Internalist Deflationism:
on the limits of ontological investigation

by

Christopher A. Vogel

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:
Paul M. Pietroski, Professor | Chair
Erin Eaker, Assistant Professor
Mathias Frisch, Professor
Steven Gross, Associate Professor | Johns Hopkins University
Alexander Williams, Associate Professor
Valentine Hacquard, Associate Professor | Dean's Representative

Preface

A version of the work below was submitted to the Graduate School at the University of Maryland in May 2015 as the final step toward fulfilling the requirements for a doctorate degree in philosophy. In its present condition, the work differs slightly from that submitted version. The differences are exclusively in the domain of typographical and formating errors, some of which were embarrassing enough to warrant a final edit of the manuscript. . . and the inclusion of this awkward preface.

–C.V.

July, 2015

Dedication

To my father—may he breathe a little easier now.

Acknowledgments

Whatever measure of accomplishment this dissertation signifies, that success is largely not mine. As with almost any long-term endeavor, an array of individuals contributed to the completion of this project that furtively bears only a single author's name. The people that provided these myriad means of support deserve recognition—much more recognition than I could hope to convey in this short passage.

Most generally, I am grateful for the support I've received from the Department of Philosophy and its faculty over my extended tenure as a graduate student at the University of Maryland. They've shown a great deal of empathy and patience toward both my stubborn refusal to pursue any particular specialization, and my well-being.

In particular, I'm most warmly appreciative of Louise Gilman, both for her uncharacteristic devotion to the department's graduate students, and her indefatigably sanguine spirit, without which the uncompromising dirge of graduate study would be difficult to bear.

I also owe much to my graduate student colleagues, for their unique brand of friendship—of the kind that persists despite numerous encounters with poorly formed ideas, ardently defended through little more than pronounced gesticulations. To Brendan Ritchie for sending me down the anti-externalist rabbit-hole, and helping me find a project in the darkness there. To Brock Rough, for our discussions on the nature of truth and metaphysical commitment, bookended by seasons of *The Wire*. To Dimiter Kirilov for the many twilight chats about my budding project in Skinner's halls. To Alexis Wellwood for serving as a willing resource of semantic theory, and as a sympathetic internalist ear. And to the members of PHLING—including Quinn Harr, Michael McCourt, Jeffery Green, and Allyson Ettinger—for keeping me engaged in areas outside of my own research.

Of course, my many thanks to the members of the dissertation committee for the numerous hours they each invested in this project, and in me. To Erin Eaker for her repeated insights, without which this project would have certainly ventured down a much longer and circuitous path. To Alexander Williams for the many meetings that left me with more questions than answers—questions whose resolution forged a stronger thesis. To Steven Gross for his considerable willingness to engage with the project, even at the sacrifice of a rather classic piece of literature. To Mathias Frisch for his guidance at the later stages of the project to bring the force

of the thesis to bear on contemporary metaphysics. And to Valentine Hacquard for heroically serving at the eleventh hour to make the formal defense possible.

I'd be remiss if I failed to recognize the dissertation's director, Paul Pietroski. It would be difficult to overstate his contribution to this project, and to my growth as a philosopher. As I re-read these many pages in the final preparation of the manuscript, I struggle to find a passage that has not benefited from his efforts. I'm deeply indebted to him not only intellectually, as the inspiration for this work quite obviously has roots in his, but also as a mentor who has offered more time and patience than I could ever have the right to expect.

I too owe a special thanks to Susan Dwyer, both for her support throughout my graduate tenure, but also quite directly for introducing me to the generative grammar tradition, which (thankfully) derailed a dissertation proposal I once had hopes of defending—a proposal that adopted, without much reflection, the very position that this dissertation seeks to undermine.

Finally, I must thank Rachel Dudley for tolerating so much from me during the course of this project. These years were some of my most difficult, and I don't wish to imagine how I might have fared without her companionship.

Contents

List of Figures

Introduction and Overview

> *We must think things not words, or at least we must constantly translate our words into the facts for which they stand, if we are to keep to the real and the true.*
>
> –Oliver Wendell Holmes (1899)

> *Semantics with no treatment of truth conditions is not semantics.*
>
> –David Lewis (1970)

Language can be used to speak the truth. A long-standing conception of linguistic meaning contends that a semantics for natural language expressions is to be given in terms of truth. But true claims are hard to come by. *Roses are red. The sky is blue. Water is wet.* These seemingly trivial claims fail to meet the scrutiny of veridicality. Despite this, the impulse to think of meanings as indications of truth persists in linguistics and philosophy of language.

This impulse has led philosophers to defend an *externalist* conception of meaning for the better part of the 20th century. Since the pioneering work of Frege (1879 [1997]), this research program has held fast to the idea that the meanings of linguistic expressions determine their truth-conditions. This tradition has since proliferated, proffering purported insights into the workings of natural languages. The central contention is that the meanings of expressions can be understood by their relation to the objects of the world, and the facts for which they supposedly stand. They accept the following externalist thesis:

($\mathcal{E}$) For any expression e (in some language L), the meaning of e determines e's truth-conditions.

That central contention also serves as the foundation for another prominent field in philosophy whose aim is to uncover the structure of the world. Metaphyscial investigation into the nature of reality proceeds under the guise that the claims we make, when true, reveal reality's structure. If the meanings of linguistic expressions

are exhaustively analyzed by the relations words bear to worldy objects, we can trace those relations, by way of the linguistic expressions that anchor them, to see the world for what it is. Since meanings are derived from truth, meanings can serve as useful guides for determining what exists.

But when we consign meaning over to truth, we do so at a cost. The price we pay for indulging our externalist impulse is to sacrifice the explanatory adequacy of our theory of meaning, or so I'll argue here. Inspired by the work of Chomsky, a growing body of evidence indicates that the externalist assumption about language fails to capture the way natural language speakers treat the meanings of linguistic expressions. Behind this critical evidence is a particular conception of language, as an object of scientific study, and as an aspect of the human mind. Motivated by the rapid acquisition of language by human children, Chomsky argues that humans have an innate component in their mental architecture responsible for the production of language as characterized by our linguistic knowledge. Thus, understanding the human linguistic capacity requires studying this proposed component of the human mind, the Language Faculty.

Chomsky's arguments for adopting this *internalist* conception of human linguistic competence serves as an avenue of argumentation against certain externalist conceptions of language. Many externalist accounts of linguistic meaning also treat languages as mind-external objects. Thinking of languages as publicly available, mind-external objects poses problems for explaining some central questions about linguistic knowledge. Insofar as the study of language needs to explain the rapid acquisition of natural language by human children, conceptualizing languages as external entities that young children come to know presents substantive challenges to forming coherent theories about the acquisition of natural language. Treating language as an *internal* component of the mind renders these otherwise intractable questions ripe for systematic investigation. These arguments from acquisition serve as one front of the assault on the externalist thesis that meanings determine truth-conditions.

A second source of arguments against the externalist thesis highlights the incongruity between the Realist's commitment to naturalism, and the ontological status of externalist meanings. Externalists maintain that the meanings of expressions are borne out by the relations that hold between linguistic expressions and objects in the world. More succinctly, the meanings of words determine their refer-

ential relations. If words bear a referential relation to objects in the world, then the ontological status of *words* requires some explication. Namely, in order for words to stand in meaningful relations to objects, words must exist. This second vein of internalist arguments challenges the idea that there is an ontologically viable category of *words*. In particular, internalist arguments highlight that there is no viable notion of 'word' to serve as the object of scientific investigation. Naturalist commitments suggest that there is no mind-independent, *sui generis* kind 'word', and as such word-objects are ontologically suspect. If there are no words, the externalist thesis that posits word-to-world relations is mistaken.

One central methodology deployed in the Chomskyan research program uses the judgments of competent language speakers, and these judgments underwrite the third line of internalist arguments—arguments that present substantive difficulties for the externalist program. Judgments about the acceptability of linguistic expressions are the product of a cognitive process, one involving the Language Faculty. If knowledge of a language is constituted by the workings of the Language Faculty, the judgments of competent speakers of a natural language can aid in constructing hypotheses about the cognitive system that plays a crucial role in generating those judgments. For this reason, the primary data a Chomskyan approach to language aims to explain is the distribution of speaker judgments about natural language expressions.

An adequate theory of meaning should capture aspects of language use, such as the entailment relations that competent speakers endorse (and fail to endorse), the interpretations such speakers are apt to accept (and reject or ignore), and the limits of interpretations in the face of shifting contexts. The primary criticism I'll suggest, is that an externalist theory of meaning will face significant obstacles in accounting for this data, and is thereby inadequate as a theory of natural language meaning.

To offer an illustrative example here, consider the following sentences:

(1) Cars have wheels.

(2) Jim owns a car.

(3) Jim's car has wheels.

Competent English speakers will endorse the inference from (1) to (3). That is, given that (1) is true, and given that Jim owns a car, competent English speakers will infer

(or accept as a good inference) that (3) is also true. The externalist explanation of these facts pertains to truth and its relationship to deduction. The inferences of competent English speakers here are modeled as logical entailment. What an English speaker understands in grasping the meanings of (1)–(3) are the truth-conditions of these expressions. Thus the inference patterns exhibited by English speakers are captured by entailment, insofar as (3) is entailed by the concomitant truth of the two preceding expressions. Because the truth-conditional contribution of "has/have wheels" in (1)–(3), the proposal is that, as a matter of logic, so long as the referent of 'Jim's car' in (2) is one of the 'Cars' referenced in (1), the human inference from (1) to (3) is licensed by the (purported) fact that (1) and (2) collectively entail (3).

But if this explanation is on the right track, then English speakers should demonstrate a similar pattern of inference for the following expressions:

(4) Unicycles have wheels.

(5) Jim owns a unicycle.

(6) Jim's unicycle has wheels.

However, English speakers will not infer (6) from (4) and (5). All the same relations that serve to ground the explanation for the inference patterns of speakers in (1)–(3), hold in (4)–(6). The only difference between this pair of triplets is the use of 'Unicycle(s)' (as opposed to 'Car(s)'). But in (4)–(6), analogous referential relations hold: so long as the referent of 'Jim's unicycle' is one of the 'Unicycles', the same entailment account that licenses the inference in (1)–(3), should equally apply in (4)–(6). That it does not is troubling for the externalist. If meanings determine truth-conditions, then the appeal to entailment that explained the inferences regarding (1)–(3) should explain the inferences regarding (4)–(6). That speakers do not infer (6) from (4) and (5) requires an explanation, and an externalist semantics will have little to appeal to in accounting for the varying way in which English speakers treat these two groups of sentences.

This example indicates the primary property that an externalist semantics is ill-suited to explain. The predicate 'has/have wheels' exhibits *lexical flexibility*, capable of indicating varied, but tightly related meanings within a single context. In Chapter 3 I'll detail other examples of lexical flexibility as a means of undermining the externalist thesis. In essence, natural languages exhibit properties that externalist theories of meaning cannot adequately explain. If the *internalist* critique is

on the right track, this undermines the foundations of metaphysical inquiry insofar as such inquiry is committed to ($\mathcal{E}$).

The thesis I defend here is that the predominant ontological methodology adopted by metaphysicians is grounded in this troubled externalist assumption of linguistic meanings. In particular, the *Quinean Realist* methodology that relies on the assumption that linguistic meaning must be couched in terms of truth, will fail to yield decisive verdicts about the structure of reality. This methodology, taking cues from Quine (1948), holds that the objects of the world serve as the meanings of linguistic expressions. Because of this, the referents of linguistic expressions, and the truth-conditional meanings that they satisfy, can be discovered by investigating linguistic meanings. Since natural language expressions have meanings that determine how the world must be if those expressions must be true, we can investigate reality by deciphering the requirements our true expressions place on the world's structure. But, if natural languages are poorly characterized by theories that tie their meanings to truth-conditions, this methodology is without a firm foundation.

Chomsky articulates this contention:

> [R]eferentially dependent elements, even the most narrowly constrained, observe some distinctions but ignore others, in ways that vary for different types of words in curious ways... What we discover is surprisingly intricate; and, not surprisingly, known in advance of any evidence, hence shared among languages. There is no a priori reason to expect that human language will have such properties; Martian could be different. The symbolic systems of science and mathematics surely are. (Chomsky 2000, p. 16)

An internalist semantics for natural language poses a challenge to doing ontology for reasons alluded to in this passage from Chomsky (2000), which I've briefly canvassed above. Even the most straightforwardly referential expressions in natural language exhibit unwieldy referential behavior, a flexibility that defies generalization. Insofar as Chomsky is correct, the ontologist's task of investigating the meaning of natural language expressions in search of their referents, seems troubled. The assumption that the meanings of natural language sentences are related in any determinate way to (the) truth (about the objects of the external world) seems unsupported and cannot merely be assumed by the ontologist.

Also suggested in this passage, however, is a response available to the Realist. If natural languages are not amenable to the Realist's aims because they do not have external meanings, the Realist can turn to languages that do. Chomsky indicates

here that the invented languages of the sciences "surely" have meanings with the referential character needed for ontological investigation. However, I'll argue that we have reason to doubt this assumption as well. The same properties exhibited by natural languages that render them infelicitous for metaphysical investigation can be found in *scientific* languages. If the invented languages used to express our best scientific theories resist externalist treatment in the way that natural languages do, then the Realist hope of investigating ontological questions by using them is frustrated.

In Chapter 1 I outline the historical roots of the externalist program. The major traditions in philosophy of language and linguistics that inform contemporary work in semantics are committed to a particular claim about linguistic meaning. In particular I show that these views about linguistic meaning, whether propositional or truth-conditional are committed to the truth of ($\mathcal{E}$). The problems that motivated Frege (1879 [1997], 1891 [1997], 1892 [1997], 1918 [1997]), Russell (1903, 1912), and Davidson (1967c,b) set the agenda for contemporary work in philosophy of language and linguistics, in the hope of defending the externalist thesis ($\mathcal{E}$).

In Chapter 2 I detail the Quinean Realist position. I argue that a metaphysician inspired by Quine (1948, 1960) is committed to four claims about the nature of meaning and truth. In particular, I show that Quine's method of *regimentation* is insufficient for the Realist's aim of investigating ontology by using linguistic expressions. Because the process of regimenting a language is speaker-relative and purpose-driven, the Realist must bolster Quine's regimenting strategy, and is thereby committed to the externalist thesis ($\mathcal{E}$) about linguistic meaning for the privileged language of ontology (or $\mathcal{L}_{\mathcal{O}}$) she hopes to use in ontological investigation.

In Chapter 3 I show, following the work of Chomsky (1986, 2000), Pietroski (2003, 2005, 2006, 2010), and others that natural languages are poor candidates for $\mathcal{L}_{\mathcal{O}}$. The Realist's naturalistic commitments compel her to adopt a Chomskyan view of natural language, as an aspect of the human mind. This view precludes certain externalist conceptions of the ontological nature of language, and counts in disfavor of accepting ($\mathcal{E}$). I'll also illustrate that natural languages exhibit a kind of *lexical flexibility* that is particularly resistant to externalist characterization, and unlike other contextual phenomena, cannot be captured by appeal to the traditional apparatus of truth-conditional semantics.

In Chapter 4 I suggest that the languages used to express our best scientific

theories also exhibit lexical flexibility. By appealing to examples in evolutionary biology (in particular, uses of the terms 'gene' and 'species'), I argue that the Realist's retreat to the invented languages of the sciences may prove no more fruitful for ontological investigation than appeal to natural languages.

In Chapter 5 I conclude by applying these findings to the prevalent use of Conceptual Analysis as a method for resolving ontological disputes. In particular, I argue that the Realist hoping to use Conceptual Analysis, as a means of enlightening ontological debates, is presented with a dilemma. They are either saddled with the burden of defending the applicability of ($\mathcal{E}$) to natural language meanings (in light of the worries presented in Chapter 3), or they must give up the use of case-driven speaker judgments that constitutes Conceptual Analysis. The implications of this dilemma for metaphysical investigation are profound, given the prevalence of Conceptual Analysis as a means for resolving metaphysical disputes in various philosophical domains. If the Realist is compelled to reject ($\mathcal{E}$) (at least with regard to natural languages) she must also vastly revise her methods of investigating ontological questions. The consequence is that many metaphysical questions must either be abandoned, or the landscape of debate must be radically re-constructed. Rather than consigning meaning over to truth, the Realist should resign herself to the limits of ontological investigation.

Chapter 1

Truth-Conditions and Semantics

A theory of meaning for a natural language should provide a means for pairing sentences of a natural language with their meanings. Various semantic theories accomplish this goal in varying ways, but externalist semantic theories insist that such a paring requires relating sentences to objects in a (worldly) domain. There are two traditional views about the nature of meanings: *propositional* semantic theories, and *truth-conditional* semantic theories. The task of semantics, on the propositional view is exhausted by pairing propositions with the natural language sentences that express them. A second approach holds that the meanings of expressions are not entities, but are truth-conditions—namely, the conditions under which sentences are satisfied by objects in the world (or domain). This chapter will survey the various approaches of these two programs, highlighting the essential role that truth-conditions play in both strategies. Given the tight relationship between the notion of truth and truth-conditions, the goal of this chapter is to show that the most plausible accounts stemming from either of these traditions requires that truth plays a prominent role in characterizing meaning. Importantly, if the notion of truth that lies at the heart of these theories is a Realist conception of truth, both approaches are committed to the following claim:

($\mathcal{E}$) For any expression e (in some language L), the meaning of e determines e's truth-conditions.

The purpose of this chapter is to show that theories of meanings beholden to traditional commitments about the nature of truth and language are committed to ($\mathcal{E}$).[1] I'll begin this largely exegetical chapter with the historical precedent for these positions at the turn of the 20th century. Propositional semantic theories find their

[1]This claim will require some slight modification, as discussed in Chapter 3. Specifically, a typical brand of externalism insists that the truth-conditions of some expressions are not determined by the meaning of the expression alone, but require *semantic* context. The paradigm case of this kind of exception are indexicals and demonstratives like 'I', 'here', and 'that'. Expressions containing these terms will fail to determine the truth-conditions of those expressions without the information contained in the semantic context of the expression (e.g. the speaker or place). The relevance of these sorts of expressions, and the role of semantic context regarding the thesis argued in this work are discussed in Ch. 3, §3.3.1.

inspiration in the work of Frege and Russell. Briefly, Russellian propositions have objects and concepts (or alternatively, properties and relations) as their constituents, while Fregean propositions take 'Sinnen' to be the constituents of propositions (or "Thoughts"). I will briefly trace the historical roots here as a means of framing the motivations for contemporary accounts of propositions. I will then detail the relationship between these accounts of propositions and truth-conditional accounts. What these various positions share is a commitment to the relationship between meaning and truth. In Chapter 3 I argue that such truth-conditional approaches are troubled, levying a serious problem for a semantics that views the meanings of natural language expressions as (or deeply indebted to) their truth-conditions.

1.1 Propositional Theories of Meaning

Propositional theories of meaning identify the meaning of a natural language sentence with a proposition. So construed these approaches hold that meanings are *entities* of a particular kind. However, what is often overlooked in discussions of such theories is that they offer (or must adopt) a particular account of truth-conditions. That is, if sentences can be true, these propositional theories of meaning are committed to an account of what conditions must hold for sentences to be true. In this section I shed light on this gap as a means of illustrating the motivations for a truth-conditional approach to meaning (outlined in section §1.2). I begin by indicating the motivations for the propositional view to see why positing these entities might do some useful philosophical work. I'll then outline the views offered by Russell and Frege as a means of indicating the relationship between contemporary propositional accounts and these historical views. The problems that emerge from this exegesis suggest a view of propositions as possible worlds, as discussed in §1.1.4.

1.1.1 Virtues of Propositions

Many philosophers of language defend a propositional semantics for natural languages (Kaplan, 1977 [1989]; King, 2007; Lewis, 1970, 1986; Salmon, 1986; Stalnaker, 1976; Soames, 1984). The motivation for these views stems from a series of pre-theoretical commitments regarding the role meanings play in the use of natural languages, and in the relationship between meaning and human thought. The basic

argument for propositional views is one of parsimony, insofar as positing propositional meanings explains, or makes clear, a series of pre-theoretical beliefs about the nature of language, meaning, and thought.[2] There are three pre-theoretical commitments that a propositional semantics purports to explain: 1) that multiple distinct sentences can convey the very same informational content, 2) that different cognizers can have mental states, (like beliefs) with the same content, 3) that the bearers of truth-values can also bear modal properties (cf. King, 2007, pp. 2–5). Beyond these uses of propositions, 4) quantifying over propositions in a semantics for natural language yields a simple semantics for otherwise complex phenomenon, like modal expressions. I will review each of these benefits in turn.

Many natural language sentences seem to communicate the same meaning. For example, 'I am here' and 'I am home' (uttered in some contexts) can have the very same meaning, despite being composed of distinct constituent expressions. Further, sentences in different natural languages can have the same meaning, despite having *no* overlap in constituents. 'The cat is on the mat' and 'Le chat est sur le tapis' seem to have the same content, insofar as they seem to express the same thought (given the appropriate context) about a cat situated on a textile. This intuition is well accounted for by a propositional semantics, insofar a these sentences (on such a view) express the same proposition. These expressions have the very same meaning because they both relate to the same entity, the particular proposition they respectively express. Likewise, for any multiple of distinct synonymous expressions in any natural language, a propositional theory can explain how these distinct expressions/sentences/utterances have identical meanings. Such sentences *express* the very same proposition(nal meaning).

However, notice the importance of the *expression* relation that must hold between a sentence and its propositional meaning. A sentence has a particular meaning insofar as it *expresses* a proposition. While a propositional semantics may do some useful work in noting the relationship between the proposition *expressed* by synonymous, though distinct, expressions in a natural language, if this theory is to adequately explain the relevant pre-theoretical intuition, the notion of 'expression' must be fairly well developed. When a competent speaker of a language under-

[2]We'll see that many of these commitments are themselves problematic, and while propositional meanings (supposing an acceptable account of propositions is on offer) might well capture these intuitions, a commitment to a naturalist methodology undermines the epistemic credentials of these intuitions (cf. Ch. 3, §3.2.1).

stands the meaning of a sentence they must, on a propositional account, be able to connect the expressed proposition with the sentence that expresses it.[3] That is, based (at least in part) on the properties attributable to the sentence (or utterance) a natural languages speaker understands, a propositional semantics must offer (the possibility of) a plausible explanation for how such a speaker arrives at the proposition expressed by the sentence (or utterance). As such, the expression relation must be of a kind that comports with how competent speakers understand sentences or utterances. Similarly, the ontological nature of a proposition must be such that a sentence can express a proposition in a way that comports with how competent speakers understand sentences. I raise this point here merely to illustrate part of the propositional semanticist's task. A semanticist fond of propositions because they can do interesting philosophical work must show that, whatever propositions are, they can actually do their required duties. I'll address this worry in Ch. 3.

We've seen that propositions can help explain our pre-theoretic commitment to the synonymy of distinct natural language expressions. Much like expressions in a natural language, distinct beliefs can (seemingly) have identical contents. If both Sally and Sarah believe that there is no largest prime number, intuitively their beliefs have the same content. That is, we want to say they have the *same* belief. But of course, Sarah and Sally have different minds (or brains), so their beliefs (in this sense) are obviously distinct. A propositional theory of meaning can explain how these distinct beliefs have the same content. What Sarah and Sally share is that they bear the same relation, the *belief* relation, to the very same entity, the propositional meaning THERE-IS-NO-LARGEST-PRIME-NUMBER.[4] Insofar as they bear the same

[3]Some deny this commitment (cf. Lewis, 1975). But this denial seems unfounded. One can say that the business of doing semantics is encompassed by the project of pairing sentences with propositions. And one can stipulate that these meanings are related to sentences, independent of the way in which humans grasp the meanings of natural language expressions. But, apart from the pre-theoretical commitment to thinking that the only possible account of meaning is one that pairs expressions with entities, this project seems somewhat strange. Whatever sentential meanings are, the best data we have in deciphering what theory best describes such meanings are those forthcoming from natural language speakers. That is, whatever meanings are, speakers grasp them. The semanticist then owes us an account of meaning that (at the very least) suggests a way in which meanings can be grasped by natural language speakers. To say that meanings have to be understood in terms of entities, irrespective of how humans are psychologically related to them is to deny the very data a theory of meaning ought to account for. This objection is explored in Ch. 3, §3.2.1.

[4]For the sake of clarification I adopt the following notational convention in this chapter regarding words, meanings, and referents: if I intend to talk about a word, I will use single quotes (e.g. 'dog', 'Frege', or 'Socrates is mortal'); if I intend to talk about the meaning of an expression I will use ⟦semantic brackets⟧ (e.g. ⟦dog⟧, ⟦Frege⟧, or ⟦Socrates is mortal⟧); if I intend to talk about the

relation to the same object, the content of their doxastic (belief) state is the same. (Notice, again that this purported benefit of adopting a propositional theory of meaning puts constraints on what can be a proposition. Whatever a proposition is ontologically speaking, it must be the sort of entity that can be believed. Thus, the notion of proposition on offer from such a theory must comport with the way in which minds have beliefs.)

Meanings also seem to have modal features that propositions can be posited to explain. For example, the following claims seem to be true: everything is self-identical, and all triangles are trilaterals. These two claims are true, but more robustly they are *necessarily* true. Accounting for this pre-theoretic commitment requires that the bearers of truth (and falsity) can also be the bearers of modal properties. Propositions (if understood as relating possible worlds) can bear these modal properties in a way sentences or utterances cannot. 'Everything is self identical' construed as an utterance, or sentence, cannot be necessarily true insofar as the English language (if there is such a thing) exists contingently. However, the meaning of 'Everything is self-identical', i.e. the proposition EVERYTHING-IS-SELF-IDENTICAL, is not tied to the contingent existence of English. Whatever a proposition is, the ontological status of that proposition is independent of anything that expresses it, like a natural language sentence or utterance. A proposition can be true, even necessarily true, even if no cognizer ever believes it, or no natural language speaker utters a sentence expressing it.

1.1.2 Russell

> Words all have meaning, in the simple sense that they are symbols which stand for something other than themselves. But a proposition [...] does not itself contain words: it contains the entities indicated by words. (Russell, 1903, Ch. V, §51)

Russell argues that sentences express propositions, and that sentential meanings are these propositions. Further, he contendss that the constituent terms of such propositional meanings are the mind-independent entities that natural language expressions denote.[5] In particular, Russell maintains that the constituents of

referent of an expression as an object I will use SMALL CAPITALS (e.g. DOG, FREGE, or SOCRATES-IS-MORTAL). This explicit usage will only apply to this chapter.

[5]Some philosophers mark a distinction between 'denotation' and 'reference' (cf. Goodman, 1981). I do not, and use the terms interchangeably throughout this work.

propositions, the building blocks of sentential meanings, are *objects* and *concepts* (or relations[6]). That is, Russell proposes that a theory of meaning for natural language is nothing more than a theory of reference (or denotation). Consider the sentence

(7) Gottlob differs from Bertrand.

The meaning of (7) on Russell's account is the proposition expressed by (7), a proposition composed of the objects and relations denoted by the constituent expressions of (7). Russell's idea here is an intuitive one. Names like 'Gottlob' and 'Bertrand' are used in natural language expressions to identify objects in the world. In this case, the relevant players are the individuals FREGE and RUSSELL. The aspects of meaning that a word contributes to the propositional meaning of a sentence then is nothing more than the object the words denotes. In this case, the names 'Gottlob' and 'Bertrand' contribute two (dead) philosophers.

The natural progression of this thought yields a semantic theory whereby the meanings of whole sentences are their referents. That is, if the meanings of words (like names) are nothing more than the objects to which they refer, then plausibly the meanings of sentences are composed of the objects to which the constituents of sentences refer. Unlike names however, the entities expressed by sentences are not all that easy to identify. Names like 'Frege' have referents that are easily accessible. There is an intuitive sense in which I can find the the object that 'Frege' denotes. However, the type of object that is a sentential referent is much less intuitively transparent. While I can point to the referent of 'Chomsky', I cannot (obviously) point to the referent of 'Parisians are ruder than New Yorkers'. But Russell's proposal treats these tasks as one in the same. The proposition expressed by (7) (the referent of (7)) is in the world in the way FREGE is. The proposition expressed by (7) has as constituents, the worldly things denoted/referenced by the words in (7). Thus the relevant proposition is exclusively composed of both FREGE and RUSSELL, in addition to whatever 'differs from' denotes. But if propositions are mind-independent entities, as the Russellian view demands, then the denotation of 'differs from' must be mind-independent as well. That is, if propositional meanings are composed of the denotations of the constituent expressions of sentences, for such composites to be wholly mind-independent, the relations denoted by expressions like 'differs from'

[6]Russell (1903) uses *concepts* as the denotations of verbs phrases, while Russell (1912) indicates that *relations* are the denotations of verb phrases. I'll ignore this difference in what follows.

must also be mind-independent. Russell argues for precisely this view.

Russell holds that verb phrases (like 'differs from') denote relations. The English words 'differs from' in (7) refers to DIFFERENCE, a relation that is realized in all pairs of differing things. This relation, on Russell's view, is in the world in much the way FREGE and RUSSELL are, in that the existence of such a relation is mind-independent:

> Consider such a proposition as 'Edinburgh is north of London'. Here we have a relation between two places, and it seems plain that the relation subsists independently of our knowledge of it. When we come to know that Edinburgh is north of London, we ... apprehend a fact which was there before we knew it ... [and] this fact involves the relation 'north of', which is a universal; and it would be impossible for the whole fact to involve nothing mental if the relation 'north of', which is a constituent part of the fact, did involve anything mental. (Russell, 1912, Ch. IX)

Russell argues here that relations must be mind-independent (in the way LONDON and EDINBURGH are). Suppose that relations depended in some way on cognition, thereby being mind-dependent. If relations were dependent in this way on mental states, the *truth* of a proposition that contains such relations as a constituent would likewise depend on such mental states. On this view, a cognizer, by entertaining a relational thought, would thereby relate the terms of the proposition *via* cognition. That is, if the existence of the relation indicated by 'north of' depended on a cognizer performing some mental operation or other, in cognizing the proposition 'Edinburgh is north of London' said cognizer would thereby *relate* LONDON and EDINBURGH.

Two absurd conclusions follow from such a view. First, the view belies the intuitive notion that whether or not Edinburgh is north of London seems independent of any cognitive process. Whether Edinburgh is north of London, as a matter of *fact*, simply does not depend on the mental processes of any particular individual. That is, the truth of the proposition expressed by 'Edinburgh is north of London' should not be held hostage by the contingency that some person or other mentally entertains a thought involving LONDON, EDINBURGH and their geographic location, as the mental-dependence view suggested here (seemingly) entails. Second, if relations were mental processes, then a cognizer could make a(n) (intuitively) false proposition true by merely having a thought in which the object terms of the (false) proposition are the *relata* for the relational constituent of the proposition. That is, if relations were nothing more than a way of thinking about objects in the world,

in so thinking about those objects a cognizer would put those objects in that relation. If those objects are thereby *related* by such a cognitive process, then the proposition containing those objects and that mental relation as constituents would be true. But surely 'Washington D.C. is north of New York City' is not made true because I have a thought about WASHINGTON-D.C. and NEW-YORK-CITY whose content is the proposition expressed by 'Washington D.C. is north of New York City'. Taken together, these two conclusions would indicate counter-intuitively that the only propositions that exist (to be meanings of sentences) are those that have been cognized (by someone or other), and that all such propositions are true.

For these reasons "we must admit that the relation, like the terms it relates, is not dependent upon thought, but belongs to the independent world which thought apprehends but does not create", making propositions compositions of mind-independent relations and the objects that are their *relata* (Russell, 1912, Ch. IX). So the proposition expressed by (7) has as constituents the mind-independent entities FREGE, RUSSELL, and DIFFERENCE. However, this picture is incomplete. The proposition expressed by (7) cannot merely be some list of these objects and relations, for two related reasons. Consider Russell's sentence

(8) Edinburgh is north of London

Given the features of the Russellian view discussed thus far, (8) expresses a *proposition* whose truth depends on a *fact* pertaining to the regions of the external world denoted by 'Edinburgh', 'London', and the relation that they stand in (denoted by 'north of'). The proposition expressed in (8) then has three constituent parts, the *objects* EDINBURGH and LONDON, and the *relation* NORTH-OF. For Russell, propositions are composed of these two kinds of terms, objects and (many-placed) relations (with properties being one-place relations). The proposition expressed by (8), then, contains three terms

(8′) <EDINBURGH, LONDON, NORTH-OF>

Suppose this triplet, so ordered, is the proposition expressed by (8). Such a view of propositions makes them a kind of entity that cannot plausibly do the philosophical work propositions ostensibly do.

First, ordered lists do not seem to be the sort of entities that can be true or false. If n-tuples like (8′) are propositions, they must be the sort of entities that can

do the work propositions are meant to do. One of the jobs propositions purportedly do is underwrite the truth-values of sentences. That is, the truth or falsity of sentences (and for that matter mental states like beliefs) are supposedly derived from the truth-values of the propositions that are their meanings (or contents). But a mere list does not seem to be the sort of entity that can be true or false. Consequently, n-tuples cannot thereby be the kind of entity from which the truth or falsity of a sentence is derived. Consider another list <1, 2, 3>. This list is structurally similar to (8′). Yet to say that '<1, 2, 3> is true' seems strange, precisely because it is a mere list of numbers. If a theory of meaning holds that propositions are numbered lists, or n-tuples, like (8′), there must be some way to determine how (8′) differs from other lists like <1, 2, 3> such that (8′) can be true or false, while <1, 2, 3> cannot be.[7]

Second, there seem to be many candidate mere-list-propositions expressed by (8) other than (8′). Both of (8″), and (8‴) seem to be equally qualified to serve as the mere-list-proposition expressed by (8):

(8″) <EDINBURGH, NORTH-OF, LONDON>

(8‴) <EDINBURGH, <LONDON, NORTH-OF>>

Because (8′), (8″), and (8‴) are ordered lists, they are different entities, insofar as their constituent terms are in different orders. What the propositional theory of meaning owes us is a way to discriminate between (8′), (8″), and (8‴) (and any number of other candidate n-tuples) in identifying the propositional meaning of (8). One of the intended chores that propositions are assumed to accomplish is the individuation the meanings (and contents) of sentences (and thoughts). As such, a propositional theory of meaning must yield distinct and unique propositions for all non-synonymous natural language expressions. The "mere-list" view fails to provide such propositions.

The third worry is related to the previous one. Consider the difference in meaning between the following:

(8) Edinburgh is north of London.

[7]One might claim that <1,2,3> *is* a proposition with truth-conditions, for say the sentence '1+2=3', and thus has truth-conditions. But surely not *all* n-tuples are propositions. So for any non-propositional n-tuple, the account must provide a means for distinguishing such lists from genuine propositions. What is not clear is what such a theorist can appeal to in distinguishing the ordered lists from meanings.

(9) London is north of Edinburgh.

The Russellian notion here is that propositional meanings are composed of the objects and relations denoted by the constituents of the sentences that express them. Sentences (8) and (9) make use of the same expressions, so the constituents of the propositional meanings they express must be the same three terms denoted by those expressions: EDINBURGH, LONDON, and NORTH-OF. Of the various n-tuples (like (8′)–(8‴)) that could do duty as the propositions for (8) and (9), our theory must explain why some particular ordering is the proposition expressed by (8), yet not (9) (and vice versa). *A fortiori*, this proposition needs to be *recoverable* from the structure of the sentence that expresses the proposition in question. That is, based on the information available in the expression, a natural language speaker must be able to somehow distinguish between the various candidate lists. But to what can a philosopher sympathetic to this account appeal in illustrating why any of (8′)–(8‴) are the particular propositional meaning of (8) (or for that matter (9))? All seem like equally good candidates for the relevant propositional meaning. Yet the theory needs to explain how, when a competent speaker of a language understands (8), they recover the propositional meaning for (8), and not (say) the propositional meaning for (9). If propositions are merely ordered n-tuples of mind-independent terms, there seems to be no resources for explaining how we understand the propositional meanings of non-synonymous sentences with the same constituents (like (8) and (9)).[8]

[8]A philosopher sympathetic to this view might insist that these lists are not n-tuples, but mereological sums. On this view the proposition represented by the brackets and commas in (8′) is the same entity as that represented in (8″) and (8‴). They are all the same mereological sum of the three terms, EDINBURGH, LONDON, and NORTH-OF. That is, (8″) and (8‴) represent the spatio-temporally dispersed object of EDINBURGH–NORTH-OF–LONDON. Even ignoring the peculiarity of such a hybrid entity, this amendment to the view will fail to generate a distinct proposition for sentences with distinct meanings. If (8) and (9) have different meanings (and surely they do), a propositional semantics must provide distinct propositions to serve as the meanings for (8) and (9). But if propositions are mereological sums of objects and relations, the mereological sum of EDINBURGH–NORTH-OF–LONDON can do duty as a the meaning of either (8) or (9), but not both. Suppose such a sum is the proposition expressed by (8). With only mereological sums at our disposal, there is no entity whose constituents are expressed by (9) to serve as the meaning of (9). The three constituents of (9) are the very same as the constituents indicated in (8), EDINBURGH, LONDON, and NORTH-OF. But mereological sums are nothing more than the amalgam of their constituents. There is *only one way* to mereologically sum EDINBURGH, LONDON, and NORTH-OF. Thus, there is only one proposition (on this view) to be generated from summing these three constituents. As such, either (8) or (9) have the same meaning (the proposition EDINBURGH–NORTH-OF–LONDON) or one is meaningless (because there is no proposition to be the meaning of the sentence). Neither option seems plausible. Unless we drop the Russellain commitment that the

These various problems for the propositional account sketched thus far can be characterized as amounting to a single concern. The reason why an ordered listing of objects and relations is inept for doing the work propositions are meant to do is that lists are not *unified* in the way propositions need to be. Russell characterizes this problem in the following passage:

> Consider, for example, the proposition A differs from B. The constituents of this proposition, if we analyze it, appear to be only A, difference, B. Yet these constituents, thus placed side by side, do not reconstitute the proposition. The difference which occurs in the proposition actually relates A and B, whereas the difference after analysis is a notion which has no connection with A and B... A proposition, in fact, is essentially a unity, and when analysis has destroyed the unity, no enumeration of constituents will restore the proposition. The verb, when used as a verb, embodies the unity of the proposition, and is thus distinguishable from the verb considered as a term, though I do not know how to give a clear account of the precise nature of the distinction. (Russell, 1903, Chapter IV, §54)

Russellian propositions are not mere lists (nor are they mereological sums[9]), but unified collections of objects and relations. The proposition expressed by (8) has the terms in (8′) as constituents, but importantly, the proposition is constituted by the object terms in (8′) *related in a particular way*. That is, the proposition expressed by (8) is the related objects terms EDINBURGH and LONDON *standing in* the NORTH-OF relation. The thing in the world that is Edinburgh bears a particular relation to the thing in the world that is London. The proposition expressed by 'Edinburgh is north of London' are just those objects bearing that (north of) relation.

Such a proposition is not merely a list, but a unified whole. Compare again the propositions expressed by (8) and (9):

(8) Edinburgh is north of London.

(9) London is north of Edinburgh.

The terms that are the constituents of the proposition expressed by (8) are the same terms that constitute the proposition expressed by (9). But again, (8) and (9) differ in meaning (and truth). The proposition expressed by (8) requires that

constituents of propositions are the mind-independent entities indicated by the words of a natural language sentence, we should abandon the view whereby propositions are simply *mereological sums* of objects and relations.

[9]Russell does not entertain this solution, though one might imagine his worries would mirror those expressed above in the previous footnote.

EDINBURGH and LONDON are related in a particular way, such the EDINBURGH is *north of* LONDON. Insofar as (9) demands the inverse, the propositions expressed by (8) and (9), while containing the same constituents, are different. The proposition expressed by (9) has the objects LONDON and EDINBURGH standing in the *north of* relation, such that London is north of Edinburgh.

But, if we understand propositions as such, then the view is saddled with a different worry. The proposition expressed by (8) involves three objects in the world (if not abstractly), EDINBURGH, LONDON, and the NORTH-OF relation (as an abstract universal). The unified proposition (what "analysis has deconstructed") is the physical object that is Edinburgh (abstractly?) related to the physical object that is London in a north-of way. But this unified proposition seems just to be the *fact* that Edinburgh is north of London. After all

> When we come to know that Edinburgh is north of London, we . . . apprehend a fact which was there before we knew it . . . [and] this fact involves the relation 'north of', which is a universal. . . (Russell, 1912, Ch. IX).

Facts are true. Propositions, however must be the sorts of things that are capable of being both true and false. Russell in appealing to the mind-independent relatedness of EDINBURGH and LONDON as the unifying aspect of the proposition expressed by (8), (seemingly) *identifies* the fact with the proposition. But propositions as facts cannot do the work propositions are meant to do.

The primary difficulty for the propositional semanticist distilled in this section is the problem of unity. Propositions must be more than mere lists of objects and relations. The constituents of propositions must hang together in a particular way that groupings of objects (and relations) do not. For example, 'Caesar died' and 'The death of Caesar' (seemingly) have the same constituents, CAESAR and DEATH—yet one expresses a proposition while the other expresses an object (presumably an event). The nature of the distinction between these two expressions embodies the kind of unification a propositional semanticist must explain. The problem however is a delicate one: a propositional semanticist must characterize propositions so as to be unified entities that can bear truth-values, but must do so without melding them inextricably to a particular truth-value. We saw this problem with the Russellian account above. The kind of unification Russell suggests blurs the distinction between propositions and the facts in the world that make them true. If propositions are the meanings of sentences, and the contents of (some) mental states, they must be

capable of being false while maintaining content. Tying propositional meanings to their (actual world) truth-values in the manner suggested by Russell makes for too tight a knot.

Having seen Russell's attempt to render the notion of a proposition intelligible, we see the basic landscape for the domain of propositional meanings. The difficulty in constructing an adequate theory is embodied in the tension between competing philosophical uses for propositions. They need to be the kinds of entities that can bear truth-values, but must also be sufficiently divorced from whatever makes them true (or false). If the referents of expressions are the only constituents one can appeal to in building propositions, one of these pressures fissures a gulf between propositions and their truth-makers, while the other forces the gulf to collapse. Propositions must be unified structured entities in order to explain how competent speakers of a language can come to know the propositional meanings of natural language expressions. But whatever binds a proposition together better not simultaneously bind the proposition to truth, without generating the implausible result that no proposition is false. Insofar as we can have false beliefs (or false doxastic states generally), if propositions are the contents of those states, some propositions better be false.

Returning briefly to the purpose of this chapter, this discussion shows quite plainly why a propositional semanticist of the Rusellian sort is committed to ($\mathcal{E}$). The meaning of a sentence for the propositional semanticist is just the proposition that sentence expresses. The difficulties for such a view addressed thus far trade on the need for an account of propositions such that they can be bearers of truth and falsity. Thus, whatever view a propositional semanticist poses as a purported solution to these problems, such a view will hold that propositions bear truth-values. As such, on any propositional semantics, a sentence will be true, just in case it expresses a proposition that is true. Put another way, the propositional meaning of any expression will determine the truth-conditions of the expression—the sentence will be true *iff* the propositional meaning of the sentence is true. This, of course, reflects a commitment to ($\mathcal{E}$).

1.1.3 Frege

In a series of papers around the turn of the 20th century, Gottlobb Frege constructed a logic aimed at expressing the Dedekind-Peano axioms of arithmetic. In so doing he suggests (or has been interpreted as suggesting[10]) a way in which this logic can do work in understanding the propositional meanings of natural language expressions. The distinct character of this approach is embodied in Frege's resolution to the problem of propositional unity explored in the previous section.

Before we see Frege's semantic vision, and how it (seemingly) resolves the problem of unity, let's see why Russellian propositions (as characterized above) fail to account for the distinct meanings of co-referential expressions.

Consider the following

(10) London is the capital of England.

The sentence expresses an identity claim, of the form 'a=b'. On the Russellian view (outlined so far), the proposition expressed here involves the referents of 'London', 'is', and 'the capital of England'. These referents seem to be the city LONDON, the relation of IDENTITY, and (again) LONDON, in so far as the capital of England is the object picked out by 'London'. Compare (10) to

(11) London is London.

The meanings of (10) and (11) seem to differ in a few ways. The claim in (11) is (seemingly) both analytic, and knowable *a priori*. I need not know anything at all about the world to recognize that (11) is true. Merely consulting the nature of the meanings of the sentence's constituents is sufficient for deriving the truth of (11). Neither of these properties holds true for (10). To know that London is the capital city of England is to know a bit about geography, politics, and history, not merely the identity relation. That is, knowing that (10) is true is informative in a

[10]There is notable disagreement as to how Frege should be interpreted regarding the applicability of his logical work to natural language semantics. Often Frege indicates that his claims about the nature of meaning apply only to an ideal language, one capable of doing the work of expressing the truths of formal scientific investigation (cf. Frege, 1879 [1997], 1891 [1997]). But elsewhere, Frege's writing at least implies that his discussion of language is meant to apply directly to natural languages (cf. Frege, 1892 [1997], 1918 [1997]). My interest here concerns the logical space of views about the nature of propositions, one of which is traditionally attributed to Frege. Whether or not Frege endorses such a view may be unclear. Whatever Frege's actual positions were, I leave the task of deciphering this to the historians. The view expressed in this subsection is a plausible theory of propositional meanings for natural languages, even if it turns out not to be Frege's.

way that knowing (11) is true simply is not. But if the meaning of sentences are propositional, and propositions are Russellian in that the constituents of propositions are the referents of sentential constituents, then the proposition expressed by (11) must be identical to the proposition expressed by (10). Both share exactly the same constituent referents (LONDON and IDENTITY). But, insofar a (10) seems to be informative in a way (11) is not, they must have different meanings. And if meanings are the propositions expressed by sentences, on the Russellian account (10) and (11) express the same proposition, and thereby cannot differ in meaning.[11] Russellian propositions cannot then account for the meanings of sentences in contexts where the referents of expressions are insufficient to distinguish the meanings of non-synonymous expressions whose constituents have the same referents.[12]

Frege resolves this problem by distinguishing between two kinds of meaning. Expressions, beyond merely having referents, also have *senses*. Thus claims like (10) and (11) have different meanings, despite the co-referential nature of 'London' and 'the capital of England', because they have different senses. But distinguishing between these two kind of meaning only avoids the puzzle of identity claims if propositions contain senses. Frege's view is committed to the view that propositions take senses as their constituents.

> A statement contains (or at least purports to contain) a Thought as its Sinn; and this Thought is in general True of False; i.e. it has in general a truth-value, which must be regarded as the Bedeutung of the sentence, just as, say, the number 4 is the Bedeutung of the expression '2+2' or London the Bedeutung of the expression 'the capital of England'. (Frege, 1891 [1997], p. 139)

Crudely, expressions have two types of related semantic value for Frege, their *Sinn* and their *Bedeutung*: the thing the expression refers to (Bedeutung), and the mode in which that thing is presented (Sinn). Thus, two expressions may *refer to* (*bedeut*) the same thing, but do so in different ways, by way of their distinct *senses (Sinnen)*. 'The author of the *Begriffsschrift*' and 'The father of modern logic' both refer to FREGE, but the mode in which they present their (shared) referent is distinct. They share a Bedeutung, but not a Sinn. They have a common referent, but distinct

[11] Identity claims are not the only logical contexts in which this worry arises (cf. Salmon, 1986).

[12] Russell, of course, famously thought that most "denoting expressions" like 'the capital of England' had no (individual as a) referent, as their underlying logical structure took the form of a quantifier expression. In doing so however, Russell thereby seems to deny the referential theory of meaning, insofar as the meaning of such expressions cannot be their referents.

senses.[13]

Frege's distinction between the sense and referent of an expression is tied to his treatment of natural language expressions as functions. The basic idea ia that the sense of an expression determines the referent of the expression. This relationship between sense and reference is analogous to the relationship between a mathematical function and the value of a function, given an argument. Extending the notion of a mathematical function to natural language expressions occupies much of Frege (1918 [1997]), where he first introduces the distinction. Consider the following expressions in arithmetic

(12) 2+3

(13) 6–1

These expressions are distinct, insofar as they have no mathematical terms in common. But, quite clearly these expressions are importantly equivalent, as they have the same value, the number five. That is

(14) 2+3 = 6–1

is true. Now compare (12) and (13) with the following related mathematical *functions*

(f_{12}) ( )+3

(f_{13}) ( )–1

where '( )' stands for a position that can be occupied by any natural number such that

$f_{12}(2) = 2+3$

The function (f_{12}) differs from (12) in that (12) has a determinate value, while (f_{12}) does not. The expression in (f_{12}) is incomplete in a way that (12) is not. This incompleteness, indicated by '( )', marks the distinction between a *saturated* function and an *unsaturated* one. What the unsaturated function (f_{12}) requires in

[13]There is considerable debate over how 'Sinn' and 'Bedeutung' are to be best interpreted. For ease of exposition, even if somewhat misleading, I have settled in 'sense' and 'reference' (and their related forms) as the English translations of the German words 'Sinn' and 'Bedeutung' respectively. For a brief discussion of the concerns involved, see (Beaney, 1997, pp. xx)

order to yield a determinate value is a number. Filling in the argument role of the expression with a number, thereby fills the gap indicated by '( )' and returning a saturated value. As such, (f_{12}) is a function from numbers to numbers. Give it a number and it returns a number. However, not any old expression can fulfill this saturating role. For example, the expression in (f_{13}) would not saturate (f_{12}). The result of (f_{13}) filling the argument position of (f_{12}) is

$f_{12}(f_{13})$ (( )–1)+3

This expression remains unsaturated, insofar as it does not name a value (or number). For an expression to saturate a function like (f_{12}) that expression must itself be saturated or complete. Thus

'$f_{12}(2)$', or '2+3'

is a saturated expression, as it names, or refers to, the number five.

With this function-argument architecture in place, we have a means of distinguishing between (12) and (13), while still accepting that they are equivalent, as expressed in (14). While they both have the same value (the number five), the functions that characterize (12) and (13) respectively, are not equivalent. That is, (f_{12}) and (f_{13}) differ, insofar as they require different saturaters to yield an expression that has the value five.

The Fregean conjecture is to treat the *semantic* value of natural language expressions as having this function-argument structure. Natural language sentences express unified propositions because the senses of the constituent expressions of natural language sentences exhibit function-argument structure of the kind just illustrated in mathematical expressions. Consider the following expression in English

(15) John's mother

The expression in (15) refers to a particular object, much like (12) has a particular value. Likewise, the sense of (15) has a functional character, with 'John' as the saturating argument. Thus

(16) ( )'s mother

represents the unsaturated nature of the sense of 'mother'. As an unsaturated function from objects to objects, (16) is saturated by 'John' in (15), and thereby

refers to some particular human woman. Because John refers to a particular person (much like '2' refers to a number), the sense of 'John' is saturated, or complete. As such this saturated sense, when in the argument position for the unsaturated (or incomplete) functionally characterized sense of (16), yields a sense composed of these saturated and unsaturated parts. Thus the expression 'John's mother' has a saturated sense, insofar as it refers to an object. This sense is composed of the senses of the constituent expressions of (15) ('John' and '( )'s mother') in accordance with a function-argument architecture, and thereby yields a saturated sense that refers to JOHN'S-MOTHER. That is, much the way (f_{12}) fails to have a value in the absence of a number term (like '2') as an argument, likewise (15) fails to refer to an object in the absence of a saturating expression (like 'John') in argument position. Thus, unsaturated senses require saturated senses to be complete—i.e., to be referring expressions.

A final component to Frege's proposal requires exposition before we can appreciate the basic structure of his solution to the unity problem. Thus far we have seen that co-referential expressions can differ in meaning by having different senses. But such senses cannot be *propositions*. Propositions, recall, must be bearers of truth-values. The kinds of expressions we have considered thus far (and the senses that present their referents) are neither true or false. 'John's mother is true' is as nonsensical as '2+3 is true'. To complete Frege's view on propositions we need to see how the function-argument structure that informs the distinction between senses and referents applies to (what Frege calls) *statements*.

Statements are expressions whose senses are Thoughts (*Gedanken*), or (to unify terminology) *propositions*. They are linguistic expressions that have truth-values. In the case of arithmetic, expressions like

(14) 2+3 = 6–1

are statements. To utter '2+3 = 6–1 is true' is felicitous in the way '2+3 is true' is not. The same functional insights just explored regarding naming expressions extends to statements. The expression in (14), after all, is the result of putting '2' in the argument position of

(f_{14}) ( )+3 = 6–1

The difference between this function and the one expressed in (say) (f_{12}) is that (f_{12}), when taking a particular number term as an argument yields a number

value. But, saturating (f_{14}) with a number term does not likewise yield a number value. Replacing '( )' in (f_{14}) with a number term yields a statement, which expresses a Thought, and refers to a truth-value. After all, '2+3 = 6–1' is true. What unifies the Thought expressed by (14) is accounted for by the unsaturated character of the function in (f_{14}). Likewise, statements in a natural language, with Thoughts as their senses, are unified by the properties intrinsic to the senses that constitute the Thought, as reflected in the expression.

(17) John's mother is brilliant.

This statement expresses a Thought composed of the senses of the constituent English words in (17). The sense of 'is brilliant' is unsaturated, as characterized by the function '( ) is brilliant'. Intrinsically unsaturated, this sense, once saturated, refers to a truth-value. The saturated sense of 'John's mother' fills the argumentative role characterized by the function '( ) is brilliant', completing a Thought that refers to a truth-value.

Treating the senses of expressions as having function-argument structure explains how we understand novel Thoughts. We can understand (17), assuming this expresses a novel thought, because we understand the sense of the constituents of (17). We arrive at the Thought that (17) expresses because we embed the senses of its constituents in the way we can embed functions. By embedding the functionally characterized sense of 'mother' (namely '( )'s mother') in the functionally characterized sense of 'is brilliant' ('( ) is brilliant'), we can build the characteristic function for (17), namely '(( )'s mother) is brilliant'. That is, the sense of the English expression 'is brilliant' characterized by the function '( ) is brilliant' that takes objects and maps them to truth-values just in case they are brilliant can be constructed by composing the senses of its constituent expressions. Taking the sense of 'mother', characterized by the function '( )'s mother' that maps objects to objects, as the argument for the 'brilliant' function yields the characteristic function of the predicate in (17), namely

(18) (( )'s mother) is brilliant.

Further saturating (18) with the sense of 'John' yields the characteristic (saturated) function (indicating the Thought) expressed by (17). This Thought is true,

just in case the characteristic function of (18) maps JOHN to TRUE.[14] In this way the senses of complex expressions, including statements, are constructed from the senses of their constituent expressions. When a sense is composed in this way, admitting to a structure characterized by a function from objects to truth-values, saturating it with an object-referring sense results in a saturated, unified propositional Thought. What then distinguishes a Thought from a mere list is that a list, unlike the characteristic function of a Thought, does not refer to a truth-value, but an ordering of objects. Because senses have function-argument structure, a Thought, like a complete mathematical function, is unified is the way a list (of numbers and operations) is not. What accounts for the unity of the Thought is that some constituents of the Thought are unsaturated, thereby made complete once they are saturated by object-referring senses.[15]

Such Thoughts also differ from Russellian propositions. For Russell, a proposition bears some relation to the world, such that the world makes a proposition true. This relation (however it is precified) brings the objects and relations of the world into a constitutive relation to form the proposition. The problem this account faces, as illustrated in the last subsection, is that accounting for this relationship between propositions and their truth-makers either reduces propositions to their truth-makers (like facts), or leaves propositions un-unified (and list-like). Frege's resolution to this dilemma is to treat Thoughts as having function-argument structure. This purportedly resolves the unity problem (and other puzzles), because a saturated Thought is unified in the way a function bearing an argument is complete, and complete in a way a list of objects and concepts is not. So too, this treatment avoids reducing propositions with their truth makers. Saturated Thoughts *refer* to their truth-makers, the truth-values TRUE and FALSE, they are not *identical* to them.[16] In the way 'John's mother' refers to a particular object (a person), 'John's

[14]I've put TRUE in small capitals here indicating that it is a referent. This accords with much of what Frege says about truth and, as we'll see toward the end of the chapter, agrees with the current model-theoretic approaches to semantics that treats the truth values ($\bot$ and $\top$) as objects in the domain.

[15]For Frege Thoughts are (in some sense) intrinsically saturated and unified wholes. They do not strictly speaking have any parts. Rather, what his *Begriffshrift* gives us is a means of *analyzing* Thoughts, imposing a structure upon them. The idea is that by taking the complete, saturated character of Thoughts (like JOHN'S-MOTHER-IS-BRILLIANT), and imagining the result of extracting from that complete unity the saturated meanings of referring expressions (like 'John'), what remains can be characterized as a function with an unsaturated meaning. Frege's logic describes these imagined parts, and the manner in which they fit together to form Thoughts.

[16]Frege seems to be of two minds on this point. In Frege (1891 [1997]) he clearly treats the TRUE

mother is brilliant' refers to a particular (logical) object (a truth-value). Thus, while the (problematically unified) Russellian proposition expressed by 'John's mother is brilliant' is just the thing in the world that is brilliant, the Thought expressed by 'John's mother is brilliant' is the function-argument structured sense of 'is brilliant', that when saturated by the sense of 'John's mother' refers to TRUE (or FALSE).

What emerges from our exegesis is an ontology of propositions that posits three distinct kinds of senses (Sinnen): those that present objects (as Bedeutung), those that present Concepts (as Bedeutung), and those that present truth-values (as Bedeutung). Names refer to objects, and their senses are thereby saturated. Likewise, Thoughts (or propositions) as complexes of the senses of the constituent expressions of statements, are the senses of statements that take truth-values (the TRUE and the FALSE) as their objective referents. Predicates however, refer to *Concepts*. Because functions require arguments to saturate them, the sense of a predicate is unsaturated, as the predicate fails to present an object (or truth-value), but instead presents a Concept. As a result, Thoughts are bound together because the senses that compose them are appropriately saturated to refer to the TRUE or the FALSE.

This view raises a number of concerns. First, there is, of course, something odd about speaking of the TRUE and the FALSE as logical objects, somehow like the objects of the physical world. Relatedly, the notion of sense (Sinn) is mysterious. Frege alludes to the ontological nature of senses in a few places, but these discussions are opaque and metaphorical. In Frege (1892 [1997]) he offers some explanation:

> [TRIANGLE]
> A difference [between identity statements with co-referential expressions] can arise only if the difference between the signs corresponds to a difference in the mode of presentation of the designated. Let a, b, and c be lines connecting the vertices of a triangle with the midpoints of the opposite sides. The point of intersection of a and b in then the same as the point of intersection of b and c. So we have different designations for the same point and these names ([(α)]'point of intersection of a and b', [(β)]'point of intersection b and c') likewise indicate the mode of presentation; and hence the statement [the intersection of a and b is the same as the intersection of b and c] contains actual knowledge. (Frege, 1892 [1997], p. 152)

and the FALSE as logical objects, extant in the Third Realm. However, in Frege (1892 [1997]) he seems to hold a view whereby the True and the False are shorthand for something like a state of affairs. States of affairs for Frege seem to be somewhat like the truth-makers for Russellian propositions.

The distinction between an expression's sense and its references is fairly clear here. The difference between (α) and (β) is not captured by their (mutual) referent. They differ in the way in which they indicate which point is up for discussion. So to speak, (α) asks us to explore the length of a, and stop when we get to b. Contrastingly, (β) asks use to explore b, and halt when we've reached c. In this way these expressions present the center point of the triangle in distinct ways that bear on the senses of the respective expressions. Senses are apparently ways or "modes" of presenting a referent. But what are modes of presentation then, such that they exhibit function-argument structure?

> [TELESCOPE]
> The [referent] of a proper name is the object itself which we designate by using it; the Idea which we have in that case is wholly subjective; in between lies the sense, which is indeed no longer subjective like the Idea, but yet is not the object itself. The following analogy will perhaps clarify these relationships. Somebody observes the Moon through a telescope. I compare the Moon itself to the [referent], it is the object of observation, mediated by the real image projected by the glass of the interior telescope, and by the retinal image of the observer. The former I compare to the sense, the latter is like the Idea... (Frege, 1892 [1997], p. 155)

This second metaphor is helpful, at least insofar as it outlines what Sinnen cannot be. The sense of an expression, cannot belong to an individual mind. The perceptions and phenomenology that populate our inner mental lives occupy the realm of Ideas for Frege.[17] If John and Sally both gaze upon a rose in a field, they do not share an Idea. That is, the imagery that characterizes their respective phenomenal states is not shared between them. These experiences are about the same object (THE-ROSE), but this is distinct from (what Frege calls) their respective Ideas of the rose. Similarly, the image of the Moon before the mind, as derived from the image on an individual's retina and distinct from the image of the Moon in the telescope, is an Idea. The senses of expressions are not Ideas, as the senses of expressions and

[17] In Frege (1918 [1997]) he speaks of Thoughts, the senses of sentences, as existing in a "Third Realm," distinct from the "Realm of Objects," and the "Realm of Ideas." At face value, Frege seems to be committed to a robust kind of Platonism about the ontological character of senses. The main worry with such a view is that, given that humans seem to understand natural language expressions, a theory about the meaning of such expressions should shed light on how we come to understand these meanings. But, if these meanings are senses, existing in some distinct realm, the prospects for an adequate explanation of the relationship between expressions in a natural language and graspable other-realmly meanings does not look promising. One of the primary challenges for Frege-inspired theories of meaning is to formulate the notion of 'sense' such that we can explain how such meanings are graspable by competent speakers of a natural language.

statements are the things that are grasped by different speakers, in the way that different viewers can make use of the image of the Moon in the telescope.

Russellian propositions face two problems that Fregean propositions resolve. First, Fregean propositions are unified, insofar a unsaturated senses are made complete by saturated senses. This function-argument structure of Fregean propositions explains how a proposition can be distinguished from a mere list of referents. Second, the senses of expressions contribute to propositions in a way that mere referents cannot. Because senses present the referents of expressions, two co-referential expressions can have different senses (and thus different meanings), insofar as they pick out their shared referent in distinct ways. However, one wants to know what senses are such that they can resolve these problems. In the next section I review one approach that characterizes senses as sets (or functions) of possible worlds (to truth values). Treating propositions as such, resolves the primary objection to Fregean senses, namely that their ontological status is nebulous.

1.1.4 Propositions as Sets of Possible Worlds

The possible world semantics that developed from the treatment of Frege in Carnap (1947) is thought to have shed light on Frege's notion of 'sense'. The worry posed in the last subsection regarding Frege's semantics is that the notion of 'sense' is mysterious. Frege tells us that senses of expressions in a natural language are

1. not Ideas (in "the head");
2. not referents;
3. modes of presenting referents;
4. entities with function-argument structure; and
5. grasped (or shared) by everyone familiar with the language to which their expressions belong.

While this list is somewhat informative, the nature of an expression's sense is still fairly opaque. On one reading, from Frege (1918 [1997]) senses exist in a Third Realm, garnering an existence distinct from that of physical individuals, and

likewise distinct from mental representations.[18] Possible world semantics, and the propositional theories of meaning that they beget, can be seen as attempts to meet this mysteriousness challenge. In this section I examine this view as a means of illustrating the problems that motivate contemporary conceptions of propositional meanings. As with the Russellian view, the problems outlined here, and the motivation to resolve them, entail a commitment to ($\mathcal{E}$).

1.1.5 Intensional Contexts

In our discussion of the historical account of propositions offered by Russell and Frege, we've focused on the problem of unity. We've seen how the tasks that propositions are meant to accomplish make this problem difficult to resolve. Russellian propositions, that take real-world objects, properties, and relations as constituents, can solve the problem of unity by reducing propositions to the facts that make them true. But propositions-as-facts are incapable of being false, since all facts are true. This result is unacceptable if propositions are to be the contents of beliefs, at least insofar a people can have *false* beliefs. If we follow Frege in thinking that the meanings of expressions are not their referents, but rather the mode in which these referents are presented, we avoid the problem of reducing propositions to the facts that make them true. Propositions, composed of the senses of the sentential constituents that express them, are a step removed from their truth-values. As (complex) senses themselves, propositions *refer* to truth-values (the things that make them true or false), they are not identical to them.

The main problem with senses however is that their nature is mysterious. In observing that expressions have a kind of semantic feature, captured by the notion of sense, Frege has not successfully offered a solution for the problems that face propositional accounts of meaning unless he can tell us what senses are. That is, while senses can, given the kind of function-argument structure Frege suggests, account for the unity of propositions, while keeping them adequately divorced from their truth-makers, for this account to be a fitting explanation it must provide a cogent formulation of the ontological nature of senses. What emerges then from the birth of modern philosophy of language is a challenge to modern (propositional)

[18]Such representations, on this reading would be distinct from states of a mind/brain, leading to a kind of Cartesian Dualism.

semanticists: give an account of a proposition that does the work senses do for Frege that is ontologically scrupulous. If we think that the meanings of natural language expressions are mind-independent, embodied in the work that propositions are meant to do, one way a semanticist can respond is to provide an account of Fregean senses. Put another way, our exploration of Russell illustrates that a purely extensional account of meaning fails in many contexts, as in the case of identity statements and propositional attitudes. So called *intensional* contexts, where the meaning of an expression cannot be identified with the referent of the expression, highlight the need for (something like) Fregean senses, or *intensions*. Treating senses as intensions, or sets of possible-world-extension pairs, is one way of meeting this challenge.

The state descriptions introduced in Carnap (1947) provide a variety of philosophical tools, namely *possible worlds*, that one might put to work in meeting this challenge. Possible worlds are (intuitively) ways in which the world[19] could have been. Seemingly, for any true (non-analytic) claim about the world, the conditions that underwrite its truth could have been different. For example, this work could have been one sentence shorter, had this sentence been omitted. Given a plenitude of possible worlds, and the mereological toolkit, we can construct an account of propositions that approximates the role of senses in Frege's logic. Propositions on this view are sets of possible worlds, and senses (or intensions) are functions that map (sets of) possible worlds to: 1) truth values, 2) (sets of) individuals, or 3) other sets of possible worlds. That is, if we take possible individuals and possible worlds as basic elements of the domain, when given the mathematical structure suggested by Frege and the collection theoretic structures used by Carnap, we can account for the distinction between the intensions and extensions of expressions. The intensions of expressions are then captured by the shift in the extension of an expression from one possible world to another.

Consider the following referring expression

[19]Two points, one of clarity, and one of substance: 1) 'world' here ought to be interpreted broadly, not to describe worlds as in "places like Earth," but rather as in "places like this universe." Possible worlds (on this conception) are *not* places we could travel, given sufficient technology. 2) The use of possible worlds bears on the ontology of such worlds. If possible worlds are meant to do duty as meanings for natural language expressions, one wants to know how expressions relate to them. And how words relate to things like possible worlds depends on what kind of thing a possible world is. If we are committed to positing possible worlds as constituents of propositions we are (at least *prima facie*) committed to their existence. This point motivates Lewis (1986) in treating possible worlds and their inhabitants much like the actual world.

(19) The President of the United States

At first blush this expression (uttered in 2012) picks out a single man, BARACK-OBAMA. But, of course, uttered seven years earlier (19) refers to a different individual, GEORGE-W-BUSH. So the meaning of (19) cannot merely be the referent of (19). One initial response to this worry might simply be to point out that (19) is imprecise, for reasons peculiar to the nature of the presidency. That is, to utter (19) is *really* to say

(20) The *current* President of the United States

But surely this is no more precise than (19), other than to note that (20) wears on its sleeve the temporal marking (19) garners covertly. But again, maybe the problem here is still one of imprecision. When I utter (20) after all, one should interpret me as saying

(21) The president of the United States *in 2012*

The expression in (21) is precise enough to avoid our worries about temporal context, for (21) will (surely) always refer to Barack Obama, whether a speaker utters (21) in 2012, in 2000, or in 2020. But, consider the following sentence

(22) John McCain might have been the president of the United States in 2012.

A competent speaker of English could accept (22) as true. After all, the happenings of election night on November 4th, 2008 could have gone differently, if (say) four million Californians made a different decision in their respective voting booths. But (22) (and (22′)) contrasts straightforwardly with (23)

(22′) John McCain might have been $[_{\alpha}$ the president of the United States in 2012].

(23) John McCain might have been Barack Obama.

If the meaning of (21) was nothing more than its referent, any competent speaker that accepted (22) as true would also accept (23) as true. After all, the referent of the definite description in $[_{\alpha}\ldots]$ is just BARACK-OBAMA. So, assuming that synonymous expressions can be substituted for one another in sentences *salva veritate*, substituting 'Barack Obama' for '$[_{\alpha}\ldots]$' in (22′) should yield a sentence (the sentence in (23)) with the very same meaning and truth value. But, whatever

one might think of McCain's chances in the 2008 election, there is no sense in which John McCain could have been the son of a Kenyan economist and a Mid-West anthropologist.

In such modal contexts, referents will not suffice as the meanings of expressions. To accept (22) as true is not to entertain the possibility that McCain might have had different parents, but rather, to entertain the possibility, that had events in the world been otherwise, John McCain (and not Barack Obama) would have satisfied the predicate '( ) is[/was] the president of the United States in 2012'. Propositions as function from possible worlds to truth-values can account for the meaning of expressions in such contexts.

On this set-theoretical treatment the extension of an expression is either an individual, a set of individuals, or a truth-value. For singular terms, like names, their extension just is their referent. The extension of 'Barack Obama' just is the man. For predicates, like 'is/was president', their extension is a set of individuals, in this case all the (past and future) presidents. Finally, following Frege, we can treat the extension of a sentence as the truth-value of the sentence (TRUE, or FALSE). The intensions of expressions are the functions that have possible worlds as their domain, and range over extensions of these types.

The intensions of expressions bring to the meaning of an expression the possible world apparatus, unifying the various extensions of expressions at distinct possible worlds. Thus the intentions of expression are also sets, but sets of the extensions of expressions, paired with the relevant possible world. Functions as intensions are sets of ordered n-tuples of (possible) world-extension pairs. Consider our example above

(21) The President of the United States in 2012

The intension of this expression is a function, a kind of set that pairs possible worlds with individuals. Using standard notation, treat $w_@$ to stand for the actual world, and w_n (for any natural number n) to stand for some possible world. Then the characteristic function (the set of world-individual pairs) would be

(21) The President of the United States in 2012

(21′)
$$\left\{\begin{array}{c} <w_{@}, Obama> \\ <w_1, McCain> \\ <w_2, Romney> \\ \ldots \\ <w_{87}, Colbert> \\ \ldots \end{array}\right\}$$

For predicates such a '( ) is a president of the United States'[20], their intensions are functions from possible worlds to predicate extensions, or sets of individuals. The characteristic function for '( ) is a president of the United States' is indicated below in (24′)

(24) ( ) is a president of the United States

(24′)
$$\left\{\begin{array}{l} <w_{@}, \left\{\begin{array}{l} Washington \\ Adams \\ Jefferson \\ \ldots \\ Obama \\ \ldots \end{array}\right\}> \\ <w_1, \left\{\begin{array}{l} Washington \\ Adams \\ Jefferson \\ \ldots \\ McCain \\ \ldots \end{array}\right\}> \\ \ldots \\ <w_{87}, \left\{\begin{array}{l} Washington \\ Adams \\ Jefferson \\ \ldots \\ Colbert \\ \ldots \end{array}\right\}> \\ \ldots \end{array}\right\}$$

Finally, the intensions of *sentences* are functions from possible world to truth-values.

[20]For explanatory purposes I ignore the tense implicit in 'is' as distinct from 'was'.

So the intension of (25), given the characteristic function in (21′) for the referring expression in (21) is

(25) Barrack Obama is the President of the United States in 2012.

(25′) $$\left\{ \begin{array}{c} < w_{@}, True > \\ < w_1, False > \\ < w_2, False > \\ \ldots \\ < w_{87}, False > \\ \ldots \end{array} \right\}$$

Thus we see the following conceptions of the intensions and extensions for (roughly) nouns, verbs, and sentences. Definite nouns have individuals as their extensions, and functions from possible worlds to individuals as their intensions. Verbs have sets of individuals as their extensions, with functions from possible worlds to sets of entities as their intensions. Finally, sentences have truth-values as their extensions with, with functions from possible worlds to truth-values as their intensions.

Such a taxonomy can address many context in which the intensions of expressions bear on the meaning of the expression. To take an earlier pair of example sentences,

(10) London is the capital of England.

(11) London is London.

Here the predicate 'is the capital of England' must be understood intensionally, insofar as substituting a coextensive expression fails to preserve the informativeness of the expression. In particular, substituting 'is London' for 'is the capital of England', yielding (11), fails to preserve the informativeness of the expression in (10). But, read intensionally, we can see why (10) is informative in a way (11) is not. At all possible worlds, London is self-identical, as indicated in (11). That is, at any possible, if there is a London at the world, (11) is true at that world. But, there are some possible worlds in which (for whatever political reasons) some other city is the capital of England. More precisely, there are possible worlds in which there is a London, yet (10) is false, if (say) Phillip II of France had defeated Henry II of England. This feature of the meaning of (10) is captured by the intension of (10),

insofar as that function will output different individuals (different cities) depending on the possible world it takes as an input. For the actual world, this function maps to LONDON, which is why (10) is true in the actual world.

Two problems undermine this view of propositions. First, such propositions are not up to one of the tasks propositions are meant to do. Recall that a theory of propositional meanings must be able to generate distinct propositions for every non-synonymous declarative sentence. Put another way, no two declarative sentences can express the same proposition unless they have the same meaning. Thus, if propositions are sets of possible-world-truth-value pairs, then each semantically distinct expression must be paired with a different set. However, many semantically distinct declarative sentences express identical propositions, so construed. Consider some classic examples, (26) and (27)

(26) All mammals are chordates.

(27) All mammals are renates.

Intuitively, these two claims express different thoughts: one about the engine of cardiovascular systems of mammals, and the other about the filtration of the cardiovascular systems of mammals. That is, (26) and (27) do not have the same meaning. But on the possible-world account of propositions (26) and (27) express the same proposition, having the same propositional meaning. On the possible world reading, (26) is paired with the function that maps a world to true just in case, for any world w_n, there are mammals at w_n and all the mammals at w_n have hearts. Of course, this function is just going to map all worlds at which there are mammals to TRUE, given that a necessary condition of mammalhood is being a chordate. Thus the intension of (26) will be the smallest function that maps all the mammal-worlds to TRUE (and the other worlds to FALSE). Turning to (27), the possible world treatment of its meaning indicates the smallest function that maps a world to TRUE just in case, for any world w_n, there are mammals at w_n and all the mammals at w_n have kidneys. But as before, this function is just going to map all worlds at which there are mammals to TRUE, given that a necessary condition of mammalhood is being a renate. But again, this will be the smallest function that maps all the mammal-worlds to TRUE (and the other worlds to FALSE). Thus, the propositional meaning of (26) and (27), understood as the set of possible-world-truth-value pairs, is the very same set. But if sentential meanings are propositions, and proposition

are (characteristic functions of) possible-world-truth-value pairs, (26) and (27) have the same propositional meaning insofar as they express the same set. Thus, the propositions-as-possible-worlds account fails to distinguish between the meanings of these two non-synonymous expressions.

The second difficulty pertains to the distinction between *rigid designators* and directly referring expressions. Discussed in depth in Kripke (1980), a rigid designator is an expression that refers to (or designates) the same individual in every possible world. The paradigm cases of these expressions are names like 'Barack Obama'. Rigid designators are often contrasted with definite descriptions, like 'the man driving the bus' or 'the President of the United States'. As we just saw, such expressions can have different referents at different worlds. However, unlike definite descriptions, whose referents can shift from world to world, rigid designators refer to the same individual at every world (where they have a referent).

Directly referring expressions are like definite descriptions, in that their referents shift from world to world. However, the distinct feature of directly referring expressions is that they are not mediated by (anything like) Fregean senses. For example, to determine the referent of

(28) The student of Frege, that was also a member of the Vienna Circle

one must be acquainted with the referents of 'Frege', 'the Vienna Circle', and the notion of 'member'. In this way, (28) refers the Carnap *indirectly* in a way 'Rudolf Carnap' does not. However, some definite descriptions are rigid designators, in that they (indirectly) refer to the same individual in every possible world. For example, 'the result of squaring 2', while requiring mediation by the notion of 'squaring', designates, at every possible world, the number FOUR.

Directly referring expressions, discussed in depth in Kaplan (1977 [1989]), find their referents without the need for mediation. Such expressions refer directly in virtue of the context of their use. Demonstratives like 'I', 'that', 'here', and 'now' refer directly to individuals given the context in which they are used.

The problem, so to speak, is that on the possible world account of propositions, the intentions of all rigid designators, be they directly or indirectly referential, will have the same form. As sets of possible-world-individual pairs, all rigid designators take the form (n):$< w_n, o >$ where o is the designated individual, and w_n ranges over all possible worlds that contain o. That is, all rigid designators, whether they

refer directly (like '4') or indirectly (like 'the result of squaring 2') will have constant functions as intensions. Insofar as our semantics should mark the distinction between directly referential and indirectly referential expressions, possible world semantics fails to accommodate this demand.

One response to these worries, expressed in Soames (1985), is to return to Russellian propositions that take individuals, properties, and relations as constituents of propositions. This account of propositions generates distinct propositions for sentences that are true in all the same possible worlds. Recall the following pair

(26) All mammals are chordates.

(27) All mammals are renates.

These sentences are true at all possible worlds (where there are mammals). Yet they make distinct claim about mammals. The expression in (26) claims that mammals have hearts, while (27) claims that mammals have kidneys. These sentences have distinct meanings, yet they have the same extension and intension (in that their respective characteristic functions have the same possible-world-extension pairs as elements). For the Neo-Russellian, the propositional meanings for (26) and (27) are not intensions, but structured entities composed of the individuals and relations expressed by the sentences in (26) and (27). The proposition for (26) will be distinct from the proposition for (27), insofar a the property of BEING-A-CHORDATE is a constituents of one, but not the other.[21]

Similarly, the worry of conflating rigid designation with directly referring expressions is resolved by taking as propositional constituents the individuals, relations, and properties denoted by these expressions. Directly referring expressions, on such a view, will have the individuals they denote (given a context) as terms in their propositional meaning ('4' simply contributes the number FOUR to the propositional meaning of any sentence for which it is a constituent). Contrastingly, indirectly referring expressions, given their complexity, will contribute some structured term to

[21]Importantly, this Neo-Russellian position is constrained on the account of properties they can endorse. One would lose the purported virtue of this position by adopting (say) a view of properties that defines them as sets of possible worlds. If such properties are the constituents on Neo-Russellian proposition, then necessarily co-instantiated properties would be identified with the same set of possible worlds. In these cases, a sentence with an expression referring to one of those properties would express the same proposition as a variant of that sentence wherein the property-denoting expression is replaced with an alternative expression denoting the other, co-instantiated property.

the propositional meanings of the sentences in which they appear. The expression 'the result of squaring 2' will contribute to the propositional meaning of any sentence for which it is a constituent, among other things, the number TWO and the property of BEING-A-SQUARE-OF. So the Neo-Russellian position seemingly has the resources to account for these problems that are not available to the possible-world view of propositional meanings. But the Neo-Russellian view is still saddled with many of the worries expressed in the first section. For example, one still wants to know what brings the constituent individuals, properties, and relations together to distinguish the kind of proposition expressed by (26) and (27), from a mere list of individual, properties, and relations therein.

Current work on propositional theories of meaning attempts to overcome these challenges, either by giving an account of Neo-Russellian propositions whereby propositions are distinct from mere lists (e.g. King, 2007), or by identifying propositions with entities that are fine-grained enough to appropriately appropriate distinct propositions for each non-synonymous natural language expression (e.g. Barwise & Perry (1983), who define propositional meanings in terms of situations, and *parts* of worlds.). But, the problems expressed here have set the agenda for contemporary theories of propositional meanings.

There is certain philosophical work we want propositions to do. Sets of possible worlds are infelicitous laborers in this regard. What the backer of a propositional theory of meaning must do is identify the ontological status of propositions such that they are fine-grained enough to account for the shifting meanings exhibited in intensional contexts. Further, the kind of entities that propositions are better be of a kind that can be grasped by competent speakers of a natural language, such that speakers can arrive at the correct proposition for a given utterance (given a context). So to speak, a theory of propositional meaning needs to define propositions so that they can not only labor in the ways we need them to, but such that they are the kind of laborers we can hire.

But, whatever moves a propositional semanticist might make to address these worries, such a theorist is committed to ($\mathcal{E}$). If sentential meanings are sets of possible-world-truth-value pairs, then the meanings of sentence will determine what worlds they are true in. Whatever additions must be made to construct a finer-grained distribution of functional meanings, such functions must pair something (e.g. situations, scenes, etc.) to truth-values. That is, developing finer-grained intensions

will surely require positing some additional ontology, given that the plenitude of possible worlds is not sufficiently large to produce distinct enough functions for all the ways in which expressions can be interpreted. However, while the domain of an intension-as-function view might require some elaborate structure (at least as complicated as Kaplanian character), the range will still be the truth-values. As such, functional meanings will determine the conditions under which an expression is true (or false). Put another way, the functional meaning of an expression will determine the truth-conditions for the expression, which implies ($\mathcal{E}$).

1.2 Truth-Conditional Semantics

In the previous sections we've seen how various propositional theories of meaning, inspired by Russell and Frege, are committed to particular conceptions about the nature of truth-conditions. These theories differ from the ones discussed in the current section, in that the proposed propositional meanings of natural language expressions, while they entail that sentences have truth-conditions of a particular sort, are distinct from the truth-conditional properties of sentences. The theories discussed in this section *identify* the meanings of expressions with their truth-conditions. These truth-conditional theories of meaning not only carry commitments about the kind of things truth-conditions are, but they further claim that the meanings of expressions are nothing more than those conditions.

These theories have a common source, in Tarski (1933 [1983]) wherein he constructs a complete semantics for formal languages that meet certain constraints. The basic idea is this: given a procedure for generating sentences in a language (a syntax for the well-formed-formulas of the language), and a domain of objects, Tarski shows how one can generate of a truth predicate (call it 'Tr') for that language in terms of satisfaction (by sequences of domain objects). Put a bit more precisely, Tarski gives us a means of classifying the (well-formed) sentences of a language into those that satisfy the Tr-predicate, and those that do not by appealing to the notion of *satisfaction*: a (well-formed) sentence is in the extensions of the Tr-predicate, just in case it is satisfied by all sequences. Given a domain of objects Tarksi shows us how to construct the two essential notions here, *satisfaction* and *sequence*. Thus, specifying the truth-conditions for a (well-formed) sentence in a formal language (containing no semantic predicate of its own) can be achieved by

specifying the conditions under which that sentence is satisfied by a sequence of ordered individuals. The conjecture championed by a truth-conditional semantics for *natural* languages is that a Tarski-style Tr-predicate can be applied to natural languages to give a complete semantics for them in terms of satisfaction. Such a predicate can serve as a formal means of characterizing truth, and on can begin to construct a theory of meaning that can be reduced to a theory of truth.

In this section I begin by broadly outlining the truth-conditional semantic method of Davidson, drawing from Tarski(1933). I'll indicate how this semantic conception, or the generalizations taken from Tarski's example, relate to the Davidsonian and Montegovian approaches to truth-conditional semantics. The difference between these approaches pertains to the metalanguage used in defining truth and its relation to a domain. Simplifying a bit, the distinction between these approaches is that one posits truth-values (and functions) as elements of the domain, and the other does not. Positing such entities expands the expressibility of the metalanguage, creating expressions of various systematic types, and enabling them to be treated as functions mapping individuals of the domain to truth.

1.2.1 Davidson

> ... meanings as entities may, in theories such as Frege's, do duty, on occasion as references, thus losing their status as entities distinct from references. Paradoxically, the one thing meanings 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 nontrivially give the meaning of every sentence in 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, 1967b, p. 307)

As we saw in the previous section, propositional theories of meaning pair sentences with propositions, as the objects that serve as the meanings of sentences. In this passage from "Truth and Meaning" we find Davidson's motivation for the truth-conditional account he sketches, namely that we have no need for propositions. I'll begin this section by detailing the qualms expressed in this passage. Doing so will help frame the goal of the Davidsonian project, by making explicit the conditions a theory of meaning must meet in order to be an adequate account of how humans come to learn a language. That is, the concerns Davidson alludes to in this passage are the very aspects of language his theory is meant to capture.

I begin this section with an explanation of this passage, explicating the shortcomings of propositional theories of meaning. In particular, I detail Davidson's complaint that treating meanings as objects fails to explain how we come to understand the meaning of a sentence from the meanings of its constituents.

1.2.1.1 Compositionality

However humans manage to assign a meaning to a natural language utterance, what is clear is that we do so compositionally. This, along with the idea that a theory of meaning ought to be rooted in a theory of human understanding, is central to the Davidsonian program, and is the focus of this section. Originally presented in Davidson (1965 [1984]), Davidson argues that an adequate theory of meaning must explain how a natural language sentence can be understood by a competent speaker of that language. In light of two facts about natural languages and natural language speakers, Davidson concludes that any theory of natural language meaning must explain how competent speakers of a natural language can determine the meaning of a complex expression in that language from a finite number of more basic expressions.

The first fact is that natural language speakers can understand (in principle) an infinite number of (non-synonymous) sentences. For example, the following schema generates an infinite series of non-synonymous expressions in English:

(29) John said [$_1$that John said [$_2$ that... [$_n$ that John said [that Mary was nice]]...].

Repeated embeddings of the "said"-clause here produces a infinite number of sentences with (similar though) distinct meanings. This toy example illustrates a significant feature of natural languages and the humans that understand them—namely that a competent speaker of a natural language can understand an infinite number of expressions with distinct meanings. Noting a second fact, that all competent language speakers manage to learn (or acquire) a language within a normal life span (in typically no more than five years), a speaker's competence cannot be understood as a pairing of *independent* meanings with expressions. That is, the various expressions generated by the schema in (29) cannot have primitive meanings, that we learn by pairing those expressions with those un-related meanings, one at a time.

What these facts show is that for a natural language to be learnable, there must be a way of combining a *finite* number of basic expression-meanings such that the

meanings of the complex expressions they produce are discernible from the meanings of those basic constituents. The characterization of this capacity to understand a series of infinite (complex) expressions from the meaning of their constituent (basic) expressions is what a semantics for natural (learnable) languages must deliver.

Another way of making this same Davidsonian point, is to highlight the finite capacity of the human mind to store the meanings of expressions. Whatever our theory of meaning is, and however humans happen to understand the meaning of expressions, presumably this is carried out by the human brain. Additionally, while the limits of human memory storage may well be vast, they are finite. Thus, if our natural language competence gives us the capacity to understand an infinite number of expressions, this capacity must be the result of being able to store only a finite number of expression meanings. If so, then a theory of natural language meaning must be compositional.

A theory of meaning, if the theory is to explain the human capacity for understanding (and learning) a natural language, must be compositional in the sense that the theory must

(a) contain only a finite number of primitive expressions;

(b) contain a finite list of rules for generating complex expressions;

(c) contain only a finite number of rules for building meanings for those complex expressions out of meanings for primitives expressions.

Davidson's conjecture is that meanings as objects (like propositions) do no additional work beyond what can be done by the truth-conditions of expressions.

1.2.1.2 Meaning as Understanding

Davidson's main conjecture is that a Tarki-style theory of truth that meets these conditions captures a competent speaker's knowledge of a natural language. *A fortiori*, such a theory, assuming it indeed characterizes the knowledge[22] possessed

[22]As noted in Chomsky (2000) the kind of knowledge that characterizes the human language faculty bears little resemblance to the folk notion of knowledge. Not only is such knowledge not of the sort that one is aware of, in the sense that one cannot explicitly entertain (say) the rules that govern Universal Grammar, but that such knowledge is likely not even tacit. Chomsky himself adopts 'cognize' (and its variants) in place of 'know' as a means of indicating the distinct kind of innate knowledge contained in UG.

by all competent speakers of a natural language, is a sufficient theory of meaning for that language, insofar as anyone who understood a Tarski-style theory of truth for a particular natural language would have learned the language. Positing propositions in addition to this account needlessly adds to our ontology, insofar as these objects serve no purpose.

Put generally, Davidson's compositional claim is that a theory of meaning for natural language ought to generate, from a finite stock of (meaningful) expressions, a finite list of rules for combining those expressions, and a finite list of rules for generating meanings for the complex expressions, a (infinite) list of meaning postulates of the following form:

(M′) S means ϕ

where S is a sentence in the language the theory of meaning is supposed to characterize, and 'means' and ϕ are expressions in the language of the theory. Making this slightly more explicit we arrive at

(M) S in L means ϕ

where L is the object language the meaning theory targets. A theory of meaning is sufficient if it delivers an M-sentence (a meaning postulate in the form of (M)) for every (possible) sentence S of language L. Expressing the aim of a theory of meaning as such satisfies conditions (a)–(c) by default, so long as the meta-language of the theory can be formally defined. That is, so long as the language used in the M-sentence (bracketing S) is well-defined (in the Tarski sense).

The further conjecture of the Davidsonian program is that (M) can be re-written to read:

(T) S is true in L *iff* ϕ

What Tarski gives us is a means of specifying what makes S true for a language in terms of another language for which we can give a formal semantics. To understand the meaning of S is to understand the conditions under which S is true (i.e. satisfies the Tr-predicate), which we can specify by pairing S with a sentence in a distinct language whose content we can explicitly specify.

In the next section, I'll reconstruct the technical details of this project. But to get an idea what is at work here before bogging ourselves down with formal matters, first assume that we have a worked out theory of truth. Now suppose that

(30) Sally runs.

is true, and is made true by the world—whatever that may mean, given the assumed worked out theory of truth (which might mean something like, (30) is true just in case the individual SALLY instantiates the property RUNNING). Given this, we would thereby know that

(T_{30}) 'Sally runs' is true in English, *iff* Sally runs.

We know this because, whatever the state of the world described by 'Sally runs' is like, if that state makes 'Sally runs' true, then we know that in using the expression 'Sally runs' (as we do on the right-hand side of the '*iff*') we thereby describe the conditions under which the sentence is true. Now, suppose we constructed a (T) sentence for every expression in English of that form. Taken as a set, such a set of T-sentences would be materially equivalent to the predicate 'is true', insofar as it would tell us how to identify the true sentences in English. Further, this predicate would be materially equivalent to whatever our worked out theory of truth was when we started. That is, such a theory would predicate 'is true' of all the sentences that our assumed theory of truth would have as true. The Tarskian point then is, if we have a predicate that is materially equivalent to truth, then we have an adequate *definition* of truth. In the next section I'll illustrate how Tarski's apparatus recursively defines truth in terms of satisfaction. Davidson's claim is that a theory of meaning, really a theory of linguistic knowledge, is characterized by a (T)-theory of this kind—a Tarskian theory of truth can do duty as a theory of meaning.

1.2.2 Tarski

Tarski's approach to the notion of truth is to define truth for a language in terms of the recursive satisfaction of metalanguage expressions. That is, Tarski defines truth by recursively specifying the satisfaction conditions of an object language sentence with the satisfaction conditions of a metalanguage translation of that object language sentence. This schema for material adequacy demands that a truth definition for an object language take the form in the metalanguage (the language in which the postulate is constructed) as follows

(T) S is true in L *iff* ϕ

where 'S' is a name for a sentence, 'in L' names the object language of which the sentence named by 'S' is a part, '*iff*' takes on the meaning of the logical bi-conditional, and 'ϕ' is replaced by an expression in the metalanguage that is synonymous with S. Abbreviating for simplicity, and to make more explicit the distinction between the object language (OL) and metalanguage (ML), we get

(T) Tr(S) *iff* ϕ

where 'Tr( )' is the truth predicate for OL, and S is (implicitly) a name for an OL sentence. Let's suppose our OL is English, and our ML is a language much like English, but slightly more expressive to include the Tr( ) predicate and the bi-conditional '*iff*'. We can write the following T-sentence in ML

(T′) Tr('snow is white') *iff* snow is white

(T′) conveys the truth-conditions for the English sentence 'snow is white'. Again, suppose we have a worked out theory of truth. If the English sentence 'snow is white' is true (in the sense required by our worked out theory of truth), then the conditions under which the sentence is true are expressed by *using* the sentence 'snow is white'. Thus, because 'snow is white' is used in ML on the right-side of the bi-conditional, (T′) pairs 'snow is white' with the correct truth-conditions (those expressed by using 'snow is white').

One of course might wonder what work we've done here in articulating (T′). After all, what we seem to have done here is merely express the truth-conditions of the sentence 'snow is white' in an uninformative way. Seemingly what we want a theory of truth to tell us is what the truth-conditions for sentences are, in that we want to know what makes those sentences true. (T′) does not seem to be informative in this way. But compare (T′) with

(T*) Tr('snow is white') *iff* schnee ist weiss.

In (T*) we've replaced ML, shifting from English (plus some) to German (plus some). In comparing (T′) and (T*), (T*) seems to be informative in a way (T′) is not. That is, a speaker of German will find (T*) informative in a way they (and we) would not find (T′) informative. After all, a competent speaker of German (if they knew (T*)) would have learned something about the meaning of the OL

expression 'snow is white'. They would know that 'snow is white' is true under the same conditions expressed by using 'schnee ist weiss'. That is, because a competent speaker of German understands the conditions under which 'schnee ist weiss' is (accepted as) true[23], they can pair that meaning with the sentence named by 'snow is white', and thereby understand what conditions must be satisfied for 'snow is white' to be true.

The central idea is this: we can give a definition of truth for OL (that is, indicate the extension of the 'Tr( )' predicate) by pairing sentences of OL with sentences in ML, and so long as we can give the satisfaction conditions for those ML sentences, we thereby specify the satisfaction conditions of the OL expressions. Once we have the satisfaction conditions for all sentences in OL, we can then organize those sentence as those that are satisfied by all sequences and those that are not. Insofar as we can do this, we can define the truth predicate 'Tr( )' for OL as applying to all and only those sentence that are satisfied by all sequences. This of course suggests a further question: what does it mean for a sentence to be satisfied by a sequence? We'll turn to this question shorty, but it may be helpful to dwell on the point just made here. Tarski's strategy in defining the truth predicate 'Tr( )' is to give a formal semantics for a metalanguage, in non-semantic terms, i.e. in terms of satisfaction. By recursively pairing the conditions of sequence satisfaction of ML sentences with OL sentences we can define 'Tr( )' (or "truth") for all OL sentences. By doing this Tarski allows us to characterize truth without appealing to any semantic notion, but by merely appealing to satisfaction, given a plenitude of sequences. In the next section I illustrate how Tarski constructs such an interpretation for ML.

[23]To repeat here, the Davidsonian conjecture espoused in the previous section: because 'schnee ist weiss' is meaningful for a native German speaker, they know what the world must be like for the sentence to be true. Assuming we can use language to say true things about the world, however this is accomplished, the meaning of the sentences we use to talk truthfully about the world must play a role. Thus, someone who understands the meaning of 'schnee ist weiss' knows what conditions must hold for this sentence to be (used to say) true (things). Thus pairing 'schnee ist weiss' with 'snow is white' identifies (for the competent German speaker at least) the truth-conditions for 'snow is white'. The stronger Davidsonian point is that the meanings of such sentences do not merely convey their truth-conditions, but that those meanings are nothing more than those truth-conditions.

1.2.3 A Tarskian Example

In giving a theory of truth Tarski does not offer the sort of explanation one might expect from a theory, namely a definition of truth. Instead, Tarski provides a method (*via* a single example) for building a truth predicate for languages meeting certain conditions. In this section I give an explanation of a similar example.[24]

As indicated in the previous section, to build a truth predicate 'Tr( )' for an OL, we need an ML. This metalanguage ML must contain OL, in the sense that all sentences in OL must be expressible in ML. Given an ML and an OL, the method roughly contains two components: one indicates procedures for generating sentences in OL and the other indicates the procedures for assigning satisfaction conditions to ML translations of those sentences, thereby specifying the satisfaction conditions for their OL counterparts. To specify these procedures, of course, ML must be outfitted with elements for these procedures to apply to, namely a domain.

Tarski took his example OL to be first order logic, whose domain is occupied by classes (or sets). But the method here generalizes enough to permit any sort of domain of individuals. So let's take an intuitive series of objects as our domain of individuals:

DOMAIN:

OBAMA
MCCAIN
BUSH
GORE

Given a domain, we now construct (or stipulate) the symbols of our OL, which we will combine with a set of procedures that define the well-formed-formulae ('*wff*'s) for OL.

I. Vocabulary of OL

(a) primitives: '$\forall_{ol}$', '$\neg_{ol}$', '$\vee_{ol}$', '$Pres$', and '$Male$'.

(b) (non-constant) variables: 'x_1' 'x_2' ... 'x_n' (for any natural number n) ranging over objects in the domain

[24] I borrow here extensively from Sher (1999).

(c) (constant) variables: 'x_{Obama}', 'x_{McCain}', 'x_{Gore}', 'x_{Bush}' ranging over objects of the domain.

As will become clear below, the variables in (c) will, in effect, be referring expressions. This bears on the nature of the structure of the sequences used in defining the satisfaction conditions for the sentences we can construct for OL. In essence each sequences must assign the same (respective) object of the domain to the items in (c). Additionally, the 'ol' subscripts for the primitive constants in OL will indicate in our ML formulations of the satisfaction conditions of OL sentences that these expressions belong to OL.

II. Definition of Well-Formed Formula (*wff*s in OL)

Let 'n' range over the natural numbers[25], and let 'ϕ', 'ψ' be arbitrary *wff* expressions in OL. The following describes the procedures for generating *wff*s in OL.

1. $\forall n[\text{Pres}(x_n)]$ is a *wff*
2. $\forall n[\text{Male}(x_n)]$ is a *wff*
3. $[\text{Pres}(x_{Obama})]$ is a *wff*
4. $[\text{Pres}(x_{McCain})]$ is a *wff*
5. $[\text{Pres}(x_{Gore})]$ is a *wff*
6. $[\text{Pres}(x_{Bush})]$ is a *wff*
7. $[\text{Male}(x_{Obama})]$ is a *wff*
8. $[\text{Male}(x_{McCain})]$ is a *wff*
9. $[\text{Male}(x_{Gore})]$ is a *wff*
10. $[\text{Male}(x_{Bush})]$ is a *wff*
11. If ϕ is a *wff*, then $[\neg_{ol}\phi]$ is a *wff*

[25]Tarski's own notation avoids the use of natural numbers, taking iterations of [′] on variables to do the work of numbers; e.g. x''' is equivalent to x_3, and likewise x'''' is equivalent to x_4. The use of numbers here is merely a means of presenting Tarski's example in more intuitive notation.

12. If ϕ is a *wff* and ψ is a *wff*, then $[\phi \vee_{ol} \psi]$ is a *wff*

13. If ϕ is a *wff*, then $\forall$i $[\forall_{ol} \mathrm{x}_i \phi]$ is a *wff*.

Definition of Sentence

ϕ is a sentence in OL *iff* ϕ is a *wff*.

Thus far we've articulated the procedures for forming expressions in OL from the symbols in OL. In essence we've given the syntax for OL. We could just have well used English as our object language (OL) and articulated the rules for generating grammatical English sentences. This of course would not only require quite a bit more effort, but would also require a complete specification of English surface grammatical structure. With these procedures in place for forming *wff*s in OL (and the definition of a sentence just provided), we can see that OL generates an infinite number of (uninterpreted or meaningless) sentences. For example, [Pres(x_1)] can be combined with [Pres(x_2)] using the procedure in (12.) above to construct ϕ such that ϕ is [Pres(x_1) $\vee_{ol}$ Pres(x_2)]. Repeated applications of this step, using the newly generated expression for ϕ and Pres(x_n) for ψ, generates an infinite number of disjuncts of the form [Pres(x_1) $\vee_{ol}$ Pres(x_2) $\vee_{ol}$... Pres(x_n)]. Thus we have a procedure for forming an infinitely large set of (meaningless) sentences.

The purpose of this exercise is to show that Tarski's procedure can indeed yield a recursively defined Tr-predicate for an infinitely large language. To assign these sentences satisfaction conditions one at a time is clearly not possible. What we need is a recursive procedure for assigning satisfaction conditions to the complex expressions of OL based on the satisfaction conditions for the atomic (or basic) expressions used to build larger ones. So we stipulate the satisfaction conditions for these atomic sentences using sequences of objects in the domain, where such sequences assign values to each open variable in any ML sentence. Doing so will allow us to then define the 'Tr( )' predicate for our OL, by reducing Tr-predication to satisfaction.

III. Recursive Definition of Satisfaction

Let σ be an ordered n-tuple of objects in the Domain, and let $\sigma(n)$ be the nth element of σ. Let σ' be a variant of σ such that σ' differs from σ in at most one element n such that $\sigma(n) \neq \sigma'(n)$. We can call σ' an n-variant of σ (or a σ-variant).

The following procedures specify the method of applying satisfaction conditions to ML sentences by way of OL sentences that make use of this sequence vocabulary.

1. σ satisfies [Pres(x_n)] *iff* $\sigma(n)$ = OBAMA $\vee$ $\sigma(n)$ = BUSH.[26]
2. σ satisfies [Male(x_n)] *iff* $\sigma(n)$ = OBAMA $\vee$ $\sigma(n)$ = MCCAIN $\vee$ $\sigma(n)$ = BUSH $\vee$ $\sigma(n)$ = GORE.
3. σ satisfies [$\neg_{ol}\phi$] *iff* $\neg$(σ satisfies ϕ).
4. σ satisfies [$\phi \vee_{ol} \psi$] *iff* σ satisfies $\phi \vee \sigma$ satisfies ψ.
5. σ satisfies [$\forall_{ol} x_n \phi$] *iff* $\forall\sigma'$ (σ' satisfies ϕ)

DEFINITION OF TRUTH

A sentence ϕ of OL satisfies Tr(ϕ) *iff* ϕ is satisfied by every sequence σ.

The predicate 'Tr( )' gets defined in terms of satisfaction. A sentence is true just in case all sequences of individuals in the domain satisfy the sentence. Sequences, as ordered n-tuples, assign individuals in the domain to (constant and non-constant) variables in sentences. For example, the procedures indicated above in II for forming *wff*s in OL could yield the following *wff*:

(31) Pres(x_1)

We know from clause III(1.) that (31) is satisfied by some sequence just in case the nth element of that sequence is either OBAMA or BUSH. A sequence is merely an ordered list of individuals in the Domain, used repeatedly. As such, even with our small domain, there are an infinite number of sequences. Of all the (infinite) ways of ordering the domain (using repeated instances of domain elements as needed), we can divide these various sequences into four types: those that begin with each element respectively.

(a) <OBAMA,. . . >

(b) <MCCAIN,. . . >

[26]For our purposes we can ignore the intuitive appeal to time in office in understanding what it means for someone to be a president.

(c) <BUSH,. . . >

(d) <GORE,. . . >

That is, for any way of arranging the domain, the first element of any sequence must be one of the four elements of the domain. Since III(1.) tells us that (31) is satisfied (were it a sentence) by any sequence whose first element is either OBAMA or BUSH, we know that each of the infinite number of sequences characterized by (a) and (c) will satisfy (31). After all, (31) contains only one variable that can receive an assignment from a sequence, which in this case will always be the first element, as indicated by the subscript. Likewise, we know that the sequences indicated by (b) and (d) will not satisfy (31) since, irrespective of the other elements of these sequences, the first element fails to satisfy (31) (i.e. it is neither OBAMA nor BUSH). Thus we know that not all sequences satisfy (31), namely those sequences indicated by (b) and (d). Thus (31) does not satisfy the Tr-predicate, and is thereby no true.

Taking this same strategy and applying it to sentences with constant variables, consider the following sentence generated by OL:

(32) $\text{Pres}(\text{x}_{Obama})$

Like (31), we know that (32) is satisfied by a sequence if the nth element of that sequence is either OBAMA or BUSH. In this case, replacing n in $\sigma(\text{x}_n)$ with the appropriate variable subscript would require that the *Obama*-th element of the sequence satisfy (32). Obviously, there is no natural number OBAMA such that some individual from the domain can be that numbered member of the sequence. But, we can build sequences however we choose. So, if the OL for which we're generating a 'Tr( )' predicate has constants, we can build these feature into the sequence, such that all sequences assign the same individual to those constants. In effect, we can stipulate an *Obama*-th member for all sequences (and likewise for our other three constants).[27] Building this feature into our sequences has the result of ensuring that *every* sequence will assign OBAMA to x_{Obama}.

Recall that $\text{Pres}(\text{x}_n)$ is satisfied by a sequence just in case that sequence assigns either OBAMA or BUSH to x_n. So, if every sequence assigns OBAMA to x_{Obama} then

[27]This is not quite right, at least not if the language has quantifiers like $\forall$. Recall that the satisfaction conditions for a sentence with a universal quantifier appeal to n-variants of sequences. If we take the n here to range over all elemental positions in a sequence, this will include the *Obama*-th element as well. But there are no *Obama*-variant sequences precisely because every sequence must have Obama at the *Obama*-th position.

every sequence satisfies (32), regardless of what the other elements of the sequence are. Given our definition of the Tr-predicate, (32) turns out to satisfy Tr-predicate and is thus true because it is satisfied by every sequence.

Sentences like (31) and (32) in OL represent atomic sentences, which *via* the procedures specified earlier permits the construction of more complex interpretable sentences. The procedure indicated by II(12.) permits the construction of

(33) $\text{Pres}(\text{x}_{Gore}) \vee_{ol} \text{Pres}(\text{x}_{McCain})$.

The satisfaction conditions for sentences in the form of (33) are given in III(4.), which indicates that (33) is satisfied by a sequence σ if either part of the disjunct (the expressions on either side of $\vee_{ol}$) are satisfied by σ. In (33) both of these expressions are in the form $\text{Pres}(\text{x}_n)$. Again, III(1.) tells us that '$\text{Pres}(\text{x}_n)$' is satisfied by a sequence σ if the nth element of that sequence is either OBAMA or BUSH. If all sequences assign x_{Gore} and x_{McCain} GORE and MCCAIN respectively (in the manner suggested above) then no sequences satisfy either disjunct in (33), since neither GORE nor MCCAIN are BUSH, and neither GORE nor MCCAIN are OBAMA (as required by III(1.)). Cases like (33) illustrate how Tarski's method can be applied to languages (meeting certain conditions) to construct a truth predicate for an infinitely large language from a finite set of procedures and definitions. Iterations of disjunction can be handled recursively in the manner (33) is, given the procedures specified above.

Tarski's method can be applied to first order quantification as well. With the notion of a *sequence variant*, which classifies a set of sequences, we can use satisfaction to capture to the truth-conditions of (some) quantifier sentences. OL permits the generation of

(34) $\forall \text{x}(\text{Male}(\text{x}_1))$

We know from definition III(5.) that (34) gets satisfied by some sequence if all of its n-variants satisfy $(\text{Male}(\text{x}_n))$. As before, we can divide the set of sequences into four groups, based on their first element:

(a) <OBAMA,. . . >

(b) <MCCAIN,. . . >

(c) <BUSH,. . . >

(d) <GORE,. . . >

So, consider any sequence σ of kind (a), where the first element of the sequence $\sigma(1)$ = OBAMA. To be an n-variant of σ, where n is the first element, is to be a sequence just like σ, but with (at most) a different individual from the domain in the first elemental positions. For example, the sequence <OBAMA, MCCAIN>, has exactly four 1(st position)-variants: <OBAMA, MCCAIN> ; <GORE, MCCAIN> ; <BUSH, MCCAIN> ; <MCCAIN, MCCAIN>. Sentence (34) is satisfied by the sequence <OBAMA, MCCAIN> just in case these four variants satisfy (34). Appealing to III(2.), 'Male(x_n)' is satisfied by a sequence σ *iff* σ(n) = OBAMA or MCCAIN or GORE or BUSH. Indeed, for the four 1-variants of σ their first element ($\sigma'(1)$) is one of these four men. We repeat this process with *every* sequence to determine whether (34) is true. And indeed, for any sequence σ, and for any n, (σ'(n)) will be either OBAMA, MCCAIN, GORE, or BUSH. The only way a sequence variant can vary is regarding the individual in the relevant elemental position. Of the individuals that can occupy those positions, all of them satisfy 'Male(x_n)'. Thus, each sequence σ satisfies (34) because each variant σ' satisfies 'Male(x_n)'. This method gives us the intuitive result we want for the meaning of '$\forall$'. After all, the intuitive meaning of (34) is something like, 'everything (in the domain) is male'. And indeed, all the individuals in the domain (our four politicians) are male.

I've spent considerable space detailing the technical aspects of Tarski's theory of truth for two related reasons. The first reason is that such a theory underwrites modern semantic theories in both linguistics and philosophy of language. Insofar as the core semantic notion for these programs is the Tarskian notion of satisfaction, these theories are committed to an externalist semantics—one in which objects in a domain serve a nearly exhaustive role in sentential meaning. The second reason for the detailed description of Tarski's proposal is to lay the foundation for the contribution of Chruch (1951) to modern semantics. In particular, Church gives us a formal means of indicating the *denotation* of sentential constituents. This matters for the purpose of natural language semantics insofar as a mere Tarskian semantic conception is not compositional (in any traditional sense). For Tarksi, the only elements of OL that have interpretations (satisfaction conditions) are sentences, or well-formed formulas. The constituents of expressions in OL have no meanings independent of their usage in a sentence. The string of symbols 'Pres( )' has no

satisfaction condition (and hence no independent meaning), even if 'Pres(x_1)' and 'Pres(x_{McCain})' do. Thus there is no (traditional) sense in which the meaning of 'Pres(x_{McCain})' is *composed* of the distinct meaningful contribution of 'Pres( )' and 'x_{McCain}'.

As noted earlier in this section, one of the primary motivations for a Davidsonian truth-conditional semantics is to account for the compositionality of natural language. Since languages contain an infinite array of expressions, accounting for the human capacity to learn a language requires a semantics by which one can understand a finite number of expression meanings and a finite number of rules that dictate the manner in which the meanings of those basic expressions *compose* to form the meanings of more complex (sentential) expressions. Tarski's semantics cannot deliver on this demand. What's required, and what Church gives us, is a means for identifying the contribution of sub-sentential expressions to the meaning of sentences, by way of the Tarskian notion of satisfaction. Given that Church's logic serves as the foundation for contemporary semantics, this connection between Church and Tarski is important for understanding the commitments of a truth-conditional theory of meaning. Both Tarksian truth, and Church's notion of denotation are defined in terms of satisfaction, and satisfaction is defined in terms of a domain of objects. Thus, the ontological status of this domain will determine both the nature of truth and denotation for contemporary semantic theories. This matters for the Realist whose methodology makes use of this connection between meaning, truth, and denotation. In particular, this is why a Realist is committed to ($\mathcal{E}$), and thereby committed to an externalist construal of a truth-conditional semantics. The notion of truth at play in contemporary semantics cannot be the relative notion of truth-in-a-model. Meanings for sentences in OL tell us about the occupants of the ML domain. For the Realist methodology to succeed, as we'll see in Chapter 2, this domain must simply be occupied by the objects of reality, not merely some components of a model. In the next chapter I flesh out this point in greater detail, but for the purposes of this chapter, highlighting the connections between Tarkian truth and Church's notion of denotation helps to illustrate why contemporary truth-conditional semanticists are committed to ($\mathcal{E}$).

1.2.4 Church

Church's expansion of Tarski's logic provides a way of making expression meanings compositional. We've already seen a semantic proposal that does this, namely Frege's. Frege's distinction between saturated and unsaturated senses provided a means for composing the meanings of sub-Thought components to form Thoughts that denote truth-values. Frege analogizes unsaturated senses to mathematical functions, both of which require denoting expressions to form (a more complex) denoting expression. Church provides us with a logic that gives a formal characterization of unsaturated senses, with the denotations of these logical counterparts being functions. The result is a formal semantics that can characterize Fregean senses as functions, making use of nothing more than objects in a domain. In this section I'll illustrate how Church is able to build a notion of denotation from the Tarskian notion of satisfaction, by adding the λ-operator, some rules governing its use, and some additions to the domain. My review here has a fairly limited purpose, to show that the notion of satisfaction is at the heart of Church's additions that make a compositional semantics for a formal language possible. For that reason, my review will not follow Church too closely, and ignores details that might be important to logicians, historians, and some semanticists. However, this abridged detailing of Church's contribution indicates the fundamental relationship between Tarki's contributions and Church's regarding meaning and truth.

The goal for this section is to show how we can build a semantic notion of *denotation* from Tarski's notion of satifaction. Building off of our Tarskian example, the first addition we need to posit to OL for building a notion of denotation powerful enough to indicate the semantic contribution of predicates (like 'Pres( )') to sentences that contain them (like 'Pres(x_1)') is to add to the domain two truth values,[28] which we'll distinguish from the rest of the domain as follows:

[28]These values need not be ontologically mysterious, housed in Frege's Third Realm. They can simply be objects in the (Tarskian) domain that we simply designate as having a special status. To prevent confusion, I avoid taking that nominalist route here.

DOMAIN
$\underline{< e >}$
Obama
McCain
Bush
Gore
$\underline{< t >}$
$\top$
$\bot$

The addition of domain items of type $< t >$ (as well as a rule that indicates that variables that range over $< e >$-entities in the domain) permits us to build a notion of denotation for sentences in OL, in terms of the satisfaction:

$\mathcal{D}_t$ $\mathcal{D}$(S, σ)=$\top$ if $\mathcal{S}$(S, σ); otherwise $\mathcal{D}$(S, σ)=$\bot$

where $\mathcal{D}$ indicates the notion of denotation, and $\mathcal{S}$ indicates the notion of satisfaction. In (something closer to) English, the above reads as follows:

> Given a sequence, a sentence S denotes the value $\top$ if, for that sequences σ, σ satisfies S; otherwise the denotation of S relative to σ is the value $\bot$.

Since sentences in OL denote $< t >$-entities in the domain, we will call them expressions of type $< t >$, indicated with subscripts when needed.

For some sentences, namely those that satisfy the Tarskian truth predicate 'Tr( )', they will always denote $\top$. After all, to satisfy the predicate 'Tr( )' a sentence must be satisfied by *all* sequences. So, for any given sequence, such sentences will always be satisfied, and thereby always denote $\top$, irrespective of the chosen sequence. To use an example, take the OL sentence 'Pres(x_{Obama})'. Plugging in 'Pres(x_{Obama})' for S yields:

> $\mathcal{D}$(Pres(x_{Obama}, σ))=$\top$ if $\mathcal{S}$(Pres(x_{Obama}), σ); otherwise
> $\mathcal{D}$(Pres(x_{Obama}), σ)=$\bot$

Because of the way we built sequences, every sequence has as the *Obama*-th element, the object OBAMA. Thus every sequence σ assigns OBAMA to 'x_{Obama}'. That is, for all sequences σ, σ(x_{Obama})=OBAMA. As such, because III(1.) indicates the satisfaction condition for expressions of the form 'Pres(x_n)'—it is satisfied by a sequence with either OBAMA or BUSH in the n-th position—we know that for any sequence σ:

$\mathcal{D}(\text{Pres}(\text{x}_{Obama}), \sigma)=\top$

As we can see, the denotation definition given in ($\mathcal{D}_t$), derived from the Tarksian notion of satisfaction, gives us a way of indicating the meaning of expressions of type $< t >$.

Additionally, this explanation suggests a means of giving the denotations for the OL analogs of directly referring (natural language) expressions. While Tarski's system does not formally distinguish variables from constants, we saw that we can treat constants as variables, whose values do not vary from sequence to sequence, by putting constraints on the construction of sequences. Thus, in much the way that 'Barack Obama' always refers to some particular individual, 'x_{Obama}' is assigned OBAMA by *every* sequence, and thereby always denotes OBAMA. We can define a denotation notion for the sub-sentential expressions (like 'x_{Obama}') as follows:

$\mathcal{D}_e$ $\quad \mathcal{D}(\text{x}_n, \sigma)=\sigma(\text{x}_n)$

Even though the notion of denotation is given relative to a particular sequence, for variables like 'x_{Obama}', every sequence assigns OBAMA to 'x_{Obama}' because the *Obama*-th element of *every* sequence is OBAMA—after all, that's how we built sequences. While the denotation of variables will vary depending on the given sequence, the values of unvarying variables (or constants) like 'x_{Obama}' will be fixed. As such:

$\mathcal{D}(\text{x}_{Obama}, \sigma)=\sigma(\text{x}_{Obama})=\text{OBAMA}$

With denotation notions defined for the basic types of expressions, that is those that denote objects in the domain of either type $< e >$ or type $< t >$, we have a complete denotational semantics for all the expressions in OL (ignoring for now those that include the symbols '$\forall_{ol}$', '$\vee_{ol}$', and '$\neg_{ol}$'). However, such a semantics is still not compositional. While we have a denotation for expressions like 'Pres(x_6)' and expressions like 'x_6', we have no way of indicating the contribution of the latter's denotation to that of the former. Put another way, we have an idea what 'x_6' contributes to the expression 'Pres(x_6)'—namely, the object in the 6th position of a given sequence. But, we have no way of indicating the contribution of the remaining symbols in 'Pres(x_6)' to the expression's denotation. What we need is a formal analog to an unsaturated sense—something that indicates the semantic work being done by 'Pres( )'. But because 'Pres( )' is not an expression in OL, any more

than '(' or 'x_3 $\forall\neg$' are, we cannot give a definition for the denotation of such an expression. What we need in order to characterize the Fregean notion of saturation is a denotation for a formal proxy of unsaturated predicates like 'Pres( )', and show how combining the denotation of that 'Pres( )' proxy with the denotation of $< e >$-expressions like 'x_6' yields the denotation of $< t >$-expressions like 'Pres(x_6)'. In order for OL to have a compositional semantics then, we need to add to the list of OL expressions analogs to 'Pres( )', and a syntactic rule for combining them with expressions like 'x_6' to form *wff*s like 'Pres(x_6)'. So we add the following to II. above:

II. Definition of Well-Formed Formula (*wff*s in OL)

14. If ϕ is a *wff*, and contains a variable v of type $< e >$, then $\lambda v.\phi$ is a *wff* of type $< e, t >$.

15. If ϕ is a *wff* of type $< e, t >$, and contains a variable v of type $< e >$, then $\phi(v')$ is a *wff*.

16. If $[\lambda v.\phi](v')$ is a *wff*, $[\lambda v.\phi](v') \equiv \phi[v|v']$

These additions accomplish three purposes. First, the rule in II.(14) introduces the λ-operator, which turns any expression of type $< t >$ into an expression of type $< e, t >$. The idea is that any λ-expression will denote a function of type $< e, t >$, that maps $< e >$-entities to $< t >$-entities. Second the rule in II.(15) gives us a syntactic rule for combining expressions of type $< e >$ with expressions of type $< e, t >$ to form *wffs* (of type $< t >$). Third, II.(16) indicates the significance of the λ-operator in characterizing the analog of Fregean saturation. This rule indicates an equivalence between a kind of λ-expression (permitted by applying both II.(14) and II.(15)) and other expressions of type $< t >$. Given an expression of type $< e, t >$ and an expression of type $< t >$, these rules indicate that those expressions can syntactically compose into an expression of type $< t >$. For any expression ϕ of type $< t >$ these three rules permit an operation, whereby the variable (v) in ϕ can be bound by the λ-operator to form an expression of type $< e, t >$ ($\lambda v.\phi$). This $< e, t >$ expression can then be combined with an expression of type $< e >$ (some variable v') to form a new unbound variable expression of type $< t >$ in which we replace all occurrences of v with v'. This process of binding and unbinding (of λ-

abstraction and reduction) gives us a formal analog of (un)saturation, that permits a compositional semantics for OL.

Recall that Frege's notion of (un)saturation is analogous to the notion of a mathematical function. A saturated Thought that denotes a truth-value is the result of taking an unsaturated sense, and saturating it with a saturated sense that denotes an object. Put crudely, if you give an unsaturated sense a saturater of the right kind it gives you a saturated sense. This unsaturated sense is like a function that takes saturated senses of one kind and maps them to saturated senses of another kind. We introduced a distinction in the domain to approximate the Fregean idea that there are two kinds of saturated senses, drawing a distinction in the domain between objects of type $< e >$ and objects of type $< t >$. We also gave definitions for the kinds of expressions that denote these domain objects. The idea then is to treat unsaturated denotations as *functions* that maps domain objects of type $< e >$ to domain objects of type $< t >$: functions of type $< e, t >$.

The process of λ-abstraction (just described) makes this possible. But now we need to articulate the notion of denotation for such a λ-expressions(abstractions):

$\mathcal{D}_{et}$ $\mathcal{D}(\lambda \mathrm{x}_n.\phi_{<e,t>}(\mathrm{x}_n), \sigma)$ = the smallest function F of type $< e, t >$ such that for each σ' such that σ' is a variant of σ, F maps $\mathcal{D}(\mathrm{x}_n, \sigma')$ to $\mathcal{D}(\phi_{<e,t>}(\mathrm{x}_n), \sigma')$

The idea here is that the λ-operator serves as a kind of prefix that changes an expression of type $< t >$ that lacks a constant denotation, into a denoting expression that has a *constant* denotation—a function. An example using OL will be illustrative here. Consider the OL expression:

(35) $\mathrm{Pres}(\mathrm{x}_1)$

The OL expression in (35) is satisfied by some sequences, and not by others (see III above). Again, given our domain there a four different varieties of sequences that matter for satisfying expressions like (35) simply because they contain only one variable:

(a) <OBAMA,. . . >

(b) <MCCAIN,. . . >

(c) <BUSH,. . . >

(d) <GORE,...>

Since there is only one variable in (35), any domain objects coming after the first element in any sequence are immaterial for purposes of satisfaction. Depending on which sequence we chose, (35) will be satisfied or not. Since the notion of denotation is relativised to satisfaction by sequences, the expression in (35) will denote $\top$ given sequences of variety (a) and (c), and denote $\bot$ given sequences of variety (b) and (d).

The problem we encounter with expressions like (35) is formally characterizing the inconstancy of the expression's denotation. Depending on which object is in the 1-th(1st) position in a sequence, (35) will denote a different object ($\top$ or $\bot$). For the meanings of expressions like (35) to be compositional, the contribution of the non-variable part of the expression must have a constant denotation that, when combined with a variable expression yields the proper denotation. The additions to OL above give us an expression that can bear this denotation. Applying II.(16) to the expression in (35), we get the following λ-expression:

(36) $[\lambda \mathrm{x}_1.\mathrm{Pres}(\mathrm{x}_1)](\mathrm{x}_1)$

Since this is equivalent to (35) we can (intuitively) tease apart the significance of the variable expression, and the non-variable portions of (35). Eliminating the variable expression in (36) results in:

(37) $\lambda \mathrm{x}_1.\mathrm{Pres}(\mathrm{x}_1)$

This expression (intuitively) captures the significance of the non-variable symbols in (35). And the denotation definition given above in ($\mathcal{D}_{et}$) provides us with a procedure for determining the denotation of this expression. Substituting '$\lambda \mathrm{x}_1.\mathrm{Pres}(\mathrm{x}_1)$' therein produces:

> $\mathcal{D}(\lambda \mathrm{x}_1.\mathrm{Pres}(\mathrm{x}_1), \sigma)$ = the smallest function F of type $< e, t >$ such that for each σ' such that σ' is a variant of σ, F maps $\mathcal{D}(\mathrm{x}_1, \sigma')$ to $\mathcal{D}(\mathrm{Pres}(\mathrm{x}_1), \sigma')$

For any sequence σ we might choose, all of the possible variants are characterized by (a)–(d) above. Since a sequence variant is defined as a sequence that varies in *at most* one position, variations in the first elemental position exhaust all the satisfaction-relevant sequence variants of any given sequence. As such, the only

four possible satisfiers are the four elements of the domain. The smallest function that maps each of these four elements of type $< e >$ (of the four sequence variants) to the type $< t >$ denotation of the sentence 'Pres(x_1)' for those four sequences is the following set:

(38) $$\left\{ \begin{array}{c} < \text{OBAMA}, \top > \\ < \text{MCCAIN}, \bot > \\ < \text{BUSH}, \top > \\ < \text{GORE}, \bot > \end{array} \right\}$$

This function serves as the constant denotation of the λ-expression '$\lambda x_1.Pres(x_1)$'. Analogous reasoning guides the construction of the denotation for the other predicate in OL, namely 'Male( )' and the formal analog 'λx_1.Male(x_1)':

$\mathcal{D}(\lambda x_1.\text{Male}(x_1), \sigma)$:

$$\left\{ \begin{array}{c} < \text{OBAMA}, \top > \\ < \text{MCCAIN}, \top > \\ < \text{BUSH}, \top > \\ < \text{GORE}, \top > \end{array} \right\}$$

Thus our definition yields an invariant denotation for each predicate analog in OL, both of the type $< e, t >$.

With a definition of denotation ($\mathcal{D}$) for each type of expression in OL (again ignoring for now those that include the symbols '$\forall$', '$\vee$', and '$\neg$'), we are in a position to define a semantic rule that characterizes a way of composing the denotations of sentences out of the denotations of their constituents. Namely:

(FA) $\mathcal{D}(\Gamma[\phi_{<e,t>}\psi_{<e>}], \sigma) = [\mathcal{D}(\phi_{<e,t>}, \sigma)](\mathcal{D}(\psi_{<e>}, \sigma))$

This semantic rule indicates how the denotations of expressions in OL can combine to form the denotation of a complex expression syntactically composed of them. In more plain language (FA) reads:

> For two expressions in OL ϕ and ψ that can syntactically compose in accordance with the grammar (Γ) of OL, the denotation of the expression that results from syntactically composing ϕ and ψ is the result of applying the denotation of ψ to the functional denotation of ϕ.

By way of example, we can show how this rule works. Take the example OL expressions

(39) $\lambda x_1.\ \text{Pres}(x_1)$

(40) x_{Obama}

of type $< e, t >$ and $< e >$, respectively. These two expressions can syntactically compose to form the expression

(41) $[\lambda x_1.\ \text{Pres}(x_1)](x_{Obama})$

According to (FA) the meaning of (41) should be composable from the meanings of the expressions (39) and (40). That is:

$$\mathcal{D}(\lambda x_1.\ \text{Pres}(x_1)(x_{Obama})), \sigma) = [\mathcal{D}(\lambda x_1.\text{Pres}(x_1)_{<e,t>}, \sigma)](\mathcal{D}((x_{Obama})_{<e>}, \sigma))$$

As we have already demonstrated, we know that for any sequence σ the denotation of 'x_{Obama}' is always OBAMA. That is:

$$\mathcal{D}((x_{Obama})) = \text{OBAMA}.$$

Simplifying a bit, we can reformulate the denotation equivalence above as:

$$\mathcal{D}([\lambda x_1.\ \text{Pres}(x_1)](x_{Obama})), \sigma) = [\mathcal{D}(\lambda x_1.\text{Pres}(x_1)_{<e,t>}, \sigma)](\text{OBAMA})$$

As detailed above we also know that the denotation of (39) is a function, as a set of ordered pairs indicated in (38). As is clear, when given OBAMA as an input that function returns $\top$. Simplifying some more, this yields:

$$\mathcal{D}([\lambda x_1.\ \text{Pres}(x_1)](x_{Obama})), \sigma) = \top$$

We know, given the rule in II.(16) that the remaining λ-expression is equivalent to '$\text{Pres}(x_{Obama})$', which yields:

$$\mathcal{D}(\text{Pres}(x_{Obama}), \sigma) = \top$$

Lastly, given the satisfaction conditions of '$\text{Pres}(x_{Obama})$' we know that it is satisfied by all sequences, and hence denotes $\top$—which is precisely what the equivalence above reflects. What we have shown then is that the denotation of the expression '$\text{Pres}(x_{Obama})$' is composed of the denotation of 'x_{Obama}' and the denotation of the result of λ-abstracting the expression '$\text{Pres}(x_1)$'. This λ-abstraction is meant to

formalize the semantic contribution of 'Pres( )' in a formally respectable way, such that it can compose with the semantic contribution of 'x_{Obama}'. This same process can be repeated for every expression in OL (again, ignoring those that include the symbols '$\forall$', '$\vee$', and '$\neg$'), giving us a compositional semantics for OL.

Of course, OL is a rather simple language. If OL were expanded to permit not merely constants (like 'x_{Obama}'), variables (like 'x_1'), and predicates (like '(Pres( )'), but also to include relations (like 'Hates(,))' our additions to OL would be insufficient for giving a compositional semantics that would account for expressions like 'Hates(x_1,x_2)'. This failure would persist even if we included clauses that detail the satisfaction conditions for the hating relation. Relations, after all, would require two objects to satisfy them, reflected in the addicity of 'Hates(x_1,x_2)'.

Thus far, our additions to OL only give a compositional semantics that tells us how to combine expressions with a single variable with its would-be satisfiers. They indicate how expressions of type $< e, t >$ compose with expressions of type $< e >$. The λ-abstraction of 'Hates(x_1,x_2)' yields an expression which can be combined with a constant expression like 'x_{Obama}', but the result ('Hates(x_{Obama},x_2)') would still have an inconstant denotation. In short, what is needed is a way of generating expressions that correspond to all the possible functions from various entities in the domain, to other entities in the domain, including the entities denoted by λ-expressions. We've given a means of constructing such expressions to denote functions that map $< e >$-entities to $< t >$-entities (and added their functional-denotations to the domain). To accommodate relations like 'Hates(,)', and any other function that maps to and from functions, requires a systematic hierarchy of functions and the expressions that denote them. This is particularly needed if the language for which we are giving a semantics is a natural language like English, with relational notions like 'give' and quantifiers like 'few'. Resolving this problem requires both a generalization of the additions to OL detailed in this section, a notion of denotation that takes cues from Schönfinkel (1924), and rules for generating expressions to denote functions at any level of the functional hierarchy.

For our purpose here the details of this expansion of OL are not anymore illustrative than the small addition detailed thus far, so I'll spare the reader any further pedantry. This more involved expansion does what our incomplete example has done, namely adopt Church's addition to Tarski's semantics by giving a notion of denotation for every sub-sentential expression in OL. This addition provides

the denotations of sentential constituents that compose to yield the denotations of the sentences they constitute. More compactly, Church shows us how to build an (expressively powerful) language with a compositional semantics given in terms of satisfaction by sequences.

At the core of Church's semantic notion is Tarskian satisfaction. Tarskian satisfaction conditions serve to ground his theory of truth. This theory of truth is taken to be the core semantic notion for Davidson and Montague, and the contemporary semantic theories that find their roots therein. Given that these theories take meanings to be truth-conditions, such theories embrace ($\mathcal{E}$). If meanings *are* truth-conditions, then expression meanings must determine the truth-conditions for expressions. Thus, contemporary theories of meaning that take their foundational cues from Davidson and Montague find themselves committed to ($\mathcal{E}$).

1.3 Conclusion

In this chapter we have seen that the historical foundations of contemporary semantics in Frege, Russell, Tarski, and Church are markedly externalist, insofar as this foundational work is committed to the idea that the meaning of a linguistic expression determines its truth-conditions. For those semanticists that hold meanings to be Russellian propositions with real-world constituents, meanings determine truth-conditions insofar as propositions must serves as the bearers of truth-values. Similarly, for semanticists that hold meanings to be Fregean propositions, composed of senses, meanings determine truth-conditions insofar as the senses of Thoughts are ways of presenting the truth-values. Lastly, for truth-conditional semanticists, the central semantic notion of satisfaction requires that meanings (as denotations) are defined in terms of objects in a domain. Not only are such theorists committed to the claim that meanings determine truth-conditions, given that they identify meanings with truth-conditions, but further the meanings of expressions determine the objects than populate the domain. This is of particular importance for the Realist who hopes to use linguistic meaning as a guide for metaphysical investigation. For such a Realist the domain of objects consists of just the objects that populate reality. Not only is the tradition of semantics in philosophy and linguistics externalist (insofar as these various views are committed to ($\mathcal{E}$)), but this externalist commitment is essential for the Realist's methodology. In the next chapter I detail this

relationship, outlining the Realist's commitments that will serve as the target of the remainder of the work.

Chapter 2

Realism, Regimentation, and Quine

Since Frege (1879 [1997]), semantics in linguistics and philosophy of language has been markedly externalist. Various specific views are on offer, treating meanings as propositions (Russell, 1912; Soames, 1984; Salmon, 1986; Stalnaker, 1984; Kaplan, 1977 [1989]; King, 2007), functions (Lewis, 1970; Thomason, 1974), or truth conditions (Tarski, 1944; Larson & Segal, 1995; Heim & Kratzer, 1998). These various views hold in common the claim that the meaning of an expression determines its truth-conditions, either by identifying meanings with truth-conditions, or by identifying the propositional content of a linguistic expression in a way that determines the truth conditions of the expression. Likewise, the meaning of a word determines the mind-independent objects that constitutes the word's extension. The previous chapter surveyed this history, drawing out the general externalist commitment between meanings and truth-conditions, as outlined in ($\mathcal{E}$).

In this chapter I will explicate the commitments of the Quinean Realist, particularly regarding her commitments to an externalist semantics for the languages deployed in her metaphysical methodology. The Quinean Realist is a metaphysician that adopts a particular metaontological view regarding the relationship between language and ontological investigation. In the next chapter I will argue that the commitments of this particular metaphysical methodology regarding natural language semantics are troubled. In Chapter 4 I'll suggest that these same troubling assumptions apply to the invented languages used to express scientific theories. In this chapter, I will simply argue that the Realist is indeed committed to one or both of the following claims: a) that natural languages have externalist semantics, and b) that the languages used to express scientific theories have externalist semantics.

Taking cues from Quine (1948), the Quinean Realist holds that the correct methodology for pursuing ontological questions involves investigating linguistic meanings. As the slogan goes, "to be is to be the value of a variable [in some language]" (Quine, 1948, p. 34). This position holds to the following tenets:

1. There is a mind-external world;
2. The mind-external world has the quantificational structure of a Tarskian logic;

3. There is a language (call it the language of ontology [$\mathcal{L}_{\mathcal{O}}$]) that mirrors the quantificational structure of the mind-external world;
4. The domain entities required to satisfy the true expressions in $\mathcal{L}_{\mathcal{O}}$ fill our ontology, and indicate "what there is."

These tenets support the following conditional: if there is a mind external world with the quantificational structure of a Tarskian logic, and there is a language that mirrors this structure ($\mathcal{L}_{\mathcal{O}}$), then a metaphysician can explore ontological questions by understanding the externalist meanings of expressions in that language ($\mathcal{L}_{\mathcal{O}}$).

I take the first tenet to be unassailable. The second and third claims are less so, and recent debates in the metaontological literature highlight this (see Hirsch, 2002, 2008; Sider, 2009). While I will not engage with these disputes, I'll say something briefly about the content of these tenets. Both mention structure. Here I take the notion of 'structure' to be quite permissive. In tenet 2, the use of 'structure' commits the Realist a fairly palatable claim. For example, Sider (2002) claims that

> ... inquiry will be guided by ... [an] assumption [that] modern logic's quantificational apparatus mirrors the structure of reality: I assume an ontology of *things*. Moreover, I assume that there is a single, objective, correct account of what things there are. (Sider, 2002, p. xvi)

I understand this as a minimal commitment, namely that the world contains objects, and that these objects can be ordered. In much the way Tarski took sequences of objects in a model theoretic domain as the satisfiers of expressions in an invented language, the Realist holds that ordered objects in the worldly domain can serve as satisfiers for $\mathcal{L}_{\mathcal{O}}$ expressions. Similarly, for tenet 3 the structure that $\mathcal{L}_{\mathcal{O}}$ must mirror is the thin notion, whereby this commits the Realist to holding that $\mathcal{L}_{\mathcal{O}}$ can be given a semantics in terms of *objects*.

Tenets 3 and 4 mention $\mathcal{L}_{\mathcal{O}}$ as *a* language that mirrors the (object) structure of reality. The proposal insists there is at least one language that requires all and only the objects of the world (as defined by Tenet 2) to serve as the elements of sequences that satisfy the (true) expressions of that language. In truth, there will be (for both trivial and non-trivial reasons) a *class* of languages for which this is true. From the perspective of the Realist, each of these languages will be equally good, for the purposes of ontological investigation. To take a trivial example, that we choose imperial, as opposed to metric, units of measurement in expressing quantities of stuff

has no ontological significance. Insofar as the units of measurement put the same conditions on the stuff being measured, our choice of units has no bearing on what is in the worldly domain. Measuring an object's extension in space *via* 'meters' as opposed to 'feet' has no impact on the kinds of objects that are extended in space. They are as long as they are. Thus a language that uses feet as opposed to meters will not differ in ontological commitment, at least as a consequence of this difference. If $\mathcal{L}_{\mathcal{O}}$ uses 'meter' as a unit of measurement, then there is an ontologically equivalent language that uses 'feet' to measure extensions. Similarly, there will be a class of such languages, all with identical ontological commitments, none of which will be preferable from the Realist's position. For ease of exposition, I'll gloss over this complication, and speak of a unique $\mathcal{L}_{\mathcal{O}}$, but nothing I say here will rely on this gloss.

The focus of this chapter is not to address the disagreement about these tenets, but rather to spell out the various commitments of the Realist regarding the relationship between natural languages, the languages used to express our scientific theories, and $\mathcal{L}_{\mathcal{O}}$.

In unveiling these commitments, I'll draw special attention to some of the tenets above and claims implicit in these tenets. First is the claim that $\mathcal{L}_{\mathcal{O}}$ has a quantificational structure that permits the application of a Tarskian semantics. Given this tenet, if the Realist hopes to make use of her understanding of some language (like English) for the purposes of ontological investigation, that language must be translatable into $\mathcal{L}_{\mathcal{O}}$. The majority of this chapter is devoted to illustrating why the mere regimentation strategy suggested by Quine (1948, 1960) is insufficient for ontological investigation through language. To that end, I'll indicate what assumptions the Realist must take on to bolster this strategy in pursuit of her metaphysical aims. Second is the claim that there is some identifiable list of (epistemic) values that distinguish $\mathcal{L}_{\mathcal{O}}$ from other candidate languages. Disputes about these values will determine the kind of Quinean Realism one adopts, and will mark the distinction, relevant for our purposes, between those that wish to use natural languages as ontological guides, and those that do not.

I'll begin by illustrating the notions of ontological commitment and regimentation to make clear the role these notions play in the ontological methodology adopted by Quine. I'll then indicate the ways in which this strategy will not suit the Realist's purpose. Bolstering this strategy commits the Realist to claims about

natural language that Quine is unwilling to accept, namely that languages have meanings, and that their meanings determine their ontological commitments. This commits the Realist to the claim that the languages used for ontological investigation have an externalist semantics. I'll conclude with a promissory note regarding the arguments in the remaining chapters.

2.1 Ontological Commitment, Translation, and Regimentation

2.1.1 Ontological Commitment

Consider the following English expression:

(42) Wilbur is a horse.

If I utter (42) in an assertive context, i.e., in a context in which my communicative intention is to aptly describe the world,[1] I seem to be committed to the following:

(43) Wilbur exists.

To put the matter slightly more technically, we might say that (42) entails (43), where what we mean by 'entails' is that someone who believes (42) to be true, ought to believe (43) is true. The intuitive explanation is that in order to say something (true or false) about Wilbur, Wilbur must exist in order for me to have said anything about him. But this thought famously runs into trouble when the relevant expressions use *empty names*, as in

(44) Pegasus is a horse.

While a speaker can truthfully endorse (44), she should not thereby believe that

(45) Pegasus exists.

As Quine (1948) explains, matters get worse if we attempt to make our ontological commitments clear, by asserting

[1] That I utter (42) in this context matters. What matters more is that natural languages are the product of a multiplicity of contextual environments. That is, humans use natural languages with many aims in mind, and only on some occasions is that aim a perspicuous description of the world (cf. Chomsky, 2000). We should not be surprised if a semantics for natural languages that attempts to reduce meanings to relations involving worldly-objects fails to capture the meanings competent speakers assign to expressions. As such, even when speakers use natural language expressions with the aim of speaking literally about the world, we should not likewise be surprised if they routinely fail, given that the tools at hand were not designed with that task in mind.

(46) Pegasus does not exists.

The reasoning above that explains the entailment between (42) and (43) seems to apply here as well. In order to predicate something of Pegasus, in this case something like the property of non-existence, Pegasus must exist. But surely Pegasus cannot both exist, and not exist.

One way out of this problem might be to appeal to ambiguity, and insist that we have subtly deployed two different words by using 'exists'. As the argument goes, there are two senses of 'exists', one that is roughly synonymous with 'is actual', and another sense that is roughly synonymous with 'is possible'. Thus when asserting (46) I am committed to the *possible* existence of Pegasus, but not his *actual* existence. The move here insists that while Pegasus possibly exists, he does not actually exist. Such a move seems to have the *prima face* unwanted admission of a new class of things into the world, namely *possibilia*. My talk of Wilbur made a commitment to concrete objects necessary. Likewise, my talk of Pegasus makes a commitment to *possible* objects necessary.

Suppose you and I disagree on this point. You hold that a commitment to *possibilia* is unwarranted, while I maintain that it is not, for the reasons just rehearsed. I indicate that while such a commitment might seem troubled, the commitment to *possibilia* is simply required if we want to make sense of the inferential relations humans make, as we saw with the case of Wilbur. Thus, if the meanings of our expressions and the intuitive entailment judgments we hope to capture require positing these odd objects, we should not balk at their existence.

However, you note that our talk is not merely limited to the domain of possibility. We talk about impossible scenarios as well. To use Quine's example, were I to assertively utter

(47) The round square table in Berkeley College does not exist

we would have to conclude, based on the reasoning I've endorsed thus far, that there is some *impossible* object of which I am predicating non-actual-existence. Thus, by assertively uttering (47) I am now committed to a third kind of object, namely *impossibilia*. Unlike the existence of *possibilia*, the existence of *impossibilia* seems much less palatable. While the thought of possible objects as existing in some sense is strained, the existence of an impossible object is incoherent. Thus, if you can explain the inference patterns we wish to capture in a way that does not

commit us to the bizarre (incoherent) existence of *impossibilia*, I should welcome that explanation.

2.1.2 Regimentation

Quine's suggestion for resolving our disagreement is that we need to *regiment* our talk into a language that is more transparent (than English) regarding ontological commitment. Briefly put, the solution is to introduce a language that is explicit regarding questions of ontology, and use that language to paraphrase ontologically opaque natural language sentences. Regimenting our natural language expressions in this way will help us to get clear about what exists, given the claims we want to make about things like Pegasus and round squares.

For Quine, regimenting our natural language talk into an ontologically transparent language will make clear that we ought not endorse these same ontological commitments to *possibilia* and *impossibilia* as we do for Wilbur. That is, the offending natural language expressions about Pegasus and round-squares, once regimented, will not have the troubling ontological commitments whose analogs we are happy to endorse given our claims about Wilbur (and their analogous regimentations). As such, we can side-step the entailments that led us to accept *possibilia* and *impossiblia* into our ontology. Put in a way closer to philosophical tradition, the idea is that natural language expressions have a logical structure not reflected in their (apparent) surface structure. The underlying *logical form* of a linguistic expression is distinct from the grammatical form of the expression. And it is this mismatch that led me astray (in our supposed disagreement) when thinking that *possibilia* and *impossibilia* exist. Consider again

(42) Wilbur is a horse.

I was led astray in thinking initially that the logical form of (42) was exemplified by the grammatical structure of the expression. That is, I assumed the logical structure of the expression was predicative, as in

(42_α) $\textsc{horse}(w)$

For (42) to be true there must be an object w in the worldly domain that satisfies the predicate 'HORSE'. Our disagreement arose when, considering the analogous form for (46), I argued that we were forced to conclude that there must be some worldly

object (the referent of 'Pegasus') that satisfies the predicate 'NOT-EXISTS'. This problem worsened when considering the implications regarding the round square table in (47).

However, if the logical form of a given linguistic expression differs from the (surface) grammatical structure of the expression, then we might be able to explain why we should infer (43) from (42), but not (45) from (46). As such, we would explain why one should believe that Wilbur exists, if one believes that Wilbur is a horse. And likewise, we would thereby explain why one should *not* believe that Pegasus exists if one believes that Pegasus does not exist.

Famously, the view that "denoting phrases" have an underlying quantificational structure that differs from their (surface) grammatical structure is defended by Russell (1905). To extend Russell's proposal, when I utter (42), the logical structure of this expression is not properly captured by a formula with simple predicate-argument structure. The thought underlying (42) more closely resembles

(42′) There is something that is Wilbur and is a horse, and it is the only (relevant) Wilbur.

This expression more plainly displays its quantificational structure, which we can regiment in the language of First-Order Logic as

(42_β) $\exists x.\textsc{wilbur}(x) \wedge \textsc{horse}(x) \wedge \neg\exists y.\ y \neq x \wedge \textsc{wilbur}(y)$

The Quinean proposal then is to make use of this language as a means of resolving our ontological disagreement. So long as you and I agree that (42) conveys what (42′) conveys, then we can also agree that (42) can be regimented with the expression in (42_β). That is, for the purpose of getting clear on what exists, we can agree to regiment the English language expression in (42) as the First-Order Logic expression in (42_β).

Treating denoting phrases in this way accounts for the felicity of the entailment from (42) to (43), and the infelicity of the entailment from (46) to (45). After all, the quantificational structure of (43), namely

(43_β) $\exists x.\textsc{wilbur}(x)$

is a conjunct in (42_β). Thus we can model the inference patterns of speakers as logical entailment, deriving (43_β) from (42_β) by conjunction elimination. The inference

from (42) to (43) is licensed by the the entailment of (42_β) to (43_β). Contrastingly, the inference from (46) to (45) is not felicitous.

(46) Pegasus does not exists.

(45) Pegasus exists.

Using the same regimentation strategy we agreed to with (42), we can regiment (46) and (45) into the language of First Order Logic. Again, with the expressed purpose of getting clear about ontology, we translate the English language expression into a language meant to capture its quantificational (and logical) structure, as follows:

(46_β) $\neg\exists x.\text{PEGASUS}(x)$

(45_β) $\exists x.\text{PEGASUS}(x)$

The explicit contradiction in the quantificational forms of these expressions accounts for the infelicity of believing that Pegasus exists as a consequence of assertively uttering (46), because (46_β) does not entail (45_β). This strategy captures the inferences we are apt to make for the constellation of sentences thus far considered, without burdening our ontology with commitments to *possibilia* and *impossibilia*.

2.2 Regimentation is Not Translation

For Quine, that we *choose* to translate (42) with (42_β) *for the purpose of ontological clarity* matters. After all, Quine denies that there are meanings. Concomitantly, he insists that for a given expression, there are no determinant facts about the merits of any particular translation of that expression in to another language (Quine, 1960, Ch. 2). As such, to think that any natural language expression simply has a single, privileged logical form, would mis-characterize Quine's view.[2] The logical form of a natural language expression *for a speaker* is nothing more than the expression in First-Order Logic which that speaker would think is true on those occasions that the to-be-regimented natural language expression is true. This is why we might regiment (42) with (42_β).

Of course we might not. The crucial difference between the English expression and the First-Order expression is the manner in which they treat 'Wilbur'. In

[2]This may overstate matters a bit, since Quine seems have been confused about this very point. See Quine (1948, pp. 24–28) where he endorses the theory of definite descriptions, insisting that *the meaning* of proper names is best analyzed by Russell's analysis.

the English expression, 'Wilbur' seems most obviously to be a denoting expression that simply picks out a certain object (a 'constant' in the terminology of First-Order Logic). In the First-Order expression, 'Wilbur' is treated as a predicate. Regimenting 'Wilbur' is this way treats it like other predicates, such as 'is mortal'. Canonically, such predicates are truthfully predicated only if the relevant object instantiates a particular property, like *mortality*. To treat 'Wilbur' is this way is to say that something is Wilbur just is case that thing instantiates a particular property, ostensibly *Wilburness* (whatever that is). One might think that this treatment of the name 'Wilbur' simply fails to convey the thought originally expressed by (42), precisely because it mis-characterizes the nature of the denoting expression as a predicate. Of course, if there is some *fact of the matter* about the meaning of denoting expressions like 'Wilbur' such that a quantificational treatment is the correct treatment, then Quine is justified in insisting that (42_β) is the correct regimentation of (42). But, as Quine understands regimentation there is *no* fact of the matter regarding the proper regimentation of any expression (in any language).

This suggests that Quine's proposal is insufficient for the aims of the Quinean Realist. For Quine, that a speaker accepts a sentence in First-Order Logic as a regimentation of some sentence in English, merely tells us that, given a situation, this speaker is disposed to accent to both expressions (or neither of them). All we learn from the process of regimentation is what an *individual speaker* takes themselves to be ontologically committed to when endorsing a natural language expression. But this could differ across speakers, and for Quine it seemingly must (Quine, 1960, p. 40). Thus on Quine's proposal, ontological commitment is speaker-relative. For the Quinean Realist, however, such a relativism will not do. Indeed, if the methodological goal of the Realist is to investigate ontological matters *via* her competence with a natural language, such a commitment requires that this competence is characterizable in non-agent-relative terms.[3] What the Realist understands must be characterizable in speaker-independent terms. That is, not only must the linguistic expressions that the speaker understands be characterizable in a speaker-independent way, but the nature of this understanding must also be characterizable in terms such that any speaker of the language has that same understanding.

[3]In Chapter 3 we'll see that this competence must be characterized in non-indexical terms as well. For example, the semantics suggested by Cappelen & Lepore (2005) and Stanley (2005) will not be sufficiently externalist for the Realist's ontological purpose.

If there is some fact of the matter about whether (42_β) is the correct regimentation of (42), and the Realist holds (*pace* Quine) that by understanding (42) she grasps this fact of the matter, then there must be some properties of (42) that makes this possible. Call those properties *semantic properties*. The Realist then is committed to the claim that the semantic properties of natural language expressions determine which expressions in First-Order Logic (or any other regimenting language) serve as viable regimentations. Positing this additional commitment, the Realist holds that (42_β) is the correct *translation* of (42), insofar as this paraphrase preserves the relations determined by the semantic properties of the natural language expression. Let us mark the distinction between regimentation and translation as follows: for a given linguistic expression, a *regimentation* of that expression into another language can be better or worse, *conditional to some purpose*; a *translation* of a linguistic expression can be better or worse as *determined by the semantic properties* of the expression.

Quine, of course, denies that there are any such properties, and seemingly denies that there are natural languages (Quine, 1960, p. 63). As such, for the Realist to adopt Quine's methodology, she must endorse two further claims: both that there are speaker-independent languages, and that those languages have semantic properties that determine the appropriateness conditions for translation (into $\mathcal{L}_\mathcal{O}$). More simply, the Realist must adopt a view that the meaning of a natural language expression must have a putative logical form. More over, given the examples above, this logical form seems to differ from the surface structure of that natural language expression.

As a means of homing in on the gap between Quine's regimentation strategy and the methodology the Realist requires, it will be helpful to compare this Quinean regimentation strategy regarding logical form with another, to see why a Realist must bolster Quine's strategy to serve her methodological demands. For Frege, thoughts have definitive logical forms, insofar as the logical relations between propositions, on his view, can only be attributed to the structure of thoughts. Frege's fundamental project was to invent a logic capable of expressing the axioms of arithmetic, and thereby ground mathematics in logic. Logic, on this conception, has as its proper subject the theories of the sciences, wherein the sciences are engaged in the task of describing the world and thereby expressing the truths about the world. The role of logic is to provide a suitably powerful language for expressing

these truths. Amongst these truths are the truths of mathematics. Of course, the truths expressed in mathematics are entailments, the consequences of derivations from accepted axioms. Thus, the form of these axioms must be expressible in a language where the truths they entail are derivable from the logical expression of those axioms, *via* the application of logical rules. Frege's contention is that the thoughts we have about mathematics must be expressible in a language in which the inferential relations between those thoughts can be derived in a proof.

Take for example Euclid's proof by *reductio* that there are infinity many primes. This truth of mathematics follows from the Peano axioms of mathematics coupled with a definition for 'number', and for 'prime'. In particular that there are infinitely many primes derives from the axiom that every number has a successor. Frege's insight is that, in order for the claim 'There are infinitely many primes' to follow from the claim that 'Every number has a successor' as a matter of logic alone, there must be a language capable of expressing both these claims. Further, this language must be such that the first of them can be shown to be derivable from the second by appeal to logical rules.

For Frege, the question of logical form pertains to the thoughts expressed by claims like 'Every number has a successor." But he is careful to distinguish, on the one hand, the manner in which the thought has come about in the individual thinking the thought, and how the truth of the knowledge contained in the thought is secured (Frege, 1879 [1997], p. 48). That is, the manner in which a human mind has come to think a particular thought may present that thought as having a particular structure. Euclid's insight that there are infinitely many prime numbers may have come to him in a dream, with the form of the dream-inspired thought having a particular kind of structure. But, there's no reason to suppose that the representation Euclid entertained has a suitable structure for derivation—in particular, for deriving it from the thought that every number has successor. We might even think that the thought Euclid entertains has a different structure for different (types of) minds. The justification for the truth of that thought, however, if it is to be secured in the most robust way, must be a consequence of the rules of logic. A justification of this kind may then require a thought with a different structure than the thought had by any particular (type of) mind. For Frege this justified distinguishing between the mere thoughts had by people like Euclid, and the mind independent Thoughts whose structure encodes the logical relations that hold between them.

Thus for Frege there is an ideal Thought expressed with 'Every number has a successor' whose correct logical form is that which captures the logical relations between that Thought and other Thoughts, like the one expressed by the claim 'There are infinitely many primes'. Euclid's thought about prime numbers, as a component of his psychology, is related to (but distinct from) this ideal Thought, whose logical form is borne out by an invented logic that expresses the relations that hold between Thoughts. Using this invented language, Frege proposed, we can illustrate the logical relations between Thoughts *via* derivation. Euclid's thought, of baroque human form, has a proper *translation* into an invented language that characterizes the structure of Thoughts. Thus, Thoughts have a particular logical from, codified in an invented language whereby the logical relations between these Thoughts can be illustrated by derivations, as applications of logical rules.

We can contrast this view with Quine's view articulated above. For Quine, regimentation of a language is an exercise for resolving disagreements between competing views about what exists. Thus the language used for regimentation is chosen with a particular purpose in mind, namely for perspicuously describing the world's ontology. This is why the regimenting language suggested by Quine is a Tarksian logic. Such languages are ontologically perspicuous, as we saw in the previous chapter. The meanings for expressions in those logics are given in terms of satisfaction conditions on sequences of domain objects. As such, expressions in these (logical) languages have meanings that directly bear on the question of what objects populate the worldly domain. Thus by regimenting a particular language into a Tarskian logic, we make clear the ontological commitments of any given natural language expression by identifying the satisfaction conditions of its synonymous Tarskian counterpart.

But for Quine, independent of that ontological purpose, there's no sense in which a natural language sentence has a particular, correct logical form. For Frege, Thoughts have definitive logical forms, and the job of logicians is to undercover these forms so that we can express the truths of sciences in a language that makes explicit, *via* derivation, the logical relations between Thoughts. But the nature of these forms is not determined by a commitment to any particular purpose, or what a particular language user holds to be a suitable translation. They are as they are. Their respective structures are just facts about the constitutions of individual Thoughts. On Frege's view, there is a fact of the matter about the logical forms

of Thoughts, ones we discover though logical investigation. To the degree that one thinks natural language sentences express Thoughts, such sentences must have semantic properties that determine which Thought the sentence expresses. Thus the meanings of natural language expressions[4] as Thoughts, have definitive logical forms.

In contrast with Frege's conception of logical form, the regimentation strategy understands logical form as dependent on the purpose of ontological investigation, and an individual speaker's disposition. This purpose not only constrains the space of possible regimenting languages to those that are ontologically transparent, but also renders those translations inept for ontological investigation—at least if we are concerned with what there is, and not merely with what a particular speaker thinks they are ontologically committed to when endorsing a claim. After all, if the Realist is interested in using natural language expressions as guides to answering questions about what exists in the mind-independent world, it is of little use for such a Realist to know that a particular speaker is disposed to assent to a pair of expressions (one in English and one in First Order Logic) in all and only the same conditions. That is, it is not helpful for the Realist to know that two such sentences are thought to be equivalent for a particular speaker. However, if the Realist knows that these two expressions *are* truth-conditionally equivalent, and not merely thought to be so by a particular speaker, then the Realist's understanding of a natural language can, *prima facie*, prove as a useful guide to ontological investigation.

This however is not a minor concession. The natural language Realist commits herself to the view that the semantic properties of a natural language expression determine the truth-conditions of that expression. If there is some fact of the matter about the truth-equivalence of (say) (42) and (42_β), and that in understanding (42) one understands this equivalence, then whatever it is that one grasps in understanding the natural language expression, that thing determines the truth-conditions of the expression. What follows is that a Realist who holds that understanding a natural language expression can yield insight into questions about "what there is"—*via* the truth-conditional equivalence of that expression and its translation—must also hold that the semantic properties of natural language expressions determine their

[4]As noted in the previous chapter, this is something of a departure from Frege, given his warnings against analyzing the meanings for non-ideal languages like English, and German in terms of Thoughts (Frege, 1891 [1997]).

truth-conditions (assuming that what one grasps in understanding a sentence are these semantic properties).

For the Quinean, we can draw three lessons from this exercise. First, to make claims about "what there is" by appealing to natural language, the Realist must posit what Quine does not; namely, semantic properties that determine an expression's putative logical form. Second, that the grammatical structure of natural language expressions must often radically mismatch the logical structure determined by their semantic properties, for the Realist. Meanings (ostensibly) have a different kind of structure than natural language expressions.[5] Third, only by positing this matter of fact mismatch can the Realist maintain the ontological methodology that one is committed to those objects that one quantifies over when assertively uttering expressions.

Applying these lessons, we can begin to investigate ontological questions. If we want to use natural languages as a guide for ontological investigation, we must first pair each expression of the target natural language with an expression in an invented language that mirrors the logical form of the natural language expression. Put another way, we must discover for each expression of a natural language what logical form is determined by its semantics properties, and then pair that expression with one in an artificial language that reflects that logical structure. We must *translate* (and not merely regiment) the relevant natural language into a suitably precise, and logically perspicuous language that preserves the semantic properties of the expressions in that natural language. From these translations we can look to see what objects we need in our domain to satisfy all the true sentences in our invented language.

However, this renders the third tenet on the Quinean Realist position a substantive claim. Recall the third tenet for the Realist:

(3) There is a unique language (call it the language of ontology [$\mathcal{L}_{\mathcal{O}}$]) that mirrors the quantificational structure of the mind-external world.

For Quine, we chose the language used in regimenting natural language expressions. Quine was free to choose this language at will, because there are no determinate facts about the suitability of a regimentation. All that was required was that the speaker

[5]See May (1978) for an attempt to relate the surface structure of natural language expressions to their logical forms.

assents to that regimentation (or not). But given the additional posits required by the Realist's aim, she is constrained in her choice of a translating language. The Realist endorses tenet 3 because she needs $\mathcal{L}_{\mathcal{O}}$ to have an externalist semantics. Again, if her method for ontological investigation is to use her understanding of natural language expressions to inform her of the inhabitants of the worldly-domain, she only succeeds if the semantics for $\mathcal{L}_{\mathcal{O}}$ can be given in terms of a domain of (worldly) objects. Thus she demands that $\mathcal{L}_{\mathcal{O}}$ has a structure that mirrors reality in order for the objects of the world to serve as satisfiers for expressions in $\mathcal{L}_{\mathcal{O}}$. Since these additional posits suppose that the semantic properties of a natural language expression determine its logical form, whether or not a particular (invented) language can express this form is an empirical matter. It may turn out that any particular natural language expression may not admit to a translation into the kind of language the Realist demands for her ontological methodology. That is, the Realist is committed to the empirical claim that natural languages are translatable into a Tarksian logic. These bolstering assumptions, of course, do not undermine the third tenet, but they do have consequences for a Realist interested in using natural language as an ontological guide. On the assumption that $\mathcal{L}_{\mathcal{O}}$ mirrors the structure of reality, the meanings of natural language expressions are only helpful in ontological investigation if they can be regimented into $\mathcal{L}_{\mathcal{O}}$. The bolstering assumptions that a Realist must adopt if they hope to use natural language as a guide for doing ontology leaves open the possibility that natural languages exhibit semantic properties that simply cannot be expressed in $\mathcal{L}_{\mathcal{O}}$.

2.3 Demands on the Language of Ontology

For the Realist hoping to use natural languages as an ontological guide, $\mathcal{L}_{\mathcal{O}}$ must meet two demands. On the one hand, $\mathcal{L}_{\mathcal{O}}$ must be capable of expressing the (semantically determined) logical forms of natural language expressions. On the other hand, $\mathcal{L}_{\mathcal{O}}$ must reflect the qualificational structure of reality. Meeting the first demand will involve empirical investigation into the meaning assignments speakers of a language assign to natural language expressions. The second demand highlights the need for a method by which we come to know (or discover) $\mathcal{L}_{\mathcal{O}}$, with its structure-reflecting properties.

On the assumption that the third tenet above is true, the Realist needs some

means by which to identify which language is the one that "cuts nature at its joints." As Quine sees the issue, the questions of ontology are questions about the "conceptual scheme" (or language) we choose to adopt. His proposal is that the language best suited for ontology is the "simplest" one:

> Our acceptance of an ontology is ... similar to our acceptance of a scientific theory, say, a system of physics: we adopt ... the simplest conceptual scheme into which the disordered fragments of raw experience can be fitted and arranged. (Quine, 1948, pp. 35–36)

As Quine explains in an example, we adopt a language whereby we talk of physical objects, as opposed to "raw" sensory experience, because the language of physical objects is simpler, while still accommodating the veracity and character of our phenomenal experience. A language of sensory experience will not posit physical objects as part of the constituents of the world, since such a language only requires phenomenological entities to satisfy expressions. But a language of physical objects will (to some suitable level of abstraction) be able to give translations for broad classes of expressions with use of a single physical object.

Consider our example sentence:

(42) Wilbur is a horse.

A language that only admits of phenomenological entities would require an infinitely large class of such entities as the viable candidate referents for the many expressions which could indicate that some phenomenal experience is horse-ish. That is, for all the perspectives one could take in seeing (smelling, tasting, hearing, etc.) Wilbur, a phenomenal language would have a sentence corresponding to these unique perspectives, with a phenomenal object corresponding to each perspective.[6] While such a language matches our phenomenal experience quite well, a language that trades in physical-object-talk translates this entire infinite class of sentences (with its vast array of phenomenological satisfiers), into a single sentence, with a single satisfier. Thus, the physical-object language is simpler, and more ontologically parsimonious.

So understood, this method is not that far from the current practice of linguists and philosophers of language in developing formal expressions that indicate

[6]Ironically, some of the more recent work in (event) semantics posits just this sort of language to account for the difference in meanings speakers assign to natural language expressions (cf. Schein, 2002). Quine's hope was that regimentation would banish talk of phenomenal experience in determining linguistic meaning, yet the consequence of his proposed methodology suggests this was just optimism.

the truth-conditions of expressions. In practice, linguists and philosophers of languages develop meta-languages, for which they can give an interpretation (given a properly populated domain) and pair a meta-language expression with a object-language expression, whereby the former conveys the meaning of the latter (in terms of objects). Theories that specify the translation from one language to another *via* procedures are better. Theories that give procedural translations and account for the meaning assignments given by natural language speakers are better still.[7]

This idea is central to the Quinean Realist position, namely that the utility of a language with regard to ontological investigation is related to some feature (like simplicity) which renders that language more or less successful at "carving nature at it's joints." Different Realists might disagree about which features make a language better suited for ontology than others.

As Sider (2002, 2009, 2011) has characterized the Quinean Realist position, such a Realist holds there is both an objective structure to reality (tenet 2. above), and a unique language that mirrors this structure:

> ... some candidate meanings 'carve nature at the joints' more than others, and it is part of the nature of reference and meaning that candidates that carve nature [closer to] its joints are more eligible to be meant. The meaning of a word, then, is the best candidate, where strength of candidacy is based on (1) fit with meaning-determining facts about the speaker or her linguistic community, and (2) intrinsic eligibility on the part of the candidate. (Sider, 2002, xxi)

Sider embraces the first of these criteria in defense of the metaphysical methodology that uses natural language meanings as guides for addressing ontological questions, holding that the uses of natural language expressions constrain the space of eligible translations for an expression's "joint-carving" meanings. The second of these criteria is less clear, but the idea seems to be that a translation that satisfies the first criteria may fail to match the structure of reality. Elsewhere Sider suggests that the languages used to express our best scientific theories are better suited for questions of ontology (Sider, 2011, p. 12). Ostensibly the criteria used to evaluate the merits of a scientific theory will do the work of determining which invented language is best suited for ontological investigation.

The two positions suggested by these remarks characterize the Realist I intend to target in the proceeding chapters. The first is the Realist that takes seriously the

[7]See Quine (1960) for such an attempt, esp. chapters 5 and 6.

judgments of competent speakers of a natural language in inventing her translating $\mathcal{L}_{\mathcal{O}}$. This Realist would, like contemporary linguists and philosophers of language, see her task a one of explaining the judgments of speakers by giving a formal account of the relations exhibited by these judgments, where the relevant relations bear on truth. Success for this Realist is measured both by the degree to which the formal model she offers explains the patterns of meaning assignment exhibited by competent speakers of a natural language, and by the degree to which this model mirrors the structure of reality.

The second position constructs $\mathcal{L}_{\mathcal{O}}$ based on the practice of theory building in the sciences. The substantive assumption endorsed by this Realist is that scientific methodology is of sufficient epistemic heft that the languages our best sciences construct are those that match the quantificational structure of the world. And because the languages used to state our best scientific theories are designed to perspicaciously describe the world, translation into a Tarskian language will ostensibly be a rather simple matter. Success for this second kind of Realist is measured by the degree to which scientific methods[8] deliver languages that mirror the structure of reality. In the proceeding chapters I hope to show that both of these positions run into trouble, insofar as the meanings of expressions in both the natural and scientific languages are not amenable to translation into $\mathcal{L}_{\mathcal{O}}$, at least not if $\mathcal{L}_{\mathcal{O}}$ has a externalist semantics.

What have we learned about $\mathcal{L}_{\mathcal{O}}$, and the how it relates to the Realist's aims? First, for the Realist, $\mathcal{L}_{\mathcal{O}}$ must have an externalist semantics if it is to mirror the quantificational structure of reality. This just is the Realist's third tenet. Second, if the Realist proposes to use her understanding of natural language expressions as guides to ontological investigation, then she is committed to an externalist semantics for natural language expressions. If $\mathcal{L}_{\mathcal{O}}$ has an externalist semantics (which it must given the second and third tenets), then in holding that there is a fact of the matter about which $\mathcal{L}_{\mathcal{O}}$ expression is the correct translations for a given natural language

[8] A Realist could of course endorse some other criteria by which we can determine which of the languages is $\mathcal{L}_{\mathcal{O}}$. She need not endorse the methodologies of science, and the languages they beget, as ontologically privileged. Of course, the burden is on such a Realist to show that their preferred methodology is some how epistemologically preferable. I take it that the reason most metaphysicians endorse the languages used to express our best scientific theories as epistemologically privileged is that the history of scientific investigation has proven fruitful in the conferring knowledge. For this reason, I take such a Realist to be (one of) the target(s) of this work.

expression, the Realist must hold that those two expressions are synonymous. Given that the meanings of expressions in $\mathcal{L}_{\mathcal{O}}$ are externalist, the natural language expressions that share those meanings thereby have an externalist semantics. Recall that treating translation a mere regimentation will not suffice for the Realist aims, since mere regimentation (of the stripe that Quine endorses) is speaker-relative. Discovering that a given speaker takes a natural language expression to be synonymous with an expression in $\mathcal{L}_{\mathcal{O}}$ tells us little about "what there is" unless we take the speaker's understanding of the natural language expression to be indicative of the *semantic properties* of that natural language expression. And, *a fortiori*, these judgments of synonymy are only helpful for the Realist if there is some fact of the matter about which $\mathcal{L}_{\mathcal{O}}$ expression is the proper translation for the given natural language expression.

For the natural language Realist, the forgoing discussion highlights two distinct conditions the Realist must place on $\mathcal{L}_{\mathcal{O}}$: *epistemic* conditions, and *semantic* conditions. The *epistemic* conditions determine the manner in which humans come to know or discover $\mathcal{L}_{\mathcal{O}}$. Granting the assumption in the third tenet (that there is a $\mathcal{L}_{\mathcal{O}}$ that mirrors the structure of reality), the Realist must commit to some criteria for determining which languages are more likely to be $\mathcal{L}_{\mathcal{O}}$. As we saw, the most compelling hypothesis is that the languages used to express scientific theories are the best candidates for $\mathcal{L}_{\mathcal{O}}$. Importantly, they are the best candidates because of the epistemic *bona fides* of scientific methodology.[9] More importantly, (as we saw above) the Realist adopts this third tenet as a consequence of her commitment to an externalist treatment of the languages used to express scientific theories. The *semantic* conditions on $\mathcal{L}_{\mathcal{O}}$ require that natural language expressions have $\mathcal{L}_{\mathcal{O}}$ translations. If, by hypothesis, the semantic properties of natural language expressions determine the space of eligible translations into other languages, then for the Realist's program to be viable, natural language expressions must be translatable into $\mathcal{L}_{\mathcal{O}}$. At a minimum, this means that the semantics for natural languages must be externalist. The Realist is thereby committed to ($\mathcal{E}$)

[9] We'll see in Chapter 4 that this burdens the Realist with providing a reason for excluding the languages of certain scientific disciplines, while including others. Note for example, that Sider (2011) indicates that the languages used in physics, math, and logic, are uniquely privileged in this way, while presumably those used in chemistry, biology, and cognitive science are not.

2.4 A Hint of What's to Come

The Realist insists that the meanings speakers of a natural language assign to expressions can be characterized by an invented language that traffics in a simpler semantics involving worldly objects. There is a tension with this project, so construed, as with any scientific endeavor. The virtues of ontological simplicity motivate a theory with fewer ontological commitments, which comes at a cost of explanatory adequacy. A simpler theory is more likely to ignore distinctions that matter for competent speakers. What we seemingly want in a theory is one that matches the granularity of speaker meaning assignments without over-simplifying, while yielding explanatory relations between the various assignments speakers give to natural language expressions. To illustrate with a toy example: English has many different pronoun words: 'he', 'she', 'they', 'we', etc. A language that has less pronoun words is simpler (in some important sense). Thus in translating the meanings of English expressions that use different pronoun words to our (simpler) meta-language, we could translate them all with a single word 'PRO'. On this theory sentences like

(48) John is fair and he is smart.

(49) John is fair and they are smart.

will both be translated into our simpler language as

$$\exists x.\textsc{john}(x) \wedge \textsc{fair}(x) \wedge \textsc{pro}(x) \wedge \textsc{smart}(x) \ \wedge \ \neg\exists y.\textsc{john}(y) \wedge \textsc{pro}(y).$$

On such a theory (48) and (49) would have the same meaning. But such a theory then fails to explain why competent speakers of English treat these sentences differently in thinking that (48) is about John, and (49) is about John and some other folks. The point of this example is merely that oversimplification for the sake of ontological parsimony comes at a cost. Further, this cost should be important to the Quinean Realist that hopes to make use of natural language speakers' competence as an ontological guide. If the meta-language expressions that purportedly convey the meanings of their natural language analogs simplify over differences that matter to the competent speaker, then speakers' judgments about natural language meanings are a less reliable guide to ontological inquiry, insofar as the purported

translations that have domain-object-based satisfaction conditions would be misleading translations.[10]

The observant reader may have noticed a worry of this kind, even from our initial characterization of regimentation. Recall our first troubling example:

(44) Pegasus is a horse.

On the proposed treatment of names as definite descriptions, and the proposed method of regimentation, the logical form of this expression would be something like

(44_β) $\exists x.\textsc{pegasus}(x) \wedge \textsc{horse}(x) \wedge \neg\exists y.\textsc{pegasus}(y)$

If our proposed regimentation is meant to capture the meanings that competent speakers assign to expressions in their natural language, we should expect that (44_β) comes out true, since one can truthfully utter (44). Applying conjunction reduction, this entails that the formula

$\exists x.\textsc{pegasus}(x)$

has a satisfier, namely Pegasus. That is, if (44_β) is true (because (44) is), then there is some object in the domain that satisfies each predicate in (44_β), including the predicate '$\textsc{pegasus}(x)$'—some object that is Pegasus. So 'Pegasus exists' must be true after all. In light of this worry, there are of course a number of ways to respond.[11] But the point here is simply that oversimplification is an easy trap to fall into, and what I hope to show in what follows is that these traps are not hidden in the special cases, but are littered throughout the natural language landscape.

I will argue in the next chapter that the oversimplification concern undermines the Realist methodology, as applied to natural languages. That is, any model with a domain-object-based semantics will have trouble accounting for the meanings that

[10]For Quine himself, this issue is not so dire, since he thought that translations between languages were always radically underdetermined. In part, this is why Quine suggests that the proper language for ontology will be the languages of our best science, which (he thought) were not beholden to natural language meanings.

[11]Lewis (1986) famously explores the idea of concrete *possibilia* not substantively different from actual concrete objects. More recently Thomasson (1999) argues for the existence of fictional objects as distinct metaphysical kinds to address such problems. Without disparaging this literature, one wonders if the vast proliferation of our ontology as a response to such cases is best seen not as an innovative strategy to deal with difficult cases, but rather as a *reductio* against the (bold) premise that meanings are cashed out in terms of truth.

competent speakers of a natural language assign to broad classes of expressions, even if we grant the Realist's other problematic assumptions about the nature of natural language. And in Chapter 4, I'll suggest that a similar problem arises for the Realist hoping to use the languages constructed to express our best scientific theories.

Chapter 3

Realism, Externalism, and Natural Language Semantics

In a previous chapter, by briefly canvassing the historical treatment of natural language semantics, I argued that the predominant views about linguistic meaning are externalist. More specifically, I argued that the externalist tradition is committed to the claim that the meaning of a natural language expression determines its truth-conditions, either by equating meanings with truth-conditions, or by identifying the propositional content of an expression in a way that determines the truth-conditions of the expression. These views maintain that linguistic meaning is a relation exhibited between words and objects. That is, the meaning of a term is determined by, or determines, the mind-independent objects that constitutes the word's extension. Throughout that chapter I highlighted some troubling natural language cases that shaped the current landscape of externalist views. The response to these cases, throughout the history of the discipline, was (naturally) to augment the machinery to accommodate their troublesome features.

As promising as this approach might seem, I will argue in that natural language meanings are hostile to such externalist treatment. Specifically, natural language expressions routinely exhibit a kind of flexibility suggested by Chomsky and others, and this flexibility is ill-captured by the rigid models available to the externalist. Accommodating this flexibility under an externalist semantics either yields implausible ontological burdens on such theories, or belies good explanations for the relevant data, lending increased credibility to an internalist[1] approach to linguistic meaning.[2]

Importantly, this finding bears on the project that views natural language meanings as suitable tools for ontological investigation. More precisely, the arguments presented in this paper undermine a Quinean Realist ontological methodology, at least when applied to natural languages. Such a Realist is committed to an

[1]While the purpose of this chapter is not to directly argue for such theories, see Pietroski (2008, 2010, forthcoming) and Hinzen (2006a, 2007) for plausible internalist alternatives.

[2]Some care will be taken regarding the terminology here, since 'meaning' has been used to identify numerous different properties associated with natural language expressions, utterances, interpretations, and the content of a bit of communication. In §3.3.1 these differences are spelled out, but for the time being, 'linguistic meaning' here is intended to identify the meaning properties of an expression that remain constant across various contexts in which that expression is used.

externalist semantics for natural language, at least if natural languages are to be good guides for ontological investigation. The purpose of this chapter is two-fold: first to undermine the basic externalist claim that the meaning of an expression determines its truth-conditions,[3] and second, thereby undermine the Realist use of natural language meanings, and the intuitive judgments that rely on them, as useful tools in metaphysical inquiry. The goal is to undermine ($\mathcal{E}$), in the following Realist argument:

1. ($\mathcal{E}$) For any expression e (in some language L), the meaning of e determines e's truth-conditions.

   Corollary: an internalist semantic theory for natural language is false.

2. If natural language meanings determine the truth-conditions and referents of their constituent expressions, then natural languages can play an important role in ontological investigation...

3. ...since the meanings of natural language expressions will pick out their real-world referents to populate the Realist's ontology.

If the meanings of terms in a natural language fail to determine their truth-conditions (and their constituents' referents), then investigating ontology by analyzing the meanings of natural language expressions will yield indeterminate answers to ontological questions.

In this chapter I will present a variety of arguments and evidence which count against an externalist semantics for natural languages. The Realist has a ready reply to the worries expressed in this chapter. She can merely deny that the privileged language of ontology ($\mathcal{L}_{\mathcal{O}}$) is a natural language. She could insist that the proper language used for ontological investigation is more carefully crafted than any natural language. The sciences, in the process of naturalistic investigation, develop languages in which they can express their theories. These invented languages, contends the Realist, are better suited for the purposes of addressing ontological questions. In the next chapter I'll explore the degree to which the proceeding arguments, which render natural languages inept for addressing metaphysical disputes,

[3] I use 'truth-conditions' from here on as short-hand for 'truth-conditions or satisfaction-conditions'. The distinction here, glossed on my stipulated usage, marks the difference between the meanings of sentences and sub-sentential expressions. The latter do not have truth-conditions because they are neither true nor false. I assume that no profound confusion will result from this gloss.

likewise apply to the artificially constructed languages used to express scientific theories. To the degree such arguments can be extended, they rebuff the Realist's retreat to the languages of science.

3.1 Taxonomy and Assumptions/Motivations

3.1.1 Taxonomy

Before we proceed, some distinctions may be helpful. For purposes of taxonomy we should distinguish between internalist and externalist views about both *meanings*, and *languages*. As I understand the externalist commitments, the *meanings* of natural language expressions are relations (of a particular sort). In particular they are understood as relations between elements of a language and publicly available objects. An internalist will deny this. Meanings on internalist theories are non-relational, at least where one of the *relata* is a publicly available thing.[4]

The internalist/externalist distinction pertaining to *language* regards the ontological nature of languages, and the elements that constitutes them. For the externalist view of language, languages are mind-independent things, while for the internalist languages are aspects of the human mind.[5] The logical space of externalist views regarding language and meaning seem to be exhausted by adopting either an internalist or externalist view about either language or meaning.[6] Thus one can be an internalist about language (IL) or an externalist about language (EL), holding that language is either in the mind or not. Similarly, one can be an internalist about meaning (IM) or an externalist about meaning (EM), holding that meanings either are relations between words and objects, or not. As such, the logical space of views is displayed in the following:

[4]Drawing the distinction along relational lines under-determines the content of internalist theories of meaning, but for our purposes here, the under-determination is immaterial. Insofar as any externalist semantics is relational wherein the needed relation binds to a publicly available object, if such a semantics cannot account for the linguistic data, this undermines ($\mathcal{E}$).

[5]This is an over-simplification. One can remain agnostic about the ontology of language, while denying an externalist conception (cf. Hinzen, 2007, §1.5). One could also *deny* that there are languages at all, a claim some Chomskyans seem to endorse (cf. Chomsky, 1986). Such theorists are considered internalists in the literature. I ignore this distinction for the purpose of simplicity, as nothing I say here trades on this distinction.

[6]For the sake of completeness, one could also deny the existence of either languages or meanings. The motivations for either position aside, the Realist (the target of this work) would not welcome such a deflationary view.

Table 3.1: Externalist and Internalist Views

	Internalism Meaning (IM)	Externalism Meaning (EM)
Internalism Language (IL)	IL-IM	IL-EM
Externalism Language (EL)	EL-IM	EL-EM

3.1.2 Assumptions/Motivations

Anti-externalist arguments target both EL and EM theories. Though these two classes of arguments can be treated distinctly, they share a series of assumptions and motivations. Most notably they are inspired by a Chomskyan approach to language generally. In part this embodies a commitment to a naturalistic methodology. On this approach language is treated as an object of scientific investigation, as a naturally occurring phenomenon, in principle no different than biological reproduction, combustion, planetary motion, or viscosity. This focus is not merely on empirical investigation, but that such investigation can make testable predictions, provide insightful explanations, and can be integrated with other scientific disciplines—most notably psychology and biology.

That humans have the ability to communicate the content of their thoughts *via* vocalization, and that we learn to do this in a short four years, are naturally occurring phenomena that beg for an (naturalistic) explanation. Explaining these phenomena in large part means characterizing what it is to understand a language—put flat-footedly, one must "know English" in order to "use English". As such, any complete study of language should seek to answer three questions:

> (i) What constitutes knowledge of language?
> (ii) How is knowledge of language acquired?
> (iii) How is knowledge of language put to use?
> (Chomsky, 1986, p. 3)

As Chomsky (1986) argues (and as we'll see in the following section), this commitment to naturalism and the guiding questions in (i)-(iii) rule out certain conceptions of language as viable candidates of study. They are excluded simply

because language on these conceptions cannot be investigated through naturalistic means. Importantly, the veracity of these arguments depends on the success of the research program that insists on investigating language by naturalistic means. If the endeavors of such a research program bear no explanatory fruit, then that failure tells against treating language (and meaning) as a natural phenomenon. There is little doubt however whether the Chomskyan tradition in linguistics has failed in this regard (cf. Baker, 2002; Boeckx, 2006; Piattelli-Palmarini et al., 2009).

3.2 Arguments Against EL Theories

3.2.1 Argument from Acquisition

There are two conceptions of language, an externalist (EL) view of language and an internalist (IL) view of language. The former view construes language as a mind independent (abstract) object. Languages on this conception are abstract structures relating mind-independent objects to terms, words, or expressions. Terms in a language (or words) are themselves mind-independent objects, artifacts in some sense, that we use to denote (other) objects. Accordingly, one understands a language when they identify the abstract structure that most sensibly coheres with the usage in their linguistic community. This conception of semantics, and language can be found in Lewis (1970) and Lewis (1975):

> What is a language? Something that assigns meanings to certain strings of types of sounds or of marks. It could therefore be a function, a set of ordered pairs of strings and meanings. (Lewis, 1975, p. 3)

A commitment to naturalism speaks against thinking of languages as the abstract objects described in this passage. In treating a language as a mind external object, as Lewis (1975, p. 19) puts it, "in complete abstraction from human affairs" one wants to know how humans can come to understand or "know" languages so construed. That is, on this Lewisian characterization, languages are *abstracta*—functions that take us from symbols to truth-values, combined with a grammar that delineates how these symbols can be combined in acceptable ways to form interpretable expressions (or strings). If a naturalistic approach to language seeks to answer (i) what knowledge of a language amounts to, a Lewisian treatment of language renders this question intractable. As an infinite-membered set of ordered

pairs of expressions and their functional meanings, this conception of language not only gives us little direction as to how to answer (i), but seemingly gives too sparse a collection of resources to answer the question at all. Put more tangibly, all English speaking children have adult-like competence with English around the age of four, when placed in a community of English speakers—they come to "know English" around the age of four. On a Lewisian view, for a child to "know English" they must first decide which set amongst an infinite array of (infinitely-membered) sets of expression-function pairs is the English set, and then second, they must bear the right kind of epistemic relation (the knowledge relation) to that set.

The argument here is *not* that, given certain metaphysical commitments to nominalism that theories committing us to *abstracta* are off base. The worry Chomsky presses pertains to the conditions for explanation, and particularly whether certain conceptions of language (whatever their metaphysical commitments) forestall viable strategies to answering fundamental questions. Treating languages as abstract entities is problematic not because they are *abstracta*, but because *abstracta*, *qua* objects of knowledge, bear mysterious epistemic relations to human minds. As such, to explain what knowledge of a language (so construed) amounts to requires an account of the sort of relation that a human can bear to an abstract entity, in particular a set of ordered pairs, such that this is the sort of relation a four-year-old child can enter. How one proceeds to answer these questions seems hopelessly unclear.[7]

Chomsky puts a related point about acquisition in terms of "legibility conditions" on a natural language. If a child is to come to have "knowledge of a language" they must come to represent that language in their mind/brain. For a given child to have "knowledge of English", they must have come to represent both the grammar of English—the algorithms by which one can combine lexical items in that language into larger expressions (or sentences)—and at the very least, some internalized list of those lexical items that combine syntactically. In whatever way this information

[7]A Lewisian might argue that such an explanation need not be directly forthcoming for the project of building an externalist semantics for externally construed languages to proceed. After all, humans some how manage to learn mathematics, and the best account of the ontological nature of mathematical language is decidedly abstract, and set-theoretical. So clearly (the thought goes) humans can stand in the relevant epistemic relation to abstract objects, and in particular abstractly construed languages. Notice however, that four-year-old children do not exhibit mastery and competence with the language of mathematics, even if they do have innate mental structures that aid with deciphering numerousity (Carey, 2009, Ch, 4). This fact requires an explanation, and that explanation is precisely what the Chomskyan challenge to EL views demands.

is encoded in the mind/brain, it must fit into the architecture of the mind/brain. If our four-year-old has figured out, or "knows", which grammar (as an abstract object on an EL theory) is the English grammar, she has represented such a grammar in her mind/brain. When she wants to utter an English sentence this grammatical knowledge must be applied (to the lexical items she also "knows") in such a way so that her articulatory systems can make the right sort of audible noise. That is, the representations she builds using her grammatical knowledge must encode information in a way that her articulatory system can make sense of—those instructions must be *legible*. Thus the structure of the abstract object that is *English*, must abide by such legibility conditions, coming not only from the articulatory system, but any other aspect of the mind/brain that the child's linguistic knowledge must interact with. But once we recognize that the structure of the abstract object *English* is beholden to the legibility conditions imposed by the human mind/brain, this defeases the motivations for thinking of a language in this abstract way (Chomsky, 2000, p. 73).[8] Taken together, considerations of acquisition and legibility pose serious, and as far as I can see, unanswered challenges for EL theories.

None of this shows that these *abstracta* do not exist. There could well be, in addition to the mental structures hypothesized (and studied) by the Chomskyan tradition, functions filling an infinite Fregean heirarchy mapping objects of one kind (say $< e >$) to objects of another kind (say $< t >$). But these functions must earn their keep. We are told that we should believe in these *abstracta* because they are essential to successful explanations of linguistic phenomena (Lewis, 1986, Ch. 1). The point here about legibility conditions is that these structures serve no explanatory purpose, and indeed present explanatory obstacles, in understanding how a child comes to know a language. Thus, the externalist owes us some indication of what these objects are meant to explain.

One possible explanatory virtue of adopting externalism about languages (and meanings) is that it yields an intuitively plausible account of successful communica-

[8]Chomsky also argues that treating languages as abstract structures, akin to the formal languages of mathematics, renders aspects of natural languages inexplicable, in principle. Chomsky uses two examples, 'imperfections' in natural languages to highlight the mismatch between the structure of natural languages and formal ones: 1) that languages have uninterpretatable features, and 2) the displacement property (Chomsky, 2000, p. 12). Even if treating natural languages like formal ones leaves room for explaining such features, the point here is that there is no good justification for stipulating at the outset of investigation that the object to be investigated must meet the (optimality) conditions of a formal language (especially if even superficial differences speak against such stipulation).

tion. If languages (or their meanings) are external entities, then successful human communication is explained by the mutual relatedness of individuals to the language they know. That is, two humans can succeed in communicating because they are related to, or come to understand, the very same *thing*—a language. I address the explanatory success of externalism on this count below (in §3.2.2). To preview, this explanatory virtue is dependent on the availability of a coherent specification of the individuation conditions for the external languages that individual speakers come to grasp. The prospects for these conditions, I contend, are grim.

3.2.2 Ontological Arguments

Commitments to naturalism also motivate *ontological* arguments against EL theories. EL theories hold that natural languages are objects-in-the-world, whose existence is independent of minds. As such, expressions of a natural language are likewise objects-in-the-world. One strain of internalist arguments presses this ontological claim, showing that there can be no such thing as a mind-independent natural language expression. Or, to avoid begging certain (meta)metaphysical questions, that there is no scientifically respectable notion of 'natural language expression'. More plainly, these arguments contend that words do not exist, and cannot thereby stand in any relation (be that either the reference or meaning relation) to any object in the world.

An analogous worry arises when investigating the nature of color, in particular regarding the human experience of color. The typical human visual experience is awash with color. We see objects as *having* colors: fire hydrants are red, bananas are yellow (when ripe), leaves change their color in fall, etc. This way of seeing the world is built into the structure of the human visual experience, is cognitively impenetrable (cf. Fodor, 1983), and mistaken. There are no mind-independent colors in the world (suited for naturalist inquiry). Suppose one was interested in studying color scientifically, they would have to begin with some object suited for naturalistic study. The following proposal for such a subject of study seems hopeless: colors are sets of colored objects. On such a conception the color red just is the set of objects that are (experienced by humans as being) red. But naturalistic studying of this (abstract) object does not look promising, at least if one hopes to learn anything novel about color: how humans come to see colors, why/whether objects possess

this property, how this phenomenon relates to electromagnetic radiation, why some humans have deficiencies experiencing and seeing colors etc.

A more promising approach might be to ask what qualifies an object as being a member of this set, thereby investigating the intrinsic properties of those objects that render them qualified candidates for red-set-membership. Notice though, an appeal to the fact that they "appear red" will not be informative, as such an appeal is painfully circular. But an appeal to other properties might well be fruitful. The naturalistic study of vision has taught us that the human eye is sensitive to (limited ranges of) electro-magnetic radiation, or wave-lengths of light. We have three types of ocular structures that react to various ranges of wave-lengths of light. Humans (typical ones at any rate) are trichromats, insofar as they have *three* such structures.

Consider now how a naturalistic investigation into colors must proceed. We can now start to identify which object-properties establish the proper criteria for membership in a particular color-set. The bio-physiology of the human visual system is helpful to this end. Humans have "red" experiences when photons of a particular constellation of wave-lengths trigger the structures in our eyes. An object appears red because humans are disposed to have a certain kind of representation when they perceive that object, in virtue of the photons that object reflects. These *spectral-reflectance properties* of the object (when illuminated by full-spectrum light) dictate which wave-lengths of light are absorbed by the surface of the object, and which wave-lengths are reflected and thereby visible. For example, because an apple's skin has certain surface physical properties, the apple-skin absorbs photons of certain wave-lengths, and reflects others. When the full spectrum of visible light is illuminating an apple, these reflected photons with various wavelengths strike the human eye and, because of the state of the human visual system, bring about a 'red' experience, or response, in the human mind/brain. This representation encodes the redness as being a property of the apple.

But the naturalistic properties inherent to the apple are not the photon wave-lengths the human eye responds to, which are the objects in the world that bring about colored visual experiences. Spectral-reflectance properties are instantiated by objects, but spectral-*reflectants* are not. Further, multiple *distinct* wavelengths of light manifest the same visual color experience. That is, there is no single *sui generis* and continuous range of the spectrum of visible light that manifests 'red' experiences in humans. Physically distinct (types) of wave-lengths of light cause

humans to see objects as 'red'. Thus, if one asks what such distinct wavelengths have in common (such that they are all 'red'), appeal to the physical properties of the relevant photons will not be helpful. Nor will appeal to the spectral-reflectance properties of the relevant objects, as those will also be of equally physically distinct types. Thus if we are to look at the world in hope of finding a mind-external object 'red' that is a property of things like apples, appeal to (a set of) wave-lengths of light not be helpful. After all, such things aren't properties of apples. But even if we ignore this worry, it turns out there is no *single* wave-length of light that could be picked out.

For explanatory reasons, appealing to the surface properties of objects (like) apples will not help identify the mind independent color properties either. If redness is simply reduced to the spectral reflectance properties of objects, then the representations had by canines (which exhibit a kind of human color blindness, as they are dichromatic) would count as representations of red, despite the fact that the canine's phenomenology of (say) a Haralson (red) apple is closer to what (most) humans call 'green'. Thus, in order to predicate colors of objects, and make the color experiences we have when viewing those objects relevant to such predication, colors must be response-dependent properties. The property red, then, is a complex response-dependent property explained in terms of: 1) human responses (mental representations), 2) object properties (spectral-reflectance properties) and 3) "normal" viewing conditions (e.g., shining a red light will make many objects appear "red" that are presumably not "red", violating the *normal* viewing conditions for objects). Thus a search for a mind-independent property 'red' will not be found by scientific investigation.

Spectral-reflectance properties are mind external, but they will not do if our hope is to explain how humans see objects as colored—and appeal to human visual representations will not meet the external condition. Further, the complex interaction between these two aspects of the world characterized by "normal" viewing conditions might not be all that explanatorally productive in the absence of them. An account of redness that makes sense of both the human visual experience of red, and appropriately connects that experience to worldly-objects requires all three elements, identifying redness as a response-dependent property. But of course, such a property will not be mind-independent (i.e. external). Thus, a conception of color that includes features about the human mind, and its capacity to build representa-

tions with color content, is necessary to a scientific investigation into the nature of color. Put another way, colors (naturalistically construed) cannot be objects in the world, much less abstract objects like sets.[9]

A similar line of argument suggests that we should abandon EL theories, insofar as there are no mind-independent objects 'languages' or 'words' that are viable candidates for naturalistic investigation. Take first the ontological distinctness of different languages, as publicly available mind-independent objects. The motivation for thinking of languages in this way is to account for communicative success. In order for two speakers to understand one another, they must know the same language, and this (so goes the argument) can only be the case if there is some single mind-external thing they both know.

That is, one of the virtues of a view of languages as mind external objects, is that this understanding of language can explain communicative success between two speakers of a given natural language. The idea is that, when a speaker uses a sentences, they are making use of an external object, that is publicly available. Knowledge of language on this view is had when a speaker bears the the right sort of relation to this external object. Thus two speakers that "know English" can make sense of each other's speech because they bear the same relation to the same mind external object. Thus when you use a sentence(-object) of English, I understand this sentence because I am related to that *same* English object in the *same* way you are, *via* knowing. Of course on this view, the explanation of communicative success that counts in favor of the view requires that languages can be differentiated ontologically, insofar as you and I must bear the same relation to the *same language*.

For these purposes, "common sense" ways of differentiating languages will not do. As Chomsky notes, common sense treats Dutch and German as distinct languages, despite that fact that people "who live near the dutch border can communicate quite well with those living on the German side..." (Chomsky, 2000, p. 48). If treating languages as mind-external objects is meant to explain communicative success, then the fact that speakers of "different languages" can communicate linguistically[10] is unexplained by such theories. The common sense notion of language

[9]See Goldman (2007) for the relationship between ontology and cognitive science; see Chomsky (2000, 1992, p. 23) for similar arguments.

[10]Humans, and other animals, communicate in many non-linguistic ways: body language, facial expression, gesture, etc. The important cases for this point are those instances of communication

gets the extension wrong in the other direction as well, insofar as the Mandarin and Cantonese "dialects" of the Chinese language are mutually unintelligible. The EL-view here fails to explain why these speakers of "Chinese" cannot communicate (verbally) with language, since the relevant populations both "know Chinese".

Since the common sense division of languages will not serve the purpose of explaining communicative success, another means of distinguish languages (as objects) is in order. One might appeal to the *elements* of languages to distinguish them. On this proposal, two languages, (say) English and French, are distinct because of the differences between the elements that constitute them—one contains words like 'photographer', 'apartment', and 'cat' while the other contains words like 'photographe', 'appartement', and 'chat'. This move requires that these elements differ along some important dimension, such that the first three belong to English and the last three belong to French. Notice that appealing to usage will not be helpful. Defining 'French' and 'English' by indicating that speakers of French use the latter, and speakers of English use the former, is viciously circular.

An EL theorist must make use of some other property that these words share that marks the boundary between English and French. But to what properties could an EL theorist appeal? The sonic properties of these words seem like bad candidates. 'Cat' and 'Chat' have more in common in this regard than 'photographer' and 'cat' do. Further, considering the variation seen in pronunciation across speakers of the "same word", appeal to such properties will not distinguish words as to cohere with the communicative motivations for EL theories:

> To take an example, why are 'fotoGRAFer' (said in Bombay) and 'foTAH-grafer' (said in Toronto) the same word, yet 'fotOgrafo' (said in Buenos Aires) is not the same word as the former two? (Stainton, 2006, pp. 918–919)

To explain communicative success and failure, the EL theorist posits the mind external object *English* that our Bombay speaker and Toronto speaker both "know", and on the current proposal this language is distinguished from others based on the elements of that language—words like 'photographer'. In this example, the EL theorists wants to say that we have two words here (as opposed to one, or three), one in English and one in Spanish. But an appeal to the sonic properties of (utterances of) words here will clearly not help such a theorist, given variations

that are clearly effected with language.

in pronunciation.[11] The point is familiar to phonologists, namely that there is no *sui generis* cluster of sonic properties that utterances of a word share in common. What 'fotoGRAFer' and 'foTAHgrafer' share in common is the manner in which they are represented by humans, which involves features of the system for creating phonological representations from environmental noise (cf. Bromberger & Halle, 1995). Just as redness cannot be identified by a mind-independent object, so too, there are no word-objects that can be differentiated without appeal to structures of the human mind/brain.

These problems are particularly trenchant for theorists committed to a Quinean Realist position. For the Realist the naturalist perspective is not negotiable. If there is no mind-external, naturalistically respectable notion of 'natural language expression' or word, and thus no worldly objects of that sort, then there can be no (semantic) relations between words and objects, much less one of denotation. Such a Realist cannot fall back on a kind of pluralism or fictionalism about words to get an externalist theory going, because their ontological methodology prohibits such a retreat.

Notice too that this purported virtue of externalism ignores a more fundamental problem, namely that any particular human will fail to *completely* know any such "language". Externally construed as an infinite set of expression-meaning pairs, no individual will come to have full knowledge of a language, at least if such knowledge is construed as knowing what the members of this set are. Given the limits on human cognition, no individual could know this infinitely long list of pairs. At best, we must have an incomplete knowledge of such a 'language". Thus if appeals to some external entity are meant to aide in explaining how humans manage to successfully communicate, such an explanation will have to succeed despite the fact that no human can fully know such a language. That is, the account of communicative success above assumes that two speakers stand in the same relation to the same object. However, if a language is an infinitely large set of sentence-meaning pairs, two speakers of that language either: (a) stand in a different (incomplete) knowing relation to that same set, or (b) stand in the knowing relation to different subsets of that set— see Dummett (1993) for discussion.

Thus a commitment to naturalism will preclude certain notions of language,

[11] Appeals to orthography will not be any more helpful, given that illiterate individuals can communicate using language quite well.

specifically those that construe natural languages as mind-independent abstract entities. However, this commitment does not preclude a semantics for an IL-language from having an externalist character. Even if language is properly construed as an aspect of the human mind, the expressions of that language, seen as mental representations, can still have contents that are cashed out in terms of mind-independent objects. Such an account of meaning is still importantly externalist. I turn next to the various arguments put forward for why (natural) IL-languages are hostile to externalist treatment.

3.3 Compositional Referential Semantics and Natural Language

Bracketing the worries in the previous section about the naturalistic credentials of the objects posited by EL theories, trouble still looms for an externalist semantics; i.e. EM theories. These arguments illustrate that an externalist semantics cannot obviously accommodate the distribution of speaker judgments, or that in accommodating the relevant linguistic data such a semantics must take on board independently implausible assumptions. But, I'll begin this section by indicating what these internalist criticism are not arguing.

3.3.1 How Not to Argue for Internalism

There is a (now well known) line of criticisms regarding a referential, or externalist semantics that treats the meanings of expressions as their denotations attributable to Austin (1962), Strawson (1950), and Wittgenstein (1953, 1972). The criticism, or the proposal from which the criticism is derived, is typically summarized by indicating that words do not refer/denote, *users* of words do. One way of explaining this point is to highlight the role of indexical and demonstrative expressions like 'I', 'you', 'here', 'now', and 'this' in determining the meaning of an expression. The meanings of indexical expressions are intimately tied to the context of their use. Put another way, even if we accept that sentences like (50) have referential meanings,[12] sentences like (51)–(53) cannot refer, because they are importantly

[12]Though this is far from obvious, despite what traditional semantics textbooks might suggest. Even if the conditions that satisfy the predicate 'is foggy' can be codified, consider the ontological status of a city that moves from one location to another location two miles to the east (Rolander, 2013). And even if those metaphysical difficulties can be met, it is far from clear how such a view can account for sentences, like "The tallest mountain in Sweden, Kebnekaise is situated 100 km

incomplete (nonsensical for Wittgenstein[13]) outside of their use:

(50) Kiruna is foggy.

(51) I think that should go here.

(52) I am here.

(53) I am a philosopher.

Sentences with indexicals, like (51), (52) and (53), have different referential meanings when uttered by different agents (at different times, using different gestures). When Scott Soames utters (53) he makes a true claim, in a way Kobe Bryant does not. Thus the indexical 'I' in sentence (53) does not have a denotation, and thus (53) is not truth-evaluable independent of a context in which it is used.

This Strawson-inspired argument, however does not undermine EM theories. Following Kaplan (1977 [1989]) we can treat such expressions as having a functional meaning, or character, that, when provided a context determines the expression's content, or referent. The sense in which Kobe Bryant and Scott Soames say the same thing in uttering (53), is that they use the same expression. The static *linguistic* meaning that all utterances of (53) share is the *character* of the expression. This character, when the expression is used in a context, determines the content, identifying Scott Soames as the reference (or content) of 'I' in some contexts, and Kobe Bryant in others. This analysis preserves both the idea that the meanings of indexicals (and demonstratives) are static (from use to use), and the externalist notion that such meanings will determine a unique referent, at least in a given context of use. More generally, such a theory maintains the externalist idea that the meanings of expressions determine their truth-conditions. We know the conditions under which any indexical-using expression is true, insofar as the character of those expressions will determine what must be the case in order for such expressions to make true claims. Context plays a role in such cases of determining the referent of the indexical constituent of these expressions. Call this the role played by *semantic context*.

---

or 62 miles from Kiruna. ... Nowadays the town [of Kiruna] is not relying solely on the mine" (girontravel.se, 2013). While a city construed as a spatio-temporal object might well be situated some distance from some other object, and might even survive relocation, such an object is not of the sort that relies on anything, much less revenue. See Chomsky (2000) for examples of this sort regarding London.

[13]See Wittgenstein (1972, §10).

There are other ways in which the context of an expression's use can impact or "change" the meaning of the expression. Consider the following expression:

(54) Barack Obama is human.

The "literal" (i.e. linguistic) meaning of this expression, on an externalist understanding, indicates that some particular individual (the US President in 2014) has a particular property (of falling into the biological kind *homo sapien*). Put another way, I can use this expression to convey a thought, the meaning of which is that some individual is a *homo sapien*, on those occasions where I intend to convey the literal meaning of the expression. The related Austin-inspired point regarding the (externalist) meaning of expressions contends that because we use language in non-literal ways, and further that such non-literal usage is pervasive, the meanings of expressions are best captured by the conditions of their use. Thus what a sentence *means* follows from the context in which a speaker utters the sentence. For example, one could use (54) in various ways:

(54′) Barack Obama is human. *Context*: Obama has made some mistake.

In uttering (54) in the context of Obama's mistake in (54′), I do not merely intend to communicate information about Obama's place on the phylogenetic tree. My usage communicates (or intends to communicate) the linguistic meaning acontextually attributable to the expression in (55):

(55) One should refrain from feelings of disapprobation toward Obama in this instance in light of the fallibility of humans.

This non-literal usage of language (expressing (55) by the use of (54)) is pervasive. This feature of natural language seems to be at the heart of Strawson (1950), which is often misunderstood as merely indicating the context sensitivity of indexical (and demonstrative) expressions. While one way of making "different use of the same sentence" is to use an indexical expression in different contexts, this is but an instance of a more general phenomenon (Strawson, 1950, p. 325). The general point for Strawson is that the proposition expressed by a speaker, and thus whether what is spoken is true, depends quite heavily on the context of utterance (and the use of the sentence)—a point Russell seems to miss[14] (Strawson, 1950, p. 336). Call the

[14]Strawson focuses on the context sensitivity of indexicals, largely embedded in definite description, because he argues that Russell's theory of descriptions, which sacrifices the connection

role played by context in ascertaining (55) from (54′) the role of *cognitive context*.[15]

However, this feature of language does little to undermine the externalist program in semantics, or at least this is not the problem I'll be pressing in the remainder of the chapter. That we can use sentences to convey thoughts that do not match the intuitive literal meanings of expressions seemingly *requires* that expressions have static linguistic meanings that are context-independent. In uttering a sentence, I intend to communicate some thought or other. The task of my audience is to infer this thought from the sentence I used in my act of uttering. Grice (1957 [1989]) notes, for this communicative act to be successful my audience must recognize my communicative intention. They must recognize that I intend to communicate some thought or other, and in this way my intention to communicate is overt.

Given my overt intention to communicate a thought, my audience must identify the intended thought, in some way or other. Contextual cues, shared biological sense modalities, a common presupposed set of knowledge, and other aspects of the uttering act all constitute the evidence available to my audience in making the correct inference about my intention. When communication is successful, they ascertain my actual intention. But paramount among the evidence considered in this inferential move is the choice of sentence used in the utterance. That I utter (54) in the context indicated in (54′) seems to matter, and it matters precisely because the sentence has a static linguistic meaning. Were sentences void of any literal meaning, it's hard to imagine that they could play as substantive a role in my audiences inference making as sentences seem to. That I cannot communicate the thought in (55) by using *any* expression I choose illustrates this point.

There are limits on what a speaker can reasonably expect his audience to infer about his communicative intention on an occasion of utterance. Part of this is constrained by the context of the utterance. Consider the following:

(56) The goose is ready to eat.

---

between grammatical and logical form for the sake of sentential truth preservation, does so needlessly. The theory is not needed because sentences are not true or false, uses of sentences in utterances are. Russell's reply seems to miss this substantive point: "As regards 'the present King of France', [Strawson] fastens upon the egocentric word 'present' and does not seem able to grasp that, if for the word 'present I had substituted the words 'in 1905', the whole of his argument would have collapsed" (Russell, 1957, p. 261).

[15]The use of 'semantic' and 'cognitive' context in this section is attributable to Bach (2004).

If sitting at the Christmas table, having just placed a dish of fowl on the table, were I to utter (56), my audience would reasonably infer my intention to communicate that Christmas supper has begun. However, I could very well *intend* to convey some other thought, the thought that the goose (just placed on the table) is ready for *its* supper. And there are many contexts in which one could use (56) to convey such a thought (e.g. in a park full of hungry geese). But this latter intention seems unreasonable in the context of the Christmas supper, since the context of the utterance includes what is quite clearly a goose that is incapable of self-nourishment.

Just as context constrains what one can reasonably expect to communicate (given the force of Gricean maxims[16]), the linguistic meaning of the sentence one utters constrains what one can reasonably expect his audience to infer about his communicative intention. After setting down the goose-platter, I *could* very well intend to convey the thought that the goose is ready to be consumed, and Christmas supper has begin, by uttering (54). But my audience would have trouble ascertaining the intended thought from such an utterance—and this trouble is most straightforwardly explained by appealing to the linguistic meaning of (54) ("That's not what 'Barack Obama in human' means!").

To summarize, the externalist can avoid the implicit worry about meaning stemming from the differences between the communicative import of (54) and (54′) by marking a distinction between the *semantic* properties of sentences and the *pragmatic* properties of utterances (or more specifically, of ascertaining communicative intentions from utterances: see Bach (1999, 2001)). For all that has been noted so far, a sentence's semantic properties (or linguistic meaning) can be construed externally, while the pragmatic processes (however they are to be understood) that govern the ways in which reasonable audiences will infer communicative intentions from speech acts, will remain independent from those considerations that bear on theories of (sentence/semantic/linguistic) meaning.[17] Put less forcefully, regardless of whether this reply available to the externalist (that distinguishes between the role of semantic and cognitive context) is successful, the worry it is meant to address is *not* the one explored in the remainder of this chapter.

[16]See Grice (1989, p. 26).

[17]In light of the preceding discussion, the use of 'meaning' from this point forward should be construed as 'linguistic meaning'; i.e., the *semantic* properties a sentence brings to the context of its use, wherein an audience (attempts) to pragmatically ascertain the communicative intention of the speaker that utters the used sentence in a given context.

The gap between linguistic meaning and the thought inferred by an audience to a speech act (or utterance) does not present a problem for truth-conditional semantics. However, some utterances closer to the semantic-pragmatic boundary seem more troubling. Consider the following examples:

(57) John is too smart [for this job].

(58) John finished [writing/playing] the sonata.

(59) John is ready [for class/to go home/. . .].

Each of these sentences, though acceptable to competent speakers of English, seem (in some sense) incomplete, as indicated by the bracketed content that completes them. For the externalist, the linguistic meaning of expressions of the form in (59) [NP–COP–XP] predicates a property to the individual(s) denoted by the noun phrase. 'John is handsome', for example, predicates a kind of attractiveness of the individual denoted by 'John'. But the meaning of (59) does not predicate some general property of *readiness* to *John*.[18] The sentence in (59) does not mean that John is ready for *everything*, as such treatment would imply. The notion of *readiness* seems to be relational, requiring an individual and an event for its satisfaction. Put another way, upon hearing an utterance of (59), one wants to know what John is ready *for*.

The interesting feature of these sentences and their relationship to their completed counterparts is that this relationship exhibits features that are typical of an expression's semantic properties, yet seem to be driven by pragmatic inferences. To start, the relationship between (59a) and (59b) is different than the relationship between (54) and (55).

(59) a. John is ready.

b. John is ready for it.

[18]Though this seems to be the view of Cappelen & Lepore (2005). This view however, fails to explain what needs explanation in these cases. Namely, competent speakers of English treat (59a) and (59b) as having the same meaning. The disquotational account of meaning fails to capture this data in any non-stipulative way. In fact, for Cappelen & Lepore (2005) the proposition expressed by a sentence is the disquotated sentence once we "disambiguate every ambiguous/polysemous expression in [the sentence]" (p. 145). However, the presumption that cases of ambiguity and polysemy can be resolved, *prior to* giving a semantics for the expression either denies that the data need explanation or denies that there are any data there to explain. For (59a) and (59b), that speakers treat these sentence as having the same meaning is ill-captured by a theory that insists they express distinct (disquotated) propositions.

(54) Barack Obama is human.

(55) One should refrain from feelings of disapprobation toward Obama in this instance in light of the fallibility of humans.

My audience infers the thought expressed by (55) from my utterance of (54) in the context of Obama's mistake, leveraging the literal meaning of (54) against the context of its use, coupled with my overt intention to communicate some (reasonable) thought. But the inference from (59a) to (59b) seems different. As indicated above, the literal meaning of (59a) is incomplete in a way (59b) is not. Thus, while the inference from (54) to (55) is a (paradigmatically) pragmatic process that (ostensibly) begins with the utterance of a semantically complete sentence, and ends with the intentionally communicated thought, the shift from (59a) to (59b) cannot be characterized in this way. To make a similar inference to my intention to communicate (say) that John is prepared for his logic exam from (59a), one must *first* arrive at the semantically complete (59b) to begin the analogous process.

But the process of getting from (59a) to (59b), does not seem semantic either. Again recall the role of semantic (*v.* cognitive) context in determining the referents of indexicals. In the case of indexicals the context determines the referent of the relevant expression based on the semantic properties of the expression (its *character* if we follow Kaplan). This determination is independent (seemingly) of what the speaker of the expression intends. As Barwise & Perry indicate "even if I am fully convinced that I am Napoleon, my use of 'I' designates me, not him" (Barwise & Perry, 1983, p. 148). Regardless of the mechanism whereby context interacts with the semantic properties of expressions, semantic phenomena that are context-dependent appear to be *determined* in (or by) a context of use.

Pragmatic phenomena, as we've seen, relate to contexts differently. In inferring (55) from an utterance of (54) the beliefs and intentions of both the speaker and the audience matter. For an audience to ascertain that I intend to communicate (55) by using (54), they must, at the very least, believe *that* I intend to communicate something. This belief plays an essential role in deciphering my communicative intention, in a way not required to connect the character of an indexical with its content.[19]

[19]For a discussion of this differing role of context, and the distinction between semantic and cognitive context see Bach (1999).

Cases like (59a) and (59b) present a problem because the role that context plays seems to be more general (or cognitive), while exhibiting (semantic) entailment patterns that pragmatic cases do not. To complete (59a) and arrive at (59b), my audience has to make use of knowledge not provided by the semantic context, specifically knowledge about *readiness*. Yet (59a) and (59b) exhibit a mutually entailing relationship that pragmatically inferred thoughts rarely have. In whatever way (semantic) context provides the expression 'it' with the relevant event that John is ready for (the taking of a logic exam, say), this determination will (for the externalist) determine the truth-conditions for the expression. But for any context in which the provided event makes (59b) true, (59a) will also be true in that context.

The difficulty with cases like (59) is that completing the expression seems to require pragmatic reasoning, commensurate with the meaning of the relevant expression. 'John is attractive' does not seem to require a prepositional phrase to express a complete thought, while the seemingly similar (59a) must be "enriched" to arrive at the thought expressed in (59b). Put another way, the information required to get from (59a) to (59b) is not part of the linguistic meaning of (59a), which indicates that this enrichment is pragmatic. But unlike paradigmatic pragmatic processes, the expression that characterizes the thought inferred from the use of (59a) is truth-conditionally equivalent to the used expression. This suggests that the move is semantic, not pragmatic.

The worry then is this: (59a) and (59b) seem to have the same truth-conditions, yet appear to have different meanings, insofar as (59a) is incomplete and (59b) is not. For the externalist, the truth-conditions of an expression is its meaning. That they come apart in these cases is troubling for the externalist view.

The anti-externalist arguments in the remainder of this section are not the (now) traditional worry associated with Strawson, Austin, and Wittgenstein that a single expression can be used in a variety of ways to express a variety of thoughts. Nor is the worry expressed by internalists that expressions with indexical constituents require context to determine their truth-conditions. Such points do not speak against an externalist semantics, or undermine ($\mathcal{E}$)—or if they do so, they are no the problems outlined in the next section. The point that internalists' worries stress is that, while the meanings of indexical expressions might be well-captured by appeals to (something like) the content-character distinction, they are a special case of a much more general phenomenon that is misrepresented by such treatment,

and exhibited by cases like (57)–(59). Natural language expressions exhibit a kind of *lexical flexibility* that is not isolated to a few problem cases to be addressed by intricate logics, but is a ubiquitous feature of natural languages—one that is importantly misrepresented by Kaplanian treatment. It is to this phenomenon that I now turn.

3.3.2 Lexical Flexibility

Chomsky (1977) marks a distinction between various ways in which the meaning of an expression can be multifarious, distinguishing between expressions that exhibit flexibility (like 'book') from expressions that exhibit "idiosyncratic ambiguity" (like 'trunk').[20] The English word 'trunk' is ambiguous, having a meaning used to denote a kind of luggage, and a meaning used to denote a part of an elephant (not to mention a part of a tree, or part of a human). This kind of ambiguity is importantly different than the semantic behavior exhibited by 'book'. Compare the following expressions:

(60) John wrote a book.

(61) This book weighs five pounds.

The use of 'book' in (61) denotes a particular, concrete book, with physical dimensions, and (as expressed) a particular weight. For (61) to be true, there must be a contextually relevant physical thing, a book, with a particular heft. However, for (60) to be true, there is no requirement of this sort. As Chomsky notes, John could have the book composed in his mind, not having ever deployed pen, paper, or word processor in composing the work. This might lead us to conclude that the multifarious meaning of 'book' is like the *ambiguity* of 'trunk': we have two distinct

[20]Specialists might recognize this distinction as the difference between *ambiguity* and *polysemy*. I hesitate to use these terms here for two reasons, one priggish, and the other substantive. The priggish reason is that the term 'polysemy' indicates (by its roots) that a polysemous word has multiple meanings. This is true for such terms of course, but ambiguous words are also ones with multiple meanings. Thus the contrast in the literature between polysemous terms, which are troublesome for mainstream semanticists, and ambiguous terms, which are not troublesome, is misrepresented by this use of terminology. The substantive reason for my usage here is that lexical flexibility is a property that applies rather broadly, and manifests with different semantic behavior in different contexts, two of which I discuss here. However, the explanation for the various manifestations of polysemy can be unified by internalist proposals (Pietroski, 2005, §3.2). See also Pietroski (forthcoming).

(homophonous) lexical items, 'book$_a$' corresponding to the *abstract* usage in (60) and 'book$_c$' corresponding to the *concrete* usage in (61).

However, if this treatment of 'book' is apt, we should expect uses of 'book' to behave like uses of 'trunk'. That they do not (as we'll see) suggests that the semantic relationship between the uses of 'book' is of a different sort than the relationship between uses of 'trunk':

(62) a. This book, which John wrote, is five pounds.

b. John wrote a book, this is it, and this book is five pounds

c. John wrote a book$_a$, this is it, and this book$_c$ is five pounds

(63) a. This trunk, which Jumbo grew, is full of clothes.

b. Jumbo grew a trunk, this is it, and this trunk is full of clothes.

c. Jumbo grew a trunk$_e$, this is it, and this trunk$_l$ is full of clothes.

The expression in (62a) is (roughly) synonymous with (62b), using 'book' in the two distinct ways discussed, as reflected in (62c). But (63a) is not synonymous with (63b), at least not if we interpret (63b) as making use of both lexical expressions of 'trunk' as in (63c). They are synonymous if we imagine Jumbo's nose full of textiles, but this interpretation is not available for (63c). This difference is not attributable to the syntax of these various phrases, as (62a) and (63a) appear in the same syntactic frames, and likewise with (62b) and (63b). Thus the difference in synonymy between these pairs cannot be attributed to the syntax of the expressions.

That these two expressions ('book' and 'trunk') behave in semantically disparate ways in relative clause constructions (as in (62a) and (63a)) counts against a semantics that treats them as formally similar. That is, this behavior suggests that we not treat the relationship between uses of 'book' as we do uses of 'trunk'. As Chomsky concludes:

> Thus [in the case of 'book'] we have a single formal element with a fixed range of meaning, and relativization is possible, despite the shift of sense. But in the case of . . . ['trunk'] (idiosyncratic ambiguity) we have two formals elements . . . with the same phonetic form. (Chomsky, 1977, p. 69)

In cases like 'book', the lexical entry has a range of interpretations, exhibiting a *flexibility* that permits the kind of mixed use as in (62a). In cases like 'trunk' there are two lexical entries that are *homophonous*, each with distinct and unrelated meanings, rendering mixed use interpretations of expressions like (63a) unacceptable. The

trenchant problem for the externalist hypothesis ($\mathcal{E}$) is that the flexibility exhibited by 'book' is pervasive in natural languages, and poorly captured by theories that treat them as cases of homophony (as seemingly EM theories must).[21]

There are two kinds of flexibility explored in the remainder of this section, one based on non-linguistic knowledge, and one based on ontological type. The truth-conditions for some natural language expressions are not determined by the referents of their constituents, and the manner of their composition. To determine their truth-conditions, competent language users must deploy non-linguistic knowledge of a general sort—too general to be considered the semantic context of utterance, yet too specific to be considered the cognitive context. The second sort of flexibility permits expressions that apply multiple predicates of different types to a single noun phrase. The would-be externalist meanings of such expressions require objects of a(n) (impossibly) bizarre sort.

3.3.2.1 Non-linguistic Knowledge

Consider the following expressions:

(64) Football games are played by jerks.

(65) Residential houses are robbed by jerks.

Sentences like (64) and (65) highlight the fact that we bring extralinguistic knowledge to bear on linguistic expressions in determining the truth-conditions for sentences. The information contained in the *linguistic* properties of expressions, including their meanings, do not determine the truth-conditions of those expressions. As such, meanings cannot be (or determine) truth-conditions, and thereby cannot be externalist.

On any EM theory the difference in the truth-conditions of any two sentences (*modulo* context) must be a consequence of the difference in either their constituents or the manner in which those constituents are syntactically related.[22] So, given that (64) and (65) appear in the same syntactic frames, and that they differ only with re-

[21] As a limiting case of this strategy, treating color terms as massively homophonous at the granularity of *use*, see Rothschild & Segal (2009).

[22] 'John likes Mary' and 'Mary likes John' have different truth-conditions, not in virtue of their constituents, but rather in virtue of the syntactic composition of those constituents.

gard the two constituents 'football-games'/'played' and 'residential-houses'/'robbed' respectively, whatever (64) indicates is true of the relationship between *football-games*, *playing*, and *jerks* should, according to (65), hold true for *residential-houses*, *robbing*, and *jerks*.

However, the expression in (64) means that *every*[23] game is played by jerks, while the expression in (65) emphatically *does not* mean that *every* house is robbed by jerks. The information essential for deriving the truth-conditions for (64) involves the tight relationship between *games* and *playing*, namely that there can be no unplayed game—a relationship that does not hold between *houses* and *robbing*. But this information is not a linguistic property of the expression, and not part of the linguistic meaning of the expression. Thus in deriving the truth-conditions for (64) a competent speaker of English relies on knowledge not present in the expression (nor even the semantic context). As such, the meanings of these expressions cannot be (or determine) their truth-conditions. Semantic properties are properties of expressions. If meanings are (or determine) truth-conditions, then the truth-conditions of any meaningful sentence will be determined by the expression. But such a theory will fail to explain the differences in meaning between (64) and (65) not captured by the difference in their constituency.

This general phenomenon is not limited to generic expressions. We bring non-linguistic knowledge to bear on linguistic expressions in other ways that do not seem to rise to the level of pragmatic inferences. The differing contribution of expressions like 'coffee' as a predicate in 'coffee drink' and 'coffee grinder' present a *prima facie* problem for the externalist. If the semantic contribution of an expression can be recovered by the semantic contribution of its constituents (plus the means of their composition) the meaning of 'coffee' in these expressions should make the same contribution across uses. But a coffee drink is one that is *composed of* coffee, while a coffee grinder is not made up of coffee at all. While these are facts that *any* theory of meaning needs to explain, the externalist will have particular difficulty dealing with this problem, insofar as the contribution of 'coffee' on such a theory is exhausted

[23]Admittedly, these expressions are generics, and speaker judgments in this domain are (seemingly) not concordant. However, even if the expression in (64) is not interpreted with a universal quantifier, (64) indicates (at a minimum) that *most* football games are played by unsavory individuals. In contrast, (nearly) no one will interpret (65) as indicating that most residential homes are robbed, even if most of those robberies are perpetrated by jerks. Since this difference is not syntactic, the externalist is burdened to explain why the relationship between the VPs and the NPs in the two expressions is different.

by its reference to a property—and in particular, a property instantiated by the indicated objects in the various expressions in which it functions as a predicate. But there seems to be no obvious single candidate for the needed property in this coffee-case. The best candidate properties for the would-be denotations of 'coffee' in these two expressions seem profoundly different: *being-composed-of-coffee* and *used-in-the-production-of-coffee* are quite different properties. The difficult task for the externalist is not only in pinpointing the relevant *single* property in such cases, but also in constructing an account of how speakers come to triangulate on such (non-obvious) properties in the many cases that exhibit these features (e.g. 'metal shears', 'home loan[/inspection]', 'rain delay[/coat]', 'blue marker', etc.). This phenomenon is pervasive in natural language, and not easily explained by the externalist.

3.3.2.2 Ontology and Satisfaction

The lexical flexibility exhibited by natural language expressions cannot be accounted for by an EM theory that demands that the meanings of expressions determine a unique referent, as a mind-independent object. Consider the following two English expressions

(66) The Hirshhorn Museum is bankrupt.

(67) The Hirshhorn Museum is a cylinder.

A competent speaker of English could well think that both of these expressions are true. An EM theory accounts for this fact by indicating that such a speaker takes the following conditions to hold in the world: 1) there is a worldly object, specifically the one 'Hirshhorn-Museum' denotes, which satisfies whatever the conditions are for 'is bankrupt'; and 2) there is a worldly object, specifically the one 'Hirshhorn-Museum' denotes, which satisfies whatever the conditions are for 'is a cylinder'. This would require that there is some single object, the thing in the world 'Hirshhorn-Museum' picks out, that can be both bankrupt and a cylinder.

The expression in (66) requires that this object be a financial institution, understood through an array of socio-economic notions. Whatever these notions demand about the nature of financial institutions, be they collections of individuals—perhaps investors, or a board of directors—or something more abstract, such objects do not seem to have a *shape*. Similarly, the kind of object that would satisfy the

predicate in (67), in this case a building, is seemingly not the sort of thing that can have financial troubles. Building projects can have fiscal crises, but buildings seemingly do not. The EM theorist then either owes us an account of the kind of object that can satisfy both (66) and (67), or they must explain why this problem does not generate in the first place.

Taking the latter strategy, an EM theorist might appeal to the presence of ambiguity in natural languages, as in the following:

(68) The geese are by the bank.

The expression in (68) is ambiguous, insofar as one can interpret 'bank' in two ways: the notion that refers to a feature instantiated by rivers, or the notion that refers to a financial institution. That is, we seem to have two distinct, though homophonous, lexical items 'bank$_r$ and 'bank$_f$'. The ambiguity of (68) is then explained by appeal to the homophony of these two distinct lexical items. The expression in (68) simply fails to determine which 'bank' is being used, and as such the expression can have different meanings based on which item is deployed.

Likewise, one could argue that we really have two lexical entries for 'Hirshhorn-Museum', one that denotes the institution, and another that denotes a building. We can represent this difference between (66) and (67) as

(66′) The Hirshhorn-Museum$_f$ is bankrupt.

(67′) The Hirshhorn-Museum$_b$ is a cylinder.

Since 'Hirshhorn-Museum' identifies two distinct (thought homophonous) lexical items, contextual information determines which item is used in (66) and (67) respectively, preserving the distinct meanings of the expressions, while assuaging the worry that an EM semantics requires a single referent for the DPs in the two expressions. There need not be some single object that satisfies 'is bankrupt' and 'is a cylinder', since 'Hirshhorn-Museum$_f$' and 'Hirshhorn-Museum$_b$' can (must!) denote different objects.

But this reply will not do. The same speaker that endorses (66) and (67) could well endorse (69)

(69) The Hirshhorn Museum is bankrupt and it is a cylinder.

Whatever the technical details are that govern anaphoric meanings, 'it' in (69) must derive its meaning and referent from 'Hirshhorn-Museum'. Whichever lexical entry

the context supplies ('Hirshhorn-Museum$_f$' or 'Hirshhorn-Museum$_b$') the meaning of the anaphoric 'it' is exhausted by the referent of whichever lexical item is demanded by the context. Thus, for the EM theorist to explain how competent speakers treat (66)–(69), there must be some single worldly object that can be both a cylinder and bankrupt. So even if we grant the EM theorist his homophonous response, the ontological concern remains.

Pietroski (2005) also notes that while a natural language speaker could endorse expressions like (66)–(69), the sentence in (70) is strange in a way (69) is not

(70) # The Hirshhorn Museum is a bankrupt cylinder.

The oddity of (70) in conjunction with the acceptability of (69) (and the plethora of sentences like them) is unexplained by a semantics that treats meanings as truth-conditions. On any EM theory, the way the world would have to be in order for (69) to be true would also make (70) true: there is some object, the referent of 'Hirshhorn Museum' that is both bankrupt and is a cylinder.[24] Insofar as such a hybrid satisfier is not to be found in the domain of worldly objects, this speaks against EM theories of natural language meanings.

Notice too, that the nature of these bizarre objects is distinct from those typically associated with Chomskyan critiques of externalism. Unlike *flaws* (Chomsky, 1981), or *the average man* (Hornstein, 1984) the objects needed to satisfy expressions like (69) are not simply *abstracta* developed for the purposes of theorizing about statistics and economics (Ludlow, 2011, pp. 135–136) intricately characterized by logics that appeal to standards (Kennedy & Stanley, 2009). Whatever hybrid objects are, they are far less familiar than mere abstractions, and should be less palatable to the externalist.

But, the Realist might bite this bullet, as some semanticists do (Ludlow, 2003, 2011). They might just stipulate that the domain contains objects that are at once

[24] While concatenating predicates does not always yield an expression with the same meaning as conjuncting them, this does not seem like such a case. Consider:

(1) This is a fake diamond.

(2) This is fake.

(3) This is a diamond.

While (1) implies (2), it does not entail (3)—in fact (3) must be false if (1) is true. This and other examples are problematic for straightforward applications of concatenating predicates, but the difficulty here is not an ontological one. The sort of objects that satisfy 'is a diamond' can also satisfy 'is fake'—namely physical objects.

both concrete and abstract. However, this bullet biting is both unmotivated by the externalist argument, and tastes far worse than the Realist might suspect. The externalist *hypothesis* contends that natural language meanings can be rendered comprehensible, without appeal to the mysterious existence of things like *Sinnen*. The externalist proposal is that, given the pedestrian objects of the world like chairs and rabbits—objects that we have good *antecedent* reasons to posit—a theory of meaning can be developed given only this domain of pre-theoretically plausible things. Thus, the theory garners intuitive support because we are not forced to accept into our domain a vast hierarchy of bizarre objects, like *Sinnen*.

In this light, the Realist's bullet biting is quite strange. As a means of avoiding the troubling consequences of viewing meanings as mysterious *Sinnen*, the Realist accepts the existence of hybrid abstract-concretia (concrete-abstracta?), trading the mysterious for the bizarre. The further point is that in accepting the existence of ontologically bizarre hybrid objects, the Realist is *not* merely acknowledging the ontological entailments of natural language usage. Rather, she is making a prediction about the kinds of things we should expect to find in the domain, if the externalist hypothesis is correct. In the absence of any pre-theorectical evidence that there are such things, she bears the burden of providing a good reason for believing that such things exist, beyond the fact that their existence supports her hypothesis.

Let me dwell on this point. The dialectic at this juncture on behalf of the externalist is that ($\mathcal{E}$) is (a part of) the best theory of meaning on offer, since the other alternatives (like accepting mysterious *Sinnen* as meanings) are less palatable. Famously, Lewis (1986) argues that our best semantics for modal expressions posits a vast plenitude of extant possible concrete particulars that stand in spatio-temporal relations to form possible worlds. The argument on offer contends that these proliferate posits are necessary because the cost of failing to accept these concrete *possibilia* is too high. Namely, we're left with a woefully inadequate explanatory theory of meaning. But, this strategy is often misconstrued. These *possibilia* are not objects we merely discover in the process of analyzing meanings. Our use of modal expressions in natural language reveals a commitment on behalf of the natural language speaker to the existence of *possibilia* (concrete or otherwise) only if one *assumes* an externalist semantics. As Ludlow rightly puts it,

> . . . ontology is tied to the demands of our scientific theory of the semantics of natural language, and not the kinds of entities and objects that members of

a particular culture might believe in" (Ludlow, 2011, p. 142).

To engage in the latter kind of investigation is to do psychology—or to connect with philosophical tradition, to do natural language (descriptive) metaphysics. But, the Realist is investigating the structure of the world, not the structure of our minds. As such, to indicate that the *theory* demands certain metaphysical posits is a burden the theory has to bear, not an analytic consequence of the fact that linguistic expressions have meanings. *If* ($\mathcal{E}$) is the correct theory, then there are such things as concrete *possibilia*. In this vein, some argue that we need even more things: situations (Barwise & Perry, 1983), perspectives (Schein, 2002), modes of presentation (Ludlow, 1995), and fictional objects (Thomasson, 1999). But we only need these things if externalism is the best theory of meaning—*a fortiori* that the theory requires a domain with such things counts against the merits of the theory. Adding to this list of posits the abstract-concretia required to address lexical flexibility, the class of entities required to support the externalist hypothesis looks less and less like the pedestrian objects of everyday experience. Concomitantly, the externalist hypothesis looks less like a plausible theory that makes use of the everyday objects we are familiar with, and more like a theory with implausible commitments.

But maybe our dislike for the bizarre is unwarranted, based on some vestigial aspect of our human conceptual machinery. Maybe these abstract-concretia exist, happily residing in the domain, and validating the Realist's predictions. Or maybe hybrid objects are more palpable than *Sinnen*.[25]

The Realist's perverse predictions, however, do not end there. The externalist hypothesis predicts the existence of other hybrids, including hybrid properties, and relations given the lexical flexibility of expressions like 'bilingual', 'cut', and 'drive'.

In a conversation about the ability to acquire a second natural language, one might use the following English expressions:

(71) The child is bilingual.

(72) John is bilingual.

Likewise, one could (roughly) conjoin the meanings of these two expressions into a single sentence using either of the following acceptable expressions:

[25]Notice how this vein of reply assumes a false dichotomy, that one either accepts mysterious, reified meanings as things, or accepts whatever things an externalist semantics requires. For other alternatives again see Hinzen (2007, 2014); Pietroski (2005, 2008, 2010, forthcoming).

(73) The child is bilingual and so is John.

(74) The child and John are bilingual.

Of course, an externalist semantics can accommodate the meanings of these expressions, whereby their truth-conditions are satisfied just in case there is a (salient) child, a John, and both of them instantiate the same particular property. To put the matter somewhat formally, (73) and (74) are true just in case:

(73′) $\iota x.\ \exists y.\ \textsc{child}(x)\ \&\ \textsc{john}(y)\ \&\ \textsc{bilingual}(x)\ \&\ \textsc{bilingual}(y)$

In the sentences (71)–(74), the expression 'bilingual' has a single, univocal meaning, as reflected in the single truth-conditional predicate 'BILINGUAL'. For the externalist this identifies some *single* property, say the property had by all things that acquire/speak/know two languages. As such the sentences in (71)–(74) are well captured by an externalist semantics.

But 'bilingual' is lexically flexible. While walking the streets of London, Ontario, Canada, I came across an empty box outside a franchise sandwich shop. Printed on the outside of this box was the expression 'bilingual napkins', which presumably identified the box's contents as napkins on which information is printed in two different languages. Sitting in this franchise with a group of friends, and noticing the features of one such napkin, a competent speaker could well say to their compatriots the following acceptable sentence:

(75) The napkin is bilingual.

Supposing this group also knew our bilingual friend John, a competent speaker could say, and the group would no doubt accept as true, the following acceptable expression:

(76) The napkin is bilingual and so is John.

However, the following truth-conditions, which for the externalist are determined by the meaning of (76) are *not* satisfied in this situation:

(76′) $\iota x.\ \exists y.\ \textsc{napkin}(x)\ \&\ \textsc{john}(y)\ \&\ \textsc{bilingual}(x)\ \&\ \textsc{bilingual}(y)$

Since both bilingual predicates in (76′) are derived from the single use of 'bilingual' in (76), they must have the same truth-conditions. As such, for the externalist they

must pick out the very same property. If that property pertains to the acquisition of multiple languages, the napkin clearly fails to instantiate such a property. And if the property pertains to the kinds of orthography printed on a thing's exterior, John does not count as having such features. For the externalist, (76) turns out to be false in the situation described—the same situation that makes both (72) and (75) true. Given that (76), on an externalist semantics, has the meaning attained by conjoining (72) and (75) (as indicated by the relationship between (71)–(73)), a theory of meaning should predict that competent speakers accept that (76) is true.

As before, the externalist could claim that 'bilingual' is homphonous, with two lexical entires $bilingual_a$ and $bilingual_o$ pertaining to the aforementioned acquisitional and orthographic properties (respectively). But just as in the example regarding the Hirshhorn Museum, this reply will not do. Whichever lexical entry the context demands, the truth-conditional predicates in (76′) will have the same satisfaction conditions as each other. The $bilingual_a$ interpretation of these predicates leaves the first unsatisfied by the napkin, while the $bilingual_o$ interpretation leaves the second unsatisfied by John. Either way, externalism makes the wrong prediction that, relativized to the situation, (76) is (treated as) false (by competent speakers). The externalist's must seemingly accept that there is some single bizarre, hybrid (or multifarious) property that admits to having shifting satisfaction conditions within a single context.

As with the flexibility exhibited by 'Hirshhorn-Museum', the Realist could bite this bullet. She can simply accept that the domain contains not only hybrid abstract-concretia, but also hybrid properties such that the very same property can be instantiated in distinct ways by disparate objects within the same context.[26] But again, as with the Hirshhorn-Museum, even this (bizarre) concession fails to explain the distribution of competent speaker judgments. Consider

[26]To be clear, this is not an injunction against multiple realizability. To take the paradigm case, mental properties like *belief* can be realized in Martian brains just as well as human ones. But what is instantiated in these distinct organisms is (say) an entity that plays a particular functional role in the mind of the organism, and indeed the *same* functional role. The worry here is not that the napkin and John embody different ways of instantiating the same property, but that they instantiate *different* properties. Contrast this difference with the manner in which distinct humans are bilingual. John Kerry and Nicolas Sarkozy are both bilingual (let's suppose), yet the bilingual property is realized in each person in distinct ways. At the very least, they differ regarding the languages in which they are fluent: Kerry is fluent in English and French, while Sarkozy is fluent in French and German. The manner in which Kerry and Sarkozy differ is quite clearly not the manner in which John and the napkin differ regarding bilingualism.

(76) The napkin is bilingual and so is John.

(77) # The napkin and John are bilingual.

The expression in (77) is strange in a way that (76) is not. This strangeness, whatever it amounts to, seems to be a fact about the meanings of the constituent expressions of the sentence, since there is no general prohibition against joint predication (as exemplified in (74)). The syntactic frames of (73) and (74) are repeated in (76) and (77), yet the latter do not bear the same semantic relations to one another as the former, insofar as they are *not* synonymous (as should be clear by the oddity of (77)).

[Keats example maybe . . .]

The externalist might respond to these problems not by masticating more and more metaphorical cartridges, but by indexicalizing flexible expressions.[27] On this view the semantics of lexically flexible expressions are treated as indexical expressions, whose extensions are indeterminate, absent a context of utterance. Much the way that 'I' needs inputs from the context of an utterance that contains it (specifically information about the *speaker* of the utterance), uses of 'bilingual' and the like require information from the context of utterance to determine their extensions. Formally, the proposal would be that lexically flexible expressions have a semantic character as their linguistic meaning. As a function from contexts to extensions, these character-meanings are what uses of a given flexible expression have in common.

In the case of 'bilingual' above, a rough gloss on the formalism provided above in (76′) for (76), once indexicalized would yield a form like

(76″) $\iota x.\ \exists y.\ \textsc{napkin}(x)\ \&\ \textsc{john}(y)\ \&\ \textsc{bilingual}_i(x)\ \&\ \textsc{bilingual}_j(y)$

The added indexing requires that the context provides, *via* the interpretation function, a mapping to the proper denotation of each BILINGUAL predicate. This enables the predicates to have different extensions, depending on the mapping provided by the interpretation. The consequence is that different mappings can reflect the differing ways in which the napkin and John might be bilingual, picking out difference properties as the extensions of the two BILINGUAL predicates. In this way the externalist can accommodate the judgments of competent speakers regarding these

[27] See Burge (1979b) and DeRose (1992) for examples of this strategy regarding 'true' and ''know' respectively.

flexible expressions, at least insofar as the truth-conditions indicated by (76″) (once contextually bolstered) can be satisfied by the sandwich shop situation.

There are many reasons to reject this strategy.[28] Many of these reasons are variations on a general point, namely that most natural language expressions do not behave like indexicals. The response offered by the externalist to the context sensitivity of "classical" indexical expressions (e.g. 'I' and 'now'), as manifest in the character-content distinction, pertains to quite general features of utterance contexts. After all, indexical expressions like 'I', 'here', and 'now' track features of contexts of utterances that are (nearly) universal: utterances have speakers (s), they are always uttered somewhere (l), and at a particular time (t) to someone (a). In proposing that the interpretation function, which is meant to model the process by which speakers understand the meaning of *any* expression, includes provisions for allocating referents to expressions in accordance with various speakers, times, locations, and audiences, the externalist can plausibly highlight the universal nature of these features as a good reason to suspect that speakers track this information quite broadly. Put another way, even in the absence of indexical expressions that require a speaker, location, or time, a competent language user will have good cause to be sensitive to these features of any uttered expression. This is a good reason to build into the interpretation function a mechanism from deploying this information in determining the content of an expression, like 'I' or 'here'.

But a similar argument for introducing analogous machinery into the interpretation function to accommodate lexically flexible expressions is far less fetching. Foremost, the information supplied by the context in which flexible expressions are used is not universal. For example, whether or not an utterance requires an abstract or concrete interpretation of the noun phrases expressed therein is not universal to all contexts of utterance. Relatedly, the required contextual information that addresses which interpretation is appropriate for lexically flexible expressions is more fine-grained than what we tend to find with indexicals. To take the proposed truth-conditions for (76) in (76′), the information needed to determine which predicate is meant is specific to the meanings of the other expressions in the sentence.

(76′) $\iota x.\ \exists y.\ \textsc{napkin}(x)\ \&\ \textsc{John}(y)\ \&\ \textsc{bilingual}(x)\ \&\ \textsc{bilingual}(y)$

[28] See Cappelen & Lepore (2005) for an extensive critique of this approach.

That the first predicate picks out a property pertaining to orthography, for example, is ascertained by the nature of the predicate's purported satisfier. In contrast, indexical expression like 'I' or 'here' do not require information of this granularity. Competent speakers can ascertain the referent of indexicals in complete ignorance of the meanings of other sentential constituents. To illustrate, consider

(78) I fendled.

(79) John likes the bilingual dax.

(80) John speffed the Hirshhorn Muesum.

An audience to my utterance of (78) need not know what it means to 'fendle' in order to flesh out the content of the indexical. Contrastingly, my audience need to know what a 'dax' is in (79) to determine which bilingual property is referenced. If daxes are like humans, the utterance demands an acquisitional property. If they are like billboards, an orthographic property is required. Likewise, if speffing is a construction method used by contractors (like John, say), the object needed for the DP in (80) is a building. But if speffing is a special kind of legal action 'Hirshhorn Museum' must refer to a social institution. The point is that, while natural language speakers can recover the intended referent in these cases, given this fine-grained context, classical indexicals do *not* require such fine-grained, idiosyncratic information for their semantic resolution. The special treatment of classical indexicals, and the global change to the interpretation function that results, is justified because *every* context contains the kind of information needed for their resolution. This special semantic status is important. But, on the indexicalist approach to lexical flexibility, the special character of classic indexicals is no longer accounted for in the model. Lexical flexibility is a pervasive phenomenon. If we must treat such expressions indexically, then the special semantic character of classic indexicals is no longer captured by the semantic theory.

The externalist, and the Realist, in order to explain the distribution of competent speaker judgments for sentences containing nouns like 'the Hirshhorn Museum' and predicates like 'bilingual' is then compelled to accept some rather bizarre entities into the domain of worldly things: hybrid objects the exhibit a concrete-abstract duality, and properties the have varying conditions for instantiation across instances within a single context. Neither of these concessions seems pleasant. Worse yet, conceding

in these ways still leave unexplained aspects of competent speakers' judgments, as can be seen by the contrast between (76) and (77).

Finally, consider the following:

(81) The tractor was easier to drive than a heard of cattle.

(82) The lawn was harder to cut than her hair.

(83) Napolean's defeat was worse than Kasperov's.

Given the acceptability of sentences like these, relational expressions like 'drive', 'cut', and 'defeat' seem to exhibit lexical flexibility as well. If the Realist is forced to accept the existence of hybrid *relations* as a result such flexibility, the pedestrian nature of the objects needed to accomplish the externalist's aims is substantively undermined—especially if this bizarre ontology remains insufficiently explanatory. Rather than viewing such ontological commitments as the price to be paid for an adequate theory of meaning, such requirements might be better seen, or so I contend, as a *reductio* against the hypothesis that requires them.

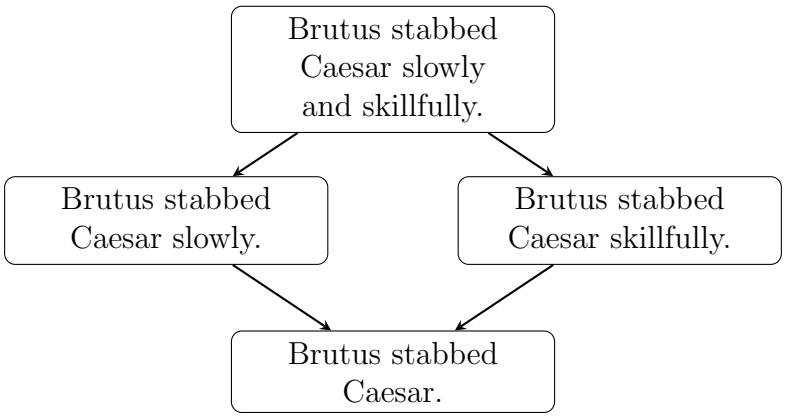
3.3.3 An Externalist Reply

The externalist might reply to these worries by leveraging the purported virtues of an externalist semantics. She might contend that while the flexibility cases rehearsed above might require bearing an ontological burden, the burden is better than the alternative. As a part of an overall externalist theory, the virtues of the theory far outweigh these burdens. One principal virtue that supporters of externalism might trumpet is that understanding meanings in this way uniquely preserves semanticists' main source of linguistic data, the truth-value judgments of competent speakers. To quote a prominent externalist:

> In short, intuitions about the truth and falsity of what is said by utterances of sentences have formed the data by which theorists have tested their hypotheses about meaning. There is no other obvious source of native speaker intuitions that are related to meaning. So if we did not have robust intuitions about the truth-conditions of our utterances, it would not be clear how to test such hypotheses; there would be no firm basis on which to construct a theory of meaning. (Stanley, 2007, p. 7)

Consider for example, the landmark insight of Davidson (1967b) in treating the logical form of actions sentences as involving quantification over *events*. The

Figure 3.1: Diamond-Shaped Inferences



sentences in (84) display a particular pattern of inference, as indicated in Figure 3.1, wherein the arrows represent the direction of inferences that speakers of English are apt to make.

(84) a. Brutus stabbed Caesar slowly and skillfully.
b. Brutus stabbed Caesar slowly.
c. Brutus stabbed Caesar skillfully.
d. Brutus stabbed Caesar.

Davidson's proposal means to capture these patterns of inference. Proposing that the logical forms in (84′) are indicative of the truth-conditions of the sentences in (84), this approach captures the inferential judgments of English speakers, by modeling these inferences as logical entailment.

(84′) a. $\exists e[\textsc{stab}(e, \text{Brutus}, \text{Caesar}) \;\&\; \textsc{slow}(e) \;\&\; \textsc{skill}(e)]$
b. $\exists e[\textsc{stab}(e, \text{Brutus}, \text{Caesar}) \;\&\; \textsc{slow}(e)]$
c. $\exists e[\textsc{stab}(e, \text{Brutus}, \text{Caesar}) \;\&\; \textsc{skill}(e)]$
d. $\exists e[\textsc{stab}(e, \text{Brutus}, \text{Caesar})]$

Because the proposed logical forms for the expressions in (84) quantify over *events*, the expression in (84′a) entails the other expressions in (84′) by way of conjunction reduction. As such, the "diamond-shaped" inference patterns of speakers are captured by an *externalist* theory that takes events as the satisfiers of expressions.

This reasoning only makes sense if the *explanada* of the hypothesis are the inferential judgments of speakers, as judgments about the concomitant truth of collections of sentences. Thus the purported justification for an externalist semantics is that it maintains the theoretical import of speaker judgments. To deny that sentences have externalist meanings is to deny the connection between meaning and truth that renders these judgments worthy of capture. Thus the externalist might contend, if meanings have nothing to do with truth, then these truth-value judgments are not indicative of expression meanings and of no use for semantic investigation.

However, denying the externalist thesis ($\mathcal{E}$) does not also require denying that the truth-value judgments of speakers are relevant data for the purposes of semantic theorizing. Externalist are committed to a *particular* relationship between truth and meaning—namely the one codified in ($\mathcal{E}$). In denying this, a semanticist need not deny that meaning is related to truth. She must simply deny that meanings *determine* truth-conditions. One can hold that natural language speakers can use sentences to make utterances that are true, and still deny the externalist thesis. And this can be done without denying that there is some "systematic" manner in which meaning is related to truth (*pace* Stanley, 2007, p. 8). Such a view merely holds that the systematic manner in which linguistic meaning relates to the external world involves the interaction of multiple non-linguistic cognitive and external systems that connect in complex ways. This complexity can even be systematic, but because many of the systems involved are extra-linguistic (and not semantic) the meaning of an expression will not, in the absence of this complex interaction, determine its truth-conditions. One can, as Stanley does, amalgamate this motley group of disparate non-linguistic systems[29] under the term 'context'. But as we'll saw in §3.3.1, the gap between what a sentence means and what a speaker communicates in conveying that meaning *via* a linguistic utterance admits to some marked complexity. Respecting, and not merely masking, the complexity of this

[29]The diversity of components that collectively make up the "context" of an utterance, so construed, is important to note. Shared human systems that recognize gaze following, emotional facial gestures, object detection, agency detection, and many others, not to mention the external physical systems that govern "normal" visual and auditory environments, all fall under the "context" that determinately links meanings to truth. But if one wants to know how meanings differentially interact with these various systems in order for a speaker to utter something true, abstracting over these differences by indicating that the context somehow fills this gap is no answer at all. Worse yet, it commits one to a theory of meaning that is thereby incapable of addressing such questions.

relationship between the meaning of a sentence and the truth of an utterance not only preserves the (nuanced) use of truth-value judgments as linguistic data, but it opens up *new* sources of data (cf. Pietroski et al., 2009; Lidz et al., 2011; Vogel et al., 2014). So, far from making the semanticist's task impossible (or without basis) denying ($\mathcal{E}$) expands the data-set for the theoretician, while preserving the utility (though augmenting the informativeness) of speaker judgments.

Another avenue of reply for the externalist is to focus of the process of pragmatic inference. The suggestion here might be that the best account of pragmatic phenomena requires that the inputs to pragmatic processes be truth-conditional. The claim might be that our best accounts of phenomena like conversational implicature rely on a distinction between the truth-conditional and non-truth-conditional aspects of an utterance's meaning (cf. Grice, 1967 [1989]). But a broadly Gricean picture need not depend on a distinction between the truth-conditional and non-truth-conditional aspects of an utterance. As we saw in §3.3.1, characterizing the phenomena does not require that literal meanings are truth-conditional. Of course that brief review does constitute a theory, but distinguishing between the literal meaning of an expression and the pragmatic inferences an audience might make as a result of hearing an utterance with such a meaning need not depend on truth-conditions. They *do* depend on a distinction between an expression's literal meaning and the conditions in which the expression can be used, but this distinction need not trade on truth. All that is needed is a means of distinguishing between the meaning of an expression and what the speaker using that sentence intends to communicate by way of such use. One can of course hold that expression meanings are truth-conditions and communicative intentions are not. But one can equally well distinguish meanings as instructions to build representations (Pietroski, 2010), while marking the communicative intent of speakers as being characterized by the representations a speaker hopes their audience to entertain as a product of implementing those instructions. The distinction between literal meanings and the thoughts an audience entertains as a consequence of an utterance need not be made by appeal to truth-conditions, even if some distinction along literal/non-literal lines must be made.

3.4 Mental Content

Arguments against an EL view push an externalist to adopt externalist views about *mental content*. If languages are not external objects, but rather aspects of the human/mind brain, then an externalist is committed to the view that the content of those mental representations can be characterized externally, as relations between representations (or concepts) and worldly objects. As such, some arguments for internalism address both externalist theories about mental content, and the relationship between mental content and linguistic meaning. In this section I present these concerns.

3.4.1 Naturalist Theories of Content

Both sentences and thoughts seem to be about the world, and thereby exhibit intentionality. The close proximity of these disciplines gives rise to a simple solution to the problem of intentionality for language. Namely, that the problem of intentionality is solved at the level of thought, not language. An enticing view about the relationship between thought and language is that the contents and structure of our thoughts are merely mirrored in language. If the structure of natural language mirrored the structure of thought, wherein an expression in a language is merely a way of making public some particular thought composed of conceptual content, then the intentionality (and meaning) of an expression simply tracks that of the concepts used to compose the expressed thought. On such a view, natural language expressions are merely labels for thoughts, and likewise, words are merely labels for concepts, as a way of making them articulatable. Call this the *label theory* of linguistic meaning.[30]

So long as the language I speak syntactically composes in a way commensurate with the structure of my thoughts, linguistic meaning would perfectly mirror conceptual content. On this proposal, the meanings of our expressions would hook up with the world *via* conceptual content, so long as conceptual content can be characterized externally. So, to the degree that our concepts align with the "fine structure" of the world, expressions of a natural language will likewise accord with

[30]Jackendoff (2002) explicitly adopts this position, though not by this name. Oddly enough he defends an internalist proposal for linguistic meaning on the basis that no naturalistic account of external mental content is plausible.

the mind-external objects of the world.

But why should one insist that linguistic meaning is mediated by our conceptual system? For one, this answers the problem of intentionality at the level of language. But more importantly for the Realist, the desire here relates to naturalistic explanation. As we saw, a substantive source of contention in thinking about linguistic meaning as externalist relates to the aims of the scientific enterprise of linguistics. Facts about the acquisition and productivity of language in humans deserve explanation, and a theory of meaning ought to add to (or at least make possible) an explanation of these facts. As we saw in §3.2.1 these considerations strongly suggest that the object of study for a naturalistic investigation of language is in the mind. As such, to the degree one thinks that language and thought are independent, a naturalistic inquiry into these matters will address the way in which these distinct mental faculties interact.

There is no shortage of literature addressing the viability of naturalistic accounts of content.[31] Whether or not naturalistic accounts of intentionality are viable is beyond the scope of this work, but the point I want to emphasize here is that the force of the arguments presented so far against externalism compel the externalist to adopt two contentious views: a labeling theory of linguistic meaning, and a naturalistic account of intentionality. I'll not take to the time to illustrate the contentiousness of the latter,[32] but the former position is worth analyzing, in part because so many philosophers seem to adopt this view without much defense.

The labeling view of the relationship between words and concepts is that words are like labels for concepts. This view is pervasive amongst philosophers. In fact, the view is often adopted as obvious, without much need to articulate that indeed adopting the view embodies a collection of commitments about the relationship between the human language faculty and the conceptual system. Burge (1979a) is a paradigmatic example. In discussing the expansiveness of his famous 'arthritis' case, Burge writes:

> On the other hand, the [arthritis] thought experiment does appear to depend on the possibility of someone's having a propositional attitude despite an incomplete mastery of some *notion* in its content ... Suppose a subject thinks falsely that all swans are white ... that 'swan' means 'white swan' (Burge,

[31] For a good survey see the introduction to Macdonald & Papineau (2006).

[32] This has been done by many, and better than I could hope to do here. See Jackendoff (2002); Loewer (1997); Boghossian (1991); Godfrey-Smith (1989); and McGinn (1982).

1979a, p. 83) (my emphasis)[33]

Burge treats the content of a concept, or *notion*, which plays an important role in determining the content of the propositional attitude someone might hold, as no different than the meaning of a natural language word: hence the notion SWAN has the same meaning as 'swan'. In his book-length critique of Burge's account of wide-content Segal (2000) commits to this same theory about the relationship of words to concepts.

> Zowie and Twin Zowie both say "My engagement ring is studded with diamonds." Are the concepts expressed by their words "diamond" the same? (Segal, 2000, p. 6)
>
> ...
>
> - Let w be the focal word
> - Let c be the concept [the subject] expresses by w.
>
> (Segal, 2000, p. 67)

In more contemporary literature, Weber (2005) writes

> The meaning of the term "gene" has changed several times in the history of twentieth-century genetics. If we distinguish between a term's sense and its reference, it is possible that the term's sense has changed, but not its reference... I have examined both the reference potential and the reference connected with different historical versions of the gene concept. (Weber, 2005, p. 228)

Linguistic meaning, on this view, is simply conceptual meaning. But this view has the following consequences: first, the extension of our words must have the same extension as their underlying conceptual meanings, and second, the syntax of natural language must be mirrored in the composition of thought. Neither of these consequences seem well supported by the way natural language speakers treat the meanings of expressions.

The flexibility of natural language expressions speaks against the first consequence. If the extension of the concept HIRSHHORN-MUSEUM is going to attempt to capture the meanings natural language users apply to the term 'Hirshhorn-Museum' then the extension of the concept better include both the concrete building that

[33]Here Burge uses 'notion' as a way of talking about the content of a concept: "Talk of notions is roughly similar to talk of concepts in an informal sense"(Burge, 1979a, p. 83).

houses artworks, and the abstract institution that employs hundreds of people. As we've seen, many natural language expressions bear meanings that do not track the domain of objects in this way. Thus whatever thought corresponds to (69), and thereby stands as the meaning of (69), it must either treat 'Hirshhorn-Museum' as labeling two distinct concepts, or have a content such that some (abstract) object (or some building) is both cylindrical and bankrupt. The former avenue belies the manifest relationship between the uses of the English term, and fails to account for the felt relatedness of these uses. The latter option has much more bizarre metaphysical commitments, since we have no other reason (other than a commitment to particular views about semantics) to postulate such an entity.

Turning to the second consequence, if language mirrored the structure of thought, then thoughts should compose much the way expressions do. More strictly, the meanings of linguistic expressions and their underlying logical forms should mirror the structure of the concepts those forms express. The deep structural syntactic frames that make up interpretable expression in a natural language must mirror the structure of their underlying concepts. In this vein, consider the following sentence:

(85) Wilbur kicked Fred.

Paying attention to the syntax of this construction, and adopting the labeling theory commitment, we ought to conclude that the KICK concept is dyadic. The word 'kicked' in the complete expression in (85) takes a subject and an object, and likewise we would expect the related concept to take two elements to form a complete thought. Thus the thought expressed must make use of a concept like:

(86) $\textsc{kick}(\_\__s, \_\__o)$

which when saturated with two elements, makes the complete thought

(87) $\textsc{kick}(\textit{Wilbur}, \textit{Fred})$

However, if this dyadic notion of KICK is supposed to underlie all meaningful uses of 'kick', as implied by the labeling theory, the following expression is an apparent counterexample:

(88) Wilbur kicked Fred with his foot.

Given the syntactic structure of the expression in (88), the underlying conceptual meaning must have a *triadic* structure, to make room for the instrument used in the kicking:

(89) KICK($\_\__s, \_\__o, \_\__i$)

For any way of differentiating concepts, surely addicity falls under the identity condition for a given concept. That is, concepts with different addicities must be *different* concepts. So, (86) and (89) cannot be the same concept. As such, the meaning of 'kicked' in (85) and (88) is different on the labeling view, insofar as 'kicked' labels concepts that are (of) different (addicities). This entailment leaves unexplained why a competent speaker would find that both (85) and (88) are felicitous descriptions of the same kicking.

A defender of the labeling view might hold that really we have only one KICK concept, with sufficient addicity to accommodate all uses of 'kick', and thereby holding their meaning constant across various uses. As such, the concept in (89) is the only KICK concept, made use of in expressions where the instrument of the kicking is unmentioned.

There are three problems with that response: first, this requires that many expressions that make use of the transitive 'kick' have implicit content of an unspoken instrument. And there seems to be no syntactic evidence that such expressions have any such implicit content. Second, such a triadic concept will not capture the meaning of expressions like

(90) Wilbur kicked Fred the ball.

(91) Wilbur kicked Fred the ball with his toe.

The underlying conceptual meaning for (91) must have a tetradic addicity, to make room for the indirect and direct objects in the syntactic structure of the expression. Insisting here that the single conceptual meaning for 'kick' is a *tetradic* concept is implausible. While one might entertain the plausibility that transitive uses of 'kick' leave some unspoken implicit content about what instrument was used in a given kicking, surely such uses do not leave the existence of (nonexistent??) indirect objects implicit, as would be required if

(92) KICK($Wilbur_s, Fred_{do}, nothing_{io}, foot_i$)

was the underlying conceptual meaning of

(85′) Wilbur kicked [nothing to] Fred [with his foot].

And plainly, (85) does not mean what (85′) means.

The third problem for the super-addicity move pushes in the opposite direction. Sentences like

(93) Wilbur kicked.

seem to require conceptual meanings that are monadic. Insisting here that the expression in (93) really contains implicit content that reflects the underlying tetradic conceptual structure in (92) strains good explanation.

Of course, the label theorist could respond to this data by treating all this as evidence that there are really multiple words 'kick' with multiple KICK concepts as their meaning. There are, on this reply, multiple homophones 'kick' each with a different concept depending on whether they have direct objects, instruments, and/or indirect objects. However, such a response treats the difference between these uses of 'kick' like the difference between 'kick' and 'punch'—they are different words, with different conceptual meanings. This of course leaves unexplained what is obvious, that the many uses of 'kick' describe quite similar actions, and are conceptually related. The events these various uses of 'kick' describe bear striking features in common—those features that make them plausible kickings in the first place. A theory about the relationship between the meaning of words and the content of concepts that leaves such basic facts unexplained is troubled.

The point then is this: the relationship between linguistic meaning and conceptual content is not nearly as simple as the labeling theory would have it, as the case of (the addicity of) 'kick' and KICK shows. And addicity is but one feature of the relationship between lexical items and concepts that admits to some *prima face* complexity.[34] For the externalist, this should be troubling news, since this means

[34]Other aspects of the content of our concepts, apart from their structure highlight the complex connection between words and the concepts that underwrite them. Consider the different ways in which the predicate 'is blue' applies to objects, and what this says about the complex application of the BLUE concept.

(1) The house is blue.

(2) The marker is blue.

(3) The iris is blue.

(4) The sky is blue.

The truth-conditions for the color predicate in (1) that would make it true of some house, would not, when applied to some marker, make (2) true, despite the fact that they appear in the same syntactic frame. So if the meaning of expressions are a result of the satisfacion conditions of their

the path from the meaning of a term, through the content of the associated concepts, to their worldly extensions is rather complex. In short, the IL-EM theorist is committed to two views about language and content that are contentious, one of which (given our discussion here) seems implausible. Not only are such theorist saddled with giving a sufficiently plausible *naturalistic* account of mental content, they are also saddled with the troubles articulated here for the label theory of linguistic meaning.

3.5 Natural Language and Ontology

The arguments thus far have been multifaceted, but direct. What they have shown is that ($\mathcal{E}$) is a difficult thesis to defend. In this closing section I'll indicate how this difficulty undermines the Realist's metaphysical methodology. Primarily, if ($\mathcal{E}$) is dubious, then appealing to the truth-conditions of natural language expressions, or the satisfaction-conditions of their purported conceptual meanings, as justification for metaphysical conclusions is without foundation.

This is especially true if the human ability to construct complex thoughts from different conceptual domains depends on the human language faculty (Carruthers, 2002; Spelke, 2003; Jackendoff, 1990, 1996; Bloom, 2000). Even if we grant that our concepts have satisfaction-conditions that accord with the structure of reality,[35] once those concepts are put to work by the language faculty in building meanings to sentences, there's no assurance that the content of the resulting construction will retain such a tight connection to the world.

The problems detailed thus far for ($\mathcal{E}$) undermine the fruitfulness of the Realist's default metaphysical methodology. Ontological investigation proceeds by analyzing natural language usage. Roughly, the Realist makes use of her competence with a given natural language, since such competence assures that she understands the meanings of natural language expressions, and under the guise that such mean-

underlying concepts, the satisfaction conditions for the thoughts BLUE(HOUSE) and BLUE(MARKER) are not merely going to be atributable to the differences int he extension of HOUSE and MARKER—*mutatis mutandis* for (3). And while many uses of (4) are considered true by competent speakers of a language, what is far from clear is what object is picked out such that it satisfies 'is blue' in any of the ways just mentioned here.

[35] And this is not obviously true. The work of Michotte (1946 [1963]) illustrates the difficulty of such certainty with regard to our judgments of causation, where clearly non-causal scenes are judged as exhibiting causation by subjects. *A fortiori* these judgments persist even when objects interact in ways nearly identical to clearly non-causal events (Scholl & Nakayama, 2002).

ings are externalist, derives from them the ontology one is committed to in accepting the truth of a given expression. In the opening sections of this paper I sketched the Realist's methodology. I'll close here with a restatement of that widely-adopted strategy and articulate the reasons to reconsider its merits.

In Chapter 2 we saw that the Realist holds that there is a unique language (an interpretation of the existential quantifier) whose quantificational structure mirrors the structure of reality. With this privileged language $\mathcal{L}_{\mathcal{O}}$ in hand metaphyisicans can proceed to answer ontological questions by investigating the meanings of expressions in that privileged language, which can be given by way of Tarskian satisfaction by sequences of domain objects. Thus the Realist holds that the objects required to account for the meanings of expressions in $\mathcal{L}_{\mathcal{O}}$ are the objects of reality, since this privileged language mirrors reality's (object-based) structure.

To highlight an example of this strategy in action, consider the following points made by Sider (2002) in arguing against certain theories of time:

> The status of tense is a second issue in the philosophy of time. Tensed sentences are those which presuppose a certain position or vantage point within the whole of time, for example:
>
> > It is *now* raining.
> >
> > It *was* the case that there existed dinosaurs.
> >
> > I *will* one day visit Utah. (Sider, 2002, p. 12)

In arguing against a presentist conception of time, Sider contends that the presentist cannot clearly account for the truth of sentences that (seemingly) refer to the non-present. Insofar as the presentist denies that there are any ontologically real past or future times, *any* sentence that requires the existence of past/future times must thereby be either meaningless, or simply false. Such sentences have no truth-makers given the presentist's ontology, and thus the presentist cannot account for the truth of tensed sentences.

The success of this argument clearly presumes that the meanings of these natural language expressions determine their truth-conditions, insofar as their meaningfulness depends on the existence of past/future times. The presentist denies that there are past/future times. If the sentences Sider presents are meaningful—which they surely are, given that competent speakers of English have no problem understanding them—then *ex(ternalist) hypothesi* they have truth-conditions. Those conditions are only satisfied if there is some future time where Sider is in Utah, and

some past time where dinosaurs are alive and well. The presentist, contends Sider, must admit then that all tensed sentences are false or meaningless, since they have no temporal satisfiers. This consequence thereby seems bad for the presentist.

Of course, a key step in this line of argumentation asserts the truth of ($\mathcal{E}$). As we have seen, this externalist hypothesis is troubled. If the meaning of tensed terms do not determine whether or not they refer to *times* (*pace* externalism), then the move from linguistic meaning to ontological commitment is without warrant. And the supposition, if not false, is (at least) difficult to defend in light of the flexibility of natural language.

However, the Realist has a ready (and plausible) response to this objection. After all, natural languages like English are awash with vexing semantic properties like vagueness, ambiguity, and (apparently) lexical flexibility. As such, there is little surprise that they are ill-suited for the purposes of ontological investigation. The language the Realist needs is one that conforms to the features of classical logics, and none of these semantic properties are tolerated by such logics. But some languages are not deficient in these ways—namely the languages invented in the process of scientific inquiry. The privileged language $\mathcal{L}_{\mathcal{O}}$ needed for ontological investigation that the Realist requires is the one proffered by our best sciences. After all, scientific inquiry is guided by the expressed purpose of perspicuously describing the world. This process involves making decisions about what terms to use. The result are languages that embody the kind of precision that natural languages like English lack. This embodiment makes these scientific languages better suited for ontological investigation, and thereby better candidates for $\mathcal{L}_{\mathcal{O}}$. The viability of this retreat to the languages of science on the part of the Realist is the topic of the next chapter, where the question is whether the arguments put forth in this chapter can be extended to include the languages used to express our best scientific theories. I contend that some of them can.

Chapter 4

Realism and Scientific Languages

The previous chapter reviewed the case against adopting (or pursuing) an externalist semantics for natural language. These arguments were recruited in the service of rejecting the Realist proposal that natural languages can serve as useful tools in settling metaphysical disputes. The arguments set forth there were of three types. The first focused on the scientific task taken up by the linguist, namely proffering an explanation for the development of human children with regard to their linguistic capacity. The second turned to ontological considerations, specifically whether particular views about the ontology of linguistic elements are commensurate with scientific methodology. The third series of arguments are (what we might call) data-driven—they highlight the discord between the meanings speakers assign to natural language expressions, and the would-be externalist treatment for broad classes of linguistic expressions, demonstrating that natural language is much more flexible than the apparatus of the externalist's semantic machinery permits. I take these arguments to constitute a formidable case for rejecting the externalist program, at least for natural language semantics.

The chapter concluded with a reply, on behalf of the Realist. Given the multi-faceted problems for the externalist hypothesis regarding natural language semantics, the Realist might still rescue her metaphysical methodology by turning to a different class of languages more amenable to externalist treatment. In particular, she might adopt the suggestion offered by Quine (1973), and more recently by Sider (2009, 2011), consulting the languages created for the purpose of expressing our best scientific theories in answering metaphysical questions. Such languages, insofar as they are invented for the purpose of perspicuously describing the world, ought to admit to externalist treatment of the kind needed for Realist metaphysical inquiry to proceed.

As such, the focus of this chapter is on the languages used to express theories in science—call them *scientific* languages.[1] The goal of this chapter is to question the

[1]Throughout this chapter the expression 'scientific language(s)' or 'languages of science' are intended to be synonymous with the much more clumsy 'the languages used to express scientific theories.'

Realist's retreat to scientific languages. In this vein, I'll make an effort to extend the criticisms detailed against the externalist hypothesis in the previous chapter, which dealt with the problems for such a thesis regarding natural languages, to the languages of the sciences. The main question motivating this chapter is:

> Are the meanings of expressions in scientific languages best captured by an externalist semantics?

I'll suggest that we have reason to be skeptical about an affirmative answer to this question.

4.1 Framing the Question

The invented languages used to express scientific theories are the focus of the chapter. We'll be investigating whether or not scientific languages are a viable option for pursuing the Realist project. In order for the Realist's methodology to be capable of yielding answers to ontological questions, the languages she uses in this pursuit must have an externalist semantics, whereby the meanings of each term in these languages are given by way of worldly objects. If the languages she utilizes are those given by our best sciences, these invented languages must not be burdened by many of the properties of natural languages—or at least not by those properties that make natural languages hostile to externalist treatment.

In Chapter 2 we outlined the Realist's methodology. Recall from that discussion, that the goal of the scientific enterprise is in developing a language that "carves nature at its joints." As we saw, one way of thinking about the language the Realist requires is suggested by Sider, in a discussion regarding the interpretation of the existential quantifier:

> Clearly there are multiple (inferentially and materially adequate) interpretations of quantifiers. As I see it, the real issue is whether any of these interpretations is metaphysically distinguished, whether any of them uniquely matches the structure of the world, whether any carves nature at the joints better than the others. (Sider, 2009, p. 392)

The picture underlying these claims is of a plenitude of (invented) languages, each with an externalist semantics, and (let's suppose) free from the pitfalls of natural language, with its ambiguities and lexical flexibility. For these languages, Frege's problem does not arise, since each term's denotational meaning picks out a unique

object. Nor is vagueness a difficulty for users of these languages, as each predicate meets with the precision commensurate with an algebraic model. Somewhere in this vast plenitude of languages there is special one, that "matches the structure of the world." This special language is the one that best describes the world, whose existential quantifier "cuts nature at its joints." Ontologists in possession of this language can pursue the Realist project, by gathering up the true sentences expressible in this language, and determining what objects are required for the satisfaction conditions of those sentences. Characterized in this way, the goal of science is to *discover* the *language of ontology* ($\mathcal{L}_{\mathcal{O}}$)—to journey through logical space, with naturalism as a guide, and arrive at the coordinates that house this special, ontological language.

The last chapter argued that this special language is not a natural language. Since $\mathcal{L}_{\mathcal{O}}$ must have an externalist semantics, and natural languages are not amenable to externalist treatment, it is doubtful that a natural language will be of the joint-carving kind the Realist finds at the end of inquiry.

Facts about human cognition constrain the kinds of semantics that explain human linguistic competence. And from within the context of these constraints the scientist begins the project of building (or discovering) the language of science. Can we, as the Realist proposes, strip away those properties of natural languages that undermine the externalist program and construct $\mathcal{L}_{\mathcal{O}}$? Put another way, is this Realist journey bound to fail given the constraints of human cognition?

Two points are worth highlighting at the outset. First, the question here is *not* whether humans are capable of constructing an algebra of sufficient complexity and precision to match the "fine structure" of reality. If we grant Sider's implicit assumption that there is a language, an interpretation of the existential quantifier, that best mirrors the structure of reality, the worry probed in this chapter is *not* whether humans are (in)capable of penning such a language. The question is whether the process of scientific investigation, in conjunction with the limits of human cognition, will keep us from discovering the language of ontology. In some sense then, the points made here reflect epistemological limits, not metaphysical truths. But given that the grounding of the Realist's methodology relies on the epistemic credentials of the sciences, the limits of epistemology bear directly on the success of metaphysical investigation.

The second related point is the role understanding plays in this investigation.

Part of the worry pressed in these pages is whether humans can *understand* the language of ontology. The scientific journey through logical space envisioned by the Realist would clearly involve the systematic replacement of old theories with new ones, and likewise, old languages will be replaced by new ones. But for this to happen, scientists must come to understand the theories they are accepting, rejecting, and revising. As such, if there are interesting limits on the human capacity to understand languages, and the scientific theories expressed by them, there may well be obstacles inherent to the Realist's task.

4.1.1 Distinguishing Scientific Languages

Since this chapter focuses on scientific languages, it might be helpful to indicate how they are distinct from natural languages. Unfortunately, a precise distinction that delineates scientific languages from natural languages is not easily found, and I suspect this is true because there simply is no clear boundary. However, the lack of a clear distinction, or the fact that one is not provided here, does not undermine the arguments set forth in this chapter. Recall, the Realist move away from natural languages, for the purpose of metaphysical investigation, is in response to the flexible character exhibited by natural language expressions. This compelled *the Realist* to adopt a distinction between natural and scientific languages. Thus the burden here is on her to show that scientific languages are indeed special sorts of languages, whose meanings are well captured by externalist theories. While one might grant, for the sake of argument, that a real distinction exists, if there is reason to think there are no marked differences between natural and scientific languages, all the worse for the Realist.

On the Realist's behalf, let me draw attention to some possible features of scientific languages that might lead one to think they are distinguished for metaphysical purposes. One feature a Realist might cite is that these languages are invented. The constituent expressions of those languages are developed with a particular purpose in mind: "the point of human inquiry . . . is to *conform* itself to the world. . . and our job is to wrap out minds around it" (Sider, 2011, p. 18). As such inquiry proceeds, terms are introduced as a means to perspicuously describing the world.

As a point of contrast, in acquiring a natural language, a four-year-old is not

driven by a desire to describe the world, much less to describe the world with a particular level of precision. In fact, it is not clear that a four-year-old has any explicit or conscious intention when acquiring a natural language. Likewise, as new expressions manifest in natural language communities, there's little reason to think that the introduction of these expressions is directed at more aptly describing reality, or that their introduction is the product of coordinated efforts toward some end. This contrasts starkly with the scientific enterprise, that devotes considerable effort to the clarification, precision, and aptness of their theoretical terms with the goal of aligning those terms to comport with worldly evidence.

A second feature of scientific languages is that learning/acquiring them is often dependent on orthography in a way that natural languages rarely are. One can be illiterate, and be a perfectly competent English speaker. In contrast, mastering the language of a scientific discipline seemingly must involve a comprehension of written language. Becoming a chemist, biologist, or physicist without the ability to read, for example, seems implausibly difficult. No doubt this is at least partially attributable to sociological facts about the scientific institution, but it is also hard to imagine a thriving scientific research program in the absence of a written form of communication.

Other features seem to be emblematic of scientific languages as well. Often scientific languages make use of mathematical formulae, for example. Put in less familiar terms, often times scientific languages are species of mathematical languages. Scientific languages also tend to be void of indexical expressions, like the English 'I', 'now', or 'here'. But the distinguishing feature of scientific languages that matter most to the Realist, is that these languages are intentionally constructed for the purpose of perspicuously describing the world. After all, this feature underwrites the epistemological claim that makes such languages apt for metaphysical investigation. The appeal to science as a means of addressing the inability of natural languages to yield metaphysical verdicts is grounded in naturalistic methodology, the aim of which (at least according to the Realist) is to describe the world as it is. Thus, in identifying the primary difference between natural languages and scientific ones, the explicit, intentional nature of their construction seems paramount to the Realist.

4.2 Comparing Natural Languages and Scientific Languages

The most natural place to begin our inquiry is by reviewing the arguments from the previous chapter, but now with an eye toward the languages of science. The goal of this section is to see if the arguments from the previous chapter still hold when our target is the externalist treatment of *scientific* languages and their semantics. We'll take each of the three argument types in turn.

4.2.1 Acquisition Arguments

At the foundation of the Chomskyan program for the study of language are two criteria that any theory about natural language must meet: explanatory adequacy and descriptive adequacy. The latter is a merely formal bench-mark—the theory must be able to produce all and only the acceptable expressions of a natural language and assign them meanings commensurate with the patterns of exhibited by competent speakers of the language. For example, a proposed theory of the grammar for a natural language, as a series of recursive procedures for generating strings of terms in the natural language, is descriptively adequate if those procedure permit the construction of the strings that competent users of the language find acceptable (*modulo* performance limitations), and prohibits the production of those that such users find unacceptable. The *explanatory* benchmark constrains the theory by requiring that a child be capable of acquiring the knowledge embodied in the theory in accordance with the typical development the human linguistic capacity. That is, for a given natural language, a child better be able to acquire the knowledge as characterized by the theory such that they reach competence with the language around the age of four.

Inherent to adopting these dual criteria is a tension.[2] The first criterion demands for complexity in the theory, in order to capture (and exclude) the vast array of complex expressions exhibited by the acceptability judgments of competent speakers. The second demands for a theory with a kind of simplicity. Since children are competent language speakers near the age of four, the explanatory criterion demands a simpler theory, that is easier for the child to acquire and implement, especially given the limited character of their linguistic input. As a crude example,

[2]For the canonical explanation of considerations of this kind, see Chomsky (1957, pp. 49–60).

a theory for the grammar for a language that simply enumerates a list of all and only the acceptable strings for the language will fail to meet the explanatory burden. Even assuming such a grammar meets the descriptive burden (which it does not), it cannot meet the explanatory burden, given that a child could not possibly learn such an (infinitely) long list of expressions in a short four years. As we saw in the previous chapter, this concern for the problem of explanatory adequacy compels the linguist to reject the EL conception of language, since EL views give the linguist no traction on one of the fundamental goals of her program.

But notice the importance of the tension between these two criteria. If, counterfactually, humans didn't acquire linguistic competence until (say) the age of twenty-five, then the explanatory problem would be much less burdensome. After all, with two decades of additional linguistic data to consider, any individual may well be able to acquire the knowledge described by a rather complex model. This longer period of acquisition invites a much broader space of possible explanations for how language could be learned by humans. Thus the focus of the linguist under this counterfactual condition would be on the descriptive adequacy of the theory, which would permit the use of a broader class of models of increased complexity.

While the acquisition of natural language does not fit with these counterfacts, the manner in which humans come to know the contents of scientific theories seemingly does. Becoming a scientist is hard, and is rarely accomplished in a short four years. Decades of schooling, rote memorization, an abundance of negative feedback, and a bounty of legal stimulants make up the typical course of scientific training—none of which holds for the young natural language learner. As such, the explanatory burden that constrains the space of theories for a natural language, has much less bearing on a theoretical account for scientific languages.

4.2.2 Ontological Arguments

An externalist semantics holds that the meanings of terms in a natural language are cashed out in terms of relations between those terms and worldly objects. Thus the externalist owes us an account of the *relata* for that meaning bearing relation. If no convincing account can be given for the ontological *bona fides* of these *relata*, this gives us warrant to reject the externalist account. The ontological arguments presented in the previous chapter focus on the elements of an externalist

conception of natural languages, in particular *words*. If meanings are word-to-world relations, there had better be an ontologically stable notion of a *word*, such that objects can be related to them (meaning-wise). The point emphasized in these arguments is that there is no clear way to proceed in addressing this ontological demand. The most plausible method would be to group words by the sounds humans produce in constructing expressions. But a point familiar to phonologists is that appeals to the characteristics of the sonic waves produced by speakers of a given language do not divide these noises into *sui generis* kinds. Put another way, any naturalistic inquiry into the nature of words shows that there is no (mind-independent) subject matter to be investigated, and no scientifically viable conception of *words*. Thus, if there is no natural kind 'cat'—i.e. no ontologically genuine category for the word 'cat'—then there can be no relation, meaning-bearing or otherwise, between 'cat' and (all of) the (possible) cats. This then precludes the possibility of an externalist semantics, since the externalist requires that there be such meaning-bearing relations.

So what to say about the ontological *bona fides* of terms in a scientific language? Two points seem salient. First, the conditions that one would appeal to in distinguishing terms for a scientific language seem to be orthographic, not phonetic (or phonological). The languages of science are primarily written languages, with the verbalizations of theories derived from those written representations. To see this, merely consider how implausibly difficult learning Schrödinger's wave equation would be if one were only able to hear it spoken:

Schrödinger's Wave Equation

$$i\hbar\frac{\partial\psi}{\partial t} = \frac{-\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\psi + V\psi.$$

The complex, and often mathematical, nature of discoveries in the sciences requires an equally complex means of articulating those discoveries. This leads to the reliance on written, as opposed to spoken, means of conveying such discoveries.

The second point: given that the use of terms in a scientific language are governed by stipulative and explicit conventions, delineating which orthographs are genuine, meanings-bearing expressions is clear. Indication of what does (and does not) count as a term denoting (say) mass is much less troubling to come by than

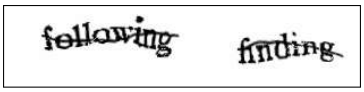
indicating the conditions for what human sounds do (and do not) count as utterances of (say) the English 'cat'.

There of course are interesting cognitive limitations on orthographic representations. The two tokens of the "broken A" appearing in the two representations below, are visually identical, yet we treat one as an 'A' and the other an 'H':

Figure 4.1: Vague Orthography

Relatedly, the use of CAPTCHA protocols, meant to distinguish human users from non-humans users of a computer system is premised on the differences between humans and non-humans in their abilities to identify distorted orthography (Von Ahn et al., 2003). Given a distorted image of some text, as in the figure below, humans can identify these expressions quite well, while computer algorithms perform rather poorly.[3]

Figure 4.2: CAPTCHA distortions



Thus whether a particular blotch on a page, or pixel pattern on a screen, counts as an 'm', might well depend on facts about the human visual/cognitive system. As such, one might argue that the differentiation of symbol-terms in a scientific language is mind-dependent, in ways similar to the differentiation of (spoken) words in a natural language.

While there might be room to level this worry about the ontology of scientific terms by appeal to the mind-dependence of these phenomena, this worry seems less substantive than priggish. Suppose the complete (and correct) theory of physics

[3]The primary difficulty in for a computer using OCR software to identify distorted text lies in the segmenting of words into individual characters, not the identification of the characters themselves (Kumar Chellapilla, 2005).

made use of the symbols '🚺' and '🚹', denoting putative natural kinds. Suppose too that this theory is written down and given to some advanced extra-terrestrial species, to be compared with their own complete and correct theory of physics. Suppose finally that these aliens could not distinguish '🚺' from '🚹'—because these aliens cannot, as a feature of their cognitive capacities, distinguish human men from human women, generally. To the alien physicists, the human physics differs from their own, with a slightly smaller ontology (and a much less descriptive physics). Thus the aliens reach the conclusion that these theories differ importantly, despite the fact that both are correct theories. This purported difference is a function of the difference in human and alien cognition, not a difference in the theory or the metaphysics it begets.

But this problem is easily surmountable, once this difference in human and alien cognition is recognized. The stipulative conventions involved in selecting the relevant terms could easily overcome this problem by simply constructing a new notation that accommodates the alien species' cognitive idiosyncrasy, thereby bringing the two theories to accord with one another. So while there may be human cognitive constraints that dictate what can (and cannot) count as a symbol in scientific notation, and likewise, what marks count as the *same* symbol, these limits do not seem as pressing, given the stipulative conventions of science.

The second point is more theoretical. The ontological criticism against an externalist semantics for natural language holds that there are no words. Rephrased, the thought expressed is that, because our best (language) science tells us there are no words, ontologically there can be no such things. This criticism against the externalist program relies on the Realist's ontological methodology. Recall, the Realist insists that ontological investigation proceeds by identifying the true expressions of a language, and determining from there what kind of objects must occupy the worldy domain in order to satisfies the truth-conditions of those expressions. When that language is the invented language of our best (and complete) science, the ontological posits of that language are "what there is." As we've indicated, this methodology only succeeds if the semantics for the relevant language is externalist. So if the criticism leveled here is that our best scientific theories says that there is no ontologically *bona fide* category 'scientific term', and there are therefore no scientific terms to stand in meaning-bearing relations to the mind-external world, the argument is self-defeating. The justification for saying that there are no scientific

terms relies on the assumption that the *scientific terms* in our best theories denote all and only the objects of ontology, because the semantics for such a language is externalist. Thus the criticism that would yield the anti-externalist conclusion about the semantics for scientific languages, assumes that such a semantics is indeed externalist. For this reason, ontological arguments against an externalist semantics do not seem to extend to scientific languages.

4.2.3 Data-Driven Arguments

The data-driven arguments of the previous chapter draw attention to the discord between the pattern of meanings assignments competent speakers of the language give to expressions and the truth-conditional properties of those expressions. Such arguments are data-driven because they show that externalist models cannot explain the distribution of meaning assignments they are meant to explain. Specifically, natural languages exhibit a kind of flexibility that externalist models cannot accommodate. The cases that speak most strongly against an externalist semantics for natural language are those that have ontological commitments which require a single object to satisfy ontologically distinct types of predicates. Externalist construals of the meanings for such expressions attribute ontologically incommensurate properties to a single object, and thereby have implausible entailments. Recall the following examples:

(62a) This book, which John wrote, is five pounds.

(69) The Hirshhorn Museum is bankrupt and it is a cylinder.

Expressions like these require metaphysically suspect objects to satisfy their purported truth-conditions. There is a reading of (62a) which does not require a metaphysically suspect object, but rather requires that John scribed a fairly long book (in the sense that might lead to a hand cramp). But there is a perfectly sensible meaning to (62a) which doesn't require this, whereby John has never interacted with the demonstrated, five pound physical object. On this reading the usage of 'book' must refer to some abstract object that John wrote, related to, but not identical with, the physical, hefty text. And while abstract objects can have properties, weight is not one of them. To be five pounds, that abstract object must have the kind of heft it cannot have. Likewise, the referent of 'the Hirshhorn Museum', on an

externalist semantics, must be of the sort that can be both bankrupt and cylindrical. And while abstract institutions can have many properties, they do not seem to have a shape. Yet, for (69) to be true there must be an object of this odd sort.

And even if there are metaphysically bizarre objects, of the sort required by an externalist semantics for these flexible expressions, such a semantics fails to explain the oddity of

(70) # The Hirshhorn Museum is a bankrupt cylinder.

On a externalist semantics (69) and (70) are truth-conditionally equivalent, and thereby have the same meanings according to such theories. Yet (70) is odd in a way that (69) is not, a fact unexplained by theories that treat meanings as truth-conditions.

As we saw, it might be tempting to treat such examples as cases of homophony, as exhibited by expressions like 'bank'. But as I argued, such treatment fails to explain both the felt relatedness between uses of the relevant noun phrases, and also fails to explain the (un)acceptability of truth-conditionally equivalent expressions.

Other cases of flexibility are less decisive evidence against an externalist semantics for languages exhibiting flexibility, but nonetheless speak against an externalist treatment of expressions that are lexically flexible. These are cases in which a term expresses a constellation of concepts with distinct satisfaction conditions, but are of the same ontological type. Color predicates exhibit this kind of flexibility. Consider the following:

(94) The house is blue.

(95) The marker is blue.

(96) The ink is blue.

(97) The iris is blue.

(98) The sky is blue.

The predicate 'is blue' attributes a similar property in all these instances, related to human phenomenological experiences of color. But the extension of this predicate will vary depending of contextual aspects of its use. For (94) to be true, the exterior of the house must be (mostly) blue, but the interior can be any color.

Contrastingly (95) is still true even if the marker's casing is white, so long as the contents of the marker are of the right sort. And while (97) can be true even if the majority of the plant is not blue, the house in (94) must be mostly blue. This seems to indicate that, while uses of the term 'blue' are deeply related, they express different concepts with conflicting satisfaction conditions. These differing conditions suggest that we should be pluralists about the concept BLUE, all expressible by the expression 'blue'. Of course, we can treat these uses as cases of homophony, or even a kind of indexicality (Rothschild & Segal, 2009). But because these uses of 'blue' are deeply related, treating these various uses as cases of homophony is disingenuous to the phenomena. Importantly, accepting a kind of pluralism about the meaning of terms like 'blue' undermines the Realist methodology, which requires that meanings of expressions determine their extensions.

The relevant inquiry for the purposes of this chapter then is whether our scientific languages exhibit these kinds of flexibility. If the languages used to express claims or generalizations in science are flexible, then these languages are bad candidates for ontological investigation. I argue that there is sufficient reason to hold that expressions used in biology exhibit this kind of flexibility. In particular, uses of terms like 'gene' and 'species', while they serve an invaluable explanatory purposes within the biological sciences, exhibit lexical flexibility. Insofar as biological explanations cannot be reduced to physiochemical explanations, the Realist methodology requires that the languages used in biological explanations be externalist. Thus, if the terms of biological languages are indeed lexically flexible this speaks against the Realist use of such languages for ontological investigation.

4.3 Lexical Flexibility of Scientific Terms

In this section I review two cases of terms in the biological sciences that exhibit lexical flexibility, and thereby cannot be assimilated into an externalist semantics for the scientific languages of biology, as the Realist requires.

4.3.1 Gene

Explanations in evolutionary biology make use of the term 'gene' in ways that are indispensable. The study of biology has, from its conception, been concerned

with reproduction, and specifically with the means by which features of an organism are passed on to their progeny (Aristotle, 1941). Aristotle's concern was to explain how an organism's form could be transferred from one generation to the next. Since the work of Darwin, interest in heritability took on a particular importance, insofar as explanations of fitness require that individual organisms can pass on their features to their kin. If a particular trait was adaptationally advantageous, that characteristic must (typically) be inheritable by the organism's offspring.[4]

Early conceptions of 'gene' in post-Darwinian biology were meant to provide such an explanation. Mendel's cross-breeding experiments with peas led to an initial conception of the gene, characterized by Mendel's followers as the "unit-character" that is both responsible for the ratios seen in Mendel's Laws, and carried in the gametes of parenting organisms. Around the same time that Bateson coined the use of 'unit-character', Johannsen made explicit the use of the term 'gene' as a means of distinguishing between the characteristics an organism exhibits, or its *phenotypic* traits, and "whatever it is that determines an organism's properties and is passed down through the gametes" (Weber, 2005, p. 195). So construed the manifestation of a gene was left free of any particular physical commitments, but the thought was that an "organism's properties are determined by special, separable and therefore autonomous [units]" (Johannsen, 1909, pp. 143-144) [as translated in Weber, 2005]. Thus, there was thought to be a one-to-one mapping between a gene and a particular organism trait.

Work on *Drosophilia* (a species of fruit fly) over the next decades, and later with *E. coli* (a rod-shaped bacteria) would prove that this relationship was actually quite complex, admitting to a many-to-many character. Not only are phenotypic traits the consequence of many genes interacting in intricate ways, but so too, a given gene can be involved in multiple phenotypic traits. Around the time of Watson and Crick's published double helix model, the identification of the gene with a sequence of DNA emerged. This conception identified genes as sequences of DNA that code for the production of proteins.

[4]This is not necessarily the case. For example, a random mutation for an individual organism can yield an advantage despite never being inherited from its parents. And given some unique circumstances, the mutation could lead to that organism surviving a plague that destroys the rest of the species. Assuming this trait is not passed on to this individual organism's progeny, the trait plays an essential role in explaining why the organism and its lineage survived while all others did not. And this is true, despite the fact that the trait was never inherited, nor heritable. But such a case is surely at the periphery.

The result of this history is that contemporary biologists make use of the term 'gene' in two distinct ways.[5] The first pertains to the form of an organism, and the manner in which this form is passed on in reproduction. Such a notion is in this sense "preformationist", which we can call Gene-P:

> To speak of a gene for a phenotype is to speak as if, but only as if, it directly determines pheonotype. It is a form of preformationism but one deployed for the sake of instrumental utility. I call this sense of the gene—Gene-P, with the P for preformationist. (Moss, 2002, p. 45)

The other use of the term 'gene' is related to the discovery of DNA, and the research that followed this discovery. On this use of the term, a 'gene' is a sequence of DNA that encodes for a protein product. Call this the Gene-D use of 'gene':

> *Gene-D is defined by its molecular sequence.* A Gene-D is a developmental resource (hence the D) which in itself is *indeterminate* with respect to phenotype. To be a Gene-D is to be a transcription unit on a chromosome within which are contained molecular template resources. These templates typically serve in the production of various gene products—directly to the synthesis of RNA and indirectly on the synthesis of a host of related polypeptides. (Moss, 2002, p. 46)

Given this distinction, Gene-P uses of the term 'gene' are meant to denote whatever is passed on generationally that manifests a particular phenotypic trait. Understood this way, Gene-Ps cannot be sequences of DNA, since most phenotypic traits are the result of the complex interaction between multiple sequences of DNA, development, and the organism's environment. The very same sequence of DNA in different environments will produce different polypeptide products (more on this below). So not only is the relationship from DNA sequences to products a one-to-many relationship, but so too is the relationship between DNA sequences and phenotypic traits. Were the relationship between DNA sequences and phenotypic traits one-to-one, then biologists could adopt a univocal conception of 'gene', which would be good news for the Realist. That they are not, coupled with the need for an explanation for how characteristics of form can be passed on from generation to generation, yields the distinction between Gene-P and Gene-D uses of 'gene'.

There are instances where the mapping of DNA sequences to phenotypic traits is well understood, even if not one-to-one. In these cases explanatory generalizations

[5]Moss (2002) presents a historicity of the "gene concept" tracing the usage of the term 'gene' and its historical analogs noting a shift is usage marked by the "phylogetic turn" stemming from the work of many, including Darwin and Mendel. There's some evidence that contemporary biologists conceptualize genes in the way marked by this distinction (Stotz & Griffiths, 2004).

and claims in biology (seem to) exhibit a kind of flexibility. For instance, when scientists talk about 'the gene for breast cancer' such usage seems to be of the type related to Gene-P, insofar as the property of having-breast-cancer is a phenotypic trait (or perhaps more precisely, the property of having-high-susceptibility-to-breast-cancer is phenotypic). The use of the term 'gene' is useful in these context, insofar a breast cancer admits to a certain degree of heritability (Pharoah et al., 1997). Biologist and oncologist are interested in studying such an entity, at the level of phenotype since discoveries in this domain might contribute to understanding environmental risk-factors, patterns of infection, and thereby new treatments. But 'the gene for breast cancer' is not a particular DNA sequence responsible for some collection of protein products. Rather, such a 'gene' is characterized by the absence of a DNA sequence responsible for the production of a particular class of proteins in breast tissue and other tissues in the human body. More specifically:

> The normal resource at the breast cancer locus (BRCA1) is not a gene for healthy breasts, but a template for a large and complex protein which is present in many different cell types ... capable of binding to DNA and influencing cell division in context specific ways. (Moss, 2002, p. 48)

So, in breast tissue cells that lack a particular DNA sequence at a particular location (or locus) on a particular chromosome, a human is more likely to acquire breast cancer. This is because the normal DNA sequence at that locus is absent, as are the multiple protein products it produces which promote cellular stability (and hence stave off cell mutation) during reproduction. Further, this absence can be filled by many divergent sequences of DNA, not some single aberrant sequence. Thus, when someone has 'the gene for breast' cancer, what they posses is not a particular sequence of DNA that is responsible for a protein product—they do not have some particular Gene-D. The sequence of DNA they have, amongst the many aberrant sequences they could have, at the relevant locus fails to produce any product at all. To say someone has the 'gene for breast cancer' is to say that they lack a particular gene. That is, the breast cancer Gene-P is the absence of a Gene-D.[6]

Two points are worth emphasizing here: first, the uses of 'the breast cancer gene' cannot be reduced to uses of 'gene' as a sequence of DNA; and second, use of 'the breast cancer gene' only makes sense by making use of the conception of

[6]Breast cancer is not unique here. Many other heritable conditions have a similar structure, where the Gene-P for those diseases is marked by the absence of a Gene-D: for example Huntington's disease, sickle-cell disease, and possibly Leprosy.

genes as DNA sequences. That is, one of these notions is not reducible to the other, yet their meanings are intimately related. Clearly, a Gene-P cannot be reduced to a Gene-D conception, precisely because in diseases like breast cancer there is no Gene-D responsible for breast cancer. None of the DNA sequences involved in explaining the manifestation of breast cancer are Gene-Ds, because they do not produce polypeptide products. Relatedly, any description of a Gene-P in cases like breast cancer requires an appeal to the "normal" Gene-D present at the appropriate place on the relevant chromosome. These uses of 'gene' then seemingly exhibit the feature that renders natural languages hostile to externalist treatment, namely *lexical flexibility*.

These uses of 'gene' have distinct, though (inextricably) related meanings, as exhibited by the following expressions in the language of biology:

(99) The breast cancer gene is multiply realizable, and it prevents the production of a class of proteins originating from the BRCA1 locus on chromosome 17 in breast tissue.

(100) Jill and Jan both inherited the breast cancer gene from their mothers, and it prevents them from producing a class of proteins originating at the BRCA1 locus on chromosome 17 in their breast tissue, making their breast tissue more susceptible to mutation.

These expressions seem to be claims biologists would accept as true (given the obvious assumptions about Jan and Jill). Further this seems to hold despite the fact that the use of 'gene' in both expressions cannot be univocal. If we take the expression 'the breast cancer gene' to refer to the absence of the Gene-D that "normally" appears at the BRCA1 locus, then clearly both notions of 'gene' are being expressed by the single use of 'gene' in these sentences. Such an absence cannot enter into a causal interaction with the production of a protein product, even if an alternate DNA sequence can. Yet, on an externalist reading of (99), this is precisely how the world must be for it to be true. There is some object denoted by 'the breast cancer gene' that interrupts the production of a protein. But as we've seen, 'the breast cancer gene' refers to some abstract object, an absence of some particular DNA sequence that is normally found at a particular locus. Just like the case of the Hirshhorn Museum, an externalist treatment of (99) would require the existence of a metaphysically suspect object.

Not only does 'gene' exhibit the flexibility of words like 'book', it exhibits a kind of flexibility that suggests a pluralist treatment, much like we saw with color terms. Gene-D uses of 'gene' suggest that genes are molecules, and that the extension of such a conception is precise—a gene is a DNA sequences that codes for polypeptide products. In the paradigm case, a sequence of DNA encodes the production of an RNA sequence, which is then used by the cell to produce a protein product. But this simple paradigm admits to exceptions, in multiple complex ways. For one, in many viruses DNA plays no role in the production of polypeptides, as RNA plays this role directly, not requiring the presence of any DNA sequence in the virus. But presumably, all the reasons biologists have for holding that DNA sequences are genes due to their role in the production of polypeptides and their ability to be passed on in reproduction, also holds for RNA in the context of a virus cell. But if we identify RNA as viral genes, there must be some good reason not to consider sequences of RNA as genes in other environments. Shifting the extension of 'gene' to only include RNA sequences would have the consequence of denying that genes are heritable for many organisms, as in the paradigm case. In most organisms RNA is not inherited in reproduction, though DNA is.

In other cases, a single DNA sequence can encode for the production of an RNA sequence that, in some cells, is only *partly* used to produce particular polypeptides. Yet in other cells, different parts of the same RNA sequence are selected by different processes to produce other polypeptides. In these environments, the RNA sequence yielded by a sequence of DNA is separated by processes further downstream in the production of the resulting polypeptide. In some cellular environments one part of the RNA sequence is used to produce the resulting protein, while in other environments different (sometimes overlapping) parts of the RNA sequence are used to produce different proteins. Thus, the same sequence of DNA is responsible for different protein products, depending on the environment in which the DNA (and its RNA pair) is utilized. In such cases, only part of the DNA sequence that encodes for an RNA chain is responsible for the production of a protein, since the other parts of the RNA sequence the DNA chain begot are not utilized in the production of the protein product. This leads to indeterminate answers as to what (part of) the DNA sequence is in fact the gene (Wilson et al., 2007, p. 203).

Further, because the expression of a gene depends on the other *regulatory*

DNA sequences in its environment, sequences that govern the order in which genes are expressed within a cellular environment, the very same sequence of DNA in one cell will produce a protein, yet fail to do so in a different type of cell in the same organism (because a regulatory sequence of DNA is present/absent). In terms of pinning down the extension of 'gene', even bracketing concerns about the difference between Gene-P and Gene-D uses of the term 'gene', whether a particular sequence of DNA is considered a gene or not will depend on the environment in which that DNA resides. In some contexts that DNA sequence produces a protein and counts as a gene, while in others that same sequence will fail to produce a protein (and concomitantly fail to count as a gene). Much like our color predicates in natural language, the predicate 'is a gene' in the language of molecular biology seems to be context dependent in ways that suggest a kind of pluralism about gene concepts.

The upshot is that 'gene' seems to exhibit both kinds of flexibility. Insofar as 'gene' is lexically flexible, not reducible, and importantly explanatory for the biological sciences, this suggests that a semantics for the scientific languages used to express biological claims is not externalist. This calls into question the Realist insistence that the languages of our best science must have an externalist semantics.

4.3.2 Species

The term 'species' plays a crucial role in biological explanation. As biologists investigate a particular organism, the usefulness of their findings only gains traction under the assumption that the individuals they study are in someway representative of a larger group. In this way species membership is informative. Knowing that an individual organism is a member of a species enables us to predict a host of other properties associated with that class (Dupré, 1999; Griffiths, 1999). For the Realist, then, the term 'species' must have an externalist meaning if the languages of our best biological sciences are to be apt for purposes of ontological investigation. Relatedly, if 'species' fails to admit to externalist treatment, this speaks against the Realist methodology that insists the languages produced by naturalist investigation are useful tools for settling ontological disputes. However, 'species' seems to be flexible, in the way natural language color predicates are, suggesting a pluralist treatment of the species concept.

One way for the term 'species' to be univocal is if individual species terms

denoted natural kinds. If there are natural divisions in the world between species, and the terms we use mark those distinctions, the term 'species' can thereby have a univocal meaning as referring to such natural kinds. The work of Kripke (1980) and Putnam (1975) invited a resurgence in essentialist thinking about species as natural kinds. The idea promoted by Kripke and Putnam is that uses of natural kind terms denote natural kinds, irrespective of speaker knowledge. Putnam in particular argues that there is an implicit convention to natural kind denoting terms that, while individual users of those terms might be ignorant or confused about their denotations, such terms track the essential properties constitutive of the denoted natural kind.

For example, all objects composed of gold share features in common, many used to identify which things are gold and which are not. Gold is shiny, yellow, and mailable. However, according to Putnam, the implicit convention pervading uses of 'gold' dictates that their uses do not merely denote objects that are shiny, yellow, and mailable. Rather, such uses track the microstuctural properties that underwrite these superficial properties, and this is true even if users are ignorant of what those microstructural properties are. Uses of 'gold' denote in accordance with the underlying (possibly unknown) *essence* of the natural kind. In the case of gold, of course, we are no longer ignorant of this essence. The atomic number of gold atoms both unifies the natural kind and determines the superficial properties often used to identify bits of gold. That is, what makes an engagement ring and The Hand of Faith both essentially gold is that both objects have certain microstrucral properties, abstractly characterized by their atomic number. And uses of 'gold' have, according to Putnam, *always* denoted those properties.

Such an essentialist conception of natural kinds is conceivably at the heart of analytic metaphysics, at least for the Realist. If natural kind terms intrinsically refer to essential natural divisions, the meaning of such terms can be fruitful for ontological investigation. That the world has a certain "fine structure" and can be "carved at its joints" implies that there are natural divisions between objects, divisions that are intrinsic to those objects. The work of Kripke and Putnam invites conceiving of metaphysical inquiry generally along Realist lines. The examples offered by both Kripke and Putnam extend beyond physics and chemistry, to purported biological natural kinds. For Kripke, 'tiger' denotes a natural kind, ostensibly some animal species, members of which share some common essential properties that mark them

as members of that species kind (Kripke, 1980, p. 121). In much the way some particular engagement ring and The Hand of Faith are members of the same natural kind in virtue of their essential atomic number properties, Sita[7] and Champawat[8] are members of the same species in virtue of their (yet unknown) essential properties. Putnam makes similar remarks regarding uses of 'lemon', holding that the underlying essential properties are genetic (Putnam, 1975, p. 240).

As I've suggested, if the Putnam-Kripke line of thought is correct, this is good news for the Realist. If species terms like 'tiger' and 'lemon' track essential properties, those that make an object a lemon or a tiger, then the meanings of such terms are fruitful for ontological investigation. Likewise, the term 'species' has a univocal meaning, namely those terms that track essential properties of biological kinds. However, this essentialist line of thought is troubled.

On essentialist contruals, membership in a natural kind is determined at the level of individuals. That an object is a piece of gold, and not a piece of silver is explained by the essential atomic number properties of the relevant bits of matter. A difference in two individuals' atomic number yields a difference in natural kind. But, as Sober (1980) argues, appeal to individual differences in codifying species kinds is incompatible with evolutionary theory. Rather, in post-Darwinian explanations in evolutionary biology, individual differences in organisms within the same species are required to make sense of selection. Intrinsic differences between the reproductive workload of the two genders in mammalian species are required to explain how traits of parents are passed on to kin, and underwrite explanations of fitness. More starkly, in many insect species, the differing roles of intrinsically different organisms explain why a *population* of insects are more adaptive than others. The role of asexual worker bees is essential for the reproductive success of the hive *via* those members of the population that do produce offspring. For this reason, species *populations* and not individuals are the units of selection in evolutionary biology. Thus, to explain why a particular species has evolved, the term 'species' must identify a population of organisms, not some set of essential properties some organisms share. The essential differences between members of the same species population explain why that species has survived, where others have not.

[7]The mother of most of the tigers that currently populate Bandhavgarh National Park in India, and is thought to be the most photographed tiger in history.

[8]A tiger thought to be the most dangerous to humans, estimated to have hunted 437 humans in the plains of Northern India before she was trapped and killed in 1907.

For the essentialist, empirical differences between organisms are facts that need to be explained away. On an essentialist understanding, differences between individual organisms are, in a relevant sense, superficial to their membership in a given species. Sober's point is that these differences are not immaterial, and indeed important to explanations of natural selection. That worker bees differ from their reproductive conspecifics is not a trivial difference for the species. This difference is *required* for the explanations of species fitness, not a feature about the individuals that, from the perspective of species membership, can (or should) be explained away. As such, the existence and membership of a species is understood *because* of the intrinsic differences in its membership, not delineated by appeal to these differences as the essentialist would conclude.[9]

While the failure of essentialism regarding biological natural kinds is not good news for the Realist, such failure does not entail that the term 'species' in necessarily pluralist (and thereby lexically flexible). For all that's been shown there might be a univocal externalist meaning to the term 'species', albeit not one that appeals to intrinsic properties of individuals. Members of a population might well all share certain *relational* properties that ground their natural kind membership. Consider the astronomical term 'moon'. Many diverse celestial objects are moons, composed varyingly of ices, gases, and metals. Even though these objects do not share any particular intrinsic properties unique to moons[10] they are all members of the same kind. Being a moon is a relational property, pertaining to the movement of a celestial body in relation a *planet*, as opposed to (say) a star. Such a property might well ground the kind 'moon' despite its relational character. For all that's been said, 'species' membership might admit to this kind of treatment.

Unsurprisingly, working evolutionary biologists have focused on different aspects of populations in understanding and explaining their evolutionary success, yielding a plurality of relational species concepts (Ereshefsky, 1992). This suggests that

[9]To maintain the essentialist line, one would have to argue that, despite the focus on populations and the difference in members of a population that there is still some intrinsic property these individuals all share, despite these differences. As Ereshefsky (1992) argues, such prospects look grim. Putnam's appeal to some underlying essential genetic properties will not do, for reasons related to the flexibility of the term 'gene'. For more on why intrinsic essentialism fails for biological kinds see Okasha (2002).

[10]They share some intrinsic properties of course: they have spherical shapes, are composed of matter, they move in geometrically predictable orbits, etc. However, none of these properties are unique to moons.

'species' is lexically flexible. These various uses of the term 'species' serve various purposes in explanation, but none of them seem up to the task of serving as the univocal meaning of the term 'species'.

One view, the so-called *Biological Species Concept* (BSC), holds that species boundaries are marked by facts about interbreeding. Most famously Mayr (1970) argues that a species is an interbreeding population that is isolated from other populations according to various mechanisms. This view maintains that an individual is a member of a species based on the individuals with which it can possibly mate. The thought behind this view is that evolutionary biology is primarily concerned with the transmission and exchange of genes throughout the living world. In a world in which there were no restrictions on gene transmission, species boundaries would be too continuous to be useful. Thankfully for biologists, there are significant barriers to gene transmission, marking discrete boundaries between species. Grouping species based on their (potential) mating partnerships maintains the spirit of this idea.

However, even ignoring worries with counterfactual[11] claims about breeding partners, such views fail to classify organisms that do not reproduce sexually. And given that the history of organism on the planet is radically skewed toward asexual reproduction, such a view will (counterintuitively) classify most organism as not members of a species.

Dupré (1999) argues that such a species concept is committed to other empirically unattractive conclusions about species taxa as well. For one, some seemingly distinct species groups exchange genetic information. In particular "different species of oaks have remained coherent and distinct vehicles of evolutionary change and continuity for long periods of time" despite the fact that these "various species of oak appear to have coexisted . . . exchanging significant amounts of genetic material through hybridization" (Dupré, 1999, pp. 7–8). Thus, not only does the BSC fail to delineate asexual species, but the view even misses the mark regarding sexually reproducing species. As such, the essentialist relation proposed by the BSC cannot serve as the univocal meaning of 'species' that the Realist requires.

Other views regarding the meaning of 'species' focus on environmental factors

[11]Other views of this kind focus on mate selection mechanisms to determine species boundaries. That is, on this view an individual falls into a specific species category based on the mechanism it deploys in finding a mate (Paterson, 1985). Such views however will likewise fail to classify asexual reproducers as members of a species, since they lack such mechanisms.

to determine species boundaries. On these *ecological* views, the selection pressures that an environment places on the development of an organism determine which species it belongs to—see for example Andersson (1990). One primary difficulty with such views is that two distinct populations, with distinct selective pressures could evolve to be identical regarding all their intrinsic properties, yet fail to be the same species simply because the set of forces guiding their selection are distinct. At the limiting cases, two organisms could be genetically identical, yet (on this view) fail to be members of the same species.

A third group of theories determine species boundaries based on ancestral relations. Put crudely, on this view an organism's species membership is determined by the species membership of its parents, or in the case of asexual reproduction, its parent organism. The primary concern with this view is, on a Darwinian conception of the evolution of life, all organisms have a common ancestor traced back to the origins of life. And while not all organism are members of the same species, such a hereditary view must posit some criteria for determining taxa changes. Put metaphorically, if the tree of life branches, there must be some criteria by which we can determine where it branches. If an organism's species membership is determined by what branch it occupies on the tree, more needs to be said as to why the tree branches as it does, and these further "diagnosable" details seem to be what a conception of 'species' requires. And often identifying so-called 'speciation events' on the phylogenic tree appeal to morphological, ecological, and reproductive facts, which represent the same features that competing views hold are the appropriate conditions for species membership.

As Ereshefsky (1992) observes, these three views are mutually incompatible.[12] Each view divides up the space of organisms into different species taxa, yielding conflicting pictures about species membership. And while none of these views seems to draw the boundaries in ways that are commensurate with how all working biologists make sense of species, they are all informative:

> A taxonomy of monophyletic taxa provides a framework for examining genealogy. A taxonomy of interbreeding units offers a framework for examining the

[12]This review of the various use of 'species' is not meant to be exhaustive. For a more complete list of such views, numbering in the dozens, see De Queiroz (2007). De Queiroz uses this diversity to argue for a more general species concept that species are "separately evolving metapopulation lineages, or more specifically, ... segments of such lineages" (*ibid*, p. 880). However, the problem with this view is that it fails to distinguish species from other kinds of lineages.

> effect of sex on evolution. A taxonomy of ecological units provides a structure for observing the effect[s] of environmental selection forces. (Ereshefsky, 1992, p. 678)

The point is that various biologists use 'species' as a means of identifying different aspects of organisms that guide selection, each with a different explanatory purposes in mind. While my treatment here is certainly not exhaustive, what is clear is that the Realist is committed to holding that one of these (or some other) conceptions is somehow better than the others, and that some single species concept is the most natural. But even our brief discussion here indicates that this commitment is troubled, since "for biologist who adopt a multidisciplinary approach, or those who can step back from their own personal investments and research interests, all of the concepts seem to have some merits" (De Queiroz, 2007, p. 880). The various lineages identified by these distinct usages of 'species' provide useful insights into the ways in which life has emerged with such manifest diversity, with no conception receiving any obvious priority. Matters look all the more bleak for the Realist if we were to consider the entire array of species concepts identified in the biology literature (see De Queiroz (2007) for a more comprehensive list).

At this juncture, the Realist might insist that our discussion merely indicates that there are a multiplicity of species terms, each denoting a unique extension. Much like the English word 'bank', the biological term 'species' is simply homophonous, used to express a variety of distinct species "concepts." This treatment, she might claim, is consistent with an externalist semantic analysis for the various homophonous terms picked out by 'species'.

But, treating 'species' as a homophonous expression implies that the various 'species' terms indicated by the orthograph bear no semantic relationship to one another. After all, 'bank_f' and 'bank_r' (on the analogous natural language proposal) bear no interesting conceptual relations. They are distinct words, denoting distinct extensions. The Realist's response to the apparent plurality of species concepts would treat the term 'species' as expressing a constellation of distinct concepts, bearing no more interesting a relationship than that posed by the two words 'bank'. For the Realist, this is somewhat odd, given that the differences between these various "concepts" can be attributed to the various explanatory aims of different biologists (Ereshefsky, 1992, p. 678). As we've seen, biologists that study sexual reproduction analyze species in terms of mating, while those interested in the en-

vironmental influences on natural selection will understand species as ecologically determined. Since the Realist methodology requires that these different (pragmatically chosen) terms track the structure of reality, the ontic divisions in the world (according to this Realist move) are dependent on human interests. This belies the Realist insistence that the nature of reality is not dependent on the way humans conceptualize the world.

For the Realist, "it is really, really hard to believe that [facts are]... merely a reflection of something about us" like the context sensitivity of the terms we use (Sider, 2011, p. 18). One of the aims of evolutionary biology is to offer explanations for the distributions of traits manifest in organisms, and the term 'species' occupies an integral role in accomplishing that aim. In particular, evolutionary biologists want to explain whether and how a particular property of an organism played a role in the organism's survival. The explanatory role of the term 'species' is to unify findings about the selection of traits across individual organisms. If there are multiple, distinct species concepts, there are multiple ways of grouping individual organisms. Likewise, there are multiple ways to generalize findings about individuals across groups of them. Because of this, these different ways of grouping individual organisms offer different explanations for the way a particular organism has the traits it does. So on this proposal the phenomena to be explained will admit to multiple, equally adequate explanations. Insofar as these explanations quantify over the exact same set of things, they yield *prima facie* competing natural divisions. They attempt to carve up the world in different ways.

Put in different terms, these various conceptions of 'species' model the world differently. But what the Realist wants is a *single* model, that cleaves to the structure of the world. To quote Sider (2011):

> According to this [Realist] picture, the point of human inquiry ... is to *conform* itself to the world, rather than to *make* the world. The world is "out there", and our job is to wrap out minds around it. (p. 18)

Competing models in evolutionary biology, with different characterizations of species, offer distinct indications about what is "out there." Assuming these different models offer successful explanations, as our review here has indicated, they are all equally viable candidates for the unique description of the world. But, to accept that they are all equally valid is to accept that they are all *true*. If that's correct (as the Realist's homophonous suggestion indicates), there is no "comprehensive theory that

provides the *single*, correct way to represent the causal structure of the world" contrary to the Realist sentiment echoed in Sider[13] above (Waters, 2005, p. 312). To deny this is just to accept a kind of pluralism about the term 'species', and while many philosophers accept this position (cf. Kitcher, 1984; Stanford, 1995; Dupré, 1999; Ereshefsky, 1998), this does not seem welcome to the Realist.

All this suggests that the term 'species' is flexible, in much the way 'book' is. Just as we might use 'book' to highlight different properties of an object based on our interests—as when writing a literary essay, instead of packing a moving box—biologists use 'species' to explain different aspects of an organism's evolution, depending on whether they are focused (say) on the impact of sexual reproduction on selection as opposed to the availability of resources. If 'species' is indeed pluralistic, then the Realist's methodology will fail to yield conclusive results about the ontology of biological kinds. And barring the possibility of reducing biological kinds to chemical or physical kinds, which is problematic for familiar reasons (cf. Griffiths, 1999; Fodor, 1975; Lycan, 1990), the Realist would have to deny that the findings of evolutionary biologists are scientific. The Realist would be forced to conclude that biologists do not speak a scientific language; i.e. that biology is not a science. I take this to be a reason to reject the Realist's methodology, not the scientific *bona fides* of evolutionary biology.

4.3.3 Realist Replies

The Realist might respond to the worries discussed here by indicating that lexically flexible terms are somehow deficient. She might claim (for example) that Gene-P uses of 'gene' will be jettisoned as biology progresses. Indeed, she might insist, there is some sense in which biology is incomplete—*a fortiori*, the flexibility of terms like 'gene' is indicative of this incompleteness. Once scientific investigation in the domain of biology is complete, the language used by biologists will be free of the flexibility seen with 'gene'. Highlighting the "loose" nature of Gene-P uses of 'gene', the Realist could claim that indeed, the *benchmark* of a completed science is the elimination of ambiguity and lexical flexibility. Sider (2011) asserts this very

[13]Sider himself might balk at the idea that the world has a single *causal* structure (cf. Sider, 2011, p. 16), but insofar as the sciences describe the structure of reality, this requires a denial of the kind of pragmatism that typifies pluralist approaches in the philosophy of science (cf. Kellert et al., 2006).

point:

> I hold that the fundamental is determinate... First, no special-purpose vocabulary that is distinctive of indeterminacy ... carves [nature] at the joints. Second, fundamental languages obey classical logic. (p. 137)

Classical languages, with a Tarskian logical structure, do not exhibit properties like lexical flexibility. Since, for the Realist, the structure of the world admits to a classical Tarskian structure, a science whose theory is written in a language not amenable to classical treatment must be an incomplete science—as it does not carve nature at the joints. As such, the theories of such a science cannot be written in the privileged language of ontology ($\mathcal{L}_{\mathcal{O}}$) since "only the propositions in [$\mathcal{L}_{\mathcal{O}}$] are cast in joint carving terms" (Sider, 2011, p. 19).

But this rejoinder to the proposed complications for the Realist's methodology, namely the complications ensuing from the flexibility of terms like 'gene', seems self-defeating. Recall, the reason the Realist embraces scientific languages for use in her ontological investigations pertains to the epistemic credentials of naturalistic inquiry. As Sider indicates

> We should believe generally what good theories say; so if a good theory makes an ontological claim, we should believe it. The ontological claim took part in a theoretical success, and therefore inherits a borrowed luster... [But] the conceptual decisions ... also took part in a theoretical success, and also inherit a borrowed luster. (Sider, 2011, p. 12)

For the Realist, the terminological (or "conceptual") decisions made by a research program, and the language that is the product of those decisions, are given a privileged joint-carving status in accordance with the success of the theory that makes use of them. But in the case of biology, these choices of terminology have lead to the lexically flexible term 'gene' described above. Insofar as the progression of biology has been successful, this success has largely been ascribable to matters related to genetics. So if the epistemic credentials of naturalistic investigation imbue languages with ontological luster, the language used by evolutionary biology should shine quite brightly. Thus to claim that, despite biology's explanatory success, the language of biology is deficient simply because it does not have a classical structure, undermines the motivation for using scientific languages *simpliciter*. Concisely, if the languages of biology lack the luster needed for ontological investigation, despite being developed using (successful!) naturalistic methods, then there is no

reason for using *any* scientific language as an ontic tool. Denying the usefulness of the language of biology in ontological investigation, because of a pre-theoretical commitment to certain metaphysical doctrines, undermines the Realist's appeal to scientific languages in the first place. The Realist owes us a reason for thinking that the biological sciences are somehow epistemologically suspect, despite what seems to be a history of explanatory successes. In the absence of such a reason, the 'book'-like flexibility of 'gene' suggests the Realist's methodology is troubled.

In light of these considerations the Realist might yield some ground. She could admit that terms like 'gene' and 'species' will not be jettisoned from the scientific languages of biology, yet still hold that flexible terms are still aberrant. She might protest that we are conflating distinct notions of scientific languages. Uses of 'gene' and 'species' as described here might be essential for the practices of biologists, but their centrality is of a purely instrumental kind, for use by scientists in the practice of their craft to aid in communication. But these terms are not canonical in the sense that they will find a place in specifications of the generalizations that constitute the theories of (say) molecular biology. That is, we have good reason to mark a distinction between *pragmatic* terms, and *theoretical* terms. Only the latter, the Realist might insist, are genuine constituents of the language of biology—likewise, only the latter find their way into $\mathcal{L}_{\mathcal{O}}$.

The first point to note about this protest is that it commits the Realist to particular (though possibly plausible) views about scientific languages, and how they might be distinguished from natural languages. More importantly though, this requires that the Realist offer a means for making the distinction between pragmatic and theoretical terms. Let's suppose *arguendo* that there is a distinction between pragmatic and theoretical terms in scientific languages. What reason do we have for thinking that 'gene' (at least in the uses outlined above) falls into the former category and not the latter? What are the criteria for determining whether a term used by scientists is theoretical (and thus ontologically trenchant) or not?

The answer most readily on offer from the Realist, that theoretical terms are those terms that can be given externalist meanings, will not do. Such a response begs the question at issue. Recall that the Realist's retreat to scientific languages is motivated by concerns about the semantics for natural languages. In response to the problems posed by the lexical flexibility of natural language expressions, the Realist (rightfully) abandoned those languages as ontological guides. She turns to scientific

languages in the hope that they will be better behaved, given that such languages are invented for the purpose of perspicuously describing the world. Essential to this move is the epistemic appeal of naturalist inquiry. To quote Sider again:

> But in trying to decide how much structure there is in the world, I can think of no better strategy than this extension of Quine's criterion: believe in as much structure as your best theory of the world posits. (Sider, 2009, p. 417)

The assumption that the world has a structure of a Tarksian logic garners support when the terms adopted by the sciences, as the result of naturalistic inquiry, can be given a semantics to cohere with an object-based structure. Thus, the reason we should think that the world has an object-based structure is that scientific methods, with their privileged epistemic credentials, produce languages whose semantics reflect that structure.

The justification for the Realist's retreat to scientific languages renders the appeal to the structure of reality as a means of marking the pragmatic/theoretical distinction circular. Naturalistic inquiry has led to the invention of the term 'gene', which given the lexically flexibility it exhibits, does not easily admit to externalist treatment. If the Realist's response to dealing with this counterexample to her general claim—namely, that we should think the world has an object-based structure because scientific languages have a semantics that reflects this structure—is to mark a distinction between pragmatic and theoretical terms, she cannot appeal to the object-based structure of reality to draw the distinction. To indicate that theoretical terms are those terms that result from naturalistic inquiry *and* have an object-based semantics undermines the epistemic credentials of naturalistic inquiry in identifying reality's structure. If the thought is that such inquiry results in terms that cleave to the structure of the world, the case of 'gene' indicates that this structure in not (obviously) object-based. If the Realist insists that 'gene' is no good for ontological investigation, despite the fact that the term is the result of successful naturalistic inquiry and central to producing explanatory generalizations in biology, there must be some property, beyond its resistance of externalist treatment, that makes 'gene' ill-suited for ontological investigation.

The Realist would be able to leverage the object-based structure of reality in marking the pragmatic/theoretical distinction if there was some independent reason to think the world has such a structure. Sider offers such a reason for assuming there is an object-based structure of the world, which amounts to a denial of the

closest alternative:

> Suppose... that a suitable stuff-ontology could be constructed. Why should we accept it? One reason for moving to a stuff-ontology is inherently unstable. This reason begins with the epistemological skepticism about metaphysics considered earlier—questions about composition are unanswerable. It then adds some sort of prohibition against questions that are, in principle, unanswerable.
>
> ...But upholding stuff-ontology just substitutes one unanswerable question for another: is a stuff- or thing-ontology correct? This question seems no more answerable than questions that face the thing-theorist. (Sider, 2002, p. xviii)

The criticism offered against "epistemological skepticism" referenced in this passage is the following, which I offer in its entirety:

> Skeptics often ask too much of metaphysical arguments. A priori metaphysical arguments should not be faulted for not being decisive. For suppose the evidential support conferred by such arguments is fairly weak, though nonzero. Then the support for a typical metaphysical theory, T, will be weak. But the only support for T's rivals will also be from a priori metaphysical arguments. Thus T may well be better supported—albeit weakly—than its rivals. One would then be reasonable in giving more credence to T than to its rivals. Metaphysical inquiry can survive if we are willing to live with highly tentative conclusions. Let's not kid ourselves: metaphysics is highly speculative! It does not follow that it is entirely without rational grounds.
>
> I will proceed assuming that reasonable belief in metaphysics is indeed possible, and that something like the [Quinean Realist] methodology is legitimate. (Sider, 2002, p. xv)

These remarks suggest an argument pertaining to the question of reality's structure, and might legitimize the pragmatic/theoretical distinction deployed to deal with lexically flexible scientific terms:

1. Reality has a structure.
2. Questions about the nature of this structure (e.g. that it is thing-based) have determinate answers.
3. Evidence for various theories that answer these questions, even if minimal, make these debates substantive.
4. Thus a wholesale skepticism about the substance of such debates is unwarranted.

An anti-Realist might take issue with any of these premises. But even if one accepts each of these premises, these considerations do not amount to a distinct argument for the thing-based structure the Realist assumes—or if they do, the evidence for the position bears the burden of scrutiny. The force of highlighting examples like 'gene' and 'species' is that they offer evidence against the proposal that reality's structure is of the Tarskian sort that the Realist proposes. Such counterexamples hold sway even if one assumes that reality has a structure and that proposals about that structure constitute a substantive area of theorization. If the Realist hopes to rebuff the force of these counterexamples by insisting there is a marked distinction between pragmatic and theoretical terms, the criteria for marking that distinction cannot merely rest on the very supposition the counterexamples undermine. The Realist cannot merely insist that the flexibility of terms like 'gene' and 'species' mark them as merely pragmatic, bereft of the metaphysical merits of theoretical terms.

If the Realist proposal puts constraints on the kinds of terms that count as genuine theoretical terms of the sciences, there must be quite powerful reasons to place such constraints on the practices of scientists prior to investigation. Reality's structure may very well be thing-based. But by insisting that any scientific language that fails to reflect this structure is somehow metaphysically deficient, the Realist is burdened with providing quite powerful reasons for injuncting the offending terms. But there seems to be little to say in defense of this position. Arguments for this view simply propound metaphysical intuitions, or recapitulate the "best theory" argument above. Apart from the (possibly pronounced) intuition that the world has a thing-based structure, or Sider's professed inability to imagine a better option, there seems little support for the injunction in the case of 'gene' and 'species'. Neither of these appeals should be particularly convincing, given that these terms play a quite central role in the success of biology as a scientific discipline. In the absence of a convincing account of the distinction between pragmatic and theoretical terms, and a means for identifying which terms fall in which camp, the trouble posed by terms like 'gene' and 'species' remains.

4.4 Implications

Our cursory examination of these terms in the scientific languages used in biology is not decisive, nor can we conclude that the terms 'gene' and 'species' will *never* have univocal meanings that determine a single precise extension. Further a reduction of biological notions like 'gene' might be possible (cf. Waters, 2000). However, both the terms 'gene' and 'species' play a valuable explanatory role in biology, and the evidence considered here suggests that the meanings of these terms exhibit a kind of lexical flexibility. Insofar as there are multiple distinct, but related uses of the term 'gene' (cf. Weber, 2005; Beurton et al., 2000) and 'species' (cf. Mayden, 1997), uses that attribute properties of ontologically distinct types, these terms seem hostile to externalist treatment. At the very least these findings suggest that we should not settle at the outset of scientific investigation that the languages needed to express scientific theories *must* have a particular semantics. Insofar as the Realist requires that we make such stipulations, we have reason to doubt that the Realist ontological methodology should be adopted. Maybe the semantics for a language of biology will prove to be externalist, but given the arguments here, this metaontological assumption is at least contentious. To the degree that other terms in the sciences are like 'gene' and 'species' metaphysicians should abandon the Realist position.

Chapter 5

Against Conceptual Analysis

Conceptual Analysis is a pervasively used methodology for adjudicating questions in metaphysics. The success of this method in settling metaphysical questions hinges on the broader Realist ontological methodology. The problems presented for the Realist bear on metaphysics writ-large, insofar as they undermine the use of Conceptual Analysis in resolving metaphysical disputes. The difficulties with accepting the externalist thesis pose a dilemma for the Realist's use of Conceptual Analysis in addressing questions in metaphysics, or so I'll argue in this closing chapter.

I'll do this by way of example, highlighting what I take to be the foundational mistakes in much metaphysical speculation. To start, I will indicate explicitly what I take the preceding arguments to have shown about the prevalent methodology adopted by metaphysicians, and those engaged in metaphysical projects in other philosophical domains. These conclusions undercut particular, recent arguments regarding the Extended Mind Hypothesis. I surmise that this diagnosis for the absurdity of positing extended minds will be welcome to most philosophers. However, an analogous argument regarding debates in the metaphysics of causation undermines the primary mode of philosophical argumentation deployed in such debates. While this result might meet with more resistance than the analogous conclusion regarding the Extended Mind Hypothesis, it is no less forceful. The broader worry for metaphysicians, and much metaphysical speculation throughout philosophy, is that this anti-metaphysical argument generalizes, applying to any domain that relies on the judgments of natural language speakers as evidence for a proposal regarding the truth-conditional definition of any concept apt for addressing metaphysical disputes. Put more succinctly, if the arguments in the preceding chapters are correct, many classic and contemporary metaphysical debates are grossly misguided.

5.1 Externalism and Conceptual Analysis

Chapters 3 and 4 argued that the Realism defined in Chapter 2 requires a semantics for natural and scientific languages that is externalist, in the manner

described in Chapter 1. This externalist thesis is deeply troubled. While these arguments fall short of showing that externalism is false, in the case of natural language they pose a serious, and recalcitrant problem for any theorist committed to the idea that natural language expressions have meanings that determine their truth-conditions. Because the Realist is committed to this externalist idea, this problem renders her methodology without a firm foundation. Though similar arguments are less decisive in the case against an externalist semantics for scientific languages, they should at least give any Realist serious pause in using the meanings of scientific claims as useful tools in metaphysical investigation. At the very least then, metaphysical investigators should refrain from using methodologies that rely on the following:

($\mathcal{E}$) For any expression e (in some language L), the meaning of e determines e's truth-conditions.

The denial of (or more strictly, agnosticism regarding) ($\mathcal{E}$) undermines the primary methodology behind many metaphysical arguments throughout philosophy. The following form of argument is quite familiar, and if I'm correct, deeply troubled:

1. L-speakers understand the meaning of expression e;
2. Theory T holds that e has truth-conditions e_{def};
3. ϕ is an expression that has e as a constituent;
4. Consider case C;
5. C describes a truth-maker of ϕ, according to T;
6. Theory T predicts that L-speakers will judge expression ϕ "True" (or "False") of C;
7. L-speakers *in fact* judge expression ϕ "True" (or "False");

$\therefore$ Theory T makes the right (or wrong) prediction.

If the reader finds this form of argument unfamiliar, that is because such arguments tend to omit an explicit endorsement of 1. (as well as ($\mathcal{E}$)), and tend to phrase 6. and 7. in terms that do not refer to speakers of a language, but simply assert that

the relevant expression *is* true/false according to the theory under evaluation and simply *is* true/false (given the case).

Readers might also find this form of argument unfamiliar for different, though related reasons. Omitting explicit reference to a language, and the judgments of speakers of that language, blurs the theoretical significance of presenting counterexamples as cases against theories that define the truth-conditions for some expression. One further obfuscation is the habit of labeling this form of argument "conceptual analysis." Arguments of this form are typically worded as involving, not expressions, but *concepts*. The purported counterexample to the disputed theory about the truth-conditions of some particular *concept* is meant to show that the particular *concept* humans bring to evaluating the counterexample has truth-conditions that diverge from those offered by the theory.

However, to indicate that a case-based counterexample sheds light on a particular concept is only viable if the relationship between linguistic expressions and concepts is fairly determinant. While human understanding of language is surely mediated by the conceptual system, the relationship between language and the conceptual system must be a one-to-one mapping between words and concepts for the form of argument above to yield straightforward verdicts about the nature of *concepts*. This would require the proponent of Conceptual Analysis to adopt the labeling theory addressed in Chapter 3. As indicated, such a theory is fairly troubled, yet seemingly assumed by many philosophers. The form of argument outlined above, which is pervasive throughout philosophy, does not aid in analyzing concepts—at best they provide evidence for semantic linguistic analysis. One can conjecture that this form of argumentation is also directly informative about the content of our concepts, but that conjecture requires some substantive defense in light of the concerns expressed in Chapter 3.

Put in other terms, and with less pace: in utilizing Conceptual Analysis the Realist wants to convince us that some theory about the nature of some (natural) kind is (in)correct. She does this by offering an argument, whereby she presents a case that supports (or belies) that theory. Her argument, and the description of the case is expressed *via* some language or other. That language, whatever it might be, does not make use of concepts. The case is described in sentences that make use of expressions in that language, not concepts of the human mind. This method of argumentation and investigation can only serve as a means of analyzing conceptual

content given additional (adventurous) assumptions about the relationship between linguistic meaning and conceptual content.

While it's difficult to find an explicit articulation of these additional assumptions in the literature, given the history of semantics (as perused on Chapter 1) one can confidently speculate that these assumptions trade on the relationship between truth and meaning. If both linguistic meaning and conceptual content are characterized by way of truth, then it is quite plausible that one can map the (truth-conditional) meanings of linguistic expressions onto the satisfaction conditions of (truth-tracking) concepts. The method of Conceptual Analysis trades on this assumptions, as will be fairly clear upon examining the method.

Conceptual Analysis leverages the truth-conditions described in some case against the proposed truth-conditions for the expression under investigation. As a speaker of the language used to pen the example, readers are asked to entertain some case description as if it were truthful, and then asked to judge the truth-value of some statement about the described scenario. Put more technically, a case C describes (in language L) some happening, wherein the expressions of that description have truth-conditions. L-speakers are asked to suppose those truth-conditions are satisfied by objects in the world. The theoretical upshot is purportedly had when L-speakers are then asked whether some sentence ϕ, which contains the expression of interest e, is also made true by the same object-satisfiers of the descriptions that constitute C. The L-speakers verdict (ostensibly) speaks as to whether the truth-conditions that are offered up for e by T are apt or not.

A toy example will make clear the methodology here:

Theory T:

$[\![\text{chair}]\!] = \lambda\text{x.FOUR-LEGGED(x) \& SEAT(x) \& BACK(x)}$

Case C (as described in English):

> Mel is made of wood. Mel has a surface that can be sat on. Mel has three legs. Mel does not have four legs. Mel has a back.

ϕ = Mel is a chair.

English-speaker judgment:

ϕ is true (of C).

The truth-conditions given in C are the conjunction of the following claims:

$[\![\text{Mel}]\!] = m$ (an object in the real-world domain)

$[\![\text{Mel is made of wood}]\!] = \top$ *iff* $[\lambda\text{x.MADE-OF-WOOD(x)}](m)$

$= \top$ *iff* $\text{MADE-OF-WOOD}(m)$

$[\![\text{Mel has a surface that can be sat on}]\!] = \top$ *iff* $[\lambda\text{x.SEAT(x)}](m)$

$= \top$ *iff* $\text{SEAT}(m)$

$[\![\text{Mel has three legs}]\!] = \top$ *iff* $[\lambda\text{x.THREE-LEGGED(x)}](m)$

$= \top$ *iff* $\text{THREE-LEGGED}(m)$

$[\![\text{Mel does not have four legs}]\!] = \top$ *iff* $\neg[\lambda\text{x.FOUR-LEGGED(x)}](m)$

$= \top$ *iff* $\neg\text{FOUR-LEGGED}(m)$

$[\![\text{Mel has a back}]\!] = \top$ *iff* $[\lambda\text{x.BACK(x)}](m)$

$= \top$ *iff* $\text{BACK}(m)$

Taken in conjunction then, C is true just in case the following is true:

$\text{MADE-OF-WOOD}(m)$ & $\text{SEAT}(m)$ & $\text{THREE-LEGGED}(m)$ &

$\neg\text{FOUR-LEGGED}(m)$ & $\text{BACK}(m)$

If this expression is true, then the *satisfiers* of the five predicates are indicated therein. That is, m satisfies the truth-conditions of the related linguistic expressions in C.

According to Theory T, the truth-conditions for 'Mel is a chair' are:

$[\![\text{Mel is a chair}]\!] = \top$ *iff* $[\lambda\text{x.FOUR-LEGGED(x)}$ & SEAT(x) & $\text{BACK(x)}](m)$

or simplifying:

$[\![\text{Mel is a chair}]\!] = \top$ *iff* $\text{FOUR-LEGGED}(m)$ & $\text{SEAT}(m)$ & $\text{BACK}(m)$

As such, Theory T predicts that English-speakers will judge 'Mel is a chair' to be false in C. The truth-conditions of the expressions used to describe C indicate that '$\neg\text{FOUR-LEGGED}(m)$' is true, rendering '$\text{FOUR-LEGGED}(m)$' false. Thus, that English-speakers judge ϕ to be true, holding that Mel is indeed a chair, serves as evidence against Theory T. This analysis is meant to tell us about our concept CHAIR, indicating that chairs need not have four legs.

Admittedly, this is a toy example focused on a sophomoric case, with quite impoverished indications of the semantics for the English expressions involved. However, what should be clear is that the theoretical upshot of this methodology, when applied to a domain of purported philosophical interest, only gets traction on the relevant conception (e.g. CHAIR) if we assume both that the meanings of English language expressions (like the ones used in C) determine their truth-conditions, and those truth-conditions can be given in terms of conceptual contents that have real world satisfiers (e.g. SEAT(m)).

In reading the case in C, English-speakers understand the meanings of the expressions used therein. On the hypothesis that what said speakers understand are the truth-conditions of those expressions, this method can (straight-forwardly) inform us about the meanings of expressions like 'chair'. On the further hypothesis that the satisfiers of those truth-conditions also serve as satisfiers of concepts (e.g. CHAIR) that have similar satisfaction-conditions, this method gives us an analysis of our human concepts. The discussions in Chapters 3 and 4 should indicate the rather bold character of these two hypotheses. In the absence of a rather substantive defense of these thesis, such a method should not be adopted by philosophers hoping to get traction on metaphysical questions.

The method of Conceptual Analysis is related to the more general Realist methodology of determining the residents of the worldly domain by semantic analysis. In an example explored below regarding causal expressions like 'Suzy caused the bottle to break', the standard counterfactual analysis demands that, in order to accurately codify the truth-conditions of causal claims, the domain must include possible worlds that admit to a weak ordering in terms of proximity. Postponing the details for the discussion below, the point here is that the Realist's methodology assumes that the meanings of linguistic expression determine their truth-conditions, and because of this, metaphysicians can infer from their own comprehension of an expression's meaning, the kinds of objects that must exist in the worldy domain to serve as satisfiers for the truth-conditions of the understood expressions. If this requires positing an infinite plenitude of possible worlds capable of standing in a similarity relation, so be it (Lewis, 1986). If encoding the truth-conditions of action sentences requires positing events (Davidson, 1967b), on which different perspective can be taken (Schein, 2002), so be it. If the semantics of proper names for fictional characters requires a domain of *fictional* worlds, so be it (Thomasson, 1999). The

point I've been arguing, and that I hope to crystallize in these closing sections, is that the motivation for populating the world with these extravagant entities is predicated on the mistaken assumption that an expression's meaning determines its truth-conditions.

In what remains, I outline the use of this strategy, and the method of Conceptual Analysis that appropriates it, in the context of two distinct metaphysical debates. Noting the troubled nature of this strategy will serve to clarify the basic metaphysical question in both domains, and suggest that in both cases the vexing philosophical question is either not that vexing—or at least not vexing in the way supposed by those in the debate.

5.2 Extended Minds

Since the publication of Clark & Chalmers (1998), the proposal that human minds have boundaries that extend well beyond our bodies has received consideration attention.[1] The central proposal suggested by Clark & Chalmers (1998) is that, counter-intuitively, the human mind extends well beyond the boundaries of the human organism.

The main contention of the Extended Mind Hypothesis is that features of the external environment constitute parts of the human mind. The introductory example given by Clark & Chalmers (1998) as a means of clarifying their thesis involves the video game *Tetris*. In this video game players are tasked with arranging two-dimensional shapes into a block formation. Critical to success in this task is the ability to quickly judge whether an individual piece will fit into an opening in the block formation. Because players can rotate the two dimensional pieces clockwise ninety degrees by pressing a button on the game's control pad, the player has some control over where each piece will go, and how it will fit into the block.

Assessing whether the piece can fit into a given place in the block-formation can be accomplished in two ways: either by the player imagining the various ways in which the piece can be orientated, rotating this imagine "in her head", or by pressing the button on the control pad to rotate the piece on the video screen, and checking the fit of the piece using her visual system. As it turns out, the latter

[1]Multiple publication indexes number the papers that are about or refer to Clark & Chalmers (1998) in the thousand; e.g Goggle Scholar indicates 2,069 references; PhilPapers cites 119 references from 2010–2014.

strategy tends to be significantly faster. Clark & Chalmers (1998) indicate that in contrasting these two strategies, the function played by aspects of the mind used in imaginatively rotating the shape (in the first case), and the function played by the gaming system's rotation algorithm (in conjunction with the player's visual system, in the second case), are the same. Yet, only in the former case do we conclude that the procedure the player engages is a mental one. To Clark and Chalmers this distinction seems to be without any basis.

Their controversial claim however is not merely that, on occasion, some processes that have all the appearance of cognitive processes in fact involve organism-external objects. The claim is that key "core" components of the mind, like beliefs and desires, can be constituted by organism-external objects. If some organism-external object plays the same functional role as an organism-internal object in paradigmatic, core mental processes like belief, then (barring some other robust reason to the contrary) organism-external objects can be components of the human mind.

The primary argument presented for this view in Clark & Chalmers (1998) involves a contrast between two cases, meant to "argue that *beliefs* can be constituted partly by features of the environment, when those features play the right sort of role in driving cognitive processes" (Clark & Chalmers, 1998, p. 12). The first case is supposed to illustrate that the nature of *belief* (of a certain sort) involves recall from memory, while the second case shows that the role played by memory in the paradigmatic case of belief can be filled by an organism-external object, i.e. a notebook. In that vein, consider:

> [Inga]
> Inga hears from a friend that there is an exhibition at the Museum of Modern Art, and decides to go see it. She thinks for a moment and recalls that the museum is on 53rd Street, so she walks to 53rd Street and goes into the museum. (Clark & Chalmers, 1998, p. 12)

Clark and Chalmers conclude from this case involving Inga, that she "clearly believes that the museum is on 53rd Street" and that (because of this) *beliefs* can be stored in memory (Clark & Chalmers, 1998, p. 12). They come to this conclusion for the same (errant) reasons that most philosophers come to metaphysical conclusions based on intuitive evidence from cases of this sort—namely that the meanings of expressions determine their extensions, and thus determine what kinds of objects populate the worldly domain.

Suppose we take their data point as given, namely that English-speakers would assent to the truth of ϕ on the assumption that all of the expressions in INGA are true:

ϕ = Inga believes that the Museum is on 53rd Street.

Such a data point does not permit the conclusion that *beliefs* can be stored in memory (even if we assume that meanings determine extensions). After all, Inga's belief that motivates her to head toward 53rd Street could simply be the *occurrent* belief she is consciously entertaining, relevantly divorced from her memory. To reach the conclusion that beliefs can be constituted by memory, even under the assumption that this case-driven methodology is justified, requires a slightly different case than the one they propose. Consider then:

> [INGA*]
> Inga hears from a friend that there is an exhibition at the Museum of Modern Art, and decides to go see it. [She stops to tie her shoes.] She thinks for a moment and recalls that the museum is on 53rd Street, so she walks to 53rd Street and goes into the museum. (Clark & Chalmers, 1998, p. 12)

Given that the expressions in INGA* are true, plus the assumption that the methodology of Conceptual Analysis is justifiable, Clark and Chalmers can reach their conclusion regarding the relationship between *memory* and *belief* if English-speakers would find the following true (of INGA*):

(ϕ^*) Prior to tying her shoes, Inga believed that the Museum is on 53rd Street.

Of course, even if we assume that English-speakers have the judgments Clark and Chalmers need, such data only justifies the claim that *beliefs* can be constituted by memory if we assume that the meaning of the English language expression 'belief' determines its extension. That is, the nature of *belief* is only enlightened by such judgments if the meanings of expressions like 'thinks' and 'recalls' reflect what must be true of the world in order for speakers to judge that the relevant claim using 'belief' is also true. Again, the idea behind Conceptual Analysis, and the externalist assumption that underwrites the methodology, is that English-speakers' judgments regarding ϕ^* will accord with the (purported) fact that the objects required to satisfy the truth-conditions of the sentences in INGA* will also satisfy ϕ^*. But again, this methodology only yields the metaphysical verdict that (say) *beliefs* are partially constituted by memory, if the externalist thesis ($\mathcal{E}$) holds. Insofar as the

language used to described INGA* is a natural language like English, Chapter 3 offers compelling reasons to think these metaphysical conclusions are unjustified.

Yet Clark and Chalmers insist that cases like INGA* offer compelling evidence about the nature of *beliefs*. In this vein, they offer up the case of OTTO′ to contrast with INGA*:

> [OTTO′]
> Otto suffers from Alzheimer's disease, and like many Alzheimer's patients, he relies on information in the environment to help structure his life. Otto caries a notebook around with him everywhere he goes. When he learns new information, he writes it down. When he needs some old information, he looks it up [in the notebook] ... Today, Otto hears about the exhibition at the Museum of Modern Art, and decides to go see it. [He stops to tie his shoes.] He consults his notebook, which says that the museum is on 53rd Street, so she walks to 53rd Street and goes into the museum.[2] (Clark & Chalmers, 1998, pp. 12–13)

Contrasting Otto's actions with Inga's, Clark and Chalmers conclude "that when it comes to belief, there is nothing sacred about the skull and skin" because, much like Inga, Otto had a belief about the museum's location prior to walking toward it (Clark & Chalmers, 1998, p. 14). Otto's notebook, they claim, plays the same explanatory (and hence functional) role in the most plausible psychological account for his actions as Inga's memory plays in an analogous explanation for her actions. Thus, if Inga has a non-occurrent belief about the museum's location, so too does Otto, despite his reliance on a notebook. Insofar as beliefs constitute a core aspect of the human mind, then Otto's notebook, as a constituent of his belief, must be considered a part of his mind.

This seems to imply that, according to Clark and Chalmers, English speakers would accept the following as true:

(ϕ') Prior to tying his shoes, Otto believed that the Museum is on 53rd Street.

Supposing that English-speakers indeed find ϕ' to be true (given OTTO′), such evidence does not yield the bold metaphysical conclusion about the nature of belief without assuming that the (troubled) externalist thesis ($\mathcal{E}$) is true.

However, Clark and Chalmers do not predict that English-speakers would assent to the truth of ϕ' given the case described in OTTO′. They indicate that

[2] As with INGA, the description of Otto's case will fail to yield the verdict Clark and Chalmers require without some additional information to distinguish occurent beliefs from stored beliefs. It should also be noted that Clark and Chalmers do not give the cases names. These have been added for the purpose of clarity.

the thesis being defended is not about "common usage; [their] broader point is that the notion of belief *ought* to be used so that Otto qualifies as having the belief in question" (Clark & Chalmers, 1998, p. 14). The Extended Mind Hypothesis then is a claim about how the expression 'belief' *ought* to be used, such that the extension of the term counts Otto as having a belief about the location of the museum (prior to consulting his notebook).

A charitable[3] reading of their normative claim here suggests that the Extended Mind Hypothesis regards, not the natural language expression 'belief' (or 'mind'), but is rather a proposal for a scientific terms 'belief$_s$' (or 'mind$_s$'), whose extension includes Otto's notebook as described in Otto′. Their proposal seems to be that our best psychological theory would be better served by making use of a term like 'belief$_s$' or 'mind$_s$' in explaining human cognition.

If this is the proposal on offer, two problems undercut the metaphysical claim they want to make about the constituents of belief. First, this suggestion presupposes that the language used to express theories in psychology has an externalist semantics. Chapter 4 offers some reasons to be hesitant in adopting such a view, using examples for the domain of the biological sciences. Second, even assuming that the scientific language of psychology has an externalist semantics, the use of speaker intuitions about cases like Inga* and Otto′ as evidence for including terms that permit the admission of organism-external objects as constituents of the mind, is fairly strange. After all, English-speakers are speakers *of English*, not the technical language used for theorizing in psychology. But more importantly, if the suggestion is that a research program engaged in studying human psychology which includes extended-mind-friendly terminology has more explanatory power than one that does not, the arguments for such a view should appeal to the typical standards for evaluating scientific theories. Surely, that the Extended Mind Hypothesis can explain the behavior of a single fictional individual more succinctly than more standard theories is (at best) weak evidence for the theory. Especially if it does so at the expense of more mundane (real-world) cases (Adams & Aizawa, 2010).

[3]Another reading of their claim is that the natural language expression 'belief' ought to have a different meaning than the one it in fact has. If this is their suggestion, the proposal is quite odd given the use of English-speaker judgments as evidence. Whatever the meaning of the English expression 'belief' might be, surely the judgments of English-speakers regarding uses of the term is the primary source of data for giving a semantics for the expression. So if the goal is to give a semantic analysis of that expression, indicating that English-speakers are just plain wrong about the meaning of 'belief' makes the use of case-based judgments as evidence entirely unconvincing.

This sort of argumentation trades on an assumption about the nature of linguistic meanings, namely that the meanings of the expressions we use determine their truth-conditions. This assumption carries with it an implicit demand for metaphysical consistency. Our judgments about the truth-conditions of the expressions we use, according to this demand, ought to remain ontologically consistent across contexts. This insistence permits leveraging cases like INGA*, that ostensibly indicate the paradigmatic truth-conditions of a particular expression (as exhibited by speaker judgments about the term of interest) to make metaphysical proclamations *via* supposedly structurally similar cases like OTTO′. The general thought behind this argumentative strategy is that there is some core, truth-conditional meaning to the natural language expressions we use. These meanings, while sometimes reflected in the judgments of competent speakers, require precisification by way of imaginative counterexamples that enable us to push the limits of those truth-conditional meanings. Of course, I think this line of reasoning is predicated on a mistaken assumption of natural language meanings.

The Realist then owes us a response to these substantive worries, or she is forced to admit that her investigation does not pertain to natural language expressions and their meanings (which many, like Clark and Chalmers seem happy to admit). But this admission presents a further problem for the Realist, given her use of language as a tool for investigation. If the language she uses to describe her intuition-pumping cases is not a natural language, then she owes us a justification for thinking meanings in this language are indicative of what exists. As we've seen in Chapter 2, a promising response to this challenge appeals to naturalistic methodology. If the language used to investigate ontological questions is a scientific language, then the epistemic credentials of naturalist methods of inquiry provide (at least some) justification for thinking scientific terms "cut nature at the joints."

But then, the relevant question for metaphysical inquiry is whether a theory with the metaphysically interesting term admits to more explanatory successes with regard to the data *in the domain of inquiry* than a theory with no such term. Unless that domain is the naturalistic study of language, the evidential import of cases meant to pump intuitions (and described in a natural language) is (at best) insignificant. What matters is whether the research program that implies the existence of the proposed entity is successful along the dimensions that matter for the purposes of science. Such success is hard fought, over decades of investigation

with tested methodologies, not the product of imaginative stories that exploit the flexibility of natural languages and the speakers that understand them.

5.3 Causation

As a topic of metaphysical investigation, study regarding the nature of *causation* has as storied a history as any, dating back to Plato's dialogues. Much more recently, reductive accounts of causation favor a counterfactual analysis indebted to Lewis (1973a). The primary goal of this research program seems to be one of modification, adding to and adjusting elements of Lewis' core insight to deal with the errant predictions of the theory. Lewis' proposal is to explain causal facts in terms of counterfactual dependence, thereby reducing such facts to *modal* facts about possible worlds. For one event to cause a later event, the latter must counterfactually depend on the former. Formally, the proposal is this:

($\mathcal{C}$) For any two distinct events c and e, c causes e *iff* there is a set of events $(d_1, d_2, \ldots d_n)$ such that if c had not occurred, then d_1 would not have occurred; and if d_1 had not occurred, then d_2 would not have occurred; ... and had d_{n-1} not occurred, then d_n would not have occurred; and had d_n not occurred, then e would not have occurred.

This analysis is counterfactual insofar as the right-hand side of the biconditional is a (series of) counterfactual conditional(s). Actualized events c and e stand in a causal relation on the condition that if (counter to fact) c had not actually taken place, then (roughly) neither would e. (This is a slight misstatement of the analysis given above, which is an *ancestral* version of direct counterfactual dependence—the reason for this difference will be explained below.)

The success of this account in analyzing causation depends on the manner in which it treats counterfactual conditionals. Counterfactual conditionals contrast with more standard conditionals, in that their truth-conditions are not obviously systematic. Contrastingly, the standard material conditional has well understood truth-conditions, and is always true whenever the antecedent of the conditional is false. If the conditionals in $\mathcal{C}$ are taken to be *material* conditionals, then any actual event c would be the cause any other event at any point in history, because the conditional 'if c had not occurred, then e would not have occurred' would never be

false (as a result of the false antecedent). The familiar point is that counterfactuals can be (but are not always) true when their antecedents are false. Thus a successful reduction of *causal* facts to counterfactual dependence requires a compelling analysis of *counter*facts (so to speak).

The account of counterfactuals given by Lewis (1973a)—and Lewis (1973b)—understands counterfactuals via possible worlds and the relations that hold between them. To quote:

> Given any two propositions A and C, we have their counterfactual $A \mathrel{\Box\!\!\rightarrow} C$: the proposition that if A were true, then C would also be true. The operator $\mathrel{\Box\!\!\rightarrow}$ is defined by a rule of truth, as follows. $A \mathrel{\Box\!\!\rightarrow} C$ is true (at world w) iff either (1) there are no possible A-worlds (in which case $A \mathrel{\Box\!\!\rightarrow} C$ is *vacuous*), or (2) some A-world where C holds is closer (to w) than any A-world where C does not hold. (Lewis, 1973b, p. 559)

The thought behind this technical account of counterfactuals is based on the idea that, given a plenitude of possible worlds, any set of these worlds constitutes a proposition (Lewis, 1973a, p. 556, note 3). Thus, $A \mathrel{\Box\!\!\rightarrow} C$ is a claim about how two sets of possible worlds are related, the A-worlds and the C-worlds. These are the worlds at which for any sentence whereby A (or C respectively) is the sentence's (propositional) meaning, that sentence is true at each world that is a member of A (or C). The claim $A \mathrel{\Box\!\!\rightarrow} C$, made at some possible world w, is true at w just in case the A-world closest to w is also a C-world. Put another way, there are a bunch of worlds where A is true, a bunch of worlds at which C is true, and a bunch of world where both are true. Given that some worlds can be "closer" to others, if one of those worlds where both A and C are true is closer to the world we care about (the world w at which $A \mathrel{\Box\!\!\rightarrow} C$ is being evaluated/uttered) than any world where A is true and C is false, the counterfactual $A \mathrel{\Box\!\!\rightarrow} C$ is true (at the world we care about, w).

Given this analysis of conditionals like $A \mathrel{\Box\!\!\rightarrow} C$, one wants to know what makes a possible world *closer* to another. Lewis suggests that this relation is one of similarity, whereby two worlds are closer to each other than some third world if they are more similar to each other than either is to that third world. For the purposes of Lewis (1973a), he leaves the notion of similarity (and closeness) as primitive, offering only a suggestive analysis of what similarity must (not) be like (Lewis, 1973a, pp. 559–560). Naturally much has been made about the nature of the similarity relation and the manner in which one is to weigh the various features of possible worlds in

determining their proximal properties (cf. Bennett, 1974; Fine, 1975). Assuming a plausible analysis of similarity is available, the truth-conditions given above for sentences like '$A\Box\!\!\rightarrow C$' offer a compelling account of counterfactuals, and thereby, causation.

To explain by way of example, consider the sentence, uttered in the actual world ($w_@$):

(101) If McCainn had chosen a different running-mate, then Obama would have lost the 2008 presidential election.

Ignoring syntactic differences, this sentence can be put in a form amenable to Lewisian analysis:

(101′) [McCain chose a different running-mate] $\Box\!\!\rightarrow$ [Obama loses the 2008 presidential election].

Let's suppose that the English expression in (101) is true *iff* the (psuedo-English) expression in (101′) is true. (This, of course, is not obvious, given that these expression have different syntactic structures, and one contains terms that are not familiar to English-speakers.) On Lewis' analysis of (101′), the meaning of 'McCain chose a different running-mate' is some proposition, as a set of worlds—call that proposition M. Likewise, the meaning of 'Obama loses the 2008 presidential election' is some proposition—call it O. As such, (101) is true just in case '$M\Box\!\!\rightarrow O$' is true (at $w_@$). If we consider the worlds at which McCain choses a different running mate (someone other than Palin), we want to find the closest-to-$w_@$ world where McCain also wins the election (i.e. where Obama losses the election). Call that world w_{mo}. Still considering just the worlds at which McCain choses someone other than Palin, we now want the closest-to-$w_@$ world where Obama nonetheless wins. Call that world $w_{m\not o}$. Given these two worlds, if w_{mo} is closer to $w_@$ than $w_{m\not o}$ is (to $w_@$), then (101) is true (and otherwise it is false). The account gets the correct result just in case the truth (or falsity) of (101) accords with the proximity of $w_@$ to w_{mo} and $w_{m\not o}$.

Further this analysis underwrites (let's suppose) the causal fact that McCain's choice caused Obama to win the election. The counterfactual account of causation indicates that some event c causes e just in case, if c did not occur, then e would not have occurred. Assuming that 'McCain chose Palin as a running-mate' and

'Obama wins the 2008 presidential election' describe different events, we can take the counterfactual analysis of (101) and apply it to an analysis to the causal fact that McCainn's choice of Palin as a running-mate caused Obama to win the 2008 presidential election. This causal fact, according to the Lewisian analysis, reduces to the modal fact that $M \mathbin{\Box\!\!\rightarrow} O$. Supposing c is the event described by 'McCain chose Palin as a running-mate', this event does not occur in all and only the M-worlds indicated earlier—those worlds in which McCain choses *someone else* as a running-mate.[4] Likewise, assuming 'Obama wins the 2008 presidential election' describes event e, this event fails to occur in all and only the O-worlds, since those are the worlds in which Obama losses the election. The counterfactual 'if c had not occurred, then e would not have occurred' is just the counterfactual $M \mathbin{\Box\!\!\rightarrow} O$. In this way, the causal fact that McCain's choice of running-mate caused Obama's victory is reduced to modal facts about the proximity of M-worlds and O-worlds to the actual world. Assuming McCain's choice of running-mate did cause Obama to win the election, the (simple) counterfactual analysis of casuations yields the correct prediction.

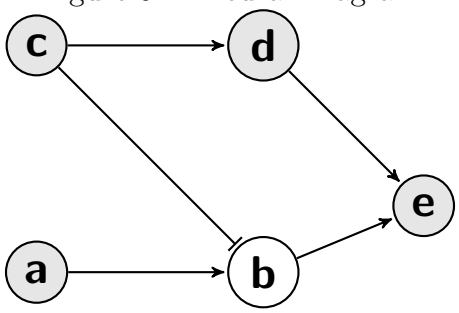
But as we noted above, the counterfactual analysis of causation offered by Lewis is slightly more complicated than the one we've rehearsed thus far. The discussion above reduces the casual relation between two events into the counterfactual dependence of *one* event on another. The (causing) event c causes the (effect) event e just in case c's non-occurrence entails e's non-occurrence. But this leads to (purportedly) counter-intuitive results.

Consider the diagram below, which represents the activation pattern of a neural network. Each circular node in the diagram represents a neuron, connected *via* the linear activation vectors passing from left to right in temporal order. The gray nodes are those that have activated, and the white nodes are those that remain unactivated. The normal arrows indicate the transmission of an activation signal stemming from an activated neuron, while the flat-headed arrow indicates a canceling signal that deactivates a node in the presence of a (distinct) activation signal. Reading the diagram, there are two causal paths, both of which bring about the activation of e. One originates from a and the other originates from c. When c activates it sends two signals: one runs down the connection with d, and the other (deactivation) signal runs down the connection to b. That deactivation signal cancels

[4]This ignores, for ease of explanation, the possibility than McCain does not choose a running-mate at all.

the signal originating at a, indicated by the whiteness of node b.

Figure 5.1: Neural Diagram



The contention is that this kind of case serves as a counterexample to the simple counterfactual account of causation. On the simple account, the fact that c's firing causes e's firing is reduced to the modal facts about the proximity of possible worlds in which c does not fire, to the actual world (where events unfold as indicated in Figure 5.1). The closest such world, the one *most similar* to the actual world, is the one where c does not fire, yet a does. The reasoning here is that such a world is closer to the actual world than any other, since it instantiates the greatest number of events that hold in the actual world. Of course, at such an a-firing-world (where c fails to fire), e fires as well. As such, the simple counterfactual, that if c does not fire, then e does not fire turns out to be false. Likewise, the causal claim that c's firing causes e's firing turns out to be false. Ostensibly, this is counterintuitive, since there is a clear causal connection between c and e in the actual world.

Lewis' *ancestral* counterfactual proposal solves this problem. Since the ancestral notion (as indicated in ($\mathcal{C}$)) requires us to evaluate *two* counterfactuals, given the intervening event of d's firing, this permits us to isolate the conditions that make the counterexample troubling, and yield the correct verdict about the causal claim that c causes e. The counterfactual $C \mathbin{\Box\!\!\rightarrow} D$ turns out true, since the closest non-c-world is also a non-d-world. Holding fixed all other events, when c fails to fire, d also fails to fire. The counterfactual $D \mathbin{\Box\!\!\rightarrow} E$ also turns out true on this proposal. Because we only need to consider the events moving forward from the time of d's firing, when we consider the worlds at which d fails to fire, we are not required to retrodict the non-firing of c. Since we hold fixed all past events, including the firing of c which would prevent the firing of b, we maintain that b does not fire. Thus, the closest non-d-world is also a non-b-world, and hence a non-e-world. Given this analysis, both counterfactuals ($C \mathbin{\Box\!\!\rightarrow} D$ and $D \mathbin{\Box\!\!\rightarrow} E$) turn out to be true, and thus

the causal fact that c's firing causes e's firing is predicted by the account.

The ancestral counterfactual account cannot, however, accommodate cases of *preemption*. These are cases that are structurally similar to the case above, but in such cases the intervening causal events are absent. Lewis (2000) offers such a case:

> [Rocks]
> Billy and Suzy throw rocks at bottles. Suzy throws first, or maybe throws harder. Her rock arrives first. The bottle shatters. When Billy's rock gets to where the bottle used to be, there is nothing but flying shards of glass. (Lewis, 2000, p. 184)

Much like the neural case above, there are two distinct causal chains, both of which lead to the same terminating event. In Rocks however, there are no intervening events on which the ancestral analysis can pivot, yielding the (ostensibly) counter-intuitive result that Suzy's throw did not cause the bottle to shatter. If Suzy fails to throw her rock, Billy's rock shatters the bottle. Thus, the closest world at which Suzy fails to throw is a world at which the bottle breaks.

The modern history of debates about causation is aptly described as offering criticisms of, and responses to, case-based counterexamples to Lewis' initial proposal, in much the way Rocks purportedly does.[5] Thus, it would be helpful at this juncture to ask what Lewis' analysis is supposed to analyze, and whether case-based intuitions aid in such an analysis. Lewis' account is clearly meant to offer up ($\mathcal{C}$) as an explanation of *something*, but what are those *explananda*?

One plausible candidate for the *explananda* of this account is the semantics for the English expression 'cause', as exhibited by the judgments of competent English-speakers. After all, cases like Rocks are meant to solicit intuitions about causal claims from the readers of those cases. Perhaps the purpose of this analysis, and the cases that inform it, is to give a semantics for the English expression 'cause'. Lewis addresses this question in a footnote, indicating that his proposal regards causal *facts*, *not* linguistic objects. Further, Collins et. al. echo this goal in their introductory contribution to a prominent volume on the metaphysics of causation. In fact, they indicate that the central misstep of a competing analysis defended by Davidson (1967a) is the focus on sentences instead of propositions (Collins et al., 2004, p. 17). They insist that the evidence brought to bear by dissecting cases, in the manner above, informs us about *propositions* and causal *facts*, not merely

[5] See Paul & Hall (2013) for discussion.

linguistic expressions that invoke the term 'cause'. However, insisting that human intuitions about cases like ROCKS inform us about the nature of propositions, and some underlying relation *cause* that constitutes (or is reduced to modal facts about) those proposition, presupposes that the English expressions used in describing these cases enable English speakers to grasp the propositions Lewis' theory is meant to explain. This assumption, as we saw in Chapter 3, is deeply troubled. More importantly, even if we accept this (troubled) externalist assumption, to deny that this methodology is engaged in semantics is odd given the work propositions are meant to do. After all, if we take the relation *cause*, as it manifests in propositions, as the *explananda* of our theory, given that one of the primary philosophical jobs of propositions is to serve as the meanings of natural language expressions, analyzing the purported modal facts about these propositions *via* cases must thereby be a semantic project.

If propositions (or their constituents) are the object of explanation, Lewis' theory gives us the truth-conditions for propositions (like c-causes-e) in terms of a proposed proximity relation that holds between possible worlds. One can *claim*, of course, that such a theory gives the truth-conditions of certain classes of propositions, and thus yields metaphysical verdicts about the nature of *causation*. But given that propositions, on the externalist proposal, are the meanings of natural language expressions, and that the cases used to ostensibly inform us about these propositions are presented using a natural language, the meanings of expressions like 'cause' play a central role in providing such information.

Cases like ROCKS are proposed to undermine the Lewisian hypothesis about the (reductive) nature of *causation*, because it yields the wrong truth-value for the proposition that Suzy caused the bottle to break, with respect to the world described in the case. Purportedly, this is because the proposition that Suzy did not throw a rock, and the proposition that the bottle did not break, do not bear the right sort of relation to the case-world. But, we only come to know this, by having an intuition of a certain sort, because we can grasp which propositions are relevant for the purpose of such inquiry—and *that* is a result of our ability to comprehend the English language expressions 'Suzy caused the bottle to break', 'Suzy did not throw a rock', and 'the bottle did not break'. Thus, it must be, that in giving the truth-conditions for the proposition that Suzy caused the bottle to break, Lewis' theory also gives us the truth-conditions for the English expression 'Suzy caused the bottle

to break' whose meanings is just the proposition the theory is meant to analyze. Otherwise, our intuitions regarding ROCKS would have no bearing on the viability of Lewis' proposal. Thus, Lewis must be offering a semantic proposal about the meaning of the English expression 'cause', under the assumption that a semantics for this expression are given by, and indicative of, worldly objects and their (modal) properties.

Notice that the Realist's methodology is in full swing here, in giving an account of causation. On the assumption that externalism holds for natural languages, investigating the meanings of causal expressions like 'Suzy's throw caused the bottle to break' has led us to a number of metaphysical conclusions. Lewis' contention is that his analysis of counterfactuals is the most parsimonious, and that this licenses some robust metaphysical conclusions (Lewis, 1986, Ch. 1). This analysis requires that we quantify over certain kinds of objects, like *events* and *possible worlds*. Since the truth-conditions for counterfactuals require a plenitude of possible worlds as their satisfiers, possible worlds must occupy the domain—a domain that, according to the Realist constitutes what there is. Unless a more metaphysically modest account of the meanings of counterfactual claims can explain the judgments of English language speakers, we can conclude that possible worlds exist. Further, that a proximity measure is required to give an adequate semantics for causal expressions, those possible worlds must stand in relations of similarity (again, in the absence of a better semantics for 'cause'). For the Realist, this is how metaphysical investigation proceeds, by consulting our capacity to understand expressions in a given language, and on the assumption that this understanding determines what the world must be like for expressions in that language to be true, proclaim what must exist if any sentences in that language are to be true. In the case we've been focusing on, for any causal expression to be true, there must be things in the domain like *events*, *possible worlds*, and a *proximity* relation that holds between triplets of worlds. But again, this method of investigating metaphysical questions is only viable if the language used for ontological investigation has an externalist semantics. To that end, natural languages seem ill-suited.

The Realist defending Lewis' proposal could insist that the conception of *causation* under investigation is *not* the one that serves to underwrite the meanings of the natural language expressions containing 'cause', holding that some more refined, and ontologically distinguished conception is the subject of investigation. This con-

ception is manifest in propositions, but these propositions do duty as meanings of expression in some privileged (non-natural) language. This is to insist that there is some ontologically privileged language which contains a term, let's call it 'cause*', whose meaning denotes the genuine notion of *causation* that these discussions of neural networks, and stone-throwings undercover.

But if metaphysicians are in the business of studying such a notion, the utility of pumping intuitions about cases like ROCKS is at best opaque. On the assumption that the (troubled) externalist thesis for natural languages holds, the purpose of consulting the case-based intuitions of English-speakers is somewhat clear, insofar as such individuals speak the language in which we ask them questions like "Is the sentence 'Suzy's throw caused the bottle to break' true given ROCKS?" But if the term whose underlying propositional-constituent meaning is part of a non-natural, ontologically privileged language, then *English*-speakers judgments about sentences like 'Suzy's throw caused* the bottle to break' give us little ontological guidance. After all, such speakers are incapable of grasping the propositional meaning of a sentence in a language they do not understand,[6] so their verdicts about cases will fail to involve the proposition up for analysis. Likewise, if the defender of this methodology wants to insist that there is some interesting scientific notion of *cause* that is useful for ontological investigation, such a philosopher must explain how stories told in a natural language like English make use of such a conception. Whether or not the languages used to express our best scientific theories need a term 'cause*' is a potentially interesting and philosophically fruitful hypothesis (though considerations discussed in Chapter 4 should temper one's enthusiasm). But the plausibility of that proposal should live or die by the explanatory benefit wrought by the theories that include such a term. And surely, whether or not (say) physics needs a term 'cause*' cannot be decided by the linguistic judgments of English-speakers regarding cases like ROCKS.

The Realist has two choices in adopting the methodology of Conceptual Analysis: 1) either she accepts that her project involves giving a semantics for natural language expressions, or 2) she insists that her investigation pertains to some on-

[6] At least if the attempted means of grasping such a proposition is simply the comprehension of sentences expressed using a language they do not understand. Assuming there are propositional meanings, as an English speaker I can surely grasp the proposition that the cat is on the mat. But I cannot grasp that proposition by comprehending the French sentence 'Le chat est sur le tapis' since I am not a French speaker.

tologically privileged language that contains the term of interest. The first option rescues her use of intuition-pumping cases in ontological investigations, but commits her to the applicability of ($\mathcal{E}$) to natural languages. As we have seen in Chapter 3 this is a risky commitment. The second option avoids the risky theoretical commitment (in exchange for a somewhat less risky theoretical commitment regarding the semantics for scientific languages as discussed in Chapter 4), but still comes at a cost. Namely, she sacrifices the methodology that pumps intuitions of natural language speakers. For metaphysical investigations of *causation* the dilemma is profound, given that much of the literature makes use of natural language intuition-pumping cases as evidence for a particular theory about *causation*, while (at least implicitly) assuming without defense that ($\mathcal{E}$) is true for natural languages.[7]

As was the case with 'belief' and the Extended Mind Hypothesis, the argument for using case-based intuitions to supply evidence for a given theory about *causation* relies on the externalist assumption. In the wake of the evidence outlined in Chapter 3, the Realist investigating *causation* might insist that while natural languages are poor investigatory tools, the language she has in mind is rigid and precise in the way natural languages are not. One wants to know then why we should think that this precise language cleaves to the structure of the world. If this language is the product of research in a scientific domain, constructed for the purpose of expressing theories in that domain, the justification for using such a language in ontological investigation is parasitic on the epistemic credentials of naturalistic methodologies. But once the Realist has accepted that the notion they are interested in investigating is not one that serves as the meaning of a natural language expression, she undermines her argument for making use of speaker intuitions as evidence for a particular hypothesis about the notion in question. There might well be some terms *mind** and *belief** that, when added to the language of psychology, help to yield novel predictions and more useful generalizations. But there's no reason to think English-speaker judgments about cases like Otto′ and Inga* are indicative of such success. Likewise, theories that use the term *cause** may yield some explanatory benefit across many scientific domains. But there is little reason to think that speaker judgments about cases like Rocks are suggestive of such a benefit. Thus the Realist that makes use of

[7]For (a sampling of) examples of authors that make use of case intuitions as counterexamples: Hall (2000); Hitchcock (1996); Mackie (1974); Paul & Hall (2013) and citations therein; Schaffer (2005); Woodward (1984).

Conceptual Analysis as a means of investigating metaphysical questions either owes us a response to the problems outlined for adopting an externalist semantics for natural language, or she must abandon her methodology in favor of a naturalistic one. Neither option bodes well for the current, prolific use of Conceptual Analysis to resolve metaphysical disputes.

5.4 Conclusion

There is an enduring tradition in philosophy of treating the meanings of natural language expressions as externalist. As with most traditions, the reasons for abiding this history often go unarticulated. As a consequences, the bold character of the externalist hypothesis is ignored, in favor of making equally bold proclamations about the nature of reality. The flexibility of natural languages, and the naturalistic commitments of the Realism that motivates these proclamations countenance the difficulty in accepting these traditional assumptions about the nature of meaning. The Realist is thereby burdened to defend her methodology, or abandon it in favor of one more amenable to her naturalist commitments. However, if this newly adopted methodology invokes the use of scientific languages, insisting that the epistemic rigor of naturalistic investigation forges languages that "cut nature at the joints," the apparent flexibility of biological terms like 'gene' and 'species' should give the Realist pause. More strikingly, such a methodological shift precludes the use of Conceptual Analysis as a tool for ontological investigation, at least insofar as the case descriptions used therein are expressed in natural languages. Given the centrality of Conceptual Analysis to ontological investigation, excising this method requires a genuine revision to the practices of philosophers in addressing metaphysical questions, and a radical reexamination of the *explananda* in many philosophical domains. The upshot to this expulsion is the re-development of a field of research where the difference between the questions that have a hope of being addressed can be clearly marked from those that do not.

Bibliography

Adams, F. R., & Aizawa, K. (2010). Defending the Bounds of Cognition. In R. Menary (Ed.) *The Extended Mind*. Mit Press.

Andersson, L. (1990). The Driving Force: Species Concepts and Ecology. *Taxon*, (pp. 375–382).

Aristotle (1941). *On the Generation of Animals*. Loeb Classical Library.

Austin, J. (1962). *How to Do Things with Words*. Oxford University Press.

Bach, K. (1999). The Semantic-Pragmatic Distinction: What It Is and Why It Matters. In K. Turner (Ed.) *The Semantic-Pragmatics Interface from Different Points of View*, (pp. 65–84). Oxford: Elsevier.

Bach, K. (2001). You Dont Say. *Sythèse*, *128*, 14–44.

Bach, K. (2004). Minding the Gap. In *The Semantics/Pragmatics Distinction*, (pp. 27–43). Stanford: CSLI Publications.

Baker, M. (2002). *The Atoms of Language*. Basic Books.

Barwise, J., & Perry, J. (1983). *Situations and Attitudes*. Bradford Books, MIT Press.

Beaney, M. (Ed.) (1997). *The Frege Reader*. Oxford: Basil Blackwell.

Bennett, J. (1974). Counterfactuals and Possible Worlds. *Canadian Journal of Philosophy*, *4*(2), 381–402.

Beurton, P. J., Falk, R., & Rheinberger, H.-J. (Eds.) (2000). *The concept of the gene in development and evolution: Historical and epistemological perspectives*. Cambridge University Press.

Bloom, P. (2000). *How Children Learn the Meaning of Words*. MIT Press.

Boeckx, C. (2006). *Linguistic Minimalism : Origins, Concepts, Methods, and Aims*. Oxford Univerisity Press.

Boghossian, P. A. (1991). Naturalizing Content. In B. M. Loewer, & G. Rey (Eds.) *Meaning in Mind: Fodor and His Critics*. Blackwell.

Bromberger, S., & Halle, M. (1995). The Ontology of Phonology. *Studies in Logic and the Foundations of Mathematics*, *134*, 725–743.

Burge, T. (1979a). Individualism and the Mental. *Midwest Studies in Philosophy*, *4*(1), 73–122.

Burge, T. (1979b). Semantical Paradox. *Journal of Philosophy*, *76*(4), 169–198.

Cappelen, H., & Lepore, E. (2005). *Insensitive Semantics: A Defence of Semantic Minimalism and Speech Act Pluralism*. Blackwell Publishing.

Carey, S. (2009). *The Origin of Concepts*. Oxford University Press.

Carnap, R. (1947). *Meaning and Necessity: A Study in Semantics and Modal Logic*. University of Chicago Press.

Carruthers, P. (2002). The Cognitive Functions of Language. *Behavior and Brain Sciences*, *25*, 261–316.

Chomsky, N. (1957). *Syntactic Structures*. Mouton.

Chomsky, N. (1977). *Essays on Form and Interpretation*. Elsevier.

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, N. (1986). *Knowledge of Language*. Praeger Publishing.

Chomsky, N. (1992). Explaining Language Use. *Philosophical Topics*, *20*(1), 205–231.

Chomsky, N. (2000). *New Horizons in the Study of Language and Mind*. Cambridge University Press.

Chruch, A. (1951). A Formulation of the Logic and Sense of Denotation. In P. Henle, H. Kallen, & S. Langer (Eds.) *Structure, Method and Meaning*. Liberal Arts Press.

Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis*, (pp. 7–19).

Collins, J., Hall, N., & Paul, L. A. (2004). Counterfactuals and causation: History, problems, and prospects. In J. Collins, N. Hall, & L. Paul (Eds.) *Causation and Counterfactuals*, (pp. 1–57). The Mit Press.

Davidson, D. (1965 [1984]). Theories of Meaning and Learnable Languages. In *Inquiries into Truth and Interpretation*. New York: Oxford University Press.

Davidson, D. (1967a). Causal relations. *Journal of Philosophy*, *64*(21), 691–703.

Davidson, D. (1967b). The Logical Form of Action Sentences. In N. Rescher (Ed.) *The Logic of Decision and Action*. University of Pittsburgh Press.

Davidson, D. (1967c). Truth and Meaning. *Synthese*, *17*, 304–323.

De Queiroz, K. (2007). Species Concepts and Species Delimitation. *Systematic Biology*, *56*(6), 879–886.

DeRose, K. (1992). Contextualism and Knowledge Attributions. *Philosophy and Phenomenological Research*, *52*(4), 913–929.

Dummett, M. (1993). *The Seas of Language*. Oxford University Press.

Dupré, J. (1999). On the Impossibility of a Monistic Account of Species. In R. A. Wilson (Ed.) *Species: New interdisciplinary Essays*. Cambridge, MA: Massachusetts Institute of Technology Press.

Ereshefsky, M. (1992). *The Units of Evolution: Essays on the Nature of Species*. The MIT Press.

Ereshefsky, M. (1998). Species Pluralism and Anti-Realism. *Philosophy of Science*, *65*(1), 103–120.

Fine, K. (1975). Critical Notice of Lewis 1973. *Mind*, *84*, 451–58.

Fodor, J. (1975). *The Language of Thought*. Harvard University Press.

Fodor, J. (1983). *The Modularity of Mind*.

Frege, G. (1879 [1997]). Begriffschrift. In M. Beaney (Ed.) *The Frege Reader*, (pp. 47–79). Oxford: Basil Blackwell.

Frege, G. (1891 [1997]). Function and Concept. In M. Beaney (Ed.) *The Frege Reader*. Oxford: Basil Blackwell.

Frege, G. (1892 [1997]). On *Sinn* and *Bedeutung*. In M. Beaney (Ed.) *The Frege Reader*. Oxford: Basil Blackwell.

Frege, G. (1918 [1997]). Thought. In M. Beaney (Ed.) *The Frege Reader*. Oxford: Basil Blackwell.

girontravel.se (2013). The town of kiruna.
URL `http://girontravel.se/en/fun-facts/the-town-of-kiruna`

Godfrey-Smith, P. (1989). Misinformation. *Canadian Journal of Philosophy*, *19*(4), 533–50.

Goldman, A. I. (2007). A Program for "Naturalizing" Metaphysics, with Application to the Ontology of Events. *The Monist*, *90*(3), 457–479.

Goodman, N. (1981). Routes of Reference. *Critical Inquiry*, (pp. 121–132).

Grice, H. P. (1957 [1989]). Meaning. In h. P. Grice (Ed.) *Studies in the Way of Words*, (pp. 213–223). Harvard University Press.

Grice, H. P. (1967 [1989]). Logic and Conversation. In H. P. Grice (Ed.) *Studies in the Way of Words*, (pp. 86–116). Harvard University Press.

Grice, P. (Ed.) (1989). *Studies in the Way of Words*. Harvard University Press.

Griffiths, P. E. (1999). Squaring the Circle: Natural Kinds with Historical Essences. In R. A. Wilson (Ed.) *Species: New interdisciplinary Essays*. Cambridge, MA: Massachusetts Institute of Technology Press.

Hall, N. (2000). Causation and the Price of Transitivity. *Journal of Philosophy*, *97*(4), 198–222.

Heim, I., & Kratzer, A. (1998). *Semantics in Generative Grammar*. Blackwell Publishers.

Hinzen, W. (2006a). Internalism and Truth. *Mind & Society*, *5*, 139–166.

Hinzen, W. (2007). *An Essay on Names and Truth*. OUP Oxford.

Hinzen, W. (2014). Recursion and truth. In T. Roeper, & M. Speas (Eds.) *Recursion: Complexity in Cognition*, vol. 43 of *Studies in Theoretical Psycholinguistics*, (pp. 113–137). Springer International Publishing.
URL `http://dx.doi.org/10.1007/978-3-319-05086-7`$_6$

Hirsch, E. (2002). Quantifier Variance and Realism. *Philosophical Issues*, *12*(1), 51–73.

Hirsch, E. (2008). Language, Ontology, and Structure. *Noûs*, *42*(3), 509–528.

Hitchcock, C. R. (1996). The Role of Contrast in Causal and Explanatory Claims. *Synthese*, *107*(3), 395–419.

Holmes, O. W. (1899). Law in Science and Science in Law. *Harvard Law Review*, *12*, 443–463.

Hornstein, N. (1984). *Logic as Grammar*. MIT Press.

Jackendoff, R. (1990). *Semantic Structures*. MIT Press.

Jackendoff, R. (1996). Semantics and Cognition. In S. Lappin (Ed.) *The Handbook of Contemporary Semantic Theory*, (pp. 539–559). Oxford: Blackwell Publishing.

Jackendoff, R. (2002). *Foundations of Language*. Oxford Univerisity Press.

Johannsen, W. (1909). *Elemente der exakten Erblichkeitslehre. Mit Grundzügen der biologischen Variationsstatistik*. Jena: Gustav Fischer.

Kaplan, D. (1977 [1989]). Demonstratives. In J. Almog, J. Perry, & H. Wettstein (Eds.) *Themes from Kaplan*, (pp. 481–563). Oxford University Press.

Kellert, S. H., Longino, H. E., & Waters, C. K. (2006). *Scientific Pluralism Vol. 19*, vol. 19. University of Minnesota Press.

Kennedy, C., & Stanley, J. (2009). On 'average'. *Mind*, *118*(471), 583–646.

King, J. (2007). *The Nature and Structure of Content*. New York: Oxford University Press.

Kitcher, P. (1984). Species. *Philosophy of Science*, *51*(2), 308–333.

Kripke, S. (1980). *Naming and Necessity*. Harvard University Press.

Kumar Chellapilla, P. S. M. C., Kevin Larson (2005). Computers Beat Humans at Single Character Recognition in Reading Based Human Interaction Proofs (HIPs). In *CEAS 2005 — The Second Conference on Email and Anti-Spam*.

Larson, R., & Segal, G. (1995). *Knowledge of Meaning: An Introduction to Semantic Theory*. MIT Press.

Lewis, D. (1970). General Semantics. *Synthese*, *22*, 18–67.

Lewis, D. (1973a). Causation. *Journal of Philosophy*, *70*(17), 556–567.

Lewis, D. (1975). Languages and Language. In *Language, Mind, and Knowledge*. University of Minnesota Press.

Lewis, D. (1986). *On the Plurality of Worlds*. Blackwell Publishing.

Lewis, D. (2000). Causation as Influence. *Journal of Philosophy*, *97*(4), 182–197.

Lewis, D. K. (1973b). *Counterfactuals*. Blackwell Publishers.

Lidz, J., Pietroski, P., Hunter, T., & Halberda, J. (2011). Interface Transparency and the Psychosemantics of Most. *Natural Language Semantics*, *19*(3), 227–256.

Loewer, B. M. (1997). A Guide to Naturalizing Semantics. In C. Wright, & B. Hale (Eds.) *A Companion to the Philosophy of Language*, (pp. 108–126). Blackwell: Oxford.

Ludlow, P. (1995). Logical Form and the Hidden-Indexical Theory: A Reply to Schiffer. *The Journal of Philosophy*, (pp. 102–107).

Ludlow, P. (2003). Referential Semantics for I-Languages? In L. M. Antony, & N. Hornstein (Eds.) *Chomsky and his Critics*, (pp. 140–161). Wiley Online Library.

Ludlow, P. (2011). *The Philosophy of Generative Linguistics*. Oxford University Press.

Lycan, W. G. (1990). The Continuity of Levels of Nature. In W. G. Lycan (Ed.) *Mind and Cognition: A Reader*, (pp. 77–96).

Macdonald, G., & Papineau, D. (Eds.) (2006). *Teleosemantics*. Oxford University Press.

Mackie, J. L. (1974). *The Cement of the Universe*. Oxford,Clarendon Press.

May, R. C. (1978). *The Grammar of Quantification.*. Ph.D. thesis, Massachusetts Institute of Technology.

Mayden, R. L. (1997). A Hierarchy of Species Concepts: the Denoument in the Saga of the Species Problem. In M. R. W. M. F. Claridge, H. A. Dawah (Ed.) *Species: The Units of Diversity*, (pp. 381–423). London: Chapman and Hall.

Mayr, E. (1970). *Populations, Species and Evolution: an Abridgement of Animal Species and Evolution*. Harvard University Press.

McGinn, C. (1982). The Structure of Content. In A. Woodfield (Ed.) *Thought and Object*. Oxford University Press.

Michotte, A. (1946 [1963]). *The Perception of Causality*. Basic Books.

Moss, L. (2002). *What Genes Can't Do*. MIT Press.

Okasha, S. (2002). Darwinian Metaphysics: Species and the Question of Essentialism. *Synthese*, *131*(2), 191–213.

Paterson, H. E. (1985). The Recognition Concept of Species. *Species and Speciation*, (pp. 21–29).

Paul, L. A., & Hall, N. (2013). *Causation: A User's Guide*. Oxford.

Pharoah, P. D., Day, N. E., Duffy, S., Easton, D. F., & Ponder, B. A. (1997). [family history and the risk of breast cancer: a systematic review and meta-analysis. *International Journal of cancer*, *71*(5), 800–809.

Piattelli-Palmarini, M., Uriagereka, J., & Salaburu, P. (2009). *Of Minds and Language: A Dialogue with Noam Chomsky in the Basque Country*. OUP Oxford.

Pietroski, P. (2003). The Character of Natural Language Semantics. In A. Barber (Ed.) *Epistemology of Language*, (pp. 217–256). Oxford University Press.

Pietroski, P. (2005). Meaning Before Truth. In G. Preyer, & G. Peter (Eds.) *Contextualism in Philosophy*, (pp. 255–302). Oxford University Press.

Pietroski, P. (2006). Character Before Content. In J. Thomson, & A. Byrne (Eds.) *Content and Modality: Themes from the Philosophy of Robert Stalnaker*. Oxford University Press.

Pietroski, P. (2008). Minimalist Meaning, Internalist Interpretation. *Biolinguistics*, *4*, 317–341.

Pietroski, P. (2010). Concepts, Meaning and Truth. *Mind & Language*, *24*(3), 247–278.

Pietroski, P. (forthcoming). *Conjoining Meanings: Semantics Without Truth-Values*. Oxford University Press.

Pietroski, P., Lidz, J., Hunter, T., & Halberda, J. (2009). The Meaning of 'Most': Semantics, Numerosity and Psychology. *Mind & Language*, *24*(5), 554–585.

Putnam, H. (1975). *Mind, Language and Reality (Philosophical Papers)*, vol. 2. Cambridge University Press.

Quine, W. (1960). *Word and Object*. MIT Press.

Quine, W. V. O. (1948). On What There Is. *Review of Metaphysics*, *2*, 21–38.

Quine, W. V. O. (1973). *The Roots of Reference*. Open Court La Salle.

Rolander, N. (2013). Swedish City of Kiruna Plans Massive Relocation. *Wall Street Journal*.
URL `http://on.wsj.com/1w3GxiL`

Rothschild, D., & Segal, G. (2009). Indexical Predicates. *Mind & Language*, *24*(4), 467–493.

Russell, B. (1903). *The Principles of Mathematics*. Cambridge University Press.

Russell, B. (1905). On Denoting. *Mind*, *14*(56), 479–493.

Russell, B. (1912). *The Problems of Philosophy*, vol. 22. Barnes & Noble Books.

Russell, B. (1957). Mr. Strawson On Referring. *Mind*, *66*(263), 385–389.

Salmon, N. (1986). *Frege's Puzzle*. Ridgeview Publishing Company.

Schaffer, J. (2005). Contrastive Causation. *Philosophical Review*, *114*(3), 327–358.

Schein, B. (2002). Events and the Semantic Content of Thematic Relations. In G. Preyer, & G. Peter (Eds.) *Logical Form and Language*, (pp. 263–344). Oxford University Press.

Scholl, B. J., & Nakayama, K. (2002). Causal Capture: Contextual Effects on the Perception of Collision Events. *Psychological Science*, *13*(6), 493–498.

Schönfinkel, M. (1924). Über die bausteine der mathematischen logik. *Mathematische Annalen*, *92*, 305–316.

Segal, G. (2000). *A Slim Book on Narrow Content*. Bradford Books, MIT Press.

Sher, G. (1999). What is Tarski's Theory of Truth. *Topoi*, *18*, 149–166.

Sider, T. (2002). *Four Dimensionalism: An Ontology of Persistence and Time*. Oxford: Clarendon Press.

Sider, T. (2009). Ontological Realism. In D. Chalmers, D. Manley, & R. Wasserman (Eds.) *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford University Press.

Sider, T. (2011). *Writing the Book of the World*. Oxford University Press.

Soames, S. (1984). What is a Theory of Truth? *Journal of Philosophy*, *81*, 411–429.

Soames, S. (1985). Lost Innocence. *Linguistics and Philosophy*, *58*, 59–71.

Sober, E. (1980). Evolution, Population Thinking, and Essentialism. *Philosophy of Science*, *47*(3), 350–383.

Spelke, E. S. (2003). What Makes Us Smart? core Knowledge and Natural Language. In *Language in Mind: Advances in the Study of Language and Thought*, (pp. 277–311). MIT Press Cambridge, MA.

Stainton, R. (2006). Meaning and Reference: Chomskian Themes. In E. LePore, & B. Smith (Eds.) *The Oxford Handbook of Philosophy of Language*, (pp. 913–940). Oxford Univerisity Press.

Stalnaker, R. (1976). Propositions. *Issues in the Philosophy of Language*, (pp. 79–91).

Stalnaker, R. (1984). *Inquiry*. MIT Press.

Stanford, P. K. (1995). For Pluralism and Against Realism About Species. *Philosophy of Science*, *62*(1), 70–91.

Stanley, J. (2005). Semantics in Context. In G. Preyer, & G. Peter (Eds.) *Contextualism in Philosophy: Knowledge, Meaning, and Truth*, (pp. 221–54). Oxford University Press.

Stanley, J. (2007). *Language in Context: Selected Essays*. Oxford University Press.

Stotz, K., & Griffiths, P. (2004). Genes: Philosophical Analyses put to the Test. *History and Philosophy of the Life Sciences*, *26*(1), 5.

Strawson, P. F. (1950). On Referring. *Mind*, *59*(235), 320–344.

Tarski, A. (1933 [1983]). Concept of Truth in Formalized Languages. In J. Woodger (Ed.) *Logic, Semantics, and Metamathematics*, (pp. 152–278). Indianapolis Hackett Publishing Company.

Tarski, A. (1944). The Semantic Conception of Truth: and Foundations in Semantics. *Philosophy and Phenomenological Research*, *4*(3), 341–376.

Thomason, R. (Ed.) (1974). *Formal Philosophy: The Selected Papers of Richard Montague*. Yale University Press.

Thomasson, A. L. (1999). *Fiction and Metaphysics*. Cambridge: Cambridge University Press.

Vogel, C. A., Wellwood, A., Dudley, R., & Ritchie, J. B. (2014). Talking about Causing Events. *Baltic International Yearbook of Cognition, Logic and Communication*, *9*.

Von Ahn, L., Blum, M., Hopper, N. J., & Langford, J. (2003). Captcha: Using hard ai problems for security. In *Advances in CryptologyEUROCRYPT 2003*, (pp. 294–311). Springer.

Waters, C. (2000). Molecules Made Biological. *Revue Internationale de Philosophie*, *4*, 539–564.

Waters, C. K. (2005). Why Genic and Multilevel Selection Theories Are Here to Stay. *Philosophy of Science*, *72*(2), 311–333.

Weber, M. (2005). *Philosophy of Experimental Biology*. Cambridge University Press.

Wilson, R. A., Barker, M. J., & Brigandt, I. (2007). When Traditional Essentialism Fails: Biological Natural Kinds. *Philosophical Topics*, *35*(1), 189.

Wittgenstein, L. (1953). *Philosophical Investigations*. Macmillan.

Wittgenstein, L. (1972). *On Certainty*. Harper Perennial.

Woodward, J. (1984). A Theory of Singular Causal Explanation. *Erkenntnis*, *21*(3), 231–262.