former is; and (c) that numbers do not exist – i.e., there are no numbers. I claim a shift in context, and hence a failure of disquotation.¹⁹⁷

Diachronic Meta-Semantics: A Genealogy of the Concept of Number

Is the position I am defending Neo-Fregean? In a sense, of course it is – I want to regard HP as the foundation of arithmetic, and it was Frege who showed that this can be done. But am I a Neo-Fregean in the sense of holding the epistemological thesis that HP can be known independently of intuition, together with the metaphysical thesis that Platonism about numbers is true?

Let me begin with the metaphysical issue. It should be clear that I am not a metaphysical Platonist – I claim that numbers do not exist! Nevertheless, it will be useful to compare my view to that of Field’s ontological reductionist discussed above. Unlike the reductionist I don’t claim that the left-hand side and right-hand side of instances of HP are *logically* equivalent; but I *do* hold that they are (metaphysically) *necessarily* equivalent – a form of equivalence Field doesn’t consider. Consequently, just like the reductionist, I must show why my ontologically sparse view is preferable to the inflationary alternative.

Obviously, the starting point for such a defence will be an appeal to Ockham’s Razor: entities are not to be multiplied beyond necessity. But what sort of necessity? It is clear that *explanatory* necessity is what’s pertinent – we should not accept the existence of things which have no role to play in the explanation of the relevant

¹⁹⁷ More precisely, I claim that the *simple* disquotation scheme fails for “existential” claims, and that it must be replaced by a suitably context sensitive alternative.

phenomenon. Yet we have seen that numbers *are* needed in giving the *semantics* of numerical expressions of the form ‘The number of Fs’ - at least if these are to be regarded as genuine descriptions, which I, unlike the reductionist, am claiming they must. It is my contention, however, that it is not semantics, but meta-semantics, which is the final arbiter of explanatory need in this connection.

Meta-semantics, recall, is that discipline which aims to explain why it is that our expressions mean what they mean. That is, it aims to answer the question, In virtue of what do our expressions possess the semantic features they do? Thus, it is not merely a *description* of the semantic behaviour of our linguistic expressions which is sought, but a metaphysical *explanation*. This is why we must look here before applying Ockham’s Razor, and determining what ontology is required for the truth of HP, and in particular the left-hand sides of its instances.

Not only do I insist on looking beyond semantic theory to meta-semantics in seeking a metaphysical account of the truth of our utterances; I also believe that we must examine the diachronic properties of languages in use in coming to understand, and metaphysically explain, the semantic properties of utterances produced by a given population.¹⁹⁸ In what follows I will give what Bernard Williams (2002) would have called a “genealogy” of the concept of number in the hope of thereby explaining the truth of HP. First, however, let me say a few words about Williams’ discussion of genealogy, in order to show why this should help with our current problem.

According to Williams, “Genealogy is intended to serve the aims of naturalism.” (Williams, 2002: 22) This of course raises the question what naturalism is.¹⁹⁹ Williams says, by way of a partial answer, “Questions about naturalism... are

¹⁹⁸ For a way of understanding languages as diachronic entities, see the Appendix.

¹⁹⁹ It is often thought that naturalism is an epistemological position; in particular, it is the view, suggested by Quine, that the methods of philosophy are continuous with those of the natural sciences. Yet there is, to my mind, also a metaphysical component involved in naturalism. Hilary Kornblith (1994) also recognizes this; and so, it seems, does Williams.

questions not about reduction but about explanation” (Williams, 2002: 23). He thinks that whether naturalism is true about a given phenomenon depends upon whether we can give an explanation of it in terms of “the rest of nature”; if so, naturalism about that phenomenon is true, and otherwise not. Of course, as Williams acknowledges, whether we can give such an explanation depends upon what we count as “the rest of nature” – and this in turn depends upon what the phenomenon is that we are trying to explain. Thus, Williams suggests that ethical naturalism succeeds, for example, if we can explain the ethical aspects of human life in terms of other (non-ethical) features of human psychology. In our case, the objective will be to explain number talk – discourse that involves reference to numbers – in terms of discourse that is not arithmetical in this way, but merely quantificational.

But what exactly is a genealogy anyway? And how will it help with such naturalistic explanation? Williams claims that a genealogy involves an historical, factually accurate aspect, and a fictional part,

an imagined developmental story, which helps to explain a concept or value or institution by showing ways in which it could have come about in a simplified environment containing certain kinds of human interests or capacities, which, relative to the story, are taken as given. (Williams, 2002: 21)

My own genealogy of the concept of number will be very sparse on factual details. More importantly, however, the mere inclusion of a fictional part raises the question how a fiction – something manifestly false – can explain anything at all. Williams suggests that a potential explanation can fail to be true in one of two ways: it can be “fact-defective”, failing to respect certain particular truths; or it can be “law-defective”, involving some aspect which is nomologically impossible.²⁰⁰ Williams claims that a fact-defective genealogy might nevertheless be explanatory by virtue of

²⁰⁰ Williams does not specify that it is nomological impossibility that he has in mind; yet it is quite important that this is so. Certainly he is not claiming that a genealogy might involve the description of a situation which is inconceivable!

showing how something is possible, even though failing to explain the actuality of that phenomenon. On the other hand, a law-defective genealogy won't be able to do this; for *ex hypothesi* such an account involves some component which is impossible!

Williams does not say explicitly how a law-defective genealogy can be explanatory; however, he does offer an example, and describes his reasons for regarding this particular case as philosophically illuminating. The example Williams discusses is

Hume's derivation of the "artificial virtue" of justice. The state of affairs from which the story starts is one in which people are self-interested and have a capacity for limited sympathy, but have no motives of justice and, correspondingly, no concept of property. Given this condition, and given some further (quite strong) conditions of common knowledge, the story tells how people converge on adopting institutions of property and develop dispositions of justice. The state of affairs from which this process starts, and hence the process itself, Hume recognizes to have been impossible. (Williams, 2002: 33)

Williams claims that this genealogy comprises three explanatory elements. (i) It gives a "functional" – that is, instrumental - account of the explanatory target (justice) in terms of elements (self-interest, limited sympathy) which must be acknowledged on independent grounds – and this despite the fact that we may not expect such an account to be available in this case. (ii) The account is rational, in the sense that the agents with the basic motives would welcome a state of affairs in which the derived motives operated – indeed, Williams claims it is this which makes the relationship between the basic motives and the derived motives functional. Finally, (iii) "it derives the functional from what is not functional or is functional only at a lower level" (2002: 34). I will say that the account is *functional*, *rational*, and *well-grounded*, by which I mean that it has the first, second, and third features mentioned here respectively.

The functionality and well-groundedness of Hume's account of justice make it naturalistically acceptable; and the rationality of the account is what makes its functionality possible. Having drawn attention to these features of Hume's genealogy

of justice, Williams then asks why we should not simply give a straight synchronic functional analysis of justice: what does the diachronic aspect of the story Hume tells add? Williams claims that the straight functional analysis is patently mistaken in this case: we simply do not regard justice as having only instrumental value. The point, I think, is that the description of rationality - under (ii) - is essentially diachronic. Thus, the merit of the genealogical approach is that it shows that a functional, well-grounded, and hence naturalistic, account of justice is conceivable - despite the failure of synchronic functionality, i.e. reductionism.

In our case – the genealogy of the concept of number - it is not the *value* of the target notion that is irreducible, but the *semantic content*. The point is that we can't give a reductive account of the meaning of "number" as it figures in HP (nor, therefore, of number terms in general); for if we attempted to do so, we would not be able to regard HP as the basis of number theory. Yet we can explain the use of the "higher-level" expressions, in terms of the use of the "lower-level" ones; that is we can explain the meaningfulness, but not the meaning itself, of these expressions in this way. Thus, we can give a naturalistic, and nominalistically acceptable, account of arithmetical discourse – and thus, a demonstration that arithmetic is supervenient upon the non-arithmetical; but again, the diachronic aspect of the account is essential.

Let me proceed to the genealogy itself.²⁰¹ Suppose that there once were a community of language users who employed numerical descriptions only in the context of sentences of the form 'The number of Fs = the number of Gs', and who drew inferences freely from any given instance of the left hand side of Hume's Principle to its right hand side, and *vice versa*. Such speakers might wish to encode

²⁰¹ I should say that my genealogy of the concept of number will not, so far as I know, involve any nomologically impossible element. But even if it should turn out that what I say *could not* constitute either an ontogeny, or a phylogeny, of the concept of number – that is, an account of how the concept develops within an individual's psychology, or how the species as a whole came to be in a position to deploy this notion – nevertheless, given the above discussion, the exercise of giving this genealogy will not have been in vain.

what they regarded as the legitimacy of this kind of inference by endorsing every *instance* of the following schema:

$$(\text{HP schema}) \#F = \#G \leftrightarrow F \approx G$$

But these speakers would not be committed by the Quinean condition, to the existence of numbers. For if they allowed as substitution instances on the right hand side only those predicates which contained no occurrences of expressions of the form $\lceil \#F \rceil$ then *every* such instance of this schema could, by Quinean criteria, be true without the existence of numbers. The reasons are twofold. First, there might be a collection of objects such that no predicate of the language had that collection as its extension, and so Boolos' argument above would not go through. But second, and more importantly, even if there were, for every collection of objects, a predicate of the language having that collection as its extension, those endorsing the schema still would not need be Quinean-committed to the existence of numbers as objects; for they would not need to view the symbol “=” in linguistic contexts of the form $\lceil \text{The number of Fs} = \text{the number of Gs} \rceil$ as expressing the identity relation - as we have seen, because the descriptions flanking it occur nowhere else in the language, they may therefore be thought to be syncategorematic.

But then, suppose further that these speakers wanted to articulate their *general* endorsement of the schema – and that, moreover, they wanted to do so *without making a meta-linguistic claim*. How would they go about this? Presumably they would write something like:

$$\forall F \forall G [\#F = \#G \leftrightarrow F \approx G]^{202}$$

But this, of course, is just HP! So it seems that HP could be introduced into the language as a correct and meaningful sentence *even if numbers did not exist*.

²⁰² In fact, Frege (1893) often wrote down sentences with free second-order variables, with the convention that sentences involving free variables were to be treated as synonymous with their universal closure. Or at least, so says Dummett (1991), and I am in no position to argue.

This, however, raises a difficulty. For how could HP thereby mean anything more than the schema which we have seen to be inadequate as a basis for number theory? That is, how could its semantic content outstrip that of the schema? I suggest that it could *provided that the community proceeded to use HP in inferences in a certain manner*; in particular, if they went on to allow substitutions for “F” and “G” that involved the newly introduced numerical expressions themselves.

If this genealogical story were true, then it seems we would have a meta-semantic account of how Hume’s Principle could come to encode the full semantic content necessary for the derivation of number theory, and be endorsed as true, even though numbers did not exist. The difficulty with the original reductionist account considered above was that it did not capture the full semantic content of HP. By contrast, this new account is able, by recognizing the distinction between semantics and meta-semantics, to do justice to the full semantic content of HP, while making no meta-semantic appeal to numbers. That is, the metaphysical explanation of the truth of HP at no point mentions numbers; by Ockham’s Razor we should deny that numbers exist.

Although I reject *metaphysical* Platonism about numbers, nevertheless I endorse a kind of *methodological* Platonism; for unlike the reductionist I deny the semantic (“logical”) equivalence of Hume’s Principle’s instances’ left hand and right hand sides. So from the synchronic, semantic perspective – the perspective we must adopt when engaging in the study of mathematics, for example – we should regard HP and number theory more generally as being *about* numbers. It is only when we turn to deeper, ontological and metaphysical investigations that we must recognize that this is a mere methodological convenience.

The genealogy I have given of the concept of number – building, obviously, on Frege’s enormous contribution – is also useful in number theoretic epistemology; that is, it is helpful in explaining our knowledge of arithmetical truths. Clearly, in order to

recognize a truth as such, one must possess the relevant concepts; yet the diachronic story I have suggested explains the origin of the key concept of number. Moreover, it does so in a normatively vindicatory way: it shows that we are perfectly justified in applying this new concept of number, since the account itself explains how claims in which this concept figures can be true.

The two principal epistemic worries surrounding the truth of HP concern (a) the existence of enough (i.e. infinitely many) things to make it true - it certainly seems we can rationally doubt that there are infinitely many things! - and (b) its consistency - after all, Basic Law V, which was similar in form, is inconsistent.²⁰³ But Boolos has shown that we can prove the consistency of HP given the assumption that there are infinitely many things. The idea is that given a domain of this size we can construct a model for HP; and, of course, any sentence which has a model is consistent. So suppose, for example, that we have a domain consisting of all and only the natural numbers. Then we take the Russellian denotation²⁰⁴ of each phrase of the form ‘#F’ to be one of the natural numbers – but not the one we would expect on the intended interpretation. Rather, each such phrase denotes the standard successor of the number which is its denotation on the intended interpretation; unless ‘F’ is true of all of the objects in the domain, in which case ‘#F’ denotes 0. Since HP comes out true in this model, (b) reduces to (a); the primary epistemic concern, really, is with the apparent (i.e. Quinean) ontology associated with HP.

Yet in any context in which HP is understood, the semantic appearance of the existence of infinitely many things is guaranteed; and so HP can be seen to be true. The reason is that, in effect, HP generates a model for itself:²⁰⁵ for we can take the

²⁰³ The following principle purportedly governing ordinals is likewise similar in form to HP, but inconsistent: $\forall R \forall S [\text{ord}(R) = \text{ord}(S) \leftrightarrow R \equiv S]$, where ‘R’ and ‘S’ range over well-ordered relations, and ‘ $\equiv$ ’ means *is isomorphic to*.

²⁰⁴ Recall that the Russellian denotation of a definite description ‘The F’ is the unique object which satisfies the open sentence ‘x is F’.

²⁰⁵ This is not the case in connection with BLV (or the inconsistent ordinal principle above)!

denotation of each description '#F' to be an equivalence class comprising just such descriptions! Indeed, we can specify a class of such models by means of the following stipulation: let '#F' denote the class of expressions of the form '#G' such that $F \approx G$. Since $F \approx F$, for any F, this equivalence class is always non-empty, including at a bare minimum '#F' itself. Moreover, we are guaranteed an infinite sequence of predicates $F_0, F_1, \dots, F_n, \dots$ such that it is not the case that $F_i \approx F_j$, for any i, j : simply let F_0 be ' $x_0 \neq x_0$ ', and for each $i \geq 1$, let F_i be the disjunction of each of the identities of the form ' $x_i = \#F_n$ ' in which $0 \leq n \leq i-1$.²⁰⁶ (A sentence with only one disjunct is to count as a degenerate disjunction.) This guarantees an infinite sequence of equivalence classes of expressions of the form '#F' all of which may be taken to belong to the domain of the model in question. HP is therefore self-verifying, and the existential doubts, articulated above as worry (a), do not arise.²⁰⁷

Do I endorse the epistemological claim (ii), then, that the truths of arithmetic can be known independently of intuition? The above argument showed simply that understanding HP puts one in a position to know that it is true. But what is required for such understanding? The genealogy of the concept of number I gave showed that we need not have some prior intuition of numbers; yet the possibility remains that simply being able to process the syntax of HP itself requires intuition. After all, it is sentences and their parts which have syntactic properties, and sentences typically require either time or space (or perhaps both) to unfold. So it does not seem to me beyond question that understanding HP, and therefore recognizing it as true, should

²⁰⁶ Thus, F_1 is ' $x_1 = \#F_0$ ', i.e. ' $x_1 = \#(x_0 \neq x_0)$ '; F_2 is ' $x_2 = \#F_0$ or $x_2 = \#F_1$ ', i.e. ' $x_2 = \#(x_0 \neq x_0)$ or $x_2 = \#(x_1 = \#(x_0 \neq x_0))$ '; and so on. The idea is Frege's.

²⁰⁷ Note that if we take the symbol '=' in HP to express not identity but mere indiscernability, we can interpret HP in such a way that it comes out true without expressions of the form '#F' having denotations at all: we simply take #F to be indiscernible from #G whenever they belong to the same equivalence class specified in the model in the main text. This, I think, provides a reasonable picture of what is going on with HP meta-semantically.

ultimately require something beyond pure thought, namely intuition. I accordingly remain agnostic about the epistemological claim (ii).

It may be worth commenting briefly on the Neo-Fregean syntactic priority thesis – that is, to be more precise, Wright’s thesis of “the priority of syntactic over ontological categories”. The basic idea underlying this thesis is that whether or not an expression may be regarded as a singular expression is to be decided on syntactic, rather than semantic grounds. It should be clear that I am sympathetic (given this gloss). It is because predicates containing expressions of the form ‘ $\#G$ ’ can be substituted for “F” in HP – in short, because HP is impredicative – that HP is trivially true. So perhaps my view is not that far removed from the Neo-Fregean position despite my rejection of (i) and my ambivalence towards (ii).

Summary and Conclusion

I began this third part of the thesis by arguing, in “Truth and Reference”, that Pluralism about truth is right: in some domains of discourse truth is coherence, while in others it consists in correspondence with reality. In “Hume’s Principle” I have argued that the Quinean approach to ontology is mistaken. First, I clarified this approach, articulating a precise condition of ontological commitment that I think captures the intentions of certain neo-Quineans who do not wish to restrict themselves to the use and investigation of first-order extensional languages. I then attempted to answer the question what logical form is in general, before going on to say what the logical form of sentences involving definite descriptions is. These investigations enabled me to conclude that sentences of the form ‘The F is G’ are committed, by the Quinean condition, to the existence of an F. I then articulated a couple of arguments in favour of the existence of numbers; both relied on the Quinean approach to ontology. In the remainder of the chapter I argued that this approach fails. I focused on the case of Hume’s Principle and showed that it can be

true despite the non-existence of numbers. In doing so I also compared my view to the standard Neo-Fregean position on arithmetic.

It remains only to provide a diagnostic of the failure of the Quinean condition of ontological commitment. Why does it go wrong? We saw briefly, in connection with names, that their semantic contribution is to refer to objects, yet that this fact is meta-semantically thought to be explained by appeal to the existence of certain causal relations between the referents and the tokens of the name. Quine's condition of ontological commitment, however, takes no account of meta-semantics. Yet my contention is that meta-semantics, and the causal theory of names in particular, points the way to a more general criterion of existence. We should regard as real only those values of variables which do some causal work to get themselves into that semantic position: to be is, therefore, to be the bearer of causal powers.

Conclusion

In this thesis I have given an extended argument against the Davidsonian method of truth in metaphysics and in particular its ontological applications. Thus, in *Part I: Truth and Truth Bearers*, I argued that the metaphysically primary truth bearers are not abstract propositions, but rather concrete representations; and I showed how to give sense to a truth predicate applying to (token) natural language sentences. Then, in *Part II: Truth and Meaning*, I investigated the question whether predicates and sentences should be regarded as having entities as meanings: I concluded that they must, but that nevertheless, the postulation of these semantic values does not metaphysically account for the meaningfulness of these expressions. Finally, in *Part III: Truth and Existence* I showed that truth is sometimes, but not always, metaphysically explained in terms of reference and satisfaction; and where it is not so explained, I argued that the Quinean condition of ontological commitment fails. Thus, to be the value of a variable, I claimed, is not yet to be. I suggested instead that to be is to be the bearer of causal powers.

It will, I think, be worth reminding the reader of the contents of this thesis in somewhat more detail before going on to compare the views I have put forward here with those of another contemporary philosopher. In chapter 1 I showed that Frege's arguments in favour of propositions as metaphysically primary truth bearers fail: a crucial part of my response to the Fregean considerations involved the recognition that truth might be a kind of relational feature of its bearers. I then gave a positive argument in favour of the alternative view that, metaphysically speaking, truth is primarily a feature of concrete representations. I showed that propositions, if they exist, are metaphysically dependent upon sentences; hence, I argued, their possession of properties (such as truth) cannot be metaphysically antecedent to the possession of analogous features by such concreta. In chapter 2, I showed that, given a truth

operator, together with quantification into sentence position, we can introduce a truth *predicate* applying to sentence tokens; and I argued that this kind of higher-order quantification does not commit us to the existence of propositions, construed as abstract objects. The principal ground for thinking this was that if propositions were objects, there would be too many objects: thus, despite the intelligibility of quantification into sentence position, inconsistency ensues from the hypothesis that there are such objects as propositions.

Of course, Davidson himself thought that truth is a feature of sentences; so I did not object to the method of truth, as he employed it, on the grounds that he misidentified the metaphysical bearers of truth. But the issue is nevertheless important in the context of a discussion of the method of truth more generally; for even if we agree that, loosely speaking, propositions are the *conceptually* primary truth bearers (as I did in *Part I* when I took “true” to figure primarily as part of a sentential operator), there is no tractable criterion of ontological commitment governing propositions. Of course, we might appeal to metaphysical necessity here, and say that anyone who holds true a given proposition is committed to whatever *must* exist if it is to be true; and this criterion will no doubt be accepted by all. The problem, however, lies in applying it: different philosophers will identify divergent commitments in the same propositions (as we saw in the introduction in connection with Williamson’s bootstrapping argument). Since “true”, *qua* predicate, applies in the first instance to concrete representations with syntactic features, this particular difficulty is overcome.

Part II was devoted to a discussion of the semantics of predicates and sentences: in particular, I was concerned to determine whether the meaningfulness of such expressions can be explained by the assignment to them of semantic values. I began, in chapter 3, by presenting Davidson’s objection to the postulation of entities to explain the meaning of predicates and sentences: such semantic values, he claimed,

are either trivial or useless. I then proceeded to examine Davidson's discussion of the "problem of predication" – that is, the puzzle that arises when we attempt to explain the semantic contribution of predicates in a way that accords with the recognition that sentences are not simply grammatically privileged lists of entities. Davidson made two principal claims in this connection: first, he offered a solution to the puzzle; and second, he attributed this solution to Tarski. Accordingly, I articulated Davidson's preferred solution to the problem – the adoption of truth theoretic semantics – which involves explaining the semantic contribution of predicates in terms of their satisfaction by sequences of objects, with the satisfaction relation itself in turn being explained by appeal to the notion of truth. Finally, I concluded chapter 3 by arguing that Davidson's attribution of this solution to Tarski was mistaken; and in doing so I drew out the difference between semantics and meta-semantics.

In chapter 4, I looked in some detail at the prospects of an inference to the best explanation, on the basis of a model theoretic semantics, to the conclusion that predicates and/or sentences have semantic values. To this end I examined Etchemendy's distinction between *representational* and *interpretational* model theoretic semantics (MTS). I argued that Etchemendy's contention that *representational* MTS can be applied directly to the empirical description of the semantic features of natural languages is mistaken; accordingly, any inference to the best explanation aiming to show that natural language predicates and/or sentences have semantic values must come from *interpretational* MTS. I then looked at two (broadly interpretational) model theoretic approaches to the semantics of modal languages – those of David Kaplan and David Lewis. I showed that the semantic values they postulate are neither trivial nor useless; but I argued that neither can adequately account for the meaning of the English word "set". This fact suggests that we should be instrumentalists about such semantic theories, and allows us to recognize that we are not ontologically committed, by our holding true various sentences (which, of

necessity, contain predicates), to the existence of such abstract objects as properties, relations, and propositions.

In *Part III* I defended a kind of anti-realism about mathematics in general, and number theory in particular. In chapter 5, I defended Pluralism about truth: in some domains, I claimed, truth consists in correspondence, while in others it is no more than coherence. I explained this difference in terms of the role of reference in accounting, metaphysically, for truth; and I argued that whereas reference has a genuine meta-semantic role of this kind in explaining the truth of empirical statements, it has no such role with respect to mathematical truths. Since reference in empirical domains may plausibly be thought to be given a causal explanation, this provides *prima facie* evidence that existence requires causal efficacy.

In chapter 6 I specifically challenged Quine's criterion of ontological commitment. I began by presenting a number of arguments designed to establish the existence of numbers: I used these to distinguish my particular species of nominalism from others. I then articulated the Quinean condition of ontological commitment, and argued, by way of general considerations regarding logical form, that sentences of the form 'The *F* is *G*' are committed, by this condition, to the existence of an *F*. Accordingly, I claimed (uncontentiously) that Hume's Principle (HP) commits one, by the Quinean condition, to the existence of numbers. In the second half of the chapter I showed that, despite this, HP can be true even if numbers do not exist. In doing so I endorsed a variant of Sider's semantic thesis of Quantifier Variance, according to which quantifier domains may vary without being mere restrictions of one maximally inclusive, most natural domain. Thus, I suggested that claims of the form 'There are *F*s' are context sensitive; though I maintained that, nevertheless, ontological disputes are not merely verbal. Lastly, I gave a genealogy of the concept of number which revealed how HP could be recognized as true; and since this

explanation made no mention of numbers, I claimed we should not admit the existence of such abstracta.

The problem with the method of truth, then, and with Quine's criterion of ontological commitment in particular, is that only semantic, and not meta-semantic (more specifically, causal) facts are appealed to in determining the truth makers of natural language sentences.

To the best of my knowledge, the only other contemporary philosopher to challenge Quine's criterion of ontological commitment is Jody Azzouni. Azzouni's (2004) book *Deflating Existential Consequence* contains a nice discussion of the ontological neutrality of semantics, showing how quantification construed as semantically objectual (rather than substitutional) need not be regarded as ontologically committing – unless the meta-language terms and quantifiers are themselves construed in this way, i.e. realistically (Azzouni, 2004: 53-62).²⁰⁸ I agree wholeheartedly, and have made similar points here in *Part III*, first by characterizing the realism/anti-realism debate in "Truth and Reference" as concerning the role of genuine (meta-semantic) reference in explaining truth,²⁰⁹ and then by showing in "Hume's Principle" that Quantifier Variance allows the semantic appearance of entities without those entities in fact existing.

In chapter 4 of his book, Azzouni articulates a criterion of existence in the hope of thereby generating a criterion for ontological commitment.²¹⁰ The candidate criteria he gives are: (1) being observable; (2) being causally efficacious; (3) being in

²⁰⁸ Azzouni gives the following, rather inelegant, "slogan: *One can't read ontological commitments from semantic conditions unless one has already smuggled into those semantic conditions the ontology one would like to read off.*" (Azzouni, 2004: 55) Despite agreeing with the content, I can't see this one catching on. Anyway, as Azzouni describes things, this recognition allows us to overcome one defence of Quine's criterion of commitment. (NB: Azzouni uses plenty of italics; so when I quote him, unless I indicate otherwise, the reader should assume they are his.)

²⁰⁹ Indeed, my distinction between reference and L-reference is similar to that (Azzouni, 2004: 61-62) between reference and reference*.

²¹⁰ As it turns out, Azzouni's approach does not provide us with a means of assessing the ontological commitments of others (Azzouni, 2004: 114).

space and time; and finally, (4) being ontologically independent of psychological and linguistic processes. The position Azzouni finally adopts is that ontological independence is the criterion we in fact employ in gauging existence. (The other criteria, Azzouni claims, are all epistemically motivated, and hence won't serve as the basis for a metaphysical distinction.) It follows that any Quinean commitment we undertake is only a genuine commitment, on Azzouni's view, if it is part of our theory, or body of belief, that the apparent entities are ontologically independent.

Azzouni goes on to argue that ontological independence (of psychological and linguistic processes) goes hand in hand with causal efficacy: so like me he thinks that to be is to be the bearer of causal powers. His argument runs approximately as follows. The Quinean commitments of our theories (or bodies of belief) that we should take seriously (from an ontological point of view) are those we take to be ontologically independent. But only those "posit" to which we require "thick" epistemic access are taken to be ontologically independent. Moreover, thick access is "grounded" in causal relations between objects and subjects. Hence, a theoretical posit is ontologically independent if, and only if, it is causally empowered.

One might be tempted to complain that epistemology should not constrain metaphysics – for this, surely, we learnt from the failure of verificationism! And so it may seem that Azzouni's argument is fundamentally flawed. However, if one embeds one's epistemology within one's metaphysics – as Azzouni does when he appeals to epistemological grounding – then the objector's suggestion is, in effect, that metaphysics should not constrain metaphysics; and this, clearly, is a mistake. So I do not object, on these grounds, to Azzouni's strategy for linking ontological independence to causal efficacy.

Nevertheless, Azzouni's attempt to show that existence requires causal powers is quite different from the one I have made here. Moreover, as Azzouni himself points

out, it is difficult to provide a characterization of ontological independence, a notion which lies at the very centre of his argument. “The problem,” as he sees it,

is that [the notion of dependence on psychological or linguistic processes] seems to include too much. Mind, mental states, language, speech acts, works of art, and artefacts are all items we’re prone to think *do* exist but that also seem to be mind dependent or ontologically dependent on us in some sense....

One suggestion... that seems not to have been previously considered in the literature, is to presuppose *something like* a causal theory of reference as already in place. We can consider a term to, at best, pick out something ontologically dependent on us in the desired sense if it fails to denote anything *causally speaking*. Ontological dependence (in *this* specific sense) then emerges as an extremely natural necessary condition for what exists because we can say, as we’d like to, that such terms fail to denote anything *at all*. We can then beef it up to a necessary and sufficient condition by generalizing to any possible language (not just ours). (Azzouni, 2004: 92-93, fn 22)

Well! The approach Azzouni describes here is, in effect, the one I have pursued in this thesis.

We might say, then, that both Azzouni and I have drawn a distinction between *real* and *nominal* existence – though not in so many words.²¹¹ That is, we may think of Quine’s condition of ontological commitment as revealing apparent existential commitment; commitment, that is, to existence at least in name – or perhaps better, in pronoun. On the other hand, the criterion of real existence is that of causal efficacy. I should say, however, that in claiming that the Quinean condition reveals nominal existential commitments, I do not intend to be disagreeing with Azzouni’s claim that “what falls outside the range of the [real] existence predicate exists *in no sense at all*” (Azzouni, 2004: 56). As Azzouni stresses, if *F*s don’t exist, then they can’t have any properties whatsoever – and this, I would add, includes

²¹¹ I thought that I was following Ockham in this respect. It turns out, however, that he, like Locke after him, distinguished real from nominal *essence*. Google reveals, nevertheless, that “real existence” and “nominal existence” are expressions already in use – they are not, as it were, neologisms.

nominal existence.²¹² This claim is consistent with what I said previously, since nominal existence is not a kind of existence, or indeed, any kind of property at all; it is simply the linguistic appearance of existence.

There is one last point where I would like to draw a comparison between Azzouni's views and my own. Azzouni thinks criteria of existence other than the one he articulates (ontological independence) are *coherent*; it's just that they aren't *ours* – that is, they aren't the one we in fact employ. In this sense he is engaging in ordinary language philosophy, drawing metaphysical conclusions on the basis of an examination of how we in fact use words. I think this method is legitimate; indeed, I have been tempted to describe my own approach to ontology as a branch of what Peter Strawson (1990) called “descriptive metaphysics”. Considered from this perspective of “descriptive ontology”, we may view Quine as having mistaken a necessary condition on existence – namely being the semantic value of a first-order (variable) term – as being both necessary and sufficient. He recognized the conceptual point that only the semantic values of terms, and not those (if any) of predicates or sentences, *exist*. (This point, I hope, came out in the course of chapters 2, 3, and 4.) On the other hand, the fact that we have a simple truth in which a first-order term figures does not (I argued in chapters 5 and 6) entail that the semantic referent (i.e. the L-referent) of that term exists.

I have, in these concluding remarks, been stressing the importance of causation in any account of existence – a constraint I derived from meta-semantic considerations. Let me end my thesis, however, with a plea for the recognition of the

²¹² Azzouni himself includes *non-existence* as being something that what does not exist can't have. He says, “E [Azzouni's existence predicate] ranges over a “subclass of what's in the domain.” Should this be described as the claim that “existence is a special property that not everything shares”? No – this *substantial* way of putting the position is misleading because (in this context) it makes it sound like there *are* things (which subsist?) but that don't exist; further these (subsisting?) things have properties – e.g., they don't exist.” (Azzouni, 2004: 52, fn 6) It is clear that Azzouni rejects this way of thinking and talking about the matter.

importance to philosophy of diachronic linguistic analysis. Since De Saussure and Frege we are accustomed to thinking of languages as synchronic systems of signs; and, of course, we should not abandon the insights this perspective affords us. Nevertheless, we saw how philosophically effective it was to give a meta-semantic genealogy of the meaningfulness and truth of Hume's Principle. I suspect that a similar approach would prove equally fruitful in other areas of metaphysical explanation; for example, in accounting for the existence of supervenience relations more generally.

Quine opened *Word and Object* with a quote from James Grier Miller: "Ontology recapitulates philology" (Quine, 1960: vii) he said. Well, perhaps. But traditionally at least, philology has been regarded as a historical discipline.

Appendix: Language Stages

I believe that we must look to the diachronic properties of languages in use in coming to understand, and metaphysically explain, the semantic properties of sentence tokens produced by a given population. My principal ground for thinking this is the existence in natural languages of pairs of sentences with similar surface syntax, yet such that one member of the pair admits of syntactically systematic, metaphysically reductive translation,²¹³ while the other does not. Consider:

(1) Not every alleged witch is a witch.

(2) Not every alleged thing is a thing.

These share a surface syntax; yet (1) is arguably necessarily and analytically equivalent to

(1') Not everything is such that if it is alleged that it is a witch then it is a witch.

On the other hand, (2) is clearly *not* in any such sense equivalent to

(2') Not everything is such that if it is alleged that it is a thing then it is a thing.

For though (2) is true, (2') is not.

Similarly we have

(3) Eight is the number of solar planets.

(4) Eight is even.

These again have a common surface syntax: each is of the form G(eight). Yet (3) is necessarily and analytically equivalent to

(3') There are eight solar planets.

²¹³ In effect, this means there is a translation *scheme* which maps metaphysically inflationary sentences onto metaphysically deflationary sentences. Given what I argued in chapter 6, the image under the translation scheme of a given sentence is not its logical form; but I imagine it as being a syntactic representation the semantic interpretation of which reveals metaphysically necessary and sufficient conditions of truth for the original sentence.

while (4) is not equivalent to any metaphysically reductive sentence that can be arrived at by translation under a scheme that is sensitive only to the shared surface structure of (3) and (4) – at least, not if this scheme maps (3) to (3')!

I propose to give an explanation of these phenomena in terms of language stages. I begin with an informal account. Thus, to explain the semantic data involved with (1) and (2), I propose an initial language stage L1, in which the word “allege” occurs only as (part of) a sentential operator – ‘S alleges that’, or “It is alleged that”²¹⁴ – and never as an adjective. In a second stage, L2, “allege” occurs in both of these surface syntactic positions, yet in the images under translation of those surface sentences the word features only in operator position. Finally, in a third stage, L3, “alleged” occurs in both positions, that is, even after translation.

Let me try to say more precisely what is meant by “language stage”. Language stages are themselves languages in a standard sense; that is, they are sets of expressions generated by a grammar. However, we can impose a partial ordering on these stages, and use this to define “language” in a new sense: a language is the ordered pair of a collection of language stages, and a partial ordering on the members of that collection. Call a form of expression of the language L an L_n-construction just in case expressions of that form belong to L_n, but not to any L_m strictly less than L_n.

I will now sketch an account of the phenomena surrounding (3) and (4) employing this terminology. The expressions of English, I claim, can be divided into language stages, in such a way that the following claims are all true. Expressions of the form (3') occur at an “early” stage, L1: (3') is an L1-construction. There exists a second stage, L2, containing such constructions as (3) (as well as (3')), but not containing expressions like (4). Finally, there is a stage, L3, containing expressions of all of these kinds. That is, instances of (3') belong to L1; instances of (3) belong to L2; and (4) and its ilk belong to L3; and L1 < L2 < L3.

²¹⁴ Which, of course, is simply the existential generalization of ‘S alleges that’, i.e. “ $\exists x (x \text{ alleges that})$ ”.

This view is subject to empirical confirmation or disconfirmation. Two forms of possible confirming evidence spring to mind: (a) one finds that children acquire the relevant expressions in stages as adumbrated; or (b) one finds that certain peoples use languages with early stage expressions, but no late stage expression, but no peoples use languages with the converse properties.²¹⁵ Evidence of type (b) is particularly important; for, to take the example at hand, even if there is no *direct* evidence for the existence of language stage L2 (or similar such stages) from either language acquisition or cross linguistic comparison; nevertheless, if one finds L1 occurring without L3, but not conversely, then one has, I believe, *indirect* evidence for the truth of my hypothesis. Barwise and Cooper (1981) suggested that we search for *semantic* universals across natural languages; that there are languages containing L1-type stages, but not L3-type stages, but no languages displaying the reverse characteristic, is just such a proposed universal. Consequently, the existence of L3-but-not-L1 speakers would disconfirm my hypothesis.

If my hypothesis is correct, then the truth or falsity of L3 sentences could be meta-semantically explained in terms of a relation of coherence with L1 sentences, rather than correspondence with an independent reality.

²¹⁵ This, of course, requires meaning preserving mappings between sentences of distinct languages that can be independently recognized.

Bibliography

- Austin, J. L., P. F. Strawson, et al. (1950). "Symposium: Truth." Proceedings of the Aristotelian Society **Supp**(24): 111-172.
- Azzouni, J. (2004). Deflating Existential Consequence: A Case for Nominalism. Oxford, Oxford University Press.
- Bach, K. (1994). "Conversational Implicature." Mind and Language **9**: 124-161.
- Bach, K. (2000). "Quantification, Qualification and Context: A Reply to Stanley and Szabo." Mind and Language **15**(2-3): 262-283.
- Barwise, J. O. N. and R. Cooper (1981). "Generalized Quantifiers and Natural Language." Linguistics and Philosophy **4**: 159-220.
- Benacerraf, P. (1965). "What Numbers Could Not Be." Philosophical Review **74**(1): 47-73.
- Boolos, G. (1997). "Is Hume's Principle Analytic?" Language, Thought, and Logic: Essays in Honour of Michael Dummett. R. G. Heck Jr. New York, Oxford University Press: 245-261.
- Boolos, G. (1998). Logic, Logic, and Logic. London, Harvard University Press.
- Bratman, M. (1984). "Two Faces of Intention." The Philosophical Review **93**(3): 375-405.
- Cartwright, R. L. (1954). "Ontology and the Theory of Meaning." Philosophy of Science **21**(4): 316-325.
- Cartwright, R. L. (1987). Philosophical Essays. Cambridge, Mass, MIT Press.
- Davidson, D. (1967a). The Logical Form of Action Sentences. The Logic of Decision and Action. N. Rescher, University of Pittsburgh Press.
- Davidson, D. (1967b). "Truth and Meaning." Synthese **17**: 304-323.
- Davidson, D. (1968). "On Saying That." Synthese **19**: 130-146.
- Davidson, D. (1969). "True to the Facts." Journal of Philosophy **66**: 748-764.
- Davidson, D. (1973a). "In Defence of Convention T." Truth, Syntax, and Modality. H. Leblanc, North-Holland Publishing Company.
- Davidson, D. (1973b). "Radical Interpretation." Dialectica **27**: 314-328.
- Davidson, D. (1977a). "The Method of Truth in Metaphysics." Midwest Studies in Philosophy **2**: 244-254.
- Davidson, D. (1977b). "Reality Without Reference." Dialectica **31**: 247-253.
- Davidson, D. (1979). "The Inscrutability of Reference." Southwestern Journal of Philosophy: 7-19.
- Davidson, D. (1983). A Coherence Theory of Truth and Knowledge. Kant oder Hegel? Stuttgart, Klett-Cotta.
- Davidson, D. (2001a). Essays on Actions and Events, 2nd edition. Oxford, Clarendon Press.

- Davidson, D. (2001b). Inquiries Into Truth and Interpretation, 2nd edition. Oxford, Clarendon Press.
- Davidson, D. (2001c). Subjective, Intersubjective, Objective, 1st edition. Oxford, Clarendon Press.
- Davidson, D. (2005). Truth and Predication. London, Belknap Press.
- Dummett, M. A. E. (1978). Truth and Other Enigmas. Cambridge, Harvard University Press.
- Dummett, M. A. E. (1981). Frege: Philosophy of Language, 2nd edition. London, Duckworth.
- Dummett, M. A. E. (1991). Frege: Philosophy of Mathematics. London, Duckworth.
- Etchemendy, J. (1988). "Tarski on Truth and Logical Consequence." The Journal of Symbolic Logic **53**(1): 51-79.
- Etchemendy, J. (1999). On the Concept of Logical Consequence. USA, CSLI Publications.
- Evans, G. (1977). "Pronouns, Quantifiers, and Relative Clauses (I)." The Canadian Journal of Philosophy **7**(3): 467-536.
- Evans, G. (1979). "Reference and Contingency." The Monist **62**(2): 161-189.
- Evans, G. (1981). "Semantic Theory and Tacit Knowledge." Wittgenstein: To Follow a Rule. L. M. Christopher. London, Routledge and Kegan Paul: 118-137.
- Evans, G. (1985). Collected Papers. Oxford, Oxford University Press.
- Field, H. (1972). "Tarski's Theory of Truth." Journal of Philosophy **69**: 347-375.
- Field, H. (1980). Science Without Numbers: a Defence of Nominalism. Oxford, Blackwell.
- Field, H. (2001). Truth and the Absence of Fact. Oxford, Clarendon Press.
- Fine, K. (2003). "The Role of Variables." Journal of Philosophy **100**(12): 605-631.
- Frege, G. and J. L. Austin (1980). The Foundations of Arithmetic: a Logico-Mathematical Enquiry into the Concept of Number, 2nd rev. edition. Oxford, Basil Blackwell.
- Frege, G., G. Gabriel, et al. (1980). Philosophical and Mathematical Correspondence. Oxford, Basil Blackwell.
- Goodman, N. (1954). Fact, Fiction, and Forecast. London, Athlone Press.
- Goodman, N. and W. V. Quine (1947). "Steps Toward a Constructive Nominalism." Journal of Symbolic Logic **12**(4): 105-122.
- Grover, D. (1972). "Propositional Quantifiers." Journal of Philosophical Logic **1**: 111-136.
- Grover, D. (1979). "Prosentences and Propositional Quantification: A Response to Zimmerman." Philosophical Studies **35**: 289-297.
- Grover, D. (1992). A Prosentential Theory of Truth. Oxford, Princeton University Press.
- Haack, S. (1978). Philosophy of Logics. Cambridge, Cambridge University Press.
- Harman, G. (1970). "Deep Structure as Logical Form." Synthese.

- Heck, R. (1997). "Tarski, Truth, and Semantics." The Philosophical Review **106**(4): 533-554.
- Heck, R. (2003). "Frege's Theorem: an Introduction." Manuscrito **26**(2): 471-503.
- Higginbotham, J. (1999). "A Perspective on Truth and Meaning." The Philosophy of Donald Davidson. L. E. Hahn. Chicago, Open Court: 671-686.
- Hodges, W. (2005). "Model Theory." The Stanford Encyclopedia of Philosophy, Winter 2005, E. N. Zalta. URL = <http://plato.stanford.edu/archives/win2005/entries/model-theory/>.
- Kaplan, D. (1989). "Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives." Themes From Kaplan. J. Almog. New York, Oxford University Press.
- King, J. C. (1994). "Can Propositions Be Naturalistically Acceptable?" Midwest Studies in Philosophy **XIX**: 53-75.
- King, J. C. (1995). "Structured Propositions and Complex Predicates." Nous **29**(4): 516-535.
- King, J. C. (1996). "Structured Propositions and Sentence Structure." Journal of Philosophical Logic **25**(5): 495-521.
- King, J. C. (2002). "Two Sorts of Claim about 'Logical Form'." Logical Form and Language. G. Preyer and G. Peter. Oxford, Clarendon Press: 118-131.
- King, J. C. (2006). "Structured Propositions." The Stanford Encyclopedia of Philosophy, Fall 2006, E. N. Zalta. URL = <http://plato.stanford.edu/archives/fall2006/entries/propositions-structured/>.
- King, J. C. (2007). The Nature and Structure of Content. Oxford, Oxford University Press.
- Kirkham, R. L. (1992). Theories of Truth: a Critical Introduction. Cambridge, Mass, MIT Press.
- Kornblith, H. (1994). "Naturalism: Both Metaphysical and Epistemological." Midwest Studies in Philosophy **XIX**: 39-52.
- Kripke, S. (1971). "Identity and Necessity." Identity and Individuation. M. Munitz. New York, New York University Press: 135-164.
- Kripke, S. (1980). Naming and Necessity. Oxford, Blackwell.
- Lewis, D. (1968). "Counterpart Theory and Quantified Modal Logic." Journal of Philosophy **65**: 113-126.
- Lewis, D. (1970). "General Semantics." Synthese **22**: 18-67.
- Lewis, D. (1975). "Languages and Language." Minnesota Studies in the Philosophy of Science **VII**: 3-35.
- Lewis, D. (1980). "Index, Context, and Content." Philosophy and Grammar. S. Kanger and S. Ohman. Dordrecht, Reidel: 79-100.
- Lewis, D. (1983). Philosophical Papers Volume I. Oxford, Oxford University Press.
- Lewis, D. (1986). "Events." Philosophical Papers. D. Lewis. Oxford, Oxford University Press. II.

- Lewis, D. K. (1998). Papers in Philosophical Logic. Cambridge, Cambridge University Press.
- Lewis, D. K. (2001). On the Plurality of Worlds. Oxford, Blackwell.
- Lowe, E. J. (1995). "Truth." The Oxford Companion to Philosophy. T. Honderich. Oxford, Oxford University Press: 881-882.
- MacFarlane, J. (2005). "Making Sense of Relative Truth." Proceedings of the Aristotelian Society **105**: 305-323.
- Martinich, A. P. (1995). The Philosophy of Language, 3rd ed edition. Oxford, Oxford University Press.
- Melia, J. (1995). "On What There's Not." Analysis **55**(4): 223-229.
- Millican, P. (1990). "Content, Thoughts and Definite Descriptions." Proceedings of the Aristotelian Society **Supp**(64): 167-203.
- Neale, S. (1990). Descriptions. Cambridge, Mass, MIT Press.
- Neale, S. (1998). "Grammatical Form, Logical Form, and Incomplete Symbols." Definite Descriptions: A Reader. G. Ostertag. Cambridge, Mass., MIT Press: 79-122.
- Peacocke, C. (1978). "Necessity and Truth Theories." Journal of Philosophical Logic **7**: 473-500.
- Predelli, S. (2005). Contexts: Meaning, Truth, and the Use of Language. Oxford, Clarendon.
- Prior, A. (1971). Objects of Thought. Oxford, Oxford University Press.
- Putnam, H. (1985). "A Comparison of Something with Something Else." New Literary History **17**(1): 61-79.
- Quine, W. V. (1948). "On What There Is." Review of Metaphysics.
- Quine, W. V. (1960). Word and Object. Cambridge, Mass., MIT Press.
- Quine, W. V. (1961). From a Logical Point of View : 9 Logico-Philosophical Essays, 2nd edition. Cambridge, Mass., Harvard University Press.
- Quine, W. V. (1970). Philosophy of Logic. Englewood Cliffs, NJ, Prentice-Hall.
- Ramsey, F. P. (1927). "Facts and Propositions." Proceedings of the Aristotelian Society **Supp**(7): 153-170.
- Rayo, A. (2002). "Word and Objects." Nous **36**(3): 436-464.
- Rayo, A. and S. Yablo (2001). "Nominalism Through Denominalization." Nous **35**(1): 74-92.
- Rumfitt, I. (2003). "Contingent Existents." Philosophy **78**: 461-481.
- Russell, B. (1905). "On Denoting." Mind **14**(56): 479-493.
- Sainsbury, R. M. (2001). Logical Forms: an Introduction to Philosophical Logic, 2nd edition. Oxford, Basil Blackwell.
- Sainsbury, R. M. (2005). Reference Without Referents. Oxford, Oxford University Press.
- Salmon, N. U. (1982). Reference and Essence. Oxford, Blackwell.

- Salmon, N. U. and S. Soames (1988). Propositions and Attitudes. Oxford, Oxford University Press.
- Shapiro, S. (1991). Foundations Without Foundationalism: a Case for Second-Order Logic. Oxford, Clarendon Press.
- Sider, T. (2007). "NeoFregeanism and Quantifier Variance." Aristotelian Society Supp(81): 201-232.
- Simons, P. (1997). "Higher-Order Quantification and Ontological Commitment." Dialectica 51(4): 255-271.
- Soames, S. (1984). "What is a Theory of Truth?" The Journal of Philosophy 81(8): 411-429.
- Soames, S. (1987). "Direct Reference, Propositional Attitudes, and Semantic Content." Philosophical Topics.
- Stalnaker, R. (1999). Context and Content: Essays on Intentionality in Speech and Thought. Oxford, Oxford University Press.
- Stalnaker, R. (2003). Ways a World Might Be: Metaphysical and Anti-Metaphysical Essays. Oxford, Clarendon.
- Stanley, J. (2002). "Nominal Restriction." Logical Form and Language, Preyer, Gerhard. Oxford, Clarendon Press: 365-388.
- Stanley, J. and Z. G. Szabo (2000). "On Quantifier Domain Restriction." Mind and Language 15(2-3): 219-261.
- Steward, H. (1997). The Ontology of Mind: Events, Processes, and States. Oxford, Clarendon Press.
- Strawson, P. F. (1950). "On Referring." Mind 59: 320-344.
- Strawson, P. F. (1990). Individuals: an Essay in Descriptive Metaphysics. London, Routledge.
- Tarski, A. (1956). Logic, Semantics, Metamathematics. Oxford, Clarendon Press.
- Williams, B. A. O. (2002). Truth & Truthfulness: an Essay in Genealogy. Oxford, Princeton University Press.
- Williamson, T. (2002). "Necessary Existents." Logic, Thought, and Language. A. O'Hear. Cambridge, Cambridge University Press: 233-251.
- Williamson, T. (2003). "Everything." Philosophical Perspectives 17(1): 415-465.
- Wittgenstein, L., D. F. Pears, et al. (1974). Tractatus Logico-Philosophicus, Pbk. edition. London, Routledge & Kegan Paul.
- Wright, C. (1983). Frege's Conception of Numbers as Objects. Aberdeen, Aberdeen University Press.
- Wright, C. (1992). Truth and Objectivity. London, Harvard University Press.